tesis doctoral

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA

METAMODELO PARA VINCULACIÓN DE RECURSOS EDUCATIVOS ABIERTOS MEDIANTE ESPECIFICACIONES LOD Y BASADO EN PRINCIPIOS DE CONFIANZA.

Jhon Francined Herrera Cubides

Énfasis en Ciencia de la información y el conocimiento



Doctorado en Ingeniería

http://doctoradoingenieria.udistrital.edu.co



METAMODELO PARA VINCULACIÓN DE RECURSOS EDUCATIVOS ABIERTOS MEDIANTE ESPECIFICACIONES LOD Y BASADO EN PRINCIPIOS DE CONFIANZA

M.Sc. Jhon Francined Herrera Cubides

Universidad Distrital “Francisco José de Caldas"
Facultad de Ingeniería
Doctorado en Ingeniería
Bogotá - Colombia
2021

METAMODELO PARA VINCULACIÓN DE RECURSOS EDUCATIVOS ABIERTOS MEDIANTE ESPECIFICACIONES LOD Y BASADO EN PRINCIPIOS DE CONFIANZA

M.Sc. Jhon Francined Herrera Cubides

Tesis presentada como requisito parcial para optar al título de:
Doctor en Ingeniería

Director:
Paulo Alonso Gaona García, PhD.

Co-Director:
Salvador Sánchez Alonso, PhD.

Grupo de Investigación GIIRA

Universidad Distrital “Francisco José de Caldas"
Doctorado en Ingeniería
Énfasis en Ciencia de la información y el conocimiento
2021

Comisión de Doctorado

Esta tesis, titulada "*Metamodelo para vinculación de recursos educativos abiertos mediante especificaciones LOD y basado en principios de confianza*", escrita por Jhon Francined Herrera Cubides, ha sido aprobada en cuanto a estilo y contenido intelectual.

Hemos leído esta tesis y la aprobamos,

Jurado 1

_____

Dr. Juan Manuel Cueva Lovelle

Jurado 2

_____

Dr. José Fernando López Quintero

Jurado 3

_____

Dr. Orlando Martínez Montoya

_____

Dr. Paulo Alonso Gaona García.
Director

_____

Dr. Salvador Sánchez Alonso.
Co-Director

Fecha de la defensa: 16 de Septiembre de 2021

© Derechos de autor 2021 de Jhon Francined Herrera Cubides.
Todos los derechos reservados.

Resumen

Actualmente se presenta una alta disponibilidad de recursos educativos publicados en la Web, una gran mayoría de estos recursos están publicados de forma propietaria. Esto ha generado que la ubicación, el acceso, la consulta y recuperación de aquellos recursos educativos que se encuentren publicados de forma abierta y se ajusten a las necesidades de los usuarios, sea una necesidad cada vez más creciente. Necesidad que se manifiesta sobre todo cuando estos recursos, junto con los escenarios de formación virtual, se transforman en elementos y estrategias claves al momento de abordar situaciones como la que hemos vivido por la presencia de la pandemia del Covid-19. En la actualidad, en el dominio de conocimiento de los Recursos Educativos Abiertos (REA), los procesos de consulta y recuperación se presentan diferentes retos y desafíos asociados a limitadas descripciones de sus propiedades, necesidades particulares de las organizaciones que publican, entre otros. Adicionalmente, las descripciones de estos REA manifiestan diferentes problemas de calidad de datos, sus modelos de datos no se encuentran disponibles de forma abierta, y por último, estas descripciones no ofrecen un enriquecimiento semántico, que contribuya a proporcionar más contexto a los usuarios en los procesos de búsqueda y recuperación de recursos, situaciones que han limitado la disponibilidad y la reutilización de los REA. Derivado de estas problemáticas, se identificó la necesidad de proveer una descripción mejorada y enriquecida de los REA, que atendiera elementos de calidad de datos desde las etapas de abstracción y modelamiento de propiedades, como primeras etapas del ciclo de vida de datos abiertos y enriquecidos.

Considerando la necesidad mencionada, y considerando la articulación de diferentes componentes teóricos que proporcionan restricciones y criterios para su implementación, en la presente tesis, se propone el diseño de un metamodelo que articula una descripción de REA, mejorada a través de la aplicación de dimensiones de calidad de datos, junto con un enriquecimiento semántico provisto por datos enlazados. Este metamodelo permite la generación de un modelo para describir REAs, el cual es mapeado en dos formatos: Dublin Core, como estándar de metadatos ampliamente reconocido; y RDF, como modelo de datos para llevar a cabo el proceso de enlace de datos. Este modelo mejorado para describir REA y sus mapeos, contribuyen a mejorar la disponibilidad y la reutilización de los REA, además de proporcionar un conjunto de propiedades más amplio que permite y facilita procesos de búsqueda y recuperación de recursos educativos por parte de los usuarios. Así mismo, permite un mayor nivel de interoperabilidad con otros modelos, al contar con un mayor número de propiedades; así como la mejora de criterios para aplicaciones que realizan procesos de búsqueda y recuperación de recursos educativos.

Palabras clave

Recursos educativos abiertos, Web Semántica, datos enlazados, metadatos, calidad de datos, disponibilidad y reutilización de metadatos, modelo de datos RDF, enriquecimiento de datos.

Abstract

At present, it is shown high availability of educational resources on the Web; a large majority of those are posted with copyright. It allows that the siting, access, query, and retrieval of those which are open and tailored to the users' requirements, will come under an increasing need. This need is to be found when these resources, along with the virtual learning environments, become key elements and strategies at the moment of addressing issues such as the one experienced due to the COVID-19 pandemic. Currently, in the knowledge domain of Open Educational Resources (OER) and the querying and retrieving processes, different challenges -associated with scarce description of their features, special needs of those organizations that post, among others- are presented. Besides, the OER descriptions unveil diverse data quality issues, their data models are not openly available, and finally, those descriptions do not offer semantic enrichment that contributes to providing more context to the users in the consulting and retrieving processes of the resources, situations that have restricted availability and reuse of OER. Arisen from the problems mentioned previously, the need to provide an improved and enriched description of OER that addresses quality data elements from the abstraction and setting stages of the features, as first steps of open and enhanced data life cycle was identified.

Taking into account the aforementioned need and the structure of various theoretical components that render restrictions and standards for its implementation, the present thesis aims at designing a metamodel that articulates an OER description with improvements through the execution of data quality dimensions, along with the semantic enhancement endowed by linked data. This metamodel allows to build a model to describe Open Educational Resources which is mapped in two formats: Dublin Core, as the simplest and most widely used metadata schema; and RDF, as standard of data interchange to develop the interconnected data process. This improved model, to describe OER and their mapping, contribute to enabling the availability and reuse of OER, also, provides a wider set of characteristics that permits and facilitates searching and retrieving processes of resources from users. Moreover, it allows a greater interoperability level with other models when having a larger number of features, as criteria upgrading in searching and retrieving processes of educational resources for other applications.

Keywords

Open educational resources, Semantic Web, Linked data, metadata, data quality, availability and reuse of metadata, RDF data model, data enrichment.

Agradecimientos

En primera instancia, agradezco a Dios, por darme la oportunidad de abordar esta tarea y llevarla a feliz término, y por estar cada día en mi camino acompañándome y guiándome.

Seguidamente, agradecer a la Universidad Distrital Francisco José de Caldas, institución que hizo posible la realización de este trabajo, alma mater que siempre ha apoyado mis esfuerzos. A los docentes, administrativos y directivos del doctorado que han permitido un proceso de formación, con un trato respetuoso y diligente, enfocado en la investigación y construcción del conocimiento.

Mis más sinceros agradecimientos a mi director Paulo Alonso Gaona García, por sus valiosas y oportunas orientaciones y aportes en mi proceso de formación. Gracias a su dirección, apoyo y compromiso, fue posible realizar este gran logro, sin importar las dificultades presentadas durante este camino. De igual forma, quiero expresar mi más profundo agradecimiento a mi codirector, Salvador Sánchez Alonso, por su dirección y sus observaciones, las cuales fueron muy acertadas y oportunas para llevar a cabo la ejecución y finalización de este proyecto. A ellos, mis más sinceros agradecimientos, porque más allá de un proceso de guía, ha sido una formación permanente, forjando un gran lazo de amistad y colaboración.

De la misma forma quisiera agradecer a todos mis colegas, compañeros y amigos del grupo de investigación GIIRA. Al Dr. Carlos Enrique Montenegro por sus acertados concejos y por facilitar las actividades de trabajo de las investigaciones articuladas. Un agradecimiento especial para el Dr. José Nelson Pérez, q.e.p.d., quien proporcionó palabras de aliento y apoyo a este proyecto en su momento.

Adicionalmente, quisiera agradecer el apoyo ofrecido por el Dr. Wilson Leandro Pardo Osorio, por su apoyo incondicional y orientaciones en la construcción de apartados valiosos de esta investigación.

Dedicatoria

Dedico este trabajo a mis padres José Isidro (q.e.p.d) y Blanca Gilma, quienes forjaron mi vida a través de sus enseñanzas y cuidados; a mis hermanas Marisol, por estar siempre a mi lado brindándome su apoyo así el horizonte se tornase gris, y Flor Helena, por darme el regalo más hermoso en la vida, mi sobrino. A Daniel Felipe, mi amado sobrino, por mostrarme a su corta edad, que las batallas que trae la vida se pueden superar, y que cada día debemos recibirlo con una gran sonrisa. A William, cuñado gracias por estar ahí apoyando mi familia.

Jhon Herrera Cubides.

Tabla de Contenido

PARTE I FORMULACIÓN DE LA INVESTIGACIÓN 1
CAPÍTULO I DESCRIPCIÓN DE LA INVESTIGACIÓN 1
1.1. Preámbulo. 1
1.2. Descripción del Problema. 2
1.2.1. Formulación del Problema. 4
1.2.2. Pregunta de Investigación. 4
1.3. Justificación. 5
1.4. Objetivos. 7
1.4.1. Objetivo General. 7
1.4.2. Objetivos Específicos. 7
1.5. Hipótesis. 7
PARTE II MARCO DE REFERENCIA Y ESTADO DEL ARTE 9
CAPÍTULO II MARCO TEÓRICO 9
2.1. Recursos Educativos Abierto (REA). 9
2.1.1. Tipos de REA. 9
2.2. Metadatos. 10
2.2.1. Enfoque para describir y recuperar REA. 11
2.2.2. Estándares de Metadatos. 12
2.2.3. Vocabularios Controlados. 14
2.3. Datos Abiertos (Open Data). 15
2.4. Web Semántica. 16
2.4.1. Modelo de Datos RDF. 17
2.4.2. Ontologías. 17
2.4.3. Datos Enlazados (Linked Data). 20
2.4.4. Datos abiertos enlazados (Linked Open Data – LOD). 21
2.4.5. Ciclo de vida de los datos enlazados. 22
2.4.6. SPARQL. 22
2.4.7. Dimensiones de calidad de datos enlazados. 22
2.5. Arquitectura Dirigida por Modelos (Model Driven Architecture - MDA). 27
CAPÍTULO III ANTECEDENTES DE LA APERTURA Y ESTADO DE LA CUESTIÓN .. 30
3.1. Antecedentes de apertura y enlace de datos. 30
3.1.1. Fuentes de Datos Abiertas. 30
3.1.2. Disponibilidad y Reutilización. 31
3.2. Estado de la Cuestión. 33
3.2.1. Plataformas para la Gestión de Datos. 33
3.2.2. Catálogos de Datos Abiertos. 34
3.2.3. Implementaciones basadas en datos enlazados. 37
3.2.4. Estudios Previos sobre Problemáticas de la Web de los Datos. 37
3.2.5. Retos de la Web de los Datos. 40
3.2.6. Estado del arte de datos abiertos en Colombia. 41
3.2.7. Estado del Arte de datos enlazados en el Sector Educativo. 42
3.2.8. Estado del Arte de LOD en el Sector Educativo Colombiano. 43
3.3. Conclusiones del Capítulo. 45
PARTE III DESARROLLO DE LA PROPUESTA 48
CAPÍTULO IV DISEÑO METODOLÓGICO 48

CAPÍTULO V DISEÑO DEL METAMODELO PARA LA DESCRIPCIÓN Y ENRIQUECIMIENTO DE LOS REA. 51
5.1. Construcción teórica del Metamodelo. 51
5.2. Nivel M2 - El Metamodelo: Condiciones Iniciales. 52
5.2.1. Propiedades que describen un REA. 53
5.2.2. Estándares de metadatos para REA. 54
5.2.3. Descriptores de recursos educativos. 56
5.2.4. Metadatos del proyecto RIUD – Universidad Distrital (Caso de Estudio). 57
5.2.5. Encuentros y entrevistas con expertos. 58
5.2.6. Definición preliminar de metadatos para recursos educativos. 59
5.2.7. Dimensiones de calidad aplicables al proceso de abstracción de datos. 60
5.2.8. Restricciones del mapeo de datos. 63
5.3. Nivel M1: Perfil de Aplicación – Recursos Educativos Abiertos (PA-REA). 66
5.3.1. Taxonomía del perfil de aplicación PA-REA. 66
5.3.2. Ontología para describir un REA. 68
5.3.3. Modelado de componentes del perfil de aplicación PA-REA. 69
5.3.4. Aplicación del modelo LOD en PA-REA. 71
5.4. Nivel M1: Mapeo del perfil de aplicación PA-REA. 75
5.4.1. Mapeo PA-REA para aplicación. 75
5.4.2. Mapeo PA-REA para Dataset. 76
5.5. Nivel M1: Formalización del Metamodelo. 80
5.5.1. Formalización del mapeo PA—REA para aplicación. 80
5.5.2. Formalización del Mapeo PA-REA para Dataset. 82
5.6. Conclusiones del Capítulo. 83
CAPÍTULO VI IMPLEMENTACIÓN DEL MODELO DE DATOS. 86
6.1. Despliegue del ambiente tecnológico. 86
6.1.1. Plataformas seleccionadas. 86
6.1.2. Despliegue de las plataformas seleccionadas. 89
6.2. Despliegue del Modelo de Datos. 89
6.2.1. Parametrización de la ingesta y almacenamiento. 90
6.2.2. Parametrización del mapeo RDF de REA. 91
6.2.3. Parametrización de la visualización y consulta. 93
6.3. Conclusiones del Capítulo. 93
CAPÍTULO VII VALIDACIÓN DEL METAMODELO. 95
7.1. Prueba de concepto del metamodelo diseñado. 96
7.1.1. Validación del mapeo para aplicación. 96
7.1.2. Validación del mapeo para dataset. 97
7.1.3. Validación de la gestión de propiedades para los mapeos de datos. 99
7.2. Validación del Despliegue del Modelo. 101
7.2.1. Pruebas de despliegue de interfaces. 103
7.2.2. Prueba de Ingesta de Datos. 104
7.2.3. Prueba de almacenamiento de recursos digitales. 104
7.2.4. Prueba de identificación de recursos. 106
7.2.5. Prueba de serialización de metadatos. 106
7.2.6. Prueba de generación de RDF Graph. 107
7.2.7. Prueba del enlace de datos. 109
7.2.8. Prueba de consultas. 111
7.2.9. Validación de grafos. 112

7.2.10. Requisitos de conexión a la nube LOD. ... 114
7.3. Validación del cubrimiento de REA. ... 115
7.4. Validación Empírica del Modelo Diseñado. ... 121
7.4.1. Enfoque de validación para evaluar el modelo de datos: Modelo GQM. ... 121
7.4.2. Definición del Objetivo. ... 122
7.4.3. Formulación de preguntas. ... 122
7.4.4. Definición de métricas. ... 124
7.4.5. Definición del componente muestral. ... 126
7.4.6. Validación del Modelo de Datos. ... 130
7.5. Prueba de Aceptación de Usuario. ... 134
7.5.1. Resultados del grupo poblacional de expertos. ... 134
7.5.2. Resultados del grupo poblacional de estudiantes. ... 136
7.6. Conclusiones del Capítulo. ... 138
CAPÍTULO VIII APORTES Y DISCUSIONES DEL METAMODELO DISEÑADO. ... 140
8.1. Aportes del Metamodelo Diseñado ... 140
8.1.1. Aportes desde la promesa de MDA. ... 140
8.1.2. Aporte de PA-REA a los problemas de calidad de datos. ... 141
8.1.3. Aporte de PA-REA a la Disponibilidad y Reutilización de los REA. ... 142
8.1.4. Interoperabilidad del modelo de datos con otros proyectos. ... 143
8.2. Otras aportes del metamodelo diseñado. ... 145
8.2.1. Aplicación del modelo de datos en la Universidad Distrital. ... 145
8.2.2. Aplicación en otras instituciones. ... 146
8.2.3. Ampliación del metamodelo a otros tipos de recursos. ... 146
8.3. Discusión Planteada ... 148
8.4. Conclusiones del Capítulo. ... 153
PARTE IV CONCLUSIONES E INVESTIGACIÓN FUTURA ... 155
CAPÍTULO IX CONCLUSIONES E INVESTIGACIÓN FUTURA. ... 155
9.1. Resumen de Objetivos y Resultados. ... 155
9.2. Conclusiones. ... 155
9.2.1. Asociadas a la descripción, apertura y enriquecimiento de los REA. ... 156
9.2.2. Asociadas a las dimensiones de calidad de datos. ... 157
9.2.3. Asociadas al metamodelo como estrategia de abstracción y modelado. ... 158
9.2.4. Asociadas a la interoperabilidad con otros modelos. ... 159
9.2.5. Asociadas al Modelo de Representación de REA. ... 160
9.2.6. Asociadas a los mapeos generados. ... 160
9.2.7. Asociadas a la disponibilidad y reutilización de los REA. ... 161
9.3. Recomendaciones. ... 162
9.3.1. Asociadas a la apertura de los REAs. ... 162
9.3.2. Asociadas a la descripción de propiedades de los REA. ... 163
9.3.3. Asociadas a los estándares de metadatos. ... 163
9.3.4. Asociadas a la calidad de datos. ... 163
9.3.5. Asociadas al enriquecimiento semántico de los metadatos. ... 164
9.3.6. Asociadas al uso de metamodelos. ... 164
9.4. Aportaciones Originales. ... 165
9.5. Trabajos Futuros. ... 165
9.6. Publicaciones derivadas de la investigación ... 166
REFERENCIAS BIBLIOGRÁFICAS ... 171

Índice de Tablas

TABLA 1. DOMINIOS DE LOD CLOUD. FUENTE: MCCRAE (2020). 5
TABLA 2. TIPO DE RECURSOS EDUCATIVOS DESDE LO EDUCATIVO. FUENTE: MEN (2012). 9
TABLA 3. TIPO DE REA SEGÚN SU FORMATO DIGITAL. FUENTE: MEN (2012). 10
TABLA 4. EJEMPLOS DE ESTÁNDARES DE METADATOS. FUENTE: RAMESH *ET AL.* (2015), EDMATRIX (2020), BODAS *ET AL.* (2004) Y BERLANGA *ET AL.* (2005). 12
TABLA 5. EJEMPLOS DE INICIATIVAS / ESTÁNDARES DE METADATOS. 13
TABLA 6. PERFILES DE APLICACIÓN DE METADATOS. 14
TABLA 7. NIVELES DE ENLACE DE DATOS. FUENTE: OPENDATA (2015). 20
TABLA 8. MÉTRICAS DE CALIDAD DE DATOS - CONTEXTUAL. FUENTE: ZAVERI *ET AL.* (2012). 24
TABLA 9. MÉTRICAS DE CALIDAD DE DATOS - CONFIANZA. FUENTE: ZAVERI *ET AL.* (2012). 24
TABLA 10. MÉTRICAS DE CALIDAD DE DATOS - ACCESIBILIDAD. FUENTE: ZAVERI *ET AL.* (2012). 26
TABLA 11. CATÁLOGOS DE DATOS ACADÉMICOS. 34
TABLA 12. CATÁLOGOS DE DATOS INSTITUCIONALES. 35
TABLA 13. CATÁLOGOS DE RECURSOS EDUCATIVOS. 35
TABLA 14. PLATAFORMAS USADAS PARA LA FORMACIÓN. 36
TABLA 15. ANÁLISIS DE ESQUEMA DE VINCULACIÓN. FUENTE: ABELLA *ET AL.* (2018). 38
TABLA 16. INICIATIVAS COLOMBIANAS PARA COMPARTIR DATOS. 42
TABLA 17. OPORTUNIDADES Y BARRERAS DE LA APERTURA DE DATOS EN REA. FUENTE: MINTIC (2015). 43
TABLA 18. EJEMPLOS DE PERFILES DE APLICACIÓN INTERNACIONALES. 55
TABLA 19. EJEMPLOS DE PERFILES DE APLICACIÓN COLOMBIANOS. 55
TABLA 20. METADATOS DEFINIDOS POR LAR. FUENTE: LAR (2019). 56
TABLA 21. ESQUEMA DE METADATOS USADOS POR RIUD. FUENTE: RIUD (2020). 57
TABLA 22. EJEMPLO DE METADATOS DUBLIN CORE USADOS POR RIUD. FUENTE: RIUD (2016). 57
TABLA 23. METADATOS SELECCIONADOS BASADO EN BIBLIOGRAFÍA. 59
TABLA 24. INFORMACIÓN CONTEXTUAL REQUERIDA POR LOS MODELOS DE DATOS. 61
TABLA 25. INFORMACIÓN DE CONFIANZA REQUERIDA POR LOS MODELOS DE DATOS. 62
TABLA 26. DETALLE DEL DICCIONARIO DE DATOS DE PA-REA. 69
TABLA 27. EJEMPLOS DE DILIGENCIAMIENTO DE METADATOS PA-REA. 72
TABLA 28. CALIFICADORES CON CONTENIDO LITERALES Y URI. 73
TABLA 29. ESQUEMA DE MAPEO PA-REA PARA APLICACIÓN. 75
TABLA 30. ALGUNAS ESPECIFICACIONES DE FUSEKI. FUENTE: BONDUEL (2019). 88
TABLA 31. EJEMPLOS DE DESCRIPCIÓN DE REA -I. 116
TABLA 32. EJEMPLOS DE DESCRIPCIÓN DE REA –II. 117
TABLA 33. EJEMPLOS DE DESCRIPCIÓN DE REA –III. 119
TABLA 34. MÉTRICAS DEFINIDAS PARA LA DIMENSIÓN DE CALIDAD CONTEXTUAL. 124
TABLA 35. MÉTRICAS DEFINIDAS PARA LA DIMENSIÓN DE CALIDAD DE CONFIANZA. 125
TABLA 36. INSTANCIAS DE DATOS CONSULTADAS. 127
TABLA 37. CANTIDAD DE RECURSOS Y PROPIEDADES DE LAS INSTANCIAS CONSULTADAS. 128
TABLA 38. INDICADORES DE INSTANCIAS CONSULTADAS. 129
TABLA 39. OBSERVACIÓN DE MÉTRICAS – INSTANCIAS 1 – 11. 130
TABLA 40. OBSERVACIÓN DE MÉTRICAS – INSTANCIAS 12 – 22. 131
TABLA 41. NIVEL DE OBLIGATORIEDAD DE ELEMENTOS DE METADATOS SEGÚN EXPERTOS. 134
TABLA 42. GRADO DE INTERÉS DE ELEMENTOS DE METADATOS SEGÚN ESTUDIANTES. 136
TABLA 43. APORTE DEL METAMODELO A LAS DIMENSIONES DE CALIDAD. 141
TABLA 44. INTEROPERABILIDAD DE METAMODELO DISEÑADO. 143
TABLA 45. RESUMEN DE OBJETIVOS Y RESULTADOS. 155
TABLA 46. CATEGORIZACIÓN DE LOS APORTES DE LA INVESTIGACIÓN. 169

Índice de Figuras

FIGURA 1. CLOUD DE LINKED OPEN DATA. FUENTE: MCCRAE (2020).....5
FIGURA 2. COHERENCIA DEL OBJETIVO, PREGUNTA E HIPÓTESIS DE INVESTIGACIÓN. 8
FIGURA 3. TRIPLETA RDF. FUENTE: NILSSON *ET AL*. (2008). 17
FIGURA 4. ARQUITECTURA LINKED DATA. FUENTE: ELABORACIÓN PROPIA BASADO EN PINKEL (2013).....21
FIGURA 5. DIMENSIONES DE CALIDAD EN DATOS ENLAZADOS. FUENTE: ZAVERI *ET AL.* (2012). 23
FIGURA 6. NODO KUJANE – PROYECTO LA REFERENCIA. FUENTE: KUJANE (2020). 44
FIGURA 7. PORTAL APRENDER DIGITAL. FUENTE: MEN (2020). 45
FIGURA 8. PROBLEMÁTICAS DE DATOS ABIERTOS Y ENLAZADOS. 46
FIGURA 9. METODOLOGÍA PARA EL METAMODELO DE DESCRIPCIÓN Y ENRIQUECIMIENTO DE REA.....48
FIGURA 10. CONTEXTOS QUE DEFINIERON EL METAMODELO. 51
FIGURA 11. CONTEXTOS QUE DEFINIERON EL METAMODELO. 52
FIGURA 12. PROPIEDADES BÁSICAS DE UN REA. FUENTE: BUTCHER *ET AL.* (2015) Y UNESCO (2015).....54
FIGURA 13. GRAFO PARA UNA TRIPLETA N-ARIA. FUENTE: RODRÍGUEZ (2009). 65
FIGURA 14. NIVEL M2 DEL METAMODELO DISEÑADO.....65
FIGURA 15. NIVEL M1 DEL METAMODELO: PA-REA. 66
FIGURA 16. TAXONOMÍA PA-REA. 67
FIGURA 17. PERFIL DE APLICACIÓN PARA REA. 68
FIGURA 18. ELEMENTOS Y CALIFICADORES PA-REA.....69
FIGURA 19. MAPEOS DEL PERFIL DE APLICACIÓN PA-REA. 75
FIGURA 20. MAPEO PA-REA PARA DATASET. 78
FIGURA 21. NIVEL M2 – MODELO DE METADATOS. 79
FIGURA 22. MAPEOS DEL PERFIL DE APLICACIÓN PA-REA. 80
FIGURA 23. NIVEL M0: INSTANCIA DE DATOS PARA APLICACIÓN. 81
FIGURA 24. ESQUEMA DE MAPEO DC PARA APLICACIÓN. 81
FIGURA 25. NIVEL M0 – INSTANCIAS DE DATOS PARA DATASET.....82
FIGURA 26. ESQUEMA DE MAPEO RDF PARA PA-REA. 83
FIGURA 27. METAMODELO PARA LA DESCRIPCIÓN Y ENRIQUECIMIENTO DE REA. 84
FIGURA 28. NIVEL M0 PARA DESPLIEGUE. 86
FIGURA 29. PLATAFORMAS DE GESTIÓN DE RECURSOS 2020. FUENTE: JIS (2020). 87
FIGURA 30. ARQUITECTURA DSPACE. FUENTE: DONOHUE (2018). 87
FIGURA 31. DESPLIEGUE DEL AMBIENTE TECNOLÓGICO. 89
FIGURA 32. MODIFICACIÓN DEL METADATA REGISTRY. 90
FIGURA 33. EJEMPLO DE VOCABULARIO CONTROLADO. 90
FIGURA 34. EJEMPLO DE INTERFAZ DE INGESTA DSPACE. 91
FIGURA 35. EJEMPLO DE CONSTRUCCIÓN DE CONSULTAS. 93
FIGURA 36. PROCESOS DE MAPEOS PA-REA PARA APLICACIÓN Y PARA DATASET.....94
FIGURA 37. ESTADIOS DE VALIDACIÓN DEL METAMODELO PROPUESTO. 95
FIGURA 38. EJEMPLO DE ESQUEMA DE MAPEO PA-REA PARA APLICACIÓN. 97
FIGURA 39. EJEMPLO DEL ESQUEMA DE MAPEO RDF PARA PA-REA. 98
FIGURA 40. FORMALIZACIÓN DEL METAMODELO AL MODELO DE DATOS. 99
FIGURA 41. INSERCIÓN DE UNA PROPIEDAD AL MAPEO PA-REA PARA APLICACIÓN. 100
FIGURA 42. INSERCIÓN DE UNA PROPIEDAD AL MAPEO PA-REA PARA DATASET. 101
FIGURA 43. PRUEBAS DEL DESPLIEGUE DEL MODELO DE DATOS. 103
FIGURA 44. DESPLIEGUE DE INTERFAZ DE INGESTA DE METADATOS. 104
FIGURA 45. PRUEBA DE INGESTA DE METADATOS.....104
FIGURA 46. PRUEBA DE ALMACENAMIENTO DE METADATOS. 105
FIGURA 47. DESPLIEGUE DEL METADATA REGISTRY DE UN RECURSO. 105
FIGURA 48. METADATOS ALMACENADOS EN FORMATO CSV**.** 106
FIGURA 49. IDENTIFICACIÓN DE LA DESCRIPCIÓN DEL REA. 106
FIGURA 50. DESPLIEGUE DE LA RDFIZACION DE UN RECURSO. 107
FIGURA 51. RDF GRAPH CARGADO EN FUSEKI. 107

FIGURA 52. RESULTADOS DEL VALIDADOR RDF TURTLE. 109
FIGURA 53. ENLACE DE DATOS DESDE LA PLATAFORMA DE GESTIÓN DE RECURSOS. 109
FIGURA 54. RESULTADOS DE MAPEO. 110
FIGURA 55. CONSULTA REALIZADA EN DSPACE. 111
FIGURA 56. CONSULTA REALIZADA EN FUSEKI. 112
FIGURA 57. EJEMPLO DE RDF GRAPH TURTLE EXPORTADO. 112
FIGURA 58. VALIDACIÓN DEL RDF GRAPH TURTLE 113
FIGURA 59. TRIPLETAS RDF DEL GRAFO EXPORTADO. 113
FIGURA 60. VISUALIZACIÓN DEL GRAFO RDF. 114
FIGURA 61. DATOS REQUERIDOS POR LOD CLOUD. 115
FIGURA 62. CUBRIMIENTO DE REAS POR PA-REA. ADAPTADO DE MEN (2012). 120
FIGURA 63. APLICACIÓN DE LOS NIVELES GQM. 122
FIGURA 64. RESULTADOS GENERALES DE APLICACIÓN DE MÉTRICAS. 130
FIGURA 65. RESULTADOS MÉTRICA 1 132
FIGURA 66. RESULTADOS MÉTRICA 2. 132
FIGURA 67. COMPARACIÓN PROPIEDADES PROCOMÚN Y MERLOT. 133
FIGURA 68. RESUMEN DE APORTES DEL METAMODELO. 140
FIGURA 69. OTROS APORTES DEL METAMODELO. 145
FIGURA 70. ADICIÓN DE LA CAPA M3 – META-METAMODELO. 147
FIGURA 71. EJEMPLO DE LA CAPA M3 – META-METAMODELO. 148
FIGURA 72. APORTES DESDE LAS MÉTRICAS DE CALIDAD. 149
FIGURA 73. METADATOS PROYECTO LA REFERENCIA. FUENTE: LAR (2019). 152
FIGURA 74. ALINEACIÓN DE LA HIPÓTESIS CON LOS RESULTADOS OBTENIDOS. 153
FIGURA 75. PRODUCTOS GENERADOS EN LA INVESTIGACIÓN. 167
FIGURA 76. APORTE DE LOS PRODUCTOS A LOS OBJETIVOS. 168
FIGURA 77. APORTE DESDE LOS PROYECTOS DE INVESTIGACIÓN. 169

Índice de Ecuaciones

ECUACIÓN 1. FÓRMULA PARA OBTENER TAMAÑO POBLACIONAL.....126
ECUACIÓN 2. CÁLCULO DE TAMAÑO POBLACIONAL.....126

Lista de símbolos, abreviaturas y otros

REA	Recurso Educativo Abierto.
LOD	Linked Open Data.
RDF	Resource Description Framework.
XML	eXtensible Markup Language.
SPARQL	SPARQL Protocol and RDF Query Language.
URI	Uniform Resource Identifier.
Skos	Simple Knowledge Organization System.
OWL	Web Ontology Language.
URL	Uniform Resource Locator.
HTTP	Hypertext Transfer Protocol.
MDA	Model-Driven Architecture.
OMG	Object Management Group.
W3C	World Wide Web Consortium.
DSL	Domain Specific Languages.
CSV	Comma-separated values.
API	Application Programming Interface.
MEN	Ministerio de Educación Nacional.
SENA	Servicio Nacional de Aprendizaje.
RIUD	Repositorio Institucional de la Universidad Distrital.
MIT	Massachusetts Institute of Technology.

PARTE I FORMULACIÓN DE LA INVESTIGACIÓN

CAPÍTULO I DESCRIPCIÓN DE LA INVESTIGACIÓN

1.1. Preámbulo.

Con la evolución de la Web de los documentos a la Web de los datos, e intrínsecamente el desarrollo y la estandarización de tecnologías como la Web Semántica apoyada en principios de "Open Access", se han venido presentando propuestas de exposición, enlace y consumo de datos abiertos (Linked Open Data - LOD). Esto ha permitido el surgimiento de iniciativas orientadas al intercambio de información de interés a partir de áreas de conocimiento. Iniciativas como DBpedia han motivado el interés de organizaciones y entidades involucradas en la gestión de grandes volúmenes de datos para compartirlos mediante especificaciones y estándares, a través del uso de conjuntos de datos – datasets. Estos datasets, expuestos a través de un punto de acceso SPARQL, contienen enlaces a recursos enlazados, formando así grafos de conocimiento basados en metadatos y gestionados por las diferentes organizaciones.

Sin embargo, el proceso de enlace y explotación de datos se ha venido enfrentando a diferentes problemáticas que han menguado el interés hacia el uso de la Web de los datos. A pesar de la gran cantidad de recursos digitales publicados en la Web y de la existencia de diferentes esfuerzos de vinculación de los mismos, algunos de estos esfuerzos no logran superar el primero de los cinco niveles de vinculación propuestos por Berners-Lee (2011); lo cual hace casi imposible la disponibilidad los recursos (niveles 1 a 3 de LOD) y la reutilización de los mismos para asociarlos a un área de conocimiento (niveles 4 y 5 de LOD).

Con relación a la disponibilidad de los datos, algunas veces presenta deficiencias sobre la calidad de los datos recuperados en las consultas sobre los puntos de acceso SPARQL. Dificultades en la integración de fuentes de datos, interoperabilidad sintáctica, esquemática o semántica, vista ésta última desde la heterogeneidad cognitiva y designativa; entre otras dificultades, gestan problemas como las inconsistencia, falta de completitud, accesibilidad a los datos, entre otros (Abella *et al.*, 2018; Smith-Yoshimura, 2018; LOD2, 2019; Klímek *et al.*, 2018; Herrera-Cubides *et al.*, 2017). Estos aspectos generan grandes desafíos para preservar la calidad de los datos en diferentes dimensiones como por ejemplo: contextuales, de accesibilidad, y de confianza.

Por otro lado, con el crecimiento de estrategias como la virtualidad, aplicada ésta como apoyo a la presencialidad o dispuesta en toda su extensión, diferentes organizaciones han generado recursos digitales educativos que han sido publicados en la Web haciendo uso de repositorios propietarios (Bozkurt *et al.* 2020; OERWM, 2020; OECD, 2007; COE, 2019). Estos recursos digitales educativos se publican como unidades individuales, con un esquema de metadatos acorde a las necesidades o conocimiento de los autores, y sin una estrategia que permita enriquecer la información que describe dicho recurso digital educativo. Con la incursión de LOD, la vinculación de recursos en el dominio de conocimiento educativo se ha proyectado como un campo poco explorado en este tipo de iniciativas, observándose en su mayoría el diseño e implementación de procesos de recolección (*harvesting*) sobre los recursos digitales expuestos de manera distribuida.

Ahora bien, dado que un gran grupo de repositorios presentan situaciones como: licenciamientos restrictivos sobre sus recursos digitales educativos, problemas de actualización, multiplicidad de formatos, entre otros aspectos; al momento de explotar estos recursos se puede incursionar en problemas de disponibilidad tales como contravención a situaciones legales, baja oportunidad de

los datos, etc.; o en su defecto, incursionar en dificultades para el reutilización de los datos que se quieren enlazar, dado que no están expuesto en formatos adecuados como el Marco de Descripción de Recursos - RDF, restringiendo la posibilidad de llevar a cabo esta labor.

A partir de este panorama, la presente investigación se enfoca en presentar un metamodelo para mejorar la descripción, apertura, disponibilidad y enriquecimiento de recursos educativos abiertos, basado en principios de calidad de datos enlazados, procurando así una mejora en los procesos de exposición y enlace de material educativo digital sobre repositorios abiertos, para su posterior reutilización y apoyo en procesos de aprendizaje.

1.2. Descripción del Problema.

Con la masificación y especialización de los servicios provistos sobre Internet, la cantidad de información y recursos disponibles en la Web ha crecido de forma exponencial. Ante una búsqueda simple, se encuentra información de todo tipo, origen y calidad, dado que cualquier persona puede publicar lo que quiera, sin ningún tipo de restricción. Un ejemplo de esta proliferación de recursos se observa en los procesos de formación apoyados por mediaciones tecnológicas digitales, usadas éstas bien sea como apoyo a la presencialidad o a la formación virtual, en donde según The National Center for Education Statistics para el 2018 aproximadamente 7 millones de estudiantes han usado o están usando al menos un curso en línea (NCES, 2019), cursos que son apoyados por diferentes tipos de recursos educativos (Bozkurt *et al.*, 2020; OERWM, 2020; OECD, 2007; COE, 2019). Según OBS (2019), se calculó que el mercado del e-Learning experimentaría un crecimiento anual entre el 7.6% y el 9.6% para el 2020. Incrementos que se han visto superados en lo transcurrido del 2020, dado el giro hacia la educación virtual generado como estrategia para responder a la pandemia de Covid-19 (Newfield, 2020).

La proliferación de recursos educativos, algunos de mejor calidad que otros, se presenta por la falta de “apertura”, disponibilidad y reutilización de los mismos, dado que éstos son diseñados y distribuidos de manera propietaria, generando entre otros aspectos duplicidad de esfuerzos en su producción, contenidos repetitivos y de calidad variante, etc. Para abordar la falta de apertura, por ejemplo, se han gestado diferentes estrategias para publicar recursos que residen en el dominio público o han sido liberados bajo una licencia de propiedad intelectual que permite su uso libre y re-uso por otros (Unesco, 2019). Proyectos tales como Europeana (Europeana Foundation, 2020), Merlot (CSU-CDL, 2020), Miríadax (Telefónica, 2020), Red CLARA (CLARA, 2020), OpenLearn (The Open University, 2020), OpenCourseWare (OpenCourseWare Consortium, 2020), Red Nacional Académica de Tecnología Avanzada (RENATA, 2020), entre otras, han logrado compartir recursos informativos y académicos, para procesos de formación y/o consulta. Estos proyectos llevaron a que se plantearan iniciativas en el contexto colombiano tales como:

- Portal Educativo Colombia Aprende (2004), cuyo propósito consistía en converger a la comunidad académica en torno a una oferta de contenidos, herramientas y servicios de oportuna respuesta a los usuarios (MEN, 2012; MEN, 2009; MEN, 2007).
- En el 2006, se plantea la necesidad de establecer un método de clasificación, que facilite el acceso a los recursos educativos. Esta actividad de catalogación es el inicio formal de la estrategia del Banco Nacional de Objetos de Aprendizaje, por parte del Ministerio de Educación Nacional de Colombia (MEN, 2012; MEN, 2009; MEN, 2007).
- En el 2012, se presenta la Estrategia Nacional de Recursos Educativos Digitales Abiertos (REDA) con la que estudiantes, docentes, investigadores y la comunidad educativa en general, podrán acceder a recursos educativos de manera libre y gratuita.

Sin embargo, muchas de estas iniciativas han ido en declive o han fracasado por factores como la falta de políticas claras de apertura, situaciones contractuales referidas a la producción de recursos y sus derechos, el temor a la pérdida de derechos e inversión realizada, entre otros factores identificados por investigaciones como las realizadas por Atenas (2014), Vila-Viñas *et al.* (2015), Miao *et al.* (2019) y Ocampo *et al.* (2015).

Para el 2020, como consecuencia de medidas tomadas para contrarrestar el del Covid-19, en casos como Colombia, se ha buscado fortalecer portales como Aprender Digital (MEN, 2020) a través de donaciones de recursos, fortaleciéndose problemas como la descripción limitada de los recursos, la falta de modelos de datos para describir adecuadamente los recursos, problemas de calidad de los datos descritos, etc.

La tendencia de la apertura de los recursos educativos, término adoptado por primera vez en el Foro 2002 de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura –UNESCO- sobre el Impacto del Material Didáctico Abierto para la educación superior en los países en desarrollo (Miao *et al.*, 2016); vio en la Web Semántica una posibilidad de desarrollo, entendida ésta como una estrategia que usa conceptos ontológicos y vocabularios para describir con precisión los contenidos de una manera legible por una máquina, y que pueden ser compartidos y recuperados en la Web (Qing Yu *et al.*, 2012). El uso de tecnologías de Web Semántica, tales como Linked Data (PORT, 2019; W3C, 2015; Bizer, 2009), buscaría enriquecer los metadatos que describe los recursos educativos abiertos, a través del enriquecimiento de sus propiedades. Con el paso de los años, la aplicación de esta iniciativa ha perdido el interés que suscitó en sus inicios, lo cual se ve reflejado en aspectos como (Martínez *et al.*, 2020; Ávila, 2020; Senso y Arroyo, 2018):

- A pesar de la existencia de mejores prácticas en Linked Data, el enlace de datos presenta interpretaciones diferentes del dominio de conocimiento generando variaciones en la abstracción de conceptos, características, e identificación de relaciones que estos conceptos tendrán entre sí (Vila-Viñas *et al.*, 2015, p. 128; Hallo *et al.*, 2012, p. 57; Sánchez, 2014, p. 4). Variaciones también pueden deberse a la falta de conocimiento del dominio a modelar, la falta de experiencia o curva de aprendizaje de las personas responsables del proceso, abstracciones particulares u acomodadas a los intereses individuales (Herrera-Cubides *et al.*, 2018b; Narváez y Piedra, 2018).
- Los modelos de datos presentan problemas de calidad en diferentes dimensiones, como por ejemplo la falta de completitud en la definición de propiedades relacionadas con los recursos que se exponen; la exactitud en dichas propiedades, la inconsistencia de la información que en estas características se almacena, entre otros (Herrera-Cubides *et al.*, 2018b; Rayón, 2016; Blaney, 2018). Como lo plantea Miller (2016), los datasets vienen a menudo sin nombre, sin descripción, sin propietario e inéditos.
- Como lo plantea Tim Berners-Lee (2011), en el enlace de recursos se debe permitir a los usuarios acceder a los datos almacenados en fuentes de datos heterogéneas, por ende, la interoperabilidad, entendida como la habilidad para interoperar o integrar diferentes datasets, se convierte en un factor fundamental para asegurar la disponibilidad, acceso, reutilización, redistribución y participación universal en los datos. Pero dado que la información siempre será heterogénea, los modelos de datos usados en cada dataset pueden ser distintos, pero si el formato de datos es común (Resource Description Framework- RDF + Uniform Resource Identifier - URIs desreferenciables + acceso a través de puntos SPARQL), la interoperabilidad es posible (Graves, 2012; Hallo *et al.*, 2012; W3C, 2015; Port, 2019; W3C, 2017; Herrera-Cubides *et al.*, 2018b). El problema suscitado en este contexto no se presenta en la vinculación de recursos de forma adecuada (pues de hecho bien usado, la tecnología permite llevar a cabo dicha vinculación), sino en el déficit en la publicación de los modelos de datos, lo cual dificulta su utilización por parte de terceros.

- En el contexto de la Web, donde cualquier persona puede publicar cualquier cosa sobre cualquier tema (Willer y Dunsire, 2013), la calidad de los datos es un asunto crucial, ya que proporciona la base para que los usuarios decidan si los datos a usar son seguros y confiables. A menudo, los usuarios deciden confiar o no en datos, basados en el valor de dimensiones de calidad de metadatos, como exactitud, precisión y otros. Debido a los problemas de calidad presentes en los datos instanciados, la generación de confianza en estos datos se torna una tarea demasiado compleja. Dado que la Web Semántica es grande, heterogénea, dinámica e incierta, la confianza será inevitablemente un problema (O'Hara *et al.*, 2004; Martínez *et al.*, 2020).
- Otro factor crítico responde a la disponibilidad y facilidad de uso de ambientes tecnológicos que sustenten la explotación de los recursos educativos enriquecidos. Factores como la curva de aprendizaje de las tecnologías, la disponibilidad de componentes como los puntos de acceso SPARQL, entre otros referentes, hacen que la labor de enlazar y enriquecer los metadatos de los Recursos Educativos Abiertos (REA) se torne una tarea cada vez más compleja.

1.2.1. Formulación del Problema.

En el contexto del enriquecimiento de datos a través del uso de Linked Data, se identifican investigaciones como las realizadas por McCrae (2020); Hogan *et al.* (2012); Herrera-Cubides *et al.* (2016); Unión Europea (2020); GODI (2019) o COL (2015), entre otras, donde se analizan problemas de enlace de datos enfocándose al manejo y curación de las instancias de datos, del mejoramiento de la calidad en su explotación, del diseño de herramientas que permitan explotar los datos y evaluar su calidad, entre otros enfoques.

Ahora bien, el trabajo en el campo de la Web Semántica se ha orientado a dominios de conocimiento con componentes muy fuertes en estandarización y parametrización, sesgando a dominios como el de los recursos educativos, al tratamiento generalizado como recursos bibliográficos.

De esta manera, se hace necesario el planteamiento de un metamodelo de datos que permita:

- Mejorar la descripción de los REA,
- Adicionar elementos de calidad de datos a la descripción de los datos,
- Contribuir a la apertura de los modelos de datos,
- Aportar en el mejoramiento de la disponibilidad, enriquecimiento y reutilización de los recursos digitales educativos, para generar confianza en los usuarios y aplicaciones que consumen estos recursos.

1.2.2. Pregunta de Investigación.

Considerando el marco ofrecido por los problemas identificados se plantea la siguiente pregunta de investigación:

¿Cómo establecer un metamodelo para la descripción y enriquecimiento de recursos educativos abiertos en Colombia, mediante especificaciones Linked Open Data, y basado en las dimensiones de calidad contextuales, de confiabilidad y accesibilidad de datos vinculados, que contribuya en mejorar la disponibilidad y reutilización de los recursos educativos abiertos, generando confianza para los usuarios y aplicaciones consumidoras de dichos recursos?.

1.3. Justificación.

El planteamiento de un metamodelo para el enlace de REA en Colombia, mediante especificaciones LOD, y basado en las dimensiones de calidad contextuales, de confiabilidad y accesibilidad de datos enlazados, se fundamenta en dos ejes principales:

a. Eje 1: Problemáticas de los Datos Enlazados:

La incursión de tecnologías semánticas como Linked Data ha incrementado el volumen de datos disponibles para realizar consultas semánticas. Estos datos publicados se encuentran dispuestos en grafos de conocimiento (Figura 1), en los cuales los datasets se ubican en diferentes dominios de conocimiento, como los que se muestran en la Tabla 1. Para el 2020 se cuenta con 1.255 datasets con 16.174 enlaces.

Tabla 1. Dominios de LOD Cloud. Fuente: McCrae (2020).

Dominio	**Subdominios**
Gobierno	Código de País, Datos Estadísticos, Ambiente, Internacional, UK, Republica Checa, Grecia, Otros
Publicaciones	Bibliografía, Otros
Lingüística	Corpora; Lexicons y Diccionarios; terminología, tesauros y bases de conocimiento; Metadatos de recursos lingüísticos; Categorías de datos lingüísticos; Base de datos tipológicas; Otros
Ciencias de la Vida	BioPortal, Bio2RDF, Otros
Contenido Generado por el Usuario	SIOC, Otros
Interdisciplinar	Wikipedia, Otros
Media	Música, Otra Media
Geografía	Geografía
Social Networking	Social Web, Otros

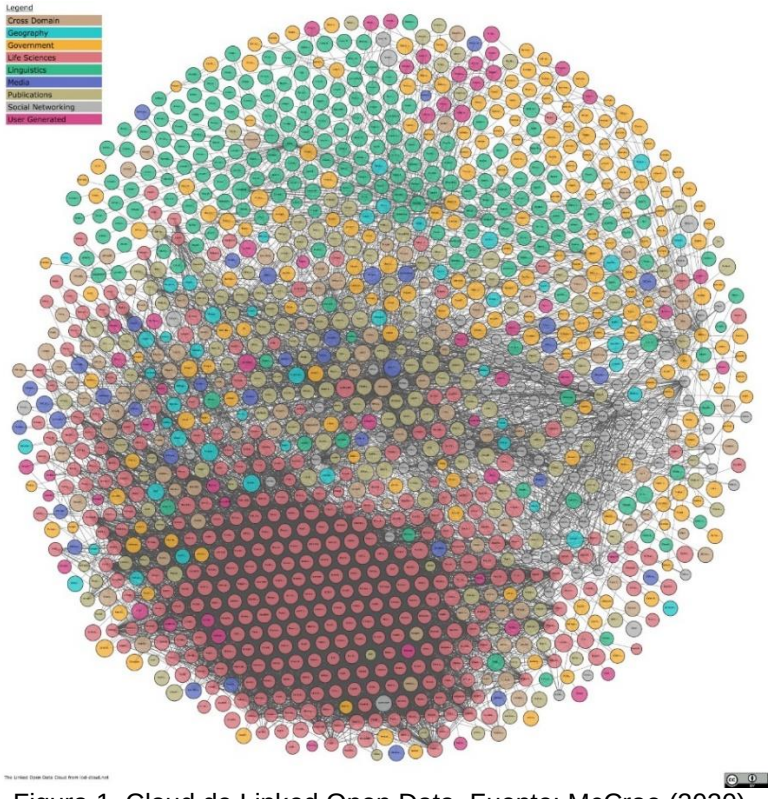


Figura 1. Cloud de Linked Open Data. Fuente: McCrae (2020).

En el marco de la Web de los Documentos existía la posibilidad de que cualquiera podía publicar lo que quisiera en la Web, generando un alto grado de exposición a información con diferentes grados de veracidad. Por su lado, en la Web de los Datos, instituciones acreditadas son quienes publican sus datos, y donde existen entidades que fungen como mediadores para la subida de los datasets solicitando el cumplimiento de ciertas condiciones de calidad, mitigando el problema de discrepancia y la veracidad de la información, pero empiezan a emerger diferentes situaciones que pueden llegar a generar problemas, tales como:

- Una adecuada abstracción del modelo de datos.
- La publicación de los modelos usados.
- El uso adecuado de vocabularios.
- El avanzar en los niveles de enlace de datos.
- Problemas de calidad tanto en la data como en los metadatos.
- El diseño y disposición de interfaces que faciliten al usuario la consulta y visualización de los recursos consultados.

Estos entre otros problemas afectan la calidad de los datos expuestos, su confiabilidad, su construcción semántica, y en general incrementan los costos asociados a la integración y el consumo que, en última instancia, afectarán a los consumidores de datos enlazados y a las aplicaciones que hacen uso de ellos. Por ende, el diseño de un metamodelo que responda a una abstracción claramente definida del dominio de conocimiento abordado, y que atienda a requerimientos de calidad de datos sugeridos, contribuirá en la generación de confianza por parte del usuario en el consumo de recursos enlazados.

b. Eje 2: Disposición de Recursos Educativos Abiertos

En cuanto a los procesos de formación que involucran recursos digitales educativos, como lo describe Rajabi *et al.* (2014), un gran número de estudiantes siguen cursos en línea y utilizan recursos de aprendizaje en sus estudios. Este hecho, junto con el rápido crecimiento en el número de datos sobre la educación en la Web, ha hecho de la exposición y acceso de los recursos de aprendizaje una cuestión importante (Bozkurt *et al.*, 2020; OERWM, 2020). La variedad de recursos de aprendizaje, que son creadas por diferentes editoriales y en diferentes formatos, motiva a las entidades de publicación hacia el intercambio de recursos y el enlace de los mismos, ya que existe un gran número de recursos educativos relacionados con temas de interés y con el contexto técnico de los aprendices.

El enlace de los recursos educativos, almacenados en repositorios en la Web de los Datos, mejora la búsqueda de recursos, independientemente de los límites geográficos y del sistema utilizado. También enriquece y hace más atractivas y útiles las colecciones, tanto para los educadores como para los educandos.

En cuanto al punto de vista técnico, el modelo de datos enlazados como solución para la exposición de datos en la Web, permite que cualquier tipo de datos sean reutilizables y accesibles para los actores involucrados en el e-Learning, ofreciendo y facilitando el enlace inteligente haciendo uso de herramientas de interconexión (interlinking). Con este fin, muchas universidades, instituciones y proyectos han seguido los principios del modelo de datos enlazados y han publicado sus recursos como parte de la nube de datos enlazados, como por ejemplo:

- M³ Matemática Multimedia, único repositorio dedicado a las matemáticas en la región. Cuenta con 350 recursos educativos en video, audio, software y experimentos bajo licencia Creative Commons (Universidad de Campiñas, 2020).

- Portal educativo de la Comunidad y Cultural Centroamericana. Portal formativo que une a ocho países centroamericanos y a sus respectivos sistemas educativos. Este portal ofrece recursos para educadores de la región que incluyen por ejemplo un centro de recursos digitales, comunidades virtuales, cursos en línea y foros, entre otros (CECC, 2020).
- Serendipity: Proyecto que, mediante sistemas de Linked Data y Web Semántica, permite acceder a recursos albergados en cursos pertenecientes al Consorcio de Educación Abierta y a repositorios, cursos en línea masivos y abiertos - MOOCs y cursos en Open Course Ware (OCW) mediante un mapa interactivo (Universidad Técnica Particular de Loja, 2020).

Considerando los diferentes esfuerzos generados y atendiendo a los principios de datos enlazados, se hace evidente la necesidad de proponer un metamodelo para mejorar la disponibilidad y reutilización de los REA para el contexto colombiano, aprovechando las ventajas del enlace de datos a través de las diferentes consultas que se pueden realizar sobre los respectivos puntos de acceso, basado en principios de calidad de datos enlazados, en lo referente a las dimensiones contextuales, accesibilidad y de confiabilidad. El objeto de este metamodelo es mejorar la experiencia de los usuarios consumidores de recursos educativos mejor descritos y enriquecidos, con el propósito de reducir el costo de sus búsquedas e incrementando el uso de estos materiales, generando un mayor nivel de confiabilidad ofrecida a los usuarios.

Es importante resaltar que no se evaluará el contenido del recurso, su contexto didáctico o pedagógico ni la calidad y suficiencia de sus componentes internos, ni tampoco los procesos necesarios para que el usuario consumidor sea validado en las plataformas de exposición.

1.4. Objetivos.

1.4.1. Objetivo General.

Proponer un metamodelo para la vinculación de REA en el contexto de LOD, basado en principios de calidad contextuales, accesibilidad y de confianza de datos vinculados.

1.4.2. Objetivos Específicos.

a. Llevar a cabo un diagnóstico de la Web de los Datos, con el fin de especificar los requerimientos funcionales y conceptuales sobre los procesos de enlace de datos y dimensiones de calidad de datos enlazados.
b. Proponer, diseñar e implementar un metamodelo para la exposición y consumo de REA, basado en principios de calidad de datos enlazados, con el fin de contribuir a la generación de confianza en los consumidores de datos enlazados y a en las aplicaciones que hacen uso de ellos.
c. Diseñar y construir un modelo instrumental que permita verificar las parametrizaciones planteadas, analizar y evaluar los resultados con el fin de establecer el grado de contribución que pueda prestar el modelo propuesto para el enlace de REA.

1.5. Hipótesis.

"*Un metamodelo para la descripción y enriquecimiento de los REA, basado en dimensiones de calidad contextuales, accesibilidad y de confianza, permitirá operar bajo parámetros de calidad conocidos, aportando mejoras en la disponibilidad y reutilización de los recursos educativos abiertos*".

En la Figura 2 se observa la coherencia entre el objetivo general, la pregunta del problema y la hipótesis para el desarrollo de la presente investigación.

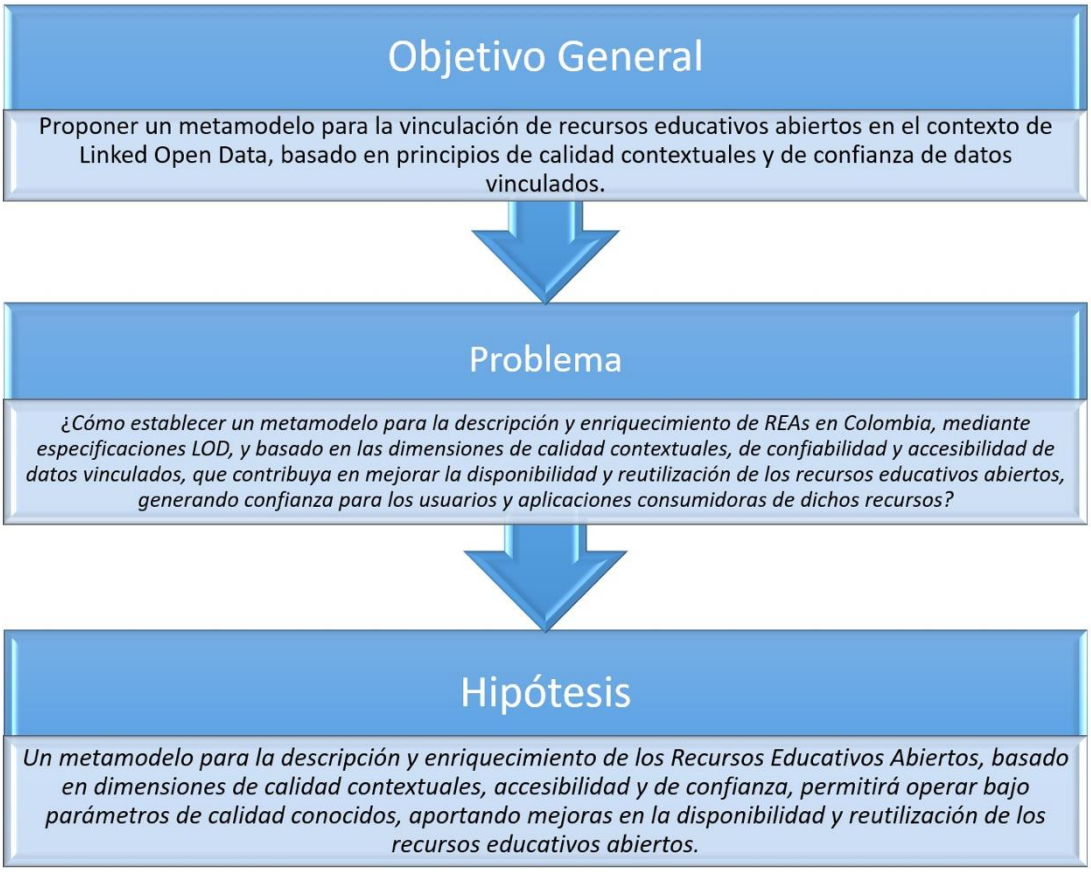


Figura 2. Coherencia del Objetivo, Pregunta e hipótesis de Investigación.

PARTE II MARCO DE REFERENCIA Y ESTADO DEL ARTE

CAPÍTULO II MARCO TEÓRICO

En Linked Open Data - LOD, la gran cantidad de datos disponibles en la Web deben describirse en un formato estándar, accesible y gestionable por herramientas de la Web Semántica; además de poder relacionar con otros datos disponibles, creando de esta forma la Web de los Datos. Para enlazar los datos haciendo uso de herramientas de Web Semántica tales como Linked Data, se debe contar con un modelo de datos que permita describir las entidades y relaciones obtenidas en el proceso de abstracción, para lo cual estándares basados en formatos RDF se convierten en una herramienta valiosa dentro de este proceso. De igual forma, se deben usar vocabularios que provean una expresividad suficiente a la hora de modelar la información requerida. Con el propósito de establecer las bases teóricas y técnicas, en el presente capítulo, se describen los aspectos teóricos que sustenta esta investigación.

2.1. Recursos Educativos Abierto (REA).

En primera instancia, se definen los REA como dominio o área de conocimiento a trabajar en esta investigación. De acuerdo con esta premisa, y considerando lo descrito en MEN (2012), la UNESCO, en su documento "*A Basic Guide To Open Educational Resources*", define los REA como:

> "*Cualquier tipo de recurso (incluyendo planes curriculares, materiales de los cursos, libros de texto, vídeo, aplicaciones multimedia, secuencias de audio, y cualquier otro material que se haya diseñado para su uso en los procesos de enseñanza y aprendizaje) que están plenamente disponibles para ser utilizados por parte de educadores y estudiantes, sin la necesidad de pago alguno por derechos o licencias para su uso*" (UNESCO, 2019).

Para el contexto colombiano y para esta investigación, los REA se definen como "*todo tipo de material que tiene una intencionalidad y finalidad enmarcada en una acción Educativa, cuya información es digital, y se dispone en una infraestructura de red pública, como internet, bajo un licenciamiento de Acceso Abierto que permite y promueve su uso, adaptación, modificación y/o personalización*" (MEN, 2012).

Teniendo claro este eje conductor, y con el fin de empezar a detallar los elementos que soportan la descripción de los REA, es necesario determinar la clasificación que enmarca cada uno de los tipos de REA, los cuales son considerados en este estudio.

2.1.1. Tipos de REA.

Como lo plantea MEN (2012), dada la heterogeneidad, multiplicidad y complejidad de los REA, éstos se clasifican bajo dos enfoques:

a. Desde lo educativo. El propósito de la clasificación de los recursos desde lo educativo se basa en características comunes entre ellos, las cuales pueden definirse desde los objetivos de aprendizaje, intencionalidades de uso, complejidad, estructura, entre otros. La Tabla 2 describe la organización de estos tipos de recursos.

Tabla 2. Tipo de Recursos Educativos desde lo Educativo. Fuente: MEN (2012).

Tipo de Formato	Descripción
Curso Virtual	Experiencia educativa mediada por un entorno tecnológico que a través de un recorrido, los estudiantes interactúan con información, conocimientos y

	actividades, buscando desarrollar capacidades, competencias y adquirir conocimientos, que permitan alcanzar los objetivos formativos propuestos.
Aplicaciones para Educación	Programas y/o piezas de software diseñados y producidos para apoyar el desarrollo y cumplimiento de un objetivo, proceso, actividad o situación que implica una intencionalidad o fin educativo; se caracterizan por brindar a los usuarios una gran funcionalidad debido a su versatilidad, nivel de interacción, portabilidad y usabilidad.
Objetos de Aprendizaje	En el contexto Colombiano, se ha definido el término Objetos de Aprendizaje como una entidad digital con un propósito educativo, constituida por, al menos, contenidos y actividades, que se dispone para ser usada y/o reutilizada.

b. Desde los formatos de información digital. Para su representación, la información digital utiliza distintos formatos, los cuales pueden ser manipulados de manera individual o en conjunto durante los procesos de producción de REA. La Tabla 3 describe los formatos más comunes.

Tabla 3. Tipo de REA según su Formato Digital. Fuente: MEN (2012).

Tipo de Formato	Descripción
Textual	Información representada en un sistema de escritura a través de caracteres, que puede apoyarse en otro tipo de representaciones visuales: esquemas, diagramas, gráficos, tablas, entre otros, cuyo uso se desarrolla a través de la lectura
Sonoro	Elementos o secuencias de información acústica, cuya oscilación y vibración puede ser percibida mayormente por el sentido del oído.
Visual	Elementos y/o secuencias de información representada en mayor porcentaje e importancia a través de imágenes, fotografías, gráficas, ilustraciones, capturas ópticas, entre otras; y cuyas oscilaciones y vibraciones pueden ser captados mayormente por el sentido de la vista.
Audiovisual	Elementos de información secuenciados sincrónicamente donde convergen, articulan y se integra lo sonoro, lo textual y lo visual; y cuyas oscilaciones y vibraciones pueden ser captadas simultáneamente por los sentidos de la vista y el oído.
Mutimediales	Elementos de información secuenciados principalmente de forma asincrónica, que articula, secuencia e integra múltiples formatos (textuales, sonoros, visuales y audiovisuales); su potencialidad reside sobre las posibilidades de interacción que ofrece.

Esta categorización proporciona los primeros elementos a ser considerados en la construcción de un modelo de datos para mejorar la disponibilidad y reutilización de los REA. Seguidamente a la identificación de los REA como dominio de conocimiento, se hace necesario caracterizar las propiedades de los recursos, para lo cual se hace uso de los metadatos.

2.2. Metadatos.

Para realizar el proceso de apertura y enriquecimiento de un recurso, inicialmente se debe proveer una descripción adecuada y completa de las características del citado recurso. Para tal fin, los metadatos adquieren un rol primordial. Según Caplan (2003), los metadatos corresponden a datos sobre los datos o información mínima que permite describir elementos de un recurso. Acorde a Cepal (2019), existen distintos modelos de metadatos, cada uno de ellos con distintos esquemas basados en descripción, en su mayoría orientados a facilitar la lectura por parte de los humanos. En los diferentes modelos, cada objeto se describe por medio de una serie de atributos y su valor puede servir para recuperar la información. Dependiendo de la clase de metadatos pueden existir: información sobre elementos de datos o atributos, información sobre la estructura

de los datos, información sobre un aspecto concreto, etc. En Cornell (2003), Caplan (2003), Gutiérrez (2019) y Cortés (2019) se identifican tres tipos de clasificación de metadatos:

a) Metadatos descriptivos, su propósito es identificar y describir un recurso de información, descubrir cómo se encuentra un recurso, y seleccionar recursos que cubran necesidades particulares. Adicionalmente sirven para formar colecciones de recursos similares, y para realizar funciones de evaluación, relación (con otros recursos) y usabilidad.
b) Metadatos administrativos, permiten anotar información que facilite la gestión de los recursos, lo que incluye información sobre cuándo y cómo fue creado el recurso, quién es el responsable del acceso o de la actualización del contenido, e información técnica, como la versión de software o hardware necesarios para ejecutar dicho recurso.
c) Metadatos estructurales, permiten identificar cada una de las partes que componen al recurso, definiendo la estructura que le da forma. Este tipo de metadatos es utilizado principalmente para el procesamiento automático de software de presentación o estilos.

Seguidamente a la construcción del esquema de descripción de características del recurso, se hace necesario identificar el enfoque utilizado para describir y recuperar dichos recursos.

2.2.1. Enfoque para describir y recuperar REA.

Como argumenta Fernández-Pampillón (2017), los vocabularios para la recuperación de REA se utilizan como un sistema de referencia para describir y recuperar esos recursos. El repositorio de recursos o el sistema de almacenamiento utiliza estas descripciones para indexar y encontrar estos recursos. Se plantean tres enfoques para describir y recuperar recursos digitales:

a) ***Metadatos***: Propiedades y valores que describen un recurso desde múltiples puntos de vista: contenido, autoría, formato, propiedad intelectual, propósito, etc. Los recursos se describen e indexan utilizando un esquema de metadatos común. Los esquemas de metadatos utilizados son generalmente estándar para facilitar la interoperabilidad entre repositorios.
b) ***Vocabularios***: Reglas sintácticas que estandarizan la combinación de términos en la indexación de recursos. Se pueden establecer dos tipos de lenguajes de indexación: (i) aquellos que usan términos de vocabularios controlados; y (ii) aquellos que usan los símbolos de clasificaciones bibliográficas y sistemas formales de representación del conocimiento, como las ontologías. En ambos casos, los lenguajes de indexación proporcionan un vocabulario compartido para expresar los términos de indexación y búsqueda y para crear menús de navegación.
c) ***Unión de metadatos con vocabularios***: describir los recursos educativos utilizando esquemas de metadatos en los que ciertas propiedades toman como valores los términos de uno o varios vocabularios establecidos o recomendados en el esquema de metadatos. Si los metadatos y los vocabularios son estándar, también resuelven el problema de la interoperabilidad y compatibilidad entre repositorios y plataformas de aprendizaje electrónico, lo que permite el intercambio, el uso compartido y la integración de colecciones de recursos. Merlot, Europeana y Ariadne (Ariadne, 2017) son ejemplos de repositorios que utilizan metadatos estándar y vocabularios controlados.

En conclusión, como enfoque para describir y recuperar REA, en esta investigación se plantea el uso de ***metadatos con vocabularios***. En contraposición, los lenguajes de indexación sólo permiten describir una propiedad o aspecto de la colección de recursos. Como evidencia de ello se identifican vocabularios como:

- Schools Online Thesaurus – ScOT: tesauro con aproximadamente 11,000 términos para describir contenido en línea, especialmente adecuado para su uso en escuelas de Australia

y Nueva Zelanda. Dentro del área educativa se enfoca en niveles, roles y currículos (http://scot.curriculum.edu.au/).

- Australian Education Vocabularies: Promovido por las agencias y autoridades educativas australianas, se enfoca en competencias, currículo y nieles de formación (http://vocabulary.curriculum.edu.au/).
- SCIS (Schools Catalogue Information Service): vocabulario controlado desarrollado para proporcionar acceso a materias en catálogos de bibliotecas en escuelas de Australia y Nueva Zelanda (https://www.scisdata.com/).
- The mEducator Resource Type Taxonomy: clasifica 45 conceptos en tres categorías: artefactos educacionales, profesionales y de referencia (Giordano *et al*., 2011).

Haciendo uso de esta tipología de metadatos, como lo describe Sánchez Yopazá (2017), se define un esquema de metadatos que parametriza los metadatos disponibles para la descripción del recurso o entidad que se está modelando, además de establecer referentes tales como: el tipo de dato usado (fecha, literal, numérico, etc.), el nivel de obligatoriedad, etc. Este esquema particular debe ser ajustado a un esquema estándar de metadatos, ampliamente reconocido (Bodoff *et al.*, 2005).

2.2.2. Estándares de Metadatos.

Corresponden a metadatos que además de tener un reconocimiento y uso ampliado, responden a una taxonomía de características clasificadas según la funcionalidad que desempeñan para un tipo de recurso particular. Estos estándares, además de establecer un conjunto uniforme de reglas básicas para el etiquetado de información, ayudan a asegurar la consistencia en la aplicación de metadatos, soportan la interoperabilidad de aplicaciones y el intercambio de recursos. En la Tabla 4, Ramesh *et al.* (2015), EdMatrix (2020), Bodas *et al.* (2004) y Berlanga *et al.* (2005) presenta algunos ejemplos de modelos desarrollados como parte de la investigación, especificaciones, perfiles de aplicación y estándares desarrollados por diferentes las instituciones, consorcios y organizaciones involucradas.

Tabla 4. Ejemplos de Estándares de Metadatos. Fuente: Ramesh *et al.* (2015), EdMatrix (2020), Bodas *et al.* (2004) y Berlanga *et al.* (2005).

Tipo	Modelos
Investigación y desarrollo	Educational Modeling Language (EML)
Especificaciones	Learning Object Metadata (LOM) IMS Learning Design (IMS LD) IMS Content Packaging (IMS CP) IMS Learning Tools Interoperability (IMS LTI) Dublin Core Metadata Initiative (DC, dcterms)
Perfiles de Aplicación	Sharable Content Reference Model (SCORM)
Estándar	Learning Object Metadata Standard (IEEE LOM) Learning Resource Metadata Initiative (LRMI) ISO/IEC 19788 Metadata for learning resources

En general, y con el fin de obtener una mayor aceptación de sus modelos de datos, se recomienda un estándar de metadatos reconocido, o en su defecto un ***Perfil de Aplicación*** de dicho esquema (Chan y Zeng, 2006), que permita identificar claramente las características de las entidades que se están describiendo en el dominio de conocimiento trabajado.

Los elementos de un perfil de aplicación se originan en recursos de uno o más esquemas de metadatos. Sin embargo, si se crean nuevos elementos no categorizados en otro esquema, se debe crear un propio esquema de elementos y asumir la responsabilidad de declarar y mantener dicho esquema (IASA, 2011). Esta estrategia será un elemento clave para mapear las ontologías diseñadas al modelo de datos RDF, con el fin de crear las instancias de datasets.

En cuanto a estándares de metadatos para REA, Hummel *et al.* (2004) argumenta que un recurso digital debe ser reproducible, y susceptible de referenciarse (gracias a sus metadatos), y debe ser empleado para llevar a cabo actividades de aprendizaje o de soporte, además de estar disponible para utilizarse por otros. De tal forma que para que un recurso digital puede ser interpretado (o reconocido) y hacerlo realmente reutilizable, dicho recurso digital debe contar con metadatos.

Según Berlanga *et al.* (2005), para las aplicaciones Web existen estándares de metadatos para describir distintas funciones. Por ejemplo, para describir un recurso se ha propagado el uso del esquema Dublin Core, que también desarrolla una línea de aplicación para la anotación específica de recursos educativos y su interoperabilidad con estándares de metadatos del sector educativo. La Tabla 5 presenta algunos estándares de metadatos que han desarrollado iniciativas para describir recursos educativos (otros estándares ver Anexo 2).

Tabla 5. Ejemplos de Iniciativas / Estándares de Metadatos.

Estándar	Descripción
Dublin Core	Compuesto por 15 elementos de metadatos que ofrecen información de catalogación ampliada y una mejor indexación de documentos para programas de motores de búsqueda. En 2008, la Dublin Core Metadata Initiative (DCMI) emitió un nuevo conjunto de términos bajo el actual DCMI namespace: http://purl.org/dc/terms/ (dcterms) para refinar los términos del Dublin Core namespace legado: http://purl.org/dc/elements/1.1/ (dc). La emisión de estos términos siguió la especificación del Modelo abstracto DCMI (DCAM) en 2007. El desarrollo del DCAM fue en parte una respuesta al desarrollo de RDFS (Baskauf, 2013).
Dublin Core Education Application Profile	Adicional a la especificación DC, incluye: a) Audience, que describe el Education Level, aplicable a Docentes y Estudiantes; y Mediator, aplicable para docentes; b) Instructional Method, aplicable para docentes.
IEEE LOM	Especifica la sintaxis y la semántica de Learning Object Metadata, definida como los atributos necesarios para describir total / adecuadamente un objeto de aprendizaje. Los objetos de aprendizaje se definen aquí como cualquier entidad, digital o no digital, que se puede usar, reutilizar o hacer referencia durante el aprendizaje con soporte tecnológico
ISO/IEC 19788-5 Metadata Learning Resource - MLR	Especifica elementos de metadatos y sus atributos para la descripción de los recursos de aprendizaje. Esto incluye las reglas que rigen la identificación de elementos de metadatos y la especificación de los atributos de metadatos.
Colombian Educational Metadata - CEM (MEN, 2013)	Conceptualización y desarrollo de la documentación de la especificación del perfil de metadatos apropiado para la descripción de los recursos educativos, propuesta por el MEN de Colombia
Learning Resource Metadata Initiative (LRMI)	Amplia la ontología de schema.org con el fin de expresar de forma más significativa las características y relaciones educativas.

En algunos escenarios puede presentarse la situación en la que se quiera modificar el conjunto base de metadatos. Este conjunto modificado, que puede tener como origen uno o más esquemas de metadatos o ***perfiles de aplicación***. La Tabla 6 muestra la descripción de algunos perfiles de aplicación.

Tabla 6. Perfiles de Aplicación de Metadatos.

Perfil de Aplicación	Características
Metadatos IMS SENA (Sena, 2017)	Usado por el Servicio Nacional de Aprendizaje – SENA, para la gestión de REA, conformado por nueve categorías con elementos obligatorios, opcionales y recomendados.
The Gateway to Educational Materials (GEM) Metadata Element Set	Basado en DC y enriquecido con elementos educativos específicos
DC-ED	Definido por DCMI en atención a las necesidades educativas. Incorpora los elementos principales de CC, calificadores recomendados y algunos elementos IEEE LOM (Learning Object Metadata)
UK Learning Object Metadata Core	Basado en LOM, optimizado para satisfacer las necesidades del sistema educativo británico
Sharable Content Object Reference Model (SCORM)	Estándares y especificaciones técnicas relacionadas, incluido el IEEE LOM. Define un modelo de contenido específico para garantizar la interoperabilidad del contenido diseñado en varios sistemas de gestión del aprendizaje.

Por último, y con el fin de evitar problemas tales como inconsistencias en la información suministrada, algunos metadatos del esquema definido deben ser alimentados a través de listas de valores controlados, para lo cual es de suma importancia establecer los vocabularios controlados a utilizar.

2.2.3. Vocabularios Controlados.

Los vocabularios controlados capturan los elementos más importantes de un dominio y controlan la forma en que deben representarse. Normalmente se usan en (meta) datos como valores de campos / elementos / celdas, etc. Algunos ejemplos de vocabularios controlados se pueden encontrar en: sistemas de clasificación denominados SKOS - Simple Knowledge Organization System, tesauros, taxonomías, diccionarios geográficos, listas de títulos de materias, entre otros (Isaac y Pastor, 2018). Como lo describe Xie y Matusiak (2016), los vocabularios controlados son esenciales para el proceso de creación de metadatos estandarizados. Si bien los esquemas ofrecen un marco estructural para la creación de registros, los vocabularios controlados son una fuente de términos autorizados para ingresar valores de ciertos elementos, como por ejemplo nombres personales, familiares o corporativos, temas y elementos de cobertura.

El uso de vocabularios controlados garantiza una descripción coherente de los recursos y sus atributos y facilita la recuperación efectiva de la información y el descubrimiento de recursos. Los vocabularios controlados permiten la identificación de relaciones y unen recursos creados por la misma persona o sobre el mismo tema. La selección de vocabularios generalmente se determina durante el proceso de diseño y personalización de metadatos y puede involucrar más de una herramienta establecida de vocabulario controlado y / o el desarrollo de listas locales de vocabulario controlado.

Algunos ejemplos de vocabularios controlados corresponden a:

- The Art & Architecture Thesaurus (AAT): contiene términos, descripciones, citas bibliográficas y otra información relacionada con bellas artes, arquitectura, artes decorativas, materiales de archivo y cultura material.
- The Union List of Artist Names (ULAN): contiene información bibliográfica sobre artistas y arquitectos, incluidos nombres de variantes, seudónimos y variantes de idioma.

- The Getty Thesaurus of Geographic Names (TGN): contiene los nombres de lugares importantes para el estudio del arte y la arquitectura.

En el campo educativo (Fernández-Pampillón, 2017), los vocabularios son un mecanismo de representación semántica que permite a los agentes humanos o al software ubicar, interpretar y recuperar contenido educativo o procesarlo con fines didácticos. Los vocabularios se aplican principalmente para resolver dos preguntas:

a) La representación y recuperación de recursos educativos abiertos: el tipo principal de vocabularios utilizados son taxonomías y tesauros, es decir, vocabularios de términos organizados en categorías y / o interconectados por hipernymy-hyponymy y otras relaciones semánticas.
b) La interoperabilidad entre las herramientas de aprendizaje electrónico y el contenido: el tipo de vocabularios utilizados son tesauros y ontologías que representan conceptualmente las dimensiones de un sistema de aprendizaje electrónico: los agentes, las herramientas, el dominio del conocimiento, las metodologías y los modelos de enseñanza, etc.

Dentro de los vocabularios controlados para el ambiente educativo se encuentran algunos como los siguientes (Nahhas *et al.*, 2018):

- AIISO - Academic Institution Internal Structure Ontology (Styles y Shabir, 2008): proporciona clases y propiedades para describir la estructura organizativa interna de una institución académica. Diseñado para trabajar en colaboración con Participation (http://purl.org/vocab/participation/schema), Friend of a Friend - FOAF (http://xmlns.com/foaf/0.1/) y aiiso-roles (http://purl.org/vocab/aiiso-roles/schema) para describir los roles que desempeñan las personas dentro de una institución.
- CURONTO: Modelo semántico de plan de estudios para la evaluación y mejora del programa.

Después de definir el esquema de metadatos que describe el recurso o entidad trabajada, se hace necesario compartir dicho esquema y permitir que se integre y enriquezca con otros componentes, surgiendo así la necesidad de la apertura de esquemas de datos y recursos, con el fin de construir sistemas complejos y grandes.

2.3. Datos Abiertos (Open Data).

Como se describe en OKF (2019): "*los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen*".

Los datos abiertos se basan en un conjunto de principios enmarcados por The Sunlight Foundation (2017), dentro de los cuales se define que:

- Completitud: Los datasets deben ser lo más completos posible, reflejando la totalidad de lo que se registra sobre un tema en particular. También se deben incluir metadatos que definan y expliquen los datos brutos, junto con fórmulas y explicaciones sobre cómo se calcularon los datos derivados.
- Primarios: Los datasets deben contener datos de fuente primaria.
- Oportunidad: Los datasets deben estar disponibles para el público de manera oportuna. Se debe dar prioridad a los datos cuya utilidad sea urgente. Las actualizaciones de información en tiempo real maximizarían la utilidad que el público puede obtener de esta información.

- Facilidad de acceso físico y electrónico: Los datasets deben ser lo más accesibles posible, permitiendo obtener información, ya sea a través de medios físicos o electrónicos.
- Legibilidad mecánica: La información debe almacenarse en formatos de archivo de uso generalizado que se presten fácilmente al procesamiento automático. Estos archivos deben ir acompañados de documentación relacionada con el formato y cómo usarlo en relación con los datos.
- No discriminación: El acceso no discriminatorio a los datos debería permitir que cualquier persona acceda a los datos en cualquier momento sin tener que identificarse ni proporcionar ninguna justificación para hacerlo.
- Estándares abiertos o de propiedad común: Los datasets deben estar en formatos de archivo disponibles gratuitamente con la mayor frecuencia posible.
- Licencias: Los datasets deben estar disponibles bajo licencias diseñadas para aumentar la apertura y minimizar las restricciones sobre el uso de los datos.
- Permanencia: La información debe estar disponible en línea en archivos a perpetuidad.
- Costos de uso: La información debe estar disponible en línea de forma gratuita.

En resumen, abrir los datos no consiste solamente en hacerlos disponibles en la Internet permitiendo su descarga. Para que sean realmente datos abiertos, además de estar disponibles en Internet, deben poseer algún tipo de licencia legal para poderlos utilizar, reutilizar y redistribuir, mezclándolos incluso con otros datos, sujetos como mínimo a la “atribución” (reconocimiento de la autoría, quién lo ha hecho), o al “*compartir igual*” (que la explotación que se haga de esos datos –incluyendo las obras derivadas– mantengan la misma licencia al ser divulgadas) (Hernández y García, 2013, citado en Ríos-Hilario, 2014).

Seguidamente a la apertura de los datos, el enriquecimiento de los metadatos a través de la adición de componentes semánticos, además de su disposición pública en formatos RDF, cobran un papel importante, rol que es suplido por la Web Semántica.

2.4. Web Semántica.

Como se describe en Laufer (2015) y Sánchez (2014), la Web Semántica, término acuñado por Berners-Lee (2009), es una red de información diseñada para que las máquinas la puedan procesar fácilmente a escala global. La idea general es crear una forma eficiente de representar datos en la World Wide Web y crear una gran base de datos global enlazada. Agregar semántica a los datos facilita la comprensión e interoperabilidad de los mismos en el universo de información heterogénea, publicada en formatos diferentes y con diferentes protocolos de acceso.

La Web Semántica busca facilitar el proceso de comunicación entre los variados participantes del ecosistema, a fin de crear un modelo mental común que minimice la posibilidad de ambigüedad y, por lo tanto, facilite el trabajo necesario para desarrollar aplicaciones que manipulen diferentes fuentes de datos. Dado que la Web Semántica se basa en información legible por máquinas, para su operación usa un pool de herramientas que le permiten trabajar. Algunas de estas herramientas operan con la capacidad de tecnologías XML para definir esquemas de etiquetado personalizados y el enfoque flexible de estándar basado en RDF para representar datos. De igual forma, proporciona un lenguaje común para registrar cómo los datos se relacionan con objetos del mundo real (Giri, 2011).

Los bloques básicos que definen la Web Semántica según Laufer (2015) y Ayllón (2012), son: a) un modelo de datos estándar; b) un conjunto de vocabularios de referencia; y c) un protocolo de consulta estándar. En primera instancia, y para facilitar la comunicación, la Web Semántica hace uso de un modelo de datos común basado en RDF, como estructura para representar la información.

2.4.1. Modelo de Datos RDF.

RDF es un estándar del World Wide Web Consortium (W3C) para el intercambio de datos y descripción de semántica. Este modelo de datos está basado en grafos de propósito general que permite la integración de datos de distintas fuentes, incluso si éstas tienen esquemas distintos (Wood *et al.*, 2014). El modelo RDF se basa en relacionar entidades (también llamados recursos), por medio de relaciones binarias, también llamadas enunciados (Cabrera, 2016). Una declaración RDF (Figura 3) constituye la forma más simple de expresión de un metadato, mediante grafos de tripletas compuestas por:

- Un sujeto: la cosa, el recurso, sobre el que se declara algo.
- Un predicado o propiedad: el aspecto que se describe.
- Un Objeto: el valor de lo que se describe sobre el sujeto.

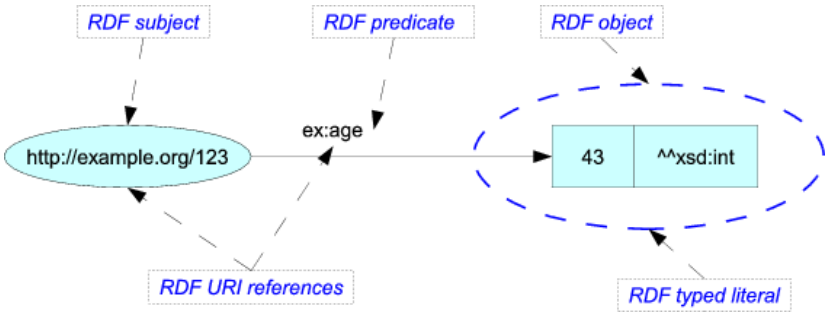


Figura 3. Tripleta RDF. Fuente: Nilsson *et al*. (2008).

Los predicados pueden ser de 2 tipos:

- Enlaces: si representan relaciones entre recursos.
- Literales: si representan propiedades de un recurso.

Como lo describe Sulé *et al.* (2016), los recursos que constituyen el sujeto y el predicado tienen que ser designados por URIs, mientras que el recurso que se sitúa en el objeto de la declaración puede ser expresado mediante URI o literales. Además, los recursos de los tres componentes de las declaraciones pueden enlazarse a ontologías, conceptualizaciones explícitas de ámbitos concretos de la realidad, expresados por vocabularios normalizados (Skos - Simple Knowledge Organization System, FOAF - Friend of a Friend, DCAT - Data Catalog Vocabulary, etc.) y formados, como componentes básicos, por conceptos y relaciones entre conceptos. Al enlazar los recursos situados en sujetos y objetos de declaraciones RDF, a los conceptos y predicados de las relaciones, pueden ser compartidos por diferentes productores y consumidores de datos, y puede ser sometido a restricciones semánticas expresadas formalmente, muy útiles en el momento que son procesadas por los sistemas automáticos.

Cuando el modelo de datos es instanciado se genera un dataset. Este dataset corresponde a un archivo estructurado LOD basado en un modelo de datos RDF, cuyas tripletas cubren un campo de interés, y contienen descripciones de entidades dentro del campo de interés definido. Estas descripciones a menudo comparten un prefijo de URI común.

2.4.2. Ontologías.

Según la definición de Gruber (1993:199), -citado en Pérez (2002)-, una ontología "*constituye una especificación explícita y formal de una conceptualización compartida*". En esta definición, sus componentes esenciales son:

- Conceptualización: hace referencia a un modelo abstracto de algún fenómeno del mundo del que se identifican los conceptos que son relevantes.
- Explícita: hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología.
- Formal: indica que la especificación debe representarse por medio de un lenguaje de representación formalizado.
- Compartida: refleja que una ontología debe, en el mejor de los casos, dar cuenta de conocimiento aceptado (como mínimo, por el grupo de personas que deben usarla).

Según Fensel (2004), citado en Aranda y Ruiz (2005), las ontologías se clasifican en:

- Ontologías Genéricas o de Sentido Común: Capturan conocimiento general acerca del mundo. Proveen nociones básicas y conceptos para cosas tales como espacio, tiempo, estado, eventos. Son válidas en varios dominios.
- Ontologías Representacionales: No se comprometen con ningún dominio en particular. Proveen entidades sin establecer qué deberían representar, por lo tanto definen conceptos para expresar conocimiento de manera orientada a objetos o a marcos de trabajo.
- Ontologías de Dominio: Capturan el conocimiento válido para un tipo particular de dominio (por ejemplo: electrónica, medicina, mecánica).
- Ontologías de Métodos y Ontologías de Tareas: Las primeras proveen términos específicos para métodos particulares de resolución de problemas, mientras que las segundas proveen términos para tareas específicas. Ambas proveen un punto de vista de razonamiento sobre conocimiento del dominio.

En resumen, una ontología describe formalmente un dominio del discurso. Típicamente, consiste en una lista finita de términos y las relaciones entre estos términos. Los términos denotan conceptos importantes (clases de objetos) del dominio (Antoniou y Van Harmelen, 2008). El vocabulario definido por la ontología es una lista cuidadosamente seleccionada de palabras y frases, que se utilizan para etiquetar unidades de información para que puedan ser recuperadas más fácilmente mediante una búsqueda. Las ontologías se convierten en recursos para mantener actualizados la estructura del modelo de datos que soporta un dataset, además de controlar las versiones del esquema usado. Según Gruber (1993), una ontología se basa en cuatro tipos de componentes:

- Conceptos: son las ideas básicas que tratan de ser formalizadas. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, entre otros.
- Relaciones: representan la interacción y enlace entre los conceptos de dominio.
- Funciones: Son un tipo especial de relación en la que un elemento se identifica mediante el cálculo de una función.
- Axiomas: son teoremas sobre las relaciones que deben cumplir los elementos de la ontología.

En cuanto al contexto educativo, como lo plantea Fernández-Pampillón (2017), las ontologías se aplican para describir los componentes de los sistemas de gestión del aprendizaje, como los tipos de usuarios, la organización y administración académica, los cursos, las actividades y los recursos. Estas descripciones semánticas para cada componente permiten crear aplicaciones para personalizar y adaptar la enseñanza y el aprendizaje a las necesidades de cada alumno, facilitando el aprendizaje colaborativo en un entorno distribuido, el intercambio y la reutilización de información y conocimiento, y la gestión y el acceso a los repositorios digitales de recursos educativos.

En cuanto al papel de las ontologías en la gestión y el acceso a los repositorios de REA, estos se utilizan básicamente como índices para indexar recursos con respecto a sus contenidos. Las ontologías también sirven para interpretar las consultas de los usuarios mediante el uso de estrategias de búsqueda basadas en conceptos. Finalmente, también se usan como mapas conceptuales y esquemas del dominio, como tesauros, a través de los cuales maestros y estudiantes pueden navegar para localizar recursos que tengan contenido, temas o conceptos relacionados.

En el contexto de los recursos educativos se identifican diferentes iniciativas en la definición de ontologías para el dominio de conocimiento, dentro de las cuales se encuentran:

- MLO Data Model (Cetis, 2020): Especificación europea que cubre una amplia gama de información sobre las oportunidades de aprendizaje. Este modelo identifica como elementos de un recurso: Ubicación, inicio, duración, costo, idioma, prerrequisitos, lugares, objetivo, evaluación, método de enseñanza y diseño de aprendizaje
- DocBook (Rocha Amorim *et al.*, 2012): ontología para describir recursos educativos, específicamente aquellos definidos como documentos y libros.
- mEducator Schema (Evangelia *et al.*, 2011) propone una ontología para describir y exponer recursos educativos médicos. El modelo conceptual utilizado como base para desarrollar el perfil de aplicación de metadatos es basado en LOM.
- Ontología OER-CC (Piedra *et al.*, 2010), que modela el conocimiento del dominio de los recursos educativos con licencia de Creative Commons Licenses.
- ISO-MLR RDF model (Paquette y Miara, 2014): ontología para el diseño y uso de cursos masivos abiertos en línea.
- OER Schema (Oer Schema, 2020): agrega un vocabulario de términos que expande el vocabulario Schema.org desarrollado por Google para incluir términos pedagógicos que cualquiera puede aplicar a su propio contenido de aprendizaje.
- AIISO - Academic Institution Internal Structure Ontology (Styles y Shabir, 2008): proporciona clases y propiedades para describir la estructura organizacional interna de una institución académica.

Dada la pluralidad de ontologías utilizadas en el contexto educativo, se identifica que:

a. Existen ontologías para el dominio de conocimiento, pero su orientación se enfoca a tipos particulares de recursos educativos.
b. Se hace evidente la generación de una ontología que permee los tipos de recursos educativos abiertos, además de proveer una mejor descripción de los mismos, y que permita ser interoperable con otros repositorios y procesos de explotación de datos.

Finalmente, para describir formalmente el dominio, tal como lo plantea Laufer (2015), es necesario establecer un conjunto de vocabularios de referencia, como modo de facilitar la comunicación de los metadatos. Para explorar colecciones de vocabularios se pueden consultar catálogos tales como LOV - Linked Open Vocabularies (LOV, 2020). Sin embargo, en caso de que ninguna ontología satisfaga las necesidades de expresión de los metadatos, puede crearse una nueva ontología, observando la posibilidad de reutilizar el mayor número posible de elementos de ontologías ya existentes, evitando de ese modo la duplicación de referencias distintas para los mismos conceptos.

Seguidamente a la descripción formal de los REA como dominio de conocimiento, se lleva a cabo el proceso de enriquecimiento de los metadatos, a través de un método de publicación de datos denominado datos enlazados, de tal forma que además de estar interconectados, ofrezcan un contenido semántico mayor.

2.4.3. Datos Enlazados (Linked Data).

Linked Data hace referencia a un conjunto de mejores prácticas para la publicación y la conexión de datos estructurados en la Web. Buenas prácticas que han sido adoptadas por diferentes proveedores de datos, dando lugar a la creación de un espacio global de datos que contiene miles de millones de recursos - la Web de Linked Data. Esta estrategia va más allá de la publicación de datos en la Web, enlazando datos a otros datos, de forma que tanto personas como máquinas puedan explorar la Web de los Datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales. La Web Semántica se construye mediante documentos en la Web, pero a diferencia del hipertexto, ésta enlaza datos arbitrarios que se describen en RDF (Berners-Lee, 2009; Berners-Lee, 2011).

Linked Data, como modelo, permite construir y enlazar datos mediante una gran base de datos interconectados y distribuidos en la Web (Bauer y Kaltenböck, 2012), haciendo uso de un conjunto de tecnologías estándar: las URI – identificador de recursos uniforme, el protocolo de transferencia de hipertexto - HTTP y los Modelos de Datos RDF. En los últimos años, fuentes de datos Web tales como Google Scholar, YouTube, Twitter, Facebook, Wikipedia, entre otras, han facilitado el acceso a sus bases de datos a través de sus interfaces de programación de aplicaciones - API Web. ProgrammableWeb.com cuenta con más de 1.300 de estas API. La riqueza de datos accesibles a través de las API Web ha llevado al desarrollo de aplicaciones web híbridas (mashups) interesantes que combinan datos de diferentes fuentes (Bizer, 2009).

Como lo describe W3C (2017) y OpenData (2015) enlazar datos, como parte de la Web Semántica, se basa en la aplicación de ciertos principios básicos y necesarios, que fomentarán el crecimiento de la Web, tanto a nivel de los documentos HTML (vista clásica de la Web), como a nivel de los datos expresados en RDF (vista de la Web Semántica). Dentro de estos principios básicos se detallan:

a) Usar URIs para identificar las cosas, proceso que ofrece una abstracción del lenguaje natural, evitando ambigüedades, además de ofrecer una forma estándar y unívoca de referencia a un recurso.
b) Usar URIs HTTP, asegurando que cualquier recurso pueda ser buscado y accedido en la Web.
c) Ofrecer información sobre los recursos usando RDF. Una vez que se busca y se accede a un recurso identificado mediante una URI HTTP, se debe obtener información útil sobre dicho recurso, representada mediante descripciones estándares en RDF.
d) Incluir enlaces a otros URIs: ofreciendo un valor añadido a la información que se provee. Los enlaces de los recursos mediante URIs, pueden hacerse tanto localmente como a través de toda la red.

En términos generales, como lo plantean Auer *et al.* (2014) la actividad de la Web Semántica ha basado su operación en la publicación generalizada de datos estructurados como RDF. Para ejecutar su trabajo, el modelo de Linked Data traduce sus principios en un esquema de 5 niveles a abordar en el proceso de enlace de datos, los cuales son mostrados en la Tabla 7.

Tabla 7. Niveles de Enlace de Datos. Fuente: OpenData (2015).

Nivel	Descripción
1	Abrir los datos a la web, en cualquier formato, pero bajo licencias abiertas
2	Hacer los datos accesibles como datos estructurados
3	Estructurar los datos bajo formatos no propietario
4	Usar URIs para identificar los datos publicados
5	Enlazar los datos publicados con otros datos para proporcionar contexto

Estos niveles se operacionalizan en la arquitectura mostrada en la Figura 4. Para realizar el enlace, se debe seguir una metodología que cubra las diferentes etapas del ciclo de vida de Linked Data.

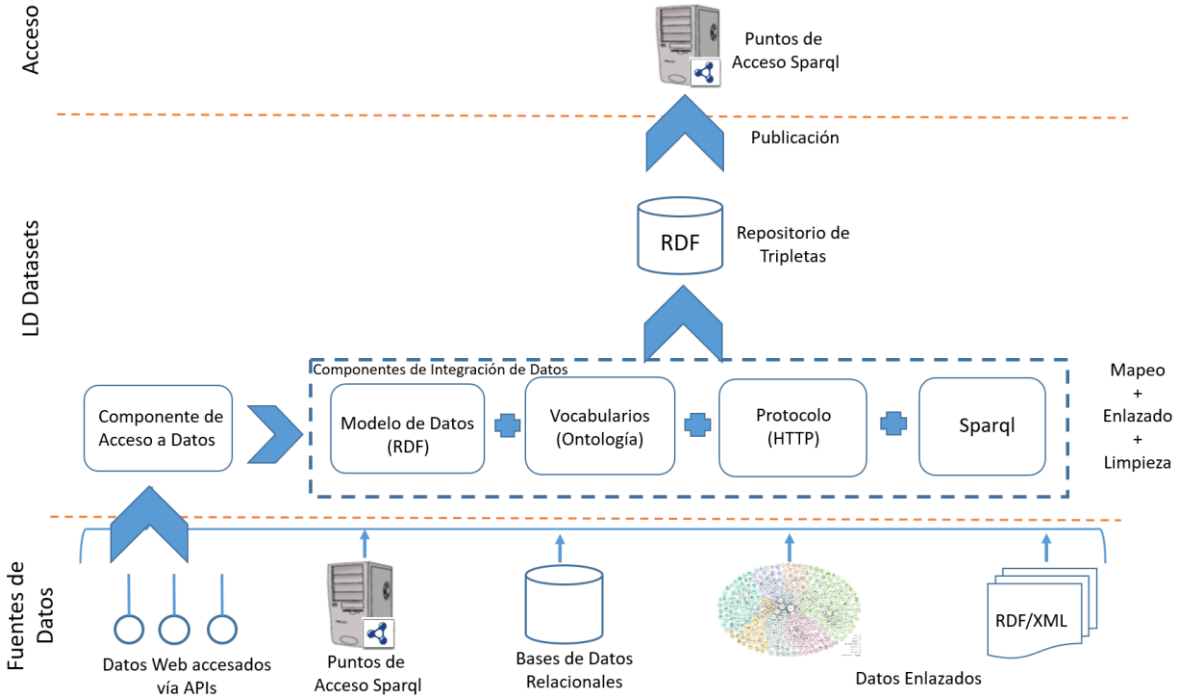


Figura 4. Arquitectura Linked Data. Fuente: Elaboración propia basado en Pinkel (2013).

Al combinarse las buenas prácticas para la publicación de datos con la utilización libre y abierta de los datos, surge el término de Datos Abiertos Enlazados.

2.4.4. Datos abiertos enlazados (Linked Open Data – LOD).

LOD es la resultante de la combinación de las buenas prácticas presentadas por el modelo de Linked Data y los lineamientos y principios de apertura de datos, es decir el enlace de datos utilizando fuentes abiertas. LOD tiene como objetivo: (i) introducir los beneficios de RDF y tecnologías de la Web Semántica al movimiento Open Data, y (ii) arrancar la Web de los Datos mediante la creación, edición e interconexión RDF de éstos datasets abiertos (Hogan, 2013).

En este sentido, es oportuna la apreciación de Zeinstra y Keller (2016) cuando manifiestan que "*los datos pueden estar abiertos pero no enlazados. Los datos pueden estar enlazados pero no abiertos. La Web Semántica puede solamente funcionar con datos que estén tanto abiertos como enlazados*". Como un ejemplo de ello, Europeana es un portal web que actúa como una interfaz de millones de libros, pinturas, películas, objetos de museo y registros de archivos digitalizados de toda Europa. Esta biblioteca es también una plataforma para el intercambio de conocimiento que promueve la colaboración entre los agentes culturales (Ríos-Hilario *et al.*, 2012).

En general, como plantea Zuiderwijk *et al.* (2012), LOD basa su trabajo en el uso de metadatos, que inicialmente son planos, y que con la articulación del modelo LOD permiten adicionar riqueza a la información en ellos almacenada. Tal articulación se lleva a cabo surtiendo cada una de las fases del ciclo de vida de datos enlazados.

2.4.5. Ciclo de vida de los datos enlazados.

Estudios realizados por autores como Hausenblas (2011), Heath y Bizer (2011), Auer *et al.* (2012), Hausenblas (2009), Nahhas *et al.* (2018), han identificado diferentes fases en el ciclo de vida de los datos enlazados. En general, las diferentes etapas del ciclo de vida de los datos enlazados (Ben Ellefi, 2017; Wood, 2011) incluyen:

- Extracción y transformación de información de la fuente de datos sin procesar a datos RDF. La fuente de datos puede estar en formato no estructurado, estructurado o semiestructurado, es decir, CSV – valores separados por coma, archivo XML, una base de datos relacional, etc. Esta etapa no incluye modelado de vocabulario, asignación de espacios de nombres ni enlaces a datasets existentes.
- Asignar espacios de nombres a todos los recursos, especialmente hacerlos accesibles a través de sus URI.
- Modelar los datos RDF reutilizando términos de vocabularios existentes relevantes siempre que sea posible.
- Alojar públicamente el dataset enlazado y sus metadatos, además de hacerlo accesible.
- Enlazar el dataset recién publicado a otros datasets ya publicados como datos enlazados en la Web.

Una vez los datos de los REA son representados en RDF y son publicados en la Web, es posible consultarlos haciendo uso de SPARQL.

2.4.6. SPARQL.

La información provista por los datasets puede ser recuperar a través de consultas. El lenguaje diseñado para consultar datasets RDF se llama SPARQL (SPARQL Protocol and RDF Query Language). SPARQL se puede utilizar para expresar consultas sobre diversas fuentes de datos, ya sea que los datos se almacenen de forma nativa como RDF o si son definidos mediante vistas RDF a través de un sistema middleware. SPARQL contiene capacidades para consultar patrones de grafos obligatorios y opcionales junto con sus conjunciones y disyunciones. SPARQL también admite la ampliación o restricción del ámbito de las consultas indicando los grafos RDF de origen. Los resultados de las consultas SPARQL pueden ser conjuntos de resultados o grafos RDF (Prud'hommeaux y Seaborne, 2008).

Para ejecutar las consultas SPARQL sobre los dataset publicados en la Web, se requiere de un punto de presencia en una red HTTP, identificado mediante una URL, que sea capaz de recibir y procesar solicitudes del protocolo SPARQL (Idehen, 2018). Este punto de presencia se denomina SPARQL Endpoint o punto de acceso SPARQL.

Si bien, los puntos de acceso se configuran como un extremo del ciclo de vida de los datos enlazados, hay un elemento clave que se debe tener en cuenta tanto a lo largo de este ciclo de vida como en los resultados de las consultas obtenidas en los puntos de acceso SPARQL. Este elemento corresponde a la calidad de los datos enlazados.

2.4.7. Dimensiones de calidad de datos enlazados.

Un factor primordial al abordar estas fases es el tratamiento dado a la calidad de los datos. Como lo describe Zaveri *et al.* (2012): "*La calidad de los datos se concibe comúnmente como una construcción multidimensional con una definición popular como la "aptitud para el uso". La calidad de los datos puede depender de varios factores como la precisión, puntualidad, integridad, relevancia, objetividad, credibilidad, comprensión, consistencia, concisión, disponibilidad y verificabilidad*".

Con la ejecución de cada fase, se deben tener en cuenta diferentes conceptos de calidad de los datos, tales como calidad del modelo, los metadatos semánticos, la noción de calidad de enlace, entre otros, con el fin de mitigar los problemas de calidad de datos que se han venido presentando tanto en los modelos como en las instancias de datos. En cuanto al problema de calidad de los datos, éstos se refieren a un conjunto de problemas que pueden afectar la potencialidad de las aplicaciones que utilizan los datos. En la revisión de literatura realizada por Zaveri *et al.* (2012), se identifican un conjunto básico de 26 dimensiones de calidad de datos diferentes que se pueden aplicar para evaluar la calidad de Datos Enlazados (Figura 5).

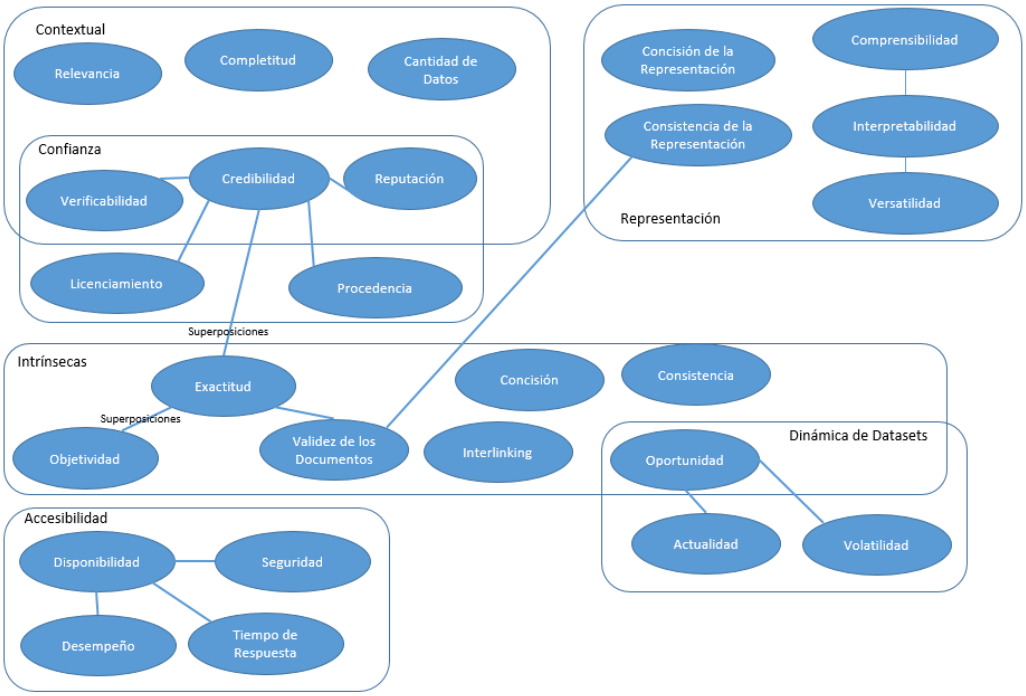


Figura 5. Dimensiones de Calidad en Datos Enlazados. Fuente: Zaveri *et al.* (2012).

La conjugación de este marco de referencia de dimensiones de calidad de datos con el ciclo de vida de datos enlazados, en los puntos focales de investigación, se convierte en un elemento relevante para identificar cuellos de botella, y poder construir elementos y estrategias que ayuden a mejorar la evaluación de calidad de los modelos de datos.

Bajo el contexto definido por Zaveri *et al.* (2012), y como objeto de esta investigación dada su incidencia sobre el enlace de recursos, se toman como base tres dimensiones de calidad de datos enlazados. Las dimensiones a trabajar se definen a continuación.

a) **Dimensiones Contextual.** Dimensiones que dependen en gran medida del contexto de la tarea en cuestión, así como de las preferencias subjetivas del consumidor de datos. Hay tres dimensiones contextuales:

- ***Completitud***: grado en que toda la información requerida está presente en un dataset particular. En general, la completitud es la medida en que los datos tienen la profundidad, la amplitud y el alcance suficientes para la tarea en cuestión.
- ***Cantidad de datos***: cantidad y el volumen de datos que son apropiados para una tarea en particular.
- ***Relevancia***: provisión de información que está de acuerdo con la tarea en cuestión y es importante para la consulta de los usuarios.

La Tabla 8 muestra las métricas aplicables para la dimensión contextual.

Tabla 8. Métricas de Calidad de Datos - Contextual. Fuente: Zaveri *et al.* (2012).

Dimensión	Métrica	Descripción	Tipo
Completitud	Grado en el cual las clases y propiedades no se omiten	Detección del grado en que se representan las clases y propiedades de una ontología.	S
	Grado en el cual los valores para una propiedad no se omite	detección de no. de valores perdidos para una propiedad especifica	O
	Grado en el cual los objetos del mundo real no se omiten	Detección del grado en que están representados todos los objetos del mundo real.	O
	Grado en el cual los enlaces no se omiten	Detección del grado en que las instancias en el dataset están relacionadas entre sí.	O
Cantidad de Datos	Apropiado volumen de datos para una tarea en particular	No. de tripletes, instancias por clase, enlaces internos y externos en un dataset.	O
	Cubrimiento	Alcance y nivel de detalle	O
Relevancia	Uso de atributos de meta información	Contando la ocurrencia de términos relevantes dentro de estos atributos o utilizando el modelo de espacio vectorial y asignando mayor importancia a los términos que aparecen dentro de los atributos de meta información	S
	Recuperación de Documentos Relevantes	Clasificar los documentos según su relevancia para una consulta determinada	S

Nota: S: Subjetiva, O: Objetiva

b) **Dimensión de Confianza.** Dimensiones centradas en la confiabilidad del dataset. Hay cinco dimensiones que forman este grupo:

- ***Procedencia:*** metadatos contextuales que se centran en cómo representar, gestionar y utilizar la información sobre el origen de la fuente. La procedencia ayuda a describir entidades para permitir la confianza, evaluar la autenticidad y permitir la reproducibilidad.
- ***Verificabilidad:*** grado en que un consumidor de datos puede evaluar la exactitud de un dataset y, como consecuencia, su confiabilidad.
- ***Credibilidad:*** grado en que se acepta que la información es correcta, verdadera, real y creíble.
- ***Reputación:*** juicio realizado por un usuario para determinar la integridad de una fuente. Se asocia principalmente con un editor de datos, una persona, organización, grupo de personas o comunidad de práctica, en lugar de ser una característica de un dataset. El editor de datos debe ser identificable para un determinado dataset (o parte de él).
- ***Licenciamiento:*** concesión de un permiso para que un consumidor reutilice un dataset en condiciones definidas.

La Tabla 9 muestra las métricas aplicables para la dimensión de Confianza.

Tabla 9. Métricas de Calidad de Datos - Confianza. Fuente: Zaveri *et al.* (2012).

Dimensión	Métrica	Descripción	Tipo
Procedencia	Indicador de metadatos sobre un dataset	Presencia del título, contenido y URI del dataset.	O

Dimensión	Métrica	Descripción	Tipo
	Informes de Confianza personalizados	Usando la procedencia de las anotaciones de confianza existentes en las redes sociales.	S
	Computando la confiabilidad de las declaraciones RDF	Valor de confianza basado en la información de procedencia.	O
	Computando la confiabilidad de las declaraciones RDF	Valor de confianza basado en calificaciones basadas en el usuario o en un método basado en la opinión	S
	Detectar la confiabilidad y credibilidad de una persona (editor).	Indicador del nivel de confianza de que se conoce una persona.	S
	Computando la confianza de una entidad	Construcción de redes de decisión informadas por gráficos de procedencia.	O
	Precisión del cómputo de la confianza entre dos entidades.	Utilizando una combinación de (1) algoritmo de propagación que utiliza técnicas estadísticas para calcular valores de confianza entre 2 entidades a través de una ruta y (2) un algoritmo de agregación basado en un mecanismo de ponderación para calcular el valor agregado de confianza en todas las rutas	O
	Adquirir confianza de contenido de los usuarios.	Basado en asociaciones que transfieren la confianza de las entidades a los recursos.	O
	Detección de confiabilidad y credibilidad de una fuente de datos	Uso de anotaciones de confianza hechas por varios individuos para derivar una evaluación de la confiabilidad y credibilidad de las fuentes	S
	Asignar valores de confianza a datos / fuentes / reglas.	Uso de ontologías de confianza que asignan valores de confianza basados en metadatos o en contenido que se pueden transferir de datos conocidos a datos desconocidos	O
	Determinar el valor de confianza para los datos	Uso de anotaciones para datos como (i) listas negras, (ii) autoridad y (iii) clasificación y uso de razonamiento para incorporar valores de confianza a los datos	O
Verificabilidad	Verificando la información del publicador	Indicando el autor y sus colaboradores, el editor de los datos y sus fuentes.	S
	Verificando la autenticidad del dataset	Si el dataset utiliza un vocabulario de procedencia, por ejemplo: el uso del Vocabulario de procedencia	O
	Verificando la correctitud del dataset	Con la ayuda de un tercero imparcial y confiable.	S
	Verificando el uso de firmas digitales	Firmar un documento que contiene una serialización RDF o firmar un gráfico RDF	O
Reputación	Reputación del Publicador	Encuesta en una comunidad cuestionada sobre otros miembros.	S
	Reputación del Dataset	Analizar referencias o rango de página o asignando un puntaje de reputación al dataset	S
Credibilidad	Meta-información sobre la identidad del proveedor de información.	Comprobando si el proveedor / colaborador está contenido en una lista de proveedores confiables	O

Dimensión	Métrica	Descripción	Tipo
Licenciamiento	Indicador legible por máquina de una licencia	Detección de la indicación de una licencia en la descripción del VoiD o en el propio dataset	O
	Indicador legible de una licencia.	Detección de una licencia en la documentación del dataset o su fuente	O
	Permisos para usar el dataset	Detección de licencia que indica si se permite la reproducción, distribución, modificación o redistribución	O
	Indicador de Atribución	Detección de si el trabajo se atribuye de la misma manera que lo especificado por el autor o licenciante	O
	Indicador de Copyleft o ShareAlike	Comprobando si los contenidos derivados se publican bajo la misma licencia que el original	O

Nota: S: Subjetiva, O: Objetiva

c) **Dimensión de Accesibilidad.** Las dimensiones que pertenecen a esta categoría incluyen aspectos relacionados con la forma en que se puede acceder y recuperar los datos. Hay cuatro dimensiones que forman parte de este grupo:

- La ***Disponibilidad*** de un dataset es la medida en que la información está presente, se puede obtener y está lista para su uso.
- ***Rendimiento:*** eficiencia de un sistema que se une a un gran dataset, es decir, cuanto más eficiente sea un origen de datos, más eficientemente un sistema puede procesar datos.
- ***Seguridad:*** medida en que se puede restringir el acceso a los datos y, por lo tanto, protegerse contra su alteración y uso indebidos ilegales. Se refiere al grado en que la información se transmite de forma segura de los usuarios a la fuente de información y viceversa.
- ***Tiempo de respuesta*** mide la demora, generalmente en segundos, entre el envío de una consulta por parte del usuario y la recepción de la respuesta completa del dataset.

La Tabla 10 muestra las métricas aplicables para la dimensión de accesibilidad.

Tabla 10. Métricas de Calidad de Datos - Accesibilidad. Fuente: Zaveri *et al.* (2012).

Dimensión	Métrica	Descripción	Tipo
Disponibilidad	Accesibilidad del Servidor	Comprobando si el servidor responde a una consulta SPARQL.	O
	Accesibilidad del SPARQL Endpoint	Comprobando si el servidor responde a una consulta SPARQL.	O
	Accesibilidad del RDF dumps	Comprobando si se proporciona un RDF dump y se puede descargar	O
	Problemas de desreferencia	Cuando un URI devuelve un error (4xx cliente error / 5xx servidor error) código de respuesta o detección de enlaces rotos	O
	Datos no estructurados disponibles	Detección de enlaces muertos o detección de un URI sin metadatos RDF compatibles o sin redirección utilizando el código de estado 303 o ningún código 200 OK	O
	Tipos de contenido mal informado	Detección de si el contenido es adecuado para el consumo y si se debe acceder al contenido	S
	No hay vínculos de regreso referenciados	Detección de todos los in-links o back-links locales: los triples disponibles localmente en los que el URI del recurso aparece como un	O

Dimensión	Métrica	Descripción	Tipo
		objeto, en el documento sin referencia devuelto para el recurso dado	
Rendimiento	Sin uso de slash-URIs	Comprobación del uso de slash-URI donde se proporcionan grandes cantidades de datos	O
	Baja latencia	Si una solicitud HTTP no se responde dentro de un tiempo promedio de un segundo, la latencia de la fuente de datos se considera demasiado baja	O
	Alto throughput	Número de solicitudes HTTP respondidas por segundo	O
	Escalabilidad de una fuente de datos	La detección de si el tiempo para responder a una cantidad de diez solicitudes dividida por diez no es mayor que el tiempo que se tarda en responder una solicitud	O
	Sin uso de las características prolix RDF	Detectar el uso de primitivas RDF, es decir, reificación RDF, contenedores RDF y colecciones RDF	O
Seguridad	El acceso a datos es seguro	Uso de credenciales de inicio de sesión o uso de SSL o SSH	O
	Los datos son de carácter propietario.	El propietario de los datos permite el acceso solo a ciertos usuarios.	O
Tiempo de Respuesta	Demora en el tiempo de respuesta	Retraso entre la presentación de una solicitud por parte del usuario y la recepción de la respuesta del sistema	O

Nota: S: Subjetiva, O: Objetiva

Por último, es importante considerar que los procesos de descripción, apertura y enriquecimiento de recursos o entidades, generan esquemas ontológicos de metadatos que alimentan plataformas de gestión de recursos, o en su defecto, son formalizados en datasets a través de un modelado RDF. Por ende, para la construcción de estos modelos se debe considerar una aproximación que permita aplicar las buenas prácticas y principios definidos por el modelo de datos abiertos enlazados, además de proveer portabilidad, interoperabilidad y reusabilidad. Esta aproximación se denomina es la Arquitectura Dirigida por Modelos.

2.5. Arquitectura Dirigida por Modelos (Model Driven Architecture - MDA).

Como lo plantea Montenegro *et al.* (2011), OMG (2019) y OMG (2014a), la arquitectura dirigida por modelos – MDA, como propuesta del Object Management Group (OMG), se orienta en reconocer que la interoperabilidad es fundamental, por lo cual el desarrollo de modelos permite la generación de otros modelos que luego al ser juntados, proveerán la solución a todo un sistema e independiza el desarrollo de las tecnologías empleadas.

MDA resuelve los retos de los sistemas actuales altamente conectados y constantemente cambiantes, tanto en reglas de negocio como en tecnología, proponiendo un marco de trabajo para una arquitectura que asegura (Armas, 2012):

- Portabilidad, aumentando el re-uso de las aplicaciones y reduciendo el costo y complejidad del desarrollo y administración de las aplicaciones.
- Interoperabilidad entre plataformas, usando métodos rigurosos para garantizar que los estándares basados en implementaciones de tecnologías múltiples, tengan todos idénticas reglas de negocio.

- Independencia de plataforma, reduciendo el tiempo, costo y complejidad asociada con aplicaciones desplegadas en diferentes tecnologías.
- Especificidad del dominio, a través de modelos específicos del dominio, que permiten implementaciones rápidas de aplicaciones nuevas, en una industria específica sobre diversas plataformas.
- Productividad, permitiendo a los desarrolladores, diseñadores y administradores de sistemas usar lenguajes y conceptos con los que se sienten cómodos, facilitando la comunicación e integración transparente entre los equipos de trabajo.

Dentro de los conceptos que contextualiza MDA se encuentran:

- Metametamodelo: especificación que define un lenguaje abstracto para especificar, construir y manejar elementos comunes a cualquier metamodelo.
- Metamodelo: herramientas que permite la creación de un modelo,
- Modelo: descripción de uno o varios elementos del dominio o mundo real.

Al interior de la arquitectura de modelado se plantean los siguientes niveles:

- M0 – Datos: información que maneja el sistema. Corresponden a los elementos del mundo real.
- M1 - modelo: descripción o especificación de un sistema realizado con un lenguaje determinado. Representación abstracta de un sistema o parte de él. Como lo plantea OMG (2019), un modelo es una representación selectiva de algún sistema cuya forma y contenido se eligen en función de un conjunto específico de inquietudes. El modelo está relacionado con el sistema mediante un mapeo explícito o implícito. Por ejemplo, podría ser un modelo UML (lenguaje unificado de modelado) de alguna parte de una aplicación, o un modelo de proceso comercial, o un modelo de simulación.
- M2 - Metamodelo: definición de los constructores y reglas necesarias para definir la semántica de los modelos. Contempla un lenguaje de modelado general como UML o específicos (DSL). Como lo plantea OMG (2019), un metamodelo especifica la sintaxis abstracta de un lenguaje de modelado y, como tal, es un tipo especial de modelo. Se puede entender como la especificación del conjunto de todos los modelos posibles expresados en ese lenguaje de modelado. Cada modelo debe cumplir con su metamodelo (s).
- M3 - Meta – metamodelo: definición de un lenguaje de modelado para metamodelos. Hay dos meta-metamodelos (Montenegro et al., 2011; García Díaz, 2012):

    - MOF (Meta Object Facility)
    - EBNF (Extended Backus-Naur Form).

La idea de generar estos niveles de abstracción tan altos es proveer un mecanismo común que permita, a través de transformación de un modelo a otro, la interoperabilidad de los sistemas.

En cuanto a los niveles de abstracción, se especifican tres modelos básicos de un sistema, que corresponden a (Ruiz, 2012; OMG-2, 2014):

- Computation Independent Model (CIM): se centra en el entorno y los procesos de negocio alrededor del sistema, sin contener información alguna respecto de la estructura del mismo.
- Platform Independent Model (PIM): describe el sistema sin entrar en detalles tecnológicos.
- Platform Specific Model (PSM): incluye información relativa a cómo el sistema va a utilizar la tecnología o plataforma necesaria. Este modelo combina las especificaciones del PIM con respecto a un modelo particular de plataforma (Java, Microsoft .NET, etc.).

En resumen, como lo describe Siegel (2014), MDA facilita un entendimiento y / o consenso común a través de tres maneras:

- Proporcionar términos, iconos y anotaciones bien definidos que ayuden a una comprensión común de un área temática.
- Proporcionar la base para que los modelos como datos semánticos sean administrados, versionados y compartidos.
- Proporcionar bibliotecas de modelos (activos) reutilizables, como vocabularios y reglas comunes, procesos reutilizables, modelos de objetos comerciales o patrones de diseño arquitectónico. Los modelos también se convierten en parte de la “memoria corporativa” de una organización, capturando los requisitos, el diseño y la intención de su organización, procesos, información y servicios.

Razones por las cuales, la implementación de MDA para mejorar la abstracción, disponibilidad y reutilización de los recursos educativos se convierte en una estrategia clave al proveer una comprensión común del dominio de conocimiento, además de proporcionar una articulación clara y flexible de los procesos inherentes al modelado de los recursos y su posterior enriquecimiento a través de parámetros establecidos por el modelo de datos enlazados.

CAPÍTULO III ANTECEDENTES DE LA APERTURA Y ESTADO DE LA CUESTIÓN

Como lo plantean Guzmán *et al.* (2011), la adopción del enfoque de datos enlazados ha llevado a la disponibilidad de grandes cantidades de datos públicos en diferentes dominios de conocimiento. Sin embargo, a pesar de que el enfoque de datos enlazados no se ha adoptado aún de manera amplia en el dominio educativo, se observan posibilidades de proporcionar una integración significativa, automatizada y personalizada de los diversos recursos educativos, por ejemplo, los formales (repositorios de datos) e informales (redes sociales) que se encuentran en la Web.

Si bien, la referenciación y el acceso de REA en la Web se han convertido en práctica común, el foco ha estado dedicado a los repositorios centralizados de recursos educativos. Sin embargo, la interoperabilidad de datos es aún más crucial, dado que el intercambio de recursos en línea podría hacerse independiente de su fuente, sea desde un repositorio de objetos de aprendizaje, un Learning Management System - LMS, redes sociales, Flickr, YouTube, SlideShare, etc. El problema es que estos tipos de intercambios requieren de formatos normalizados. El uso de RDF, en particular, ha ayudado a resolver este problema y permitir la referenciación y la integración de información de las diversas plataformas.

El desarrollo de los recursos educativos que soportan los procesos de aprendizaje ha venido creciendo, de tal forma que se encuentran recursos de diferentes tipos, categorías, etc., además de su vinculación a procesos de inclusión, accesibilidad y usabilidad (Herrera-Cubides *et al.*, 2014). Estos recursos son contenidos educativos utilizados para la formación y aprendizaje, que se acompañan con un archivo descriptor de metadatos, para facilitar su acceso.

Pero, como lo describen Zuiderwijk *et al.* (2012), los modelos de metadatos desarrollados han venido sufriendo de diversas problemáticas tales como: i) su construcción es específica para un contexto, ii) se trabaja con un reducido número de propiedades, iii) carecen de la riqueza requerida, iv) no existe un acuerdo sobre qué metadatos enriquecer a través del enlace de datos, v) aunque para algunos datasets LOD se encuentran disponibles metadatos detallados, generalmente hay pocas e insuficientes formas de administrar metadatos. Adicionalmente, Zuiderwijk *et al.* (2012) señalan que agregar metadatos a menudo se ve como una actividad adicional que solo consume recursos, dado que aquellos que podrían beneficiarse de los metadatos no son las mismas personas u organizaciones que proporcionan los metadatos.

Por otro lado, las organizaciones que han abordado el proceso de enlace de datos, se han enfrentado a una variedad de problemas en las diferentes etapas del ciclo de vida de datos enlazados. A continuación se presentan los principales problemas identificados dentro del contexto de los datos abiertos y datos abiertos enlazados, con base en procesos de experimentación realizados dentro de la presente investigación, así como de una revisión y análisis de literatura realizado con el fin de explorar el estado de la Web de los Datos.

3.1. Antecedentes de apertura y enlace de datos.

3.1.1. Fuentes de Datos Abiertas.

Como parte de los antecedentes de esta investigación se examinó un grupo de fuentes de datos denominadas fuentes abiertas. De acuerdo con Carcaño (2018), las fuentes abiertas son fuentes públicas e independientes de si su contenido es comercializado o libre. Para evaluar las fuentes abiertas, en estudio realizado de manera previa (Herrera-Cubides *et al.*, 2020), se exploraron 51 fuentes de datos asociadas a: noticias, académicas, investigativas, repositorios de recursos educativos, entre otros, basados en publicación de recursos Open-source Intelligence (OSINT)

en el área de la diseminación de investigación, como en la producción de recursos educativos. Dentro de las fuentes consultadas se tomaron en cuenta: Google Scholar Science Direct, Zbmath, Latindex, SciELO, Udemy, Open Yale, MIT Open Courseware, entre otros. Dentro de los hallazgos identificados en esta explotación de fuentes abiertas, relacionados con el tema de investigación, se encontraron:

- Un alto uso de materiales audio-visuales, tales como videos e imágenes, utilizados para la transferencia de conocimiento en la temática de OSINT.
- Una amplia discrepancia en el manejo de metadatos usados para describir los recursos provistos sobre OSINT.
- Limitada cantidad de metadatos proporcionados por las fuentes abiertas para describir los recursos, restringiendo posibles comparaciones de modelos de datos o vocabularios contralados usados para describir los recursos.
- Una gran mayoría de fuentes abiertas han usado datos elementales para describir sus recursos, acorde a requerimientos propios y específicos, dejando de lado metadatos importantes para complementar la descripción de los recursos.
- Aunque una gran mayoría de recursos OSINT han sido usados en procesos de entrenamiento sobre aspectos como la tecnología, su uso o sus aplicaciones, éstos no son catalogados como recursos educativos.
- Los cursos online masivos y abiertos (Massive Online Open Courses – MOOCs) identificados no han conservado un esquema de metadatos común, ni poseen metadatos necesarios para complementar la información básica que un recurso educativo debería tener, tales como datos completos de procedencia, contexto educativo, competencias, etc.

Estos hallazgos permitieron identificar como necesidad primordial la implementación y uso de un modelo de datos más completo para describir recursos educativos. Dicha descripción debe estar orientada más allá de las necesidades particulares, además de proveer información de calidad, suficiente y oportuna, identificada como clave en la descripción de recursos educativos. Adicionalmente, tanto la estructura de descripción (el modelo de datos) como su contenido, deben ser cada vez más disponibles y reutilizables por los usuarios. Por último, este tipo de recursos deben ser susceptibles para enriquecer como para ser enriquecidos con información enlazada a partir de su descripción.

3.1.2. Disponibilidad y Reutilización.

Para este caso se realizó una exploración de datos sobre el dominio gubernamental, como base de análisis, teniendo en cuenta sus avances en procesos de apertura y transparencia de datos. Considerando lo expuesto por Abella *et al.* (2018), quienes argumentan que el modelo de datos enlazados fija cinco niveles de calidad (cinco estrellas), para este experimento fueron considerados dos referentes relevantes:

a) Disponibilidad (Niveles 1, 2 y 3 del modelo de Linked Data): Apoyado en los principios de datos abiertos (Tauberer, 2014) y en los análisis realizados por Riley (2017) y Atenas y Havemann (2013), fueron identificadas como variables objeto de este estudio las siguientes: Dominio de Publicación, Licenciamiento de Recursos, Formato de Publicación y Actualización de las publicaciones.
b) Capacidad de reutilización (Niveles 4 y 5 del modelo de Linked Data): Haciendo uso de la estructura RDF proporcionada por los datasets, específicamente donde se define si los datasets se comportan como sujetos u objetos de vinculación; este experimento se orientó en analizar en primer lugar si los datasets manejaban recursos enlazados, y en segundo lugar, si esta información de reutilización era registrada sobre los datasets publicados en la Web.

Con base en las variables definidas, se exploraron 226.393 datasets, provenientes de 40 repositorios que usaban Comprehensive Knowledge Archive Network - Ckan (Ckan, 2020) como plataforma de gestión de datos. Dentro de los hallazgos obtenidos se identificaron:

- La explotación de datos enlazados en el dominio gubernamental, permitió identificar un conjunto de problemáticas en las diferentes etapas de su ciclo de vida. Para el caso de la disponibilidad de los recursos enlazados – niveles 1 al 3, las variables exploradas evidenciaron que aunque han habido esfuerzos en publicación de datos (nivel 1), éstos manifestaron diferentes barreras y limitantes como por ejemplo:

  - No se disponía de un licenciamiento para una porción de datasets, o en su defecto, éste licenciamiento era demasiado específico. Dado que a través de la licencia se manifiestan diferentes restricciones, derechos y usos sobre los datos; la falta de una declaración de derechos de autor no autoriza la reutilización de los datos. Adicionalmente, la ausencia de licenciamiento tampoco permitiría verificar los requerimientos de atribución y de Share-Alike de los datos.
  - El formato de valores separados por comas - CSV ha sido uno de los formatos más usados para la publicación de datos en las instancias consultadas. Aunque CSV, como formato no propietario, se ajusta a los requerimientos del esquema de datos enlazados, la ausencia de formatos RDF restringe el alcance de los datos enlazados. Por otro lado, se observó el uso de una multiplicidad de formatos de datos, lo que dificulta su interoperabilidad.
  - Aunque se identificaron datasets que se habían actualizado en periodos de tiempo inferiores a 6 meses, se determinó que estas actualizaciones respondían a que solo un recurso dentro del dataset había sido actualizado. Por otro lado, se encontraron datasets que distaban más de 2 años en fechas de actualización. Estos rangos de periodos de tiempos tan largos en las actualizaciones realizadas, afectan el criterio de oportunidad de los datos para una adecuada reutilización y toma de decisiones basadas en sus contenidos.
  - Se identificó una falta de uniformidad de las denominaciones usadas por los datasets para referenciar el dominio de conocimiento al cual pertenecían. Problemática generada por abstracciones particulares del dominio de conocimiento representado.
  - Una gran proporción de datasets no tenían datos de procedencia, lo que incide en la construcción de estrategias para la generación de confianza en los usuarios. En general, la omisión de información relevante en los metadatos dificulta determinar si el dataset encaja dentro del propósito de información requerido por el usuario.
  - La descripción de los recursos provistos por los datasets difería de una instancia a otra. Problemática generada por abstracciones diferentes y particulares de las entidades del dominio representado, además de una limitada expresividad de los vocabularios utilizados.

- Para el caso de la capacidad de reutilización de los recursos enlazados, las instancias consultadas además de proporcionar datasets identificables, ofrecían servicios de consulta a estos datasets. Por otro lado, los datasets consultados contaban con URIs que vinculaban información a sus tripletas. Sin embargo, se identificó que a pesar de que la estructura RDF del dataset permitía registrar el comportamiento como sujeto u objeto de enlace, este comportamiento no era registrado. Problemática generada por errores humanos o por la falta de conocimiento detallado de la estructura de los datasets.

Estos hallazgos han afianzado la hipótesis del declive que presenta actualmente la Web de los Datos, dado que la calidad de los metadatos publicados manifiesta diferentes problemas asociados con aspectos como por ejemplo completitud, disponibilidad, confianza, credibilidad. Factores que han sido identificados por autores como Ruckhaus *et al.* (2014), Behkamal *et al.*

(2014), Mendes (2012), Király (2015), Farias-Lóscio *et al.* (2019), Zaveri *et al.* (2012), IGI (2018), Herrera-Cubides *et al.* (2018a), entre otros. Problemáticas que han ido en detrimento tanto de la información almacenada en los dataset, como en la información obtenida de las consultas realizadas, considerando que la calidad de los datos es crucial cuando se trata de tomar decisiones de largo alcance basadas en los resultados de consultar múltiples datasets (Zaveri *et al.*, 2012). Hallazgos que también se han visto reflejados en índices como el Open Data Barometer (WWWF, 2019) y el Open Data Maturity (EDP, 2019), los cuales han mostrado que existen múltiples países donde la apertura de datos aún presenta muchos problemas en su implementación.

Por lo tanto, es muy importante abordar los problemas de calidad de datos en una escala Web como parte de la pila de tecnología de Web Semántica central. La suma de estos factores ha generado una serie de barreras que han restado importancia de forma gradual la evolución de la Web Semántica, encontrando en la actualidad datasets publicados con URL's que no son alcanzables, o que simplemente han optado por trabajar con formatos fáciles de gestionar como los CSV. Derivado de ello, el mundo de los datos se ha centrado en el tratamiento de grandes volúmenes de datos, y en buscar estrategias que permitan compartir datos sin entrar a fondo en el tratamiento de los mismos.

Adicionalmente, con base en las falencias detectadas a través de la ejecución de estos experimentos, tales como el uso de diferentes modelos de datos para describir los recursos, la falta de uniformidad en el tratamiento de las propiedades, la falta de apertura de los modelos de datos, entre otros factores, se visualizó que, además del requerimiento de incluir elementos de calidad de datos a los procesos de abstracción, se hizo evidente la necesidad de estructurar una estrategia que mejore la disponibilidad y accesibilidad de los recursos. En primer lugar se visualizó la necesidad de plantear un modelo que permee las descripciones de los recursos, además de saltar la barda de las descripciones particulares y ajustadas a contextos propios, planteando un esquema general, consistente y que respondiera de forma generalizada al dominio de conocimiento. En segundo lugar, se evidenció la necesidad de que este modelo de datos abierto, debía proveer una dinámica que facilite tanto la gestión de sus propiedades como su aplicación hacia otros dominios de conocimiento. Derivado de ello, se vio en los metamodelos, una alternativa viable que además de permitir la construcción de un modelo para la descripción de recursos, brinda elementos de flexibilidad, interoperabilidad y ampliación a otros dominios a través del trabajo con modelos.

3.2. Estado de la Cuestión.

3.2.1. Plataformas para la Gestión de Datos.

Para la gestión de colecciones de datos se encuentran diferentes plataformas, algunas de ellas gratuitas, las cuales prestan diferentes servicios y funcionalidades. Sin embargo, es prudente considerar la compatibilidad que estas plataformas ofrezcan versus los estándares a trabajar en los proyectos de explotación de datos adelantados. Dentro de estas plataformas se encuentran:

- ***Comprehensive Knowledge Archive Network – CKAN*** *(Ckan, 2020)*: Utilizada por gobiernos, instituciones de investigación y otras organizaciones para la gestión de sus datos. Permite búsqueda y facetado, pre visualización a través de mapas, gráficos y tablas.
- ***Drupal-based open data portal and catalog – DKAN*** *(Drupal, 2020)*: plataforma Open Source basada en Drupal con funciones de catalogación, edición y visualización que permite a los Gobiernos, las organizaciones no gubernamentales - ONGs y las Universidades publicar datos fácilmente.

- ***Junar*** *(Junar, 2020)****.*** Plataforma propietaria de Datos Abiertos, con servicios de recolección, publicación y análisis de datos, que facilita la publicación de datos por parte de gobiernos, empresas u otras organizaciones.
- ***Plataforma de Gobierno Abierto OGoov*** *(*OGoov, 2020*)*. Ofrece una serie de funcionalidades combinables entre si según la orientación o iniciativas relacionadas con el gobierno abierto que se deseen priorizar: Datos Abiertos, Transparencia y Participación.
- ***Dspace*** *(*Lyrasis, 2020a*)****.*** Plataforma abierta para la preservación y acceso a todo tipo de contenido digital, incluido texto, imágenes, imágenes en movimiento, mpegs y conjuntos de datos.

En general, existen plataformas para la gestión de datos que se orientan netamente a la preservación y consulta de la información. Otras plataformas han incursionado en el mapeo de los metadatos a un modelo de datos RDF, buscando aportar en la reutilización de los recursos almacenados.

3.2.2. Catálogos de Datos Abiertos.

A través del uso de instancias de diferentes plataformas de gestión de datos, las organizaciones han implementado estrategias para la conservación, el acceso, la consulta de documentos y datos generados en su quehacer cotidiano. En cuanto a este tema, se identifican:

Catálogos de Datos Académicos: Catálogos de instituciones de educación que gestionan tanto los datos producidos por los procesos de investigación y creación académica o científica, como su acervo bibliográfico. La Tabla 11 muestra algunos ejemplos de estos catálogos.

Tabla 11. Catálogos de Datos Académicos.

Nombre	Descripción	URL
BID Recursos Educativos Abiertos	Cursos y materiales abiertos del área económica y social de Latinoamericana y del Caribe.	https://indesvirtual.iadb.org/course/index.php?categoryid=18&lang=es
Ideals - Illinois Digital Environment for Access to Learning and Scholarship	Investigación de la University of Illinois at Urbana-Champaing.	https://www.ideals.illinois.edu/
RIUD - Repositorio Institucional de la Universidad Francisco José de Caldas	Objetos digitales de la Universidad Distrital Francisco José de Caldas.	http://repository.udistrital.edu.co/
PANGAEA (Data Publisher for Earth & Environmental Science)	Biblioteca de datos digitales para la ciencia del sistema Tierra.	https://www.pangaea.de/
Purdue University Research Repository (PURR)	Datos de investigación de la Purdue University y sus colaboradores.	https://purr.purdue.edu/
Repositorio Universidad Javeriana	Repositorio de la Pontificia Universidad Javeriana.	https://repository.javeriana.edu.co/
DataCite	Localiza, identifica y cita datos de investigación.	https://datacite.org/
UK Data Archive	Colección de datos digitales de investigación de ciencias sociales y humanidades, del Reino Unido.	https://www.data-archive.ac.uk/

Catálogo de Datos Institucionales: Catálogos de instituciones públicas y privadas, para gestionar datos tanto producto de su trabajo como los exigidos por jurisprudencias que los obligan a abrir sus datos. En otro escenario, corresponden a repositorios donde diferentes entidades almacenan y gestionan sus datos. La Tabla 12 muestra algunos ejemplos de estos catálogos.

Tabla 12. Catálogos de Datos Institucionales.

Nombre	Descripción	URL
Observatorio ESVI-AL	Buscador de contenidos abiertos Serendipity, vía navegación facetada.	http://www.esvial.org/observatorio/?page_id=266
Library of Congress	Acceso a estándares y vocabularios usados por la Biblioteca del Congreso.	http://id.loc.gov/
Datahub	Repositorio abierto que gestiona datos de diferentes organizaciones.	https://datahub.io/
Catálogo de Datos Abiertos - Uruguay	Repositorio de datos abiertos del Uruguay.	https://www.gub.uy/datos-abiertos
UTPL – RiUTPL	Repositorio Institucional de la UTPL – RiUTPL	https://www.utpl.edu.ec/riutpl/
RRAAE	Red de Repositorios de Acceso Abierto del Ecuador – RRAAE.	http://rraae.org.ec/
Datos Abiertos Colombia	Portal Colombiano de datos abiertos.	https://www.datos.gov.co/
European Union Open Data Portal	Gestiona datos elaborados por las instituciones y otros organismos de la Unión Europea.	http://data.europa.eu/euodp/en/data
USPTO Open Data Portal	Portal de datos abiertos de la United States Patent and Trademark Office.	https://developer.uspto.gov/
Dados Abertos Legislativos (Brasil)	Catálogo de Datos Abiertos del Senado Federal de Brasil.	https://www12.senado.leg.br/dados-abertos
Datos El Salvador	Portal de datos abiertos de El Salvador.	http://www.datoselsalvador.org/
Datos.gov.es	Catálogo de iniciativas de Datos Públicos en España.	http://datos.gob.es/es/iniciativas
Escucha. Participa. Conversa Gobierno Abierto de Navarra	Catálogo de iniciativas de Datos Públicos del Gobierno de Navarra.	http://www.gobiernoabierto.navarra.es/es/open-data/datos/catalogo-datasets
RosarioDatos	Catálogo de la Videoteca del Centro Audiovisual Rosario.	http://datos.rosario.gob.ar/dataset/cat%C3%A1logo-de-la-videoteca-del-centro-audiovisual-rosario

Catálogos de Recursos Educativos: Catálogos gestionados por instituciones académicas públicas y privadas quienes ofrecen recursos educativos en diferentes niveles de formación, bajo acceso público o de manera restringida. Es importante resaltar que la mayoría de estos repositorios ofrecen un enlace al recurso en mención y metadatos básicos para describir el citado recurso. La Tabla 13 muestra algunos de estos ejemplos de estos catálogos.

Tabla 13. Catálogos de Recursos Educativos.

Nombre	Descripción	URL
Merlot (Liderado por California State University)	Gestiona materiales y estrategias de aprendizaje. Cuenta con un portal de recursos educativos multimedia para el aprendizaje y la enseñanza en línea.	http://www.merlot.org/merlot/
Utpl Linked Open Data	Presenta el resultado del estudio y aplicación de tecnología Linked Data	http://data.utpl.edu.ec/?q=es/node/16

Nombre	Descripción	URL
	para el acceso en línea a datos Universitarios Abiertos y Enlazados como una de las principales iniciativas de la región de la Loja.	
Open Learning Initiative	Iniciativa de Carnegie Mellon University, que ofrece cursos introductorios basados en nuevas teorías de aprendizaje, desarrollados por científicos de la Universidad.	https://www.cmu.edu/
Open.Michigan	Repositorio de productos de investigación, enseñanza y trabajo creativo, de la Universidad de Michigan.	https://open.umich.edu/
MIT Open Course Ware	Gestiona materiales pedagógicos de los cursos de The Massachusetts Institute of Technology.	https://ocw.mit.edu/index.htm
Open Yale Courses	Proyecto de Yale University que gestiona cursos introductorios de pregrado de la universidad impartidos por sus distinguidos profesores y académicos.	https://oyc.yale.edu/

Otros tipos de Catálogos: Dentro de estos catálogos se encuentran diferentes plataformas que ofrecen herramientas o material complementario para los procesos de formación, a través de servicios o APIs. La Tabla 14 muestra otras plataformas usadas en el proceso de formación.

Tabla 14. Plataformas usadas para la formación.

Nombre	Descripción	URL
Google Dataset Search	Sitio web para la búsqueda de datasets.	https://toolbox.google.com/datasetsearch
TED Ideas worth spreading	Congresos (TED Conference) y charlas (TED Talks) en diferentes dominios de conocimiento.	https://www.ted.com/#/
Youtube Education	Provee vídeos educativos, conferencias académicas, discursos, lecciones y cursos ofrecidos por universidades de amplio reconocimiento.	https://www.youtube.com/channel/UCSSlekSYRoyQo8uQGHvq4qQ
Khan Academy	Proporciona educación gratuita de calidad para cualquier persona en cualquier lugar.	https://www.khanacademy.org/
iTunes U	Contenido educativo gratuito de instituciones educativas líderes del mundo.	https://apps.apple.com/app/itunes-u/id490217893?ls=1
Edu Tube	Plataforma web de videos educativos en múltiples áreas de interés.	https://www.edutube.org/

En general, los catálogos de datos ofrecen servicios de explotación de sus datos. Los datasets almacenados en estos catálogos están descritos y categorizados conforme a estándares propios, pero en algunas ocasiones los archivos referenciados no se encuentran almacenados en estas plataformas. Las entidades que participan en estos catálogos cuentan con mecanismos manuales o de federación para publicar datasets y actualizar la información ya publicada sobre ellos. Por último, una gran mayoría de estos catálogos de datos gestionan recursos almacenados, pero no ofrecen estrategias de enlace y enriquecimiento de sus metadatos, no gestionan información expresada en formato RDF ni ofrecen servicios de consulta a través de puntos de acceso SPARQL.

3.2.3. Implementaciones basadas en datos enlazados.

En cuanto a datos enlazados, existen variados ejemplos de iniciativas que han implementado estas tecnologías. Propuestas como Health Care and Life Sciences Interest Group (HCLS, 2019) quienes manejan el proyecto Linking Open Drug Data (LODD), enlazan diversas fuentes de datos de medicamentos (W3C, 2012). Es así como el proyecto LODD ha desarrollado diferentes datasets de evaluación (W3C, 2009) dentro de los cuales se encuentran:

- Chemical Entities of Biological Interest (ChEBI): diccionario gratuito de entidades moleculares centrado en compuestos químicos pequeños (EBI, 2019).
- DailyMed: listados de medicamentos autorizados por la Food and Drug Administration - FDA (NLM, 2019).
- DrugBank: Combina los datos de los medicamentos (por ejemplo, químicos, farmacológicos y farmacéuticos) con información del objeto integral del medicamento (por ejemplo: secuencia, estructura, etc.) (DrugBank, 2018).

En otros campos se identificaron esfuerzos en uso e implementación de datos enlazados tales como:

a. ***Medios***: British Broadcasting Corporation (BBC, 2020), provee datos acerca de episodios de radio y programas de televisión, que son interconectados con MusicBrainz (MusicBrainz, 2020) y DBpedia (DBpedia, 2020).
b. ***Publicaciones***: The British National Bibliography (BNB, 2020), la Biblioteca Nacional de España (BNE, 2020), la Biblioteca Nacional de Francia (BnF, 2020), la Biblioteca Británica (Bl, 2020), la Biblioteca Nacional de Alemania (DNB, 2020), la Biblioteca del Congreso de los Estados Unidos (LOC, 2020).
c. ***Ciencia***: El proyecto Bio2RDF (Bio2RDF, 2020) enlaza datasets científicos usados en el desarrollo científico. Incluyendo UniProt (Fuente Universal de Proteínas), KEGG (Enciclopedia Kyoto de los genes y el genoma), CAS (Servicio de Abstracción Química), PubMed y la Gene Ontology.
d. ***Datos Geográficos***: GeoNames (GeoNames, 2020), base de datos geográfica que publica datos enlazados de ubicaciones. LinkedGeoData (LinkedGeoData, 2018) publica datos enlazados de Open StreetMap. En Colombia se empezó a trabajar la información geográfica a través del seminario taller “Web Semántica, Ontologías y Linked Data: nuevas tendencias para la gestión de la información geográfica”, promovido por la oficina de Infraestructura de Datos Espaciales para el Distrito Capital -IDECA (IDECA, 2015).
e. ***Contenido generado por el usuario***: Plataformas como Flickr, los exportadores SIOC de WordPress, gestores de contenido como Drupal y los bulletin boards de phpBB implementan estrategias de vinculación de datos. Zemanta (Zemanta, 2019) proporciona herramientas para enriquecimiento semi automatizado de blog posts con enlaces a nivel de datos a DBpedia, MusicBrainz, y Semantic Crunch-Base (Crunchbase, 2020).

3.2.4. Estudios Previos sobre Problemáticas de la Web de los Datos.

Con relación al estado de la Web de los Datos, estudios realizados por McCrae (2020), Hogan *et al.* (2012), Herrera-Cubides *et al.* (2016), Unión Europea (2020), GODI (2019), COL (2015), Abella *et al.* (2018), Smith-Yoshimura (2018), Herrera-Cubides *et al.* (2017), LOD2 (2019), Klímek *et al.* (2018) y Dietze *et al.* (2012), entre otros, han permitido identificar diversas problemáticas en este contexto. A continuación se relacionan las más relevantes:

- Pocos datasets proporcionan información de licencias como parte de sus metadatos, y otros datasets usan licencias muy específicas o diseñadas para su contexto local. Lo cual evidencia una necesidad palpable de (i) una definición de licenciamiento apropiado, y (ii) un acuerdo de licenciamiento común.

- Metadatos poco legibles por el ser humano o con dificultad para llevar a cabo procesos de extracción.
- Al interior de los datasets no se evidencia un extendido uso de enlaces hacia diferentes dominios externos. Los dominios de conocimiento que presentan enlaces a dominios RDF externos, suelen ser plataformas colaborativas que ofrecen algún tipo de servicio centralizado. Los enlaces externos estables, construidos entre proveedores de datos remotos, son fundamentales para realizar la visión de datos enlazados. Sin embargo, la creación de enlaces de alta calidad a proveedores de RDF externos suele ser una tarea desafiante para los editores.
- Falta de actualización de los datasets.
- Dumps y puntos de acceso SPARQL con errores o inestables. No se cuenta con direcciones Web de descarga, y se presenta baja fiabilidad del servicio.
- No existen herramientas sencillas para apoyar el proceso de consumo de datos enlazados. Adicionalmente se presentan problemas para extender los frameworks a repositorios adicionales o articular las API de las capas de integración de datos y servicios, con aplicaciones adicionales de terceros. En la actualidad no existe una herramienta que pueda ser utilizada para apoyar todo el proceso de consumo de datos enlazados. Derivado de ello, los consumidores a menudo siguen consumiendo datos nivel 3, como los archivos CSV, más cómodos de trabajar.
- Los editores de datos tienen dificultades para convencer a sus superiores de aceptar el esfuerzo adicional requerido para que sus datos alcancen las 5 estrellas del modelo de Linked Data, dado que por ejemplo los datos en nivel 3 ofrecen un conjunto de bondades para su explotación demandando menos esfuerzo, y por consiguiente inversión.
- La curva de aprendizaje para la publicación de datos es pronunciada para el personal. Adicionalmente se presentan inconsistencias de datos legados, problemas en la selección apropiada de ontologías para representar los datos, poca documentación acerca de cómo construir los sistemas, etc.
- Problemas de emparejamiento, desambigüedad y alineación de datos de origen y recursos de datos enlazados. Lo que se publica como datos enlazados no siempre es reutilizable o carece de URI, los mapeos de vocabulario no son adecuados, se presenta dificultad para entender cómo se estructuran los datos antes de usarlos. Finalmente, se presenta una deficiente selección de ontologías para representar los datos.

Adicionalmente, como lo plantean Dietze *et al.* (2012), si bien los trabajos de investigación han abordado varios desafíos como la interoperabilidad de los metadatos, el descubrimiento de servicios o la mediación de datos; aún persisten problemas asociados a los requerimientos tecnológicos tales como: extender los frameworks a repositorios adicionales o articular las API de las capas de integración de datos y servicios mediante aplicaciones adicionales de terceros.

Por su parte, Abella *et al.* (2018) plantearon un conjunto de ventajas e inconvenientes relacionados con el esquema de 5 estrellas para el enlace de datos (Tabla 15).

Tabla 15. Análisis de Esquema de Vinculación. Fuente: Abella *et al.* (2018).

Estrellas	Significado y ejemplos	Ventajas (V) e inconvenientes (I)
1	Datos disponibles en la Red independientemente del formato utilizado. Ej.: se publica como una tabla dentro del propio portal.	V: Datos accesibles en un documento, con cualquier licencia abierta. I: Dificultad para extraer datos del documento.
2	Datos publicados en la Web en un formato estructurado. Ej.: los datos son exportables en formato Excel	V: Datos accesibles de forma estructurada. I: Se requiere un software propietario para extraer datos.

Estrellas	Significado y ejemplos	Ventajas (V) e inconvenientes (I)
3	Datos publicados bajo un formato estructurado no propietario. Ej.: los datos son exportables en formato CSV.	V: Datos disponibles y utilizables en la Web. I: Datos no integrados en la Web.
4	Datos identificados mediante URIs de manera que sean fácilmente interpretables. Ej.: los datos son exportables en formato HTML y otros.	V: Los datos tienen una URI y pueden compartirse en la Web. I: Los datos no están enlazados con otros datos.
5	Datos enlazados con otros datos de manera que se encuentran contextualizados. Ej.: datos exportables en una variedad de formatos y accedidos a través de enlaces a diferentes localizaciones.	V: Datos integrados y relacionados con otros datos en la Web. I: Sólo se relacionan con datos con los que se ha preestablecido enlaces.

En términos generales, los referentes analizados han evidenciado que los problemas de calidad de los datos son un reto persistente en los procesos de enlace de datos. Estos problemas se pueden observar tanto a nivel de principios de Linked Data, como en la abstracción de los diferentes aspectos que describen una identidad o recurso direccionable, y que posteriormente serán representados en RDF, como modelo de datos adaptado a la arquitectura Web. Por otro lado, se han identificado estudios que abordan las dimensiones de calidad de datos enlazados como alternativas para solventar los problemas a los que se enfrenta la Web de los Datos, pero estos estudios se han remitido al planteamiento teórico de posibles métricas para llevar a cabo un análisis sobre el estado actual de la calidad de los datos enlazados.

En cuanto a los diferentes problemas de calidad de datos que afronta el modelo de Linked Data, el Anexo 1 identifica desarrollos para diversas áreas y dominios, donde se han aplicado estrategias para abordar escenarios puntuales. Como resultado de las fuentes revisadas, se puede resaltar:

a) La mayoría de las propuestas de datos enlazados han planteado procesos de recolección de metadatos, con almacenamientos centralizados.
b) Se ha desarrollado la evaluación de las dimensiones de calidad de forma independiente, en su mayoría orientadas a factores como ranking, interlinking, procedencia, relevancia. Algunas de las iniciativas expuestas han buscado mejorar las consultas SPARQL adicionando elementos como justificaciones, anotaciones, etc.
c) La mayoría de los referentes expuestos han trabajado sobre dominios de conocimiento de tipo estadísticos, social media, biológico, médico, etc. Dominios que poseen una amplia estandarización y disponibilidad de vocabularios controlados.
d) En cuanto a las iniciativas identificadas en el dominio educativo, éstas se enfocaron en procesos de recolección y centralización de metadatos, dejando de lado aspectos como el enriquecimiento de los metadatos, la implementación de puntos de acceso SPARQL o la aplicación de dimensiones de calidad de los datos enlazados para mejorar la abstracción de los datos.

Esta revisión permitió identificar que se ha desarrollado un trabajo bastante amplio sobre el manejo y mejora de las instancias de datos publicadas, dejando de lado las mejoras que pueden ser incluidas sobre la abstracción y modelamiento de las entidades del dominio de conocimiento trabajado. De igual forma, los esfuerzos realizados se han enfocado sobre dominios de conocimiento bastante estructurados que contienen vocabularios controlados para manejar las abstracciones realizadas.

3.2.5. Retos de la Web de los Datos.

Para permitir que las aplicaciones Linked Data descubran datasets, además de facilitar la integración de datos de diferentes fuentes, se deben cumplir un conjunto de mejores prácticas tales como las expuestas en Schmachtenberg *et al.* (2014):

- Establecer enlaces RDF que apunten a instancias en otras fuentes de datos.
- Reutilizar términos de vocabularios de uso generalizado.
- Hacer que las definiciones de términos propietarios sean desreferenciables.
- Vincular términos de vocabularios con términos de otros vocabularios.
- Publicar información de procedencia legible por las máquinas.
- Publicar información de licencias legible por las máquinas.
- Publicar metadatos sobre métodos de acceso alternativos.

A pesar de la existencia de estas buenas prácticas, el mundo de los datos enlazados se enfrenta a desafíos, los cuales en su mayoría se sustentan en la calidad de los datos. Retos que parten desde la abstracción de conceptos y relaciones, que se describen posteriormente en el modelo de datos planteado, hasta la falta de cumplimiento de las recomendaciones dadas por cada uno de los niveles del esquema de enlace de datos.

Diferentes estudios han demostrado que la mayoría de los datasets sufren de varios problemas de calidad de datos tales como problemas de abstracción, representación, inconsistencia o interoperabilidad. Estos problemas han obstaculizado significativamente la adopción de estos datasets en casos de uso particulares, afectando los resultados de las consultas a medida que la mala calidad se propaga en los datasets agregados (Zaveri, 2015; Zaveri *et al.*, 2012; Cappiello *et al.*, 2016; Song *et al.*, 2017). Por su parte, los portales de datos abiertos no han proporcionado metadatos en un formato suficientemente preciso para ser inmediatamente útiles (Holst y Höfig, 2013).

Como retos a trabajar en la Web de los Datos, Fayyaz *et al.* (2018), Nahhas *et al.* (2018) y Farias-Lóscio *et al.* (2019) han expuesto un compilado de retos a los que se enfrenta un usuario al explotar datos en la Web. Dentro de estos retos se encuentran situaciones referidas a cómo hacer disponible los metadatos, qué tipo de licenciamiento usar en los datasets, cómo generar confianza en los usuarios consumidores, qué formatos usar en la publicación de los datos, cómo mejorar la interoperabilidad de los datos a través de los vocabularios utilizados para describir las entidades, cómo proveer acceso a los datos, cómo adicionar valor a los datos publicados y enlazados, entre otros. Estos autores han identificado retos principales para abordar en el proceso de enlace de datos, desde los enfoques de disponibilidad de los datos y capacidad de reutilización, tales como:

- Manejar la apertura y enlace de datos como prioridad en las organizaciones (a nivel central como descentralizado), dado que ofrece muchas oportunidades de negocio a los usuarios consumidores. Por otro lado, debe proveerse recurso humano suficientemente capacitado, además de presupuesto necesario, para llevar a cabo la tarea de apertura y enlace de datos. Además, se debe proveer de una legislación de apertura de datos que se extienda a diferentes niveles de los gobiernos y de las organizaciones, sean estas públicas o privadas.
- Empoderar a los usuarios, incrementando su percepción sobre las posibilidades de reutilización de la información, además de ofrecerles herramientas y procedimientos ágiles y sencillos para ejecutar los procesos de apertura y enlace de datos.
- Mejorar el componente técnico a través de herramientas que permitan: alivianar las tareas de publicar y enlazar los datos, gestionar los metadatos de los recursos publicados, proveer servicios estables que le permitan a los usuarios explotar los datos de una forma más amigable.

- Mejorar la curva de aprendizaje en temas de estándares y vocabularios usados para la representación del conocimiento.
- Establecer y unificar flujos de publicación y enlace de datos, fáciles de abordar por los editores de datos y los usuarios consumidores.
- Sustentar el modelamiento de los datos en una adecuada abstracción, no solamente de las entidades y atributos, sino de las relaciones que entre ellas se establecen. Por ende, es vital una adecuada contextualización del dominio de conocimiento, además del uso de vocabularios adecuados y con una expresividad suficiente.
- Proveer formatos de datos RDF que suplan todos los requerimientos del modelo de datos enlazados, con el fin de complementar la disponibilidad y reutilización.
- Proveer estrategias de enlace de datos que permitan superar no solamente las dificultades actuales del linking, sino también la búsqueda de objetos para las tripletas RDF que contribuyan al enriquecimiento de los datos.
- Fortalecer la abstracción y diseño de los modelos de datos basado en requerimientos de calidad de datos enlazados, dado que la calidad de los datos enlazados se ha venido evaluando sobre las instancias de datos generadas.
- Avanzar hacia la apertura y reutilización de recursos, salvaguardando los debidos derechos, dado que se han construido diseños particulares de los objetos del mundo real, sumado a la falta de apertura de los modelos de datos y al uso de estándares propios.
- Disminuir las transformaciones sintácticas (lenguajes) y semánticas (significados) que se deban realizar para usar los datos.
- Establecer licenciamientos y derechos de autor: la promulgación de los derechos de publicación y la respectiva autorización de los datos enlazados, además de permitir la reutilización, permiten el uso de los datos de forma legal acatando las restricciones provistas para su manipulación.
- Diseñar y aplicar políticas y estrategias de actualización de datos, que permitan la reutilización de los datos, aprovechando la oportunidad de los mismos.

3.2.6. Estado del arte de datos abiertos en Colombia.

Los datos abiertos en Colombia se han venido trabajando desde hace cerca de 10 años, periodo en el cual a través del Decreto 2693 de 2012, se establecieron los lineamientos para la estrategia de Gobierno en Línea de la República de Colombia. El Gobierno Nacional creó el proyecto “Datos Abiertos”, bajo la estrategia de Gobierno en Línea. Con este proyecto, el Gobierno colombiano pretende promover la transparencia, el acceso a la información pública, la competitividad, el desarrollo económico, y la generación de impacto social a través de la apertura, la reutilización de los datos públicos, y el uso y apropiación de las TIC (Mintic, 2016). La iniciativa de datos abiertos busca que todas las entidades del sector público publiquen la información pertinente y de calidad, en formatos estructurados a disposición de los usuarios, para que pueda ser utilizada de diferentes maneras: generar informes, reportes, estadísticas, investigaciones, control social, oportunidades de negocio (ej. aplicaciones), entre otros temas.

Bajo este proyecto, se han desarrollado los siguientes procesos:

- Vincular a las entidades públicas del orden nacional y territorial, para que publiquen datos abiertos en el Catálogo de Datos Abiertos de Colombia. A 2020, este catálogo cuenta con 13.168 datasets abiertos disponibles para su uso, y ha vinculado a 1209 entidades.
- Se han construido instrumentos normativos que promueven la apertura de datos, como por ejemplo:

    - La Ley de Transparencia y Acceso a la Información (Ley 1712 de 2014), que establece procedimientos para garantizar el derecho de acceso a la información pública.

- Decreto Único Reglamentario del Sector TIC - 1078 de 2015, que define las obligaciones de la entidades públicas para la implementación del gobierno electrónico en Colombia, y en particular, define como componente estratégico TIC y Gobierno Abierto, dentro del cual ese establece la obligación de abrir, divulgar y promover la reutilización de datos abiertos.
- Guía para el uso y aprovechamiento de datos abiertos en Colombia (2016), y la Guía de datos abiertos en Colombia (2016): Dirigidas a las entidades sujeto de aplicación de la Ley 1712 de 2014 de transparencia y acceso a la información pública.
- Guía de estándares de calidad e interoperabilidad de los datos abiertos del Gobierno de Colombia (2016).

Atendiendo a las directrices del Gobierno Nacional, diferentes entidades nacionales territoriales y distritales, entre otros, han creados portales para compartir sus datos abiertos. La Tabla 16 muestra algunos ejemplos de estas iniciativas:

Tabla 16. Iniciativas colombianas para compartir datos.

Iniciativa	URL
Datos Abiertos - Bogotá	https://datosabiertos.bogota.gov.co/
Datos Abiertos – Colombia	https://www.datos.gov.co/
Centro de Memoria Histórica	http://centrodememoriahistorica.gov.co/observatorio/
Datos Abiertos en Medellín	https://geomedellin-m-medellin.opendata.arcgis.com/, http://medata.gov.co/
Colombia compra eficiente	https://www.colombiacompra.gov.co/transparencia/gestion-documental/datos-abiertos

El estado colombiano usa el estándar de metadatos Data Catalog Vocabulary – DCAT (W3C, 2020) para describir los datasets publicados en los catálogos públicos. Aunque las propuestas gubernamentales conservan una descripción medianamente uniforme de los datasets cargados, se observan problemas de calidad de datos abiertos, como por ejemplo completitud de la información que se diligencia.

Por otro lado, se han identificado iniciativas privadas que proveen datos abiertos para fines específicos. Algunas de estas iniciativas son:

- Datasketch (Datasketch, 2020): Portal de periodismo de investigación y de datos, que permite a periodistas, científicos de datos, científicos sociales y la ciudadanía en general, aprender y consultar sobre visualizaciones de datos, herramientas, software e investigaciones profundas sobre diversos temas coyunturales.
- Postdata (CO, 2020): Plataforma de datos abiertos habilitada por la Comisión de Regulación de Comunicaciones (CRC), para permitirle a la industria, academia, gobierno, usuarios y medios de comunicación, acceder fácilmente a los datos relevantes del sector de comunicaciones.
- Esri (Esri, 2019): Plataforma para el uso de datos arGIS.

3.2.7. Estado del Arte de datos enlazados en el Sector Educativo.

Estudios como los realizados por Navarrete y Luján (2015), Alyahibi y Alsharif (2019), Nahhas *et al.* (2018), Guy *et al.* (2014), Mouromtsev y d'Aquin (2016) y Pereira *et al.* (2018), han realizado un barrido de propuestas de enlace de datos sobre el dominio de conocimiento educativo, detectando que este dominio no es ajeno a las problemáticas de apertura y datos enlazados. Sus resultados evidencian que existen diferentes esfuerzos para aplicar los principios y tecnologías de Linked Data para tratar de afrontar problemas tales como la interoperabilidad de los datos y

recursos educativos, el acceso a los datos, enriquecimiento de material, entre otros, en torno a las tecnologías educativas.

Dentro de los desafíos en el campo educativo, para lograr la interoperabilidad de los REA y realizar procesos de enlace de datos, se han identificado:

a. Integrar datos distribuidos entre repositorios educativos heterogéneos;
b. Tratar con cambios continuos existentes en los repositorios, dado que el contexto cambia constantemente a lo largo de las actividades de aprendizaje;
c. Usar metadatos mediadores y transformadores que describen recursos;
d. Enriquecer e interconectar metadatos no estructurados;
e. Tratar los problemas de privacidad y propiedad con respecto a los datos educativos;
f. Disminuir el esfuerzo requerido para publicar datos enlazados, así como los costos computacionales involucrados en su consumo;
g. Mejorar la calidad de los datos, la actualización y la disponibilidad de los datasets enlazados con el fin de aportar en la confiabilidad y disponibilidad de los mismos;
h. Construir ontologías de dominio para diferentes dominios objetivo;
i. Diseñar herramientas y sistemas que permitan a los usuarios sin experiencia previa en tecnologías de Linked Data, contribuir a hacer el contenido;

Por otro lado, según Pereira *et al.* (2018) -citando a Dietze *et al.* (2009)-, dado el aumento significativo de los recursos educativos disponible en la Web, se han venido presentando desafíos para la reutilización tales como: la falta de calidad de los recursos, la diversidad de estándares y vocabularios de metadatos, la accesibilidad de los datos y el mantenimiento a menudo deficiente de las descripciones de metadatos, además de plantearse inquietudes con respecto a la confianza al intentar reutilizar datos y recursos de terceros, entre otros.

Este tipo de problemáticas influye tanto en la disponibilidad como en la reutilización que se pueda hacer de los recursos educativos abiertos, y por ende, limita su enriquecimiento a través de técnicas como Linked Data.

3.2.8. Estado del Arte de LOD en el Sector Educativo Colombiano.

En el "*Estudio del grado de preparación para la apertura de datos del sector educativo en Colombia*" (MinTic, 2015), se identificaron barreras (elementos que conviene mejorar o superar para ayudar a la iniciativa de datos abiertos), y oportunidades (espacios de posibilidad que surgen en torno a la dinámica educativa en el país) para la política de apertura y reutilización de datos. La Tabla 17 presenta un resumen de estas barreras y oportunidades.

Tabla 17. Oportunidades y Barreras de la Apertura de Datos en REA. Fuente: MinTic (2015).

Dentro de las Barreras se destacan:	**Dentro de las Oportunidades se destacan:**
• Dificultad de gestionar la diversidad de actores y de visiones. • Baja implementación de la política Open Data. • Incentivos perversos de los sistemas de medición. • Estricta aplicación de la normativa de protección de datos. • Faltan sistemas que relacionen la educación y el empleo. • Desequilibrio en dotación tecnológica y cobertura de comunicaciones. • Ausencia de un expediente educativo unificado.	• Diversidad de actores y complejidad del campo educación. • Existencia de una política nacional de apertura de datos. • Manejo de incentivos de los sistemas de medición. • Buenas prácticas locales. • Se han iniciado trabajos para una mejor relación con la información de empleo. • Programas intensivos de evaluación de la calidad educativa. • Renovación de sistemas de información.

Dentro de las Barreras se destacan:	Dentro de las Oportunidades se destacan:
• Baja calidad de los datos en origen. • Falta de estándares en la información educativa. • Dificultad para la sociedad civil de acceder a datos. • Carencias en la información sobre infraestructuras. • El mercado no conoce bien la problemática del sector educativo. • Carencia de mecanismos de colaboración para la mejora de datos.	• Existencia de aplicaciones a partir de datos. • Sistemas para la mejora del reporte de datos. • Muchos datos ya disponibles. • Riqueza de datos estadísticos. • Acercamiento a las organizaciones de la sociedad civil. • Presencia de una sociedad civil activa en educación.

Aunque este proyecto no prosperó, la especificación de metadatos aún se conserva y sirve como soporte para diferentes instituciones que quieran documentar sus recursos educativos. Por otro lado, diferentes instituciones, entre las cuales se cuentan bibliotecas universitarias, han adaptado LOM para describir sus recursos. Por otro lado, últimamente han orientado sus esfuerzos para alinearse con esquemas de datos propuestos por proyectos como LA Referencia (LAR, 2019), cuyo nodo en Colombia se denomina KUJANE (Figura 6). Este nodo está compuesto por el Ministerio de Ciencia Tecnología e Innovación - MinCiencias, Instituciones de Educación Superior colombianas públicas y privadas, Institutos o Centros de Investigación y otras organizaciones que tienen y manifiestan un compromiso expreso con el fomento a la producción del conocimiento en acceso abierto.

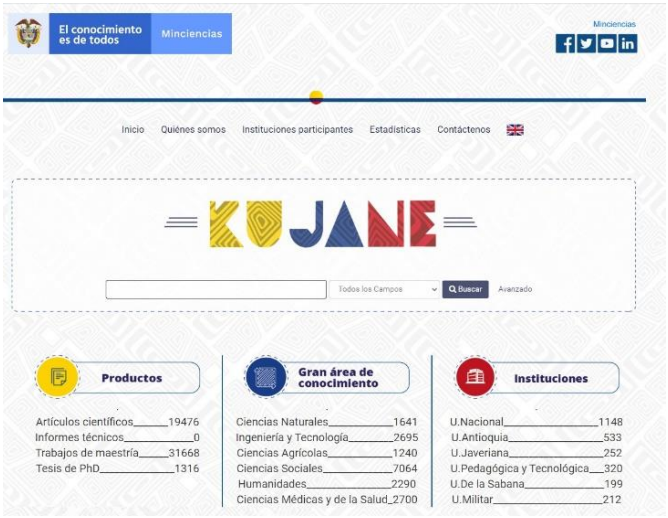


Figura 6. Nodo Kujane – Proyecto LA Referencia. Fuente: Kujane (2020).

En la actualidad, y contando con un motivamente bastante fuerte como lo es la pandemia de Covid-19, Colombia ha buscado fortalecer el portal "*Aprender Digital*" (MEN, 2020), a través de donaciones de recursos o de enlaces a recursos abiertos que diferentes instituciones han puesto a disposición del Ministerio de Educación Nacional (Figura 7).

Sin embargo, y aunque existen elementos básicos para enriquecer los recursos, este avance no se ha dado por diferentes circunstancias tales como la falta de políticas de apertura en campos diferentes al público, reglamentaciones laborales para la producción de recursos, etc.

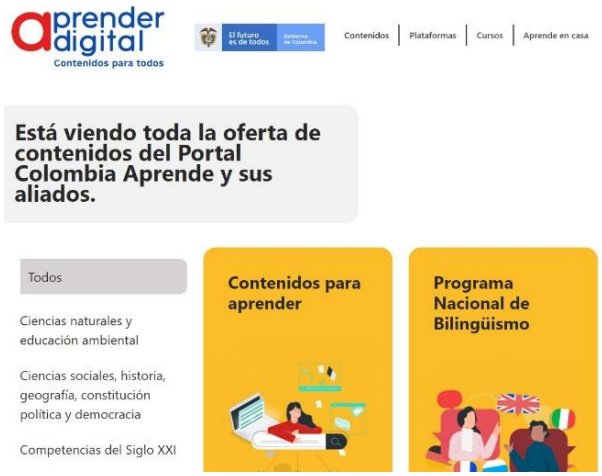


Figura 7. Portal Aprender Digital. Fuente: MEN (2020).

3.3. Conclusiones del Capítulo.

Los retos descritos enfocan su quehacer hacia los procesos de publicación y explotación de datos. Sin embargo, la calidad de los datos se ve afectada desde etapas previas en el ciclo de vida de datos enlazados. Por tal razón, y considerando que la mayoría de las investigaciones se han orientado en solucionar problemas sobre las instancias de los datos, el enfoque de esta investigación se orientó en abordar los retos presentes en las etapas previas a la generación de instancias. Retos dentro de las cuales se identifican:

- Mejorar la abstracción y modelado de datos: El modelamiento de los datos se sustenta en una adecuada abstracción no solamente de las entidades y atributos, sino de las relaciones que entre ellas se establecen. Por ende, es vital una adecuada contextualización del dominio de conocimiento, además del uso de vocabularios adecuados y con una expresividad suficiente.
- Mejorar la calidad de los datos: Dado que la publicación de datos se basa en un modelo que responde a las abstracciones realizadas del mundo real, y que por su parte, el consumo se orienta a la consulta de las instancias generadas sobre los modelos construidos; la calidad de los datos vinculados se evalúa en su mayoría sobre las instancias generadas. En otras palabras, se identifica la necesidad de fortalecer la abstracción y diseño de los modelos de datos, basado en los requerimientos de las dimensiones de calidad de datos vinculados, con el fin de que la generación de las instancias de publicación, lleven consigo unas referencias base de calidad de datos.
- Superar inconsistencias semánticas: Dado los diseños particulares de los objetos del mundo real, sumado a la no apertura de los modelos de datos y al uso de estándares propios, es necesario avanzar hacia la apertura y reutilización, salvaguardando los debidos derechos. Por otro lado, se debe buscar disminuir las transformaciones sintácticas (lenguajes) y semánticas (significados) que se deban realizar para poder usar los datos.

Por lo tanto, con base en estos retos y en el bajo desarrollo del dominio de recursos educativos en el campo de los datos enlazados, el enfoque de esta investigación se orientó a identificar y establecer un mecanismo que permitiera modelar recursos educativos abiertos con fines de enriquecimiento a través de datos enlazados, aportando sobre las problemáticas de calidad de datos identificadas en los procesos de abstracción y modelamiento de datos.

En cuanto al estado de la cuestión, se observó una gran oportunidad de mejora en los procesos de abstracción y modelado de los datos que soportan las instancias de datos generadas. De igual

forma, se identificó la necesidad de trasladar el trabajo de las dimensiones de calidad, salvando las debidas proporciones, a los procesos de abstracción y modelado de datos, con el fin de que los modelos de datos contengan elementos de calidad antes de ser instanciados.

Por último, con relación a dominios de conocimiento como los recursos educativos, se identificó la necesidad de consolidar una taxonomía de metadatos, dado que éstos son definidos como perfiles de aplicación de acuerdo a las necesidades propias del actor de investigación. Por otro lado, se consideró importante el manejo de vocabularios controlados para el trabajo con los recursos educativos, con el fin de brindar elementos que, además de enriquecer los recursos publicados, éstos puedan estar disponibles y ser reutilizados.

La Figura 8 presenta un resumen de las diferentes problemáticas identificadas en temas relacionados con apertura de datos, datos enlazados, recursos educativos y apertura y enriquecimiento de REA en Colombia. Problemáticas que además de irse sumando a los niveles siguientes de apertura o enlace de datos, afectan la disponibilidad (niveles 1 al 3) y reutilización (niveles 4 y 5) de los datos enlazados. Adicionalmente, en el dominio de conocimiento educativo, estas problemáticas son trasladadas a la apertura, disponibilidad, enriquecimiento y reutilización de los REA.

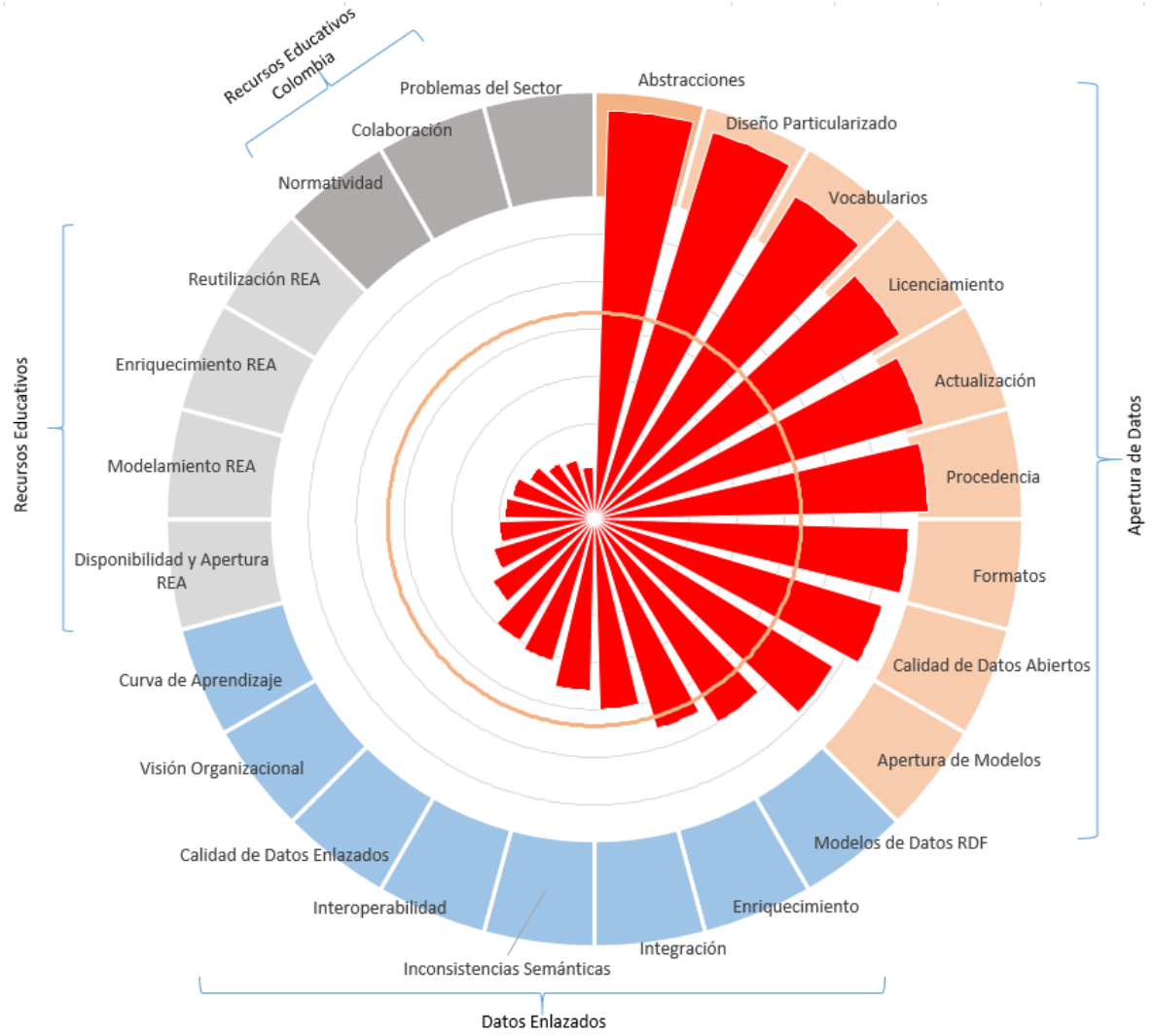


Figura 8. Problemáticas de Datos Abiertos y Enlazados.

Por ende, se requirió el prestar atención a las primeras etapas del ciclo de vida de datos enlazados para la descripción y enriquecimiento de los REA. Como se verificó en los antecedentes y en el estado de la cuestión, con el fin de aportar en la mejora de la calidad, la disponibilidad y la reutilización de los REA, se planteó trabajar las primeras fases del ciclo de vida delos datos enlazados, con el fin de proveer abstracciones de REA, que además de incluir elementos de calidad, respondieran a un dominio de conocimiento modelado desde la generalidad y pueda ser enriquecido a través del uso de datos enlazados.

Para tal fin, se proyectó el uso de una estrategia que permitiera un modelamiento flexible e interoperable, que disminuyera la rigidez, inmovilidad y la viscosidad ante posibles cambios o escalamientos de las abstracciones realizadas. Con base en estos requerimientos, surgió el trabajo con metamodelos como pilar de trabajo para el diseño de un modelo de datos para describir y enriquecer REA.

PARTE III DESARROLLO DE LA PROPUESTA

CAPÍTULO IV DISEÑO METODOLÓGICO

Con base en la revisión de literatura realizada, se identificaron publicaciones como Bozkurt *et al.* (2020), donde los autores describen recomendaciones acerca de los REA como por ejemplo:

- Apertura a la EdTech (Tecnología educativa): Es fundamental realizar una adecuada selección de tecnología educativa que priorice el uso de software libre y formatos abiertos, así como la privacidad y gestión de los datos personales de los usuarios, evitando que los estudiantes creen nuevas cuentas.
- REA: Con el fin de ofrecer materiales y cursos que permitan a los docentes obtener pautas didácticas, pedagógicas y comunicacionales para llevar a cabo el rediseño de la docencia, se propone la creación de un banco abierto de recursos educativos para la formación docente. También es una gran oportunidad para abrir todos los recursos educativos de las instituciones, permitiendo a los estudiantes un acceso completo a ellos.

Aunque en la revisión de literatura se identificaron recomendaciones acerca de la apertura y disponibilidad de los recursos educativos, no se identificaron investigaciones que involucren una mejora sustancial en la descripción y modelado de los recursos educativos, ni la inclusión de componentes de calidad de datos en la descripción de los REA, que hayan contribuido a superar los problemas de calidad de datos en las instancias de sus metadatos. Adicionalmente, no se identificaron contribuciones concretas al enriquecimiento de los metadatos de los REA, ni su articulación y reutilización para ser enlazados a través de la nube LOD.

En este sentido, los siguientes capítulos presentan el diseño e implementación de un metamodelo para la descripción de REA, que contribuye a potencializar la apertura, disponibilidad, reutilización y enriquecimiento de los REA mediante especificaciones LOD y principios de confianza.

Para tal fin, y dada la naturaleza de la investigación se definió una metodología compuesta por cuatro fases para construir el metamodelo: contextualización, diseño, despliegue y validación. La Figura 9 presenta los aspectos que se tuvieron en cuenta en este planteamiento metodológico.

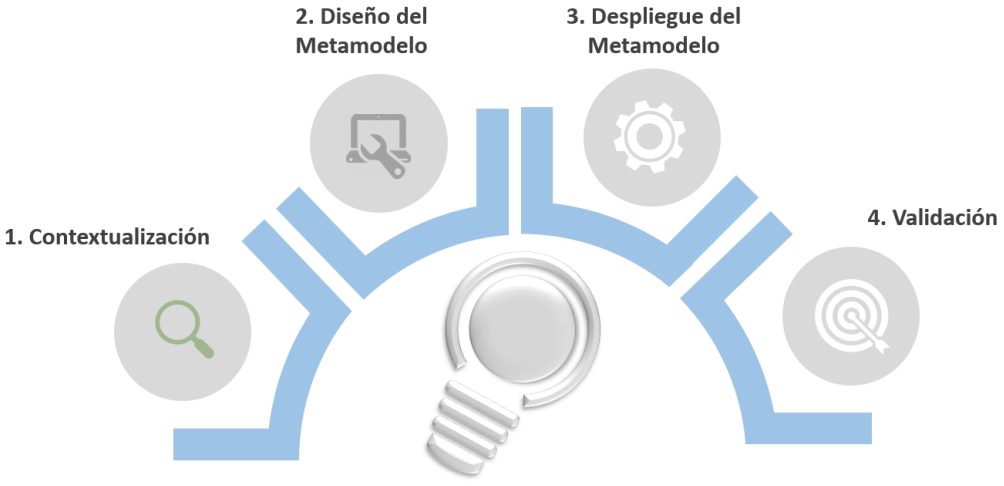


Figura 9. Metodología para el metamodelo de descripción y enriquecimiento de REA.

A continuación se detalla brevemente el alcance de cada fase:

- En la fase de *Contextualización de Información* se definieron tanto la construcción teórica requerida como las condiciones iniciales para la contextualización del metamodelo. En esta fase se revisaron y construyeron componentes tales como:

    - Dominio de Conocimiento: Recursos Educativos.
    - Enlace de Datos.
    - Calidad de Datos Enlazados.
    - Arquitectura Dirigida por Modelos.
    - Antecedentes.
    - Estado de la Cuestión.

    En esta fase se examinaron los componentes teóricos definidos para esta investigación, proporcionando elementos claves para la construcción del metamodelo. Estos elementos esbozaron parámetros y restricciones consideradas al momento de establecer cada una de las capas del metamodelo diseñado en esta investigación.

    En este sentido, los requerimientos a satisfacer en los procesos de enlace de datos, calidad de datos enlazados y arquitectura dirigida por modelos, fueron tomados como insumos a partir de iniciativas e investigaciones realizadas por diversos autores tales como Berners-Lee (2009), Berners-Lee (2011), Bauer y Kaltenböck (2012), Bizer (2009), W3C (2017), Auer *et al. (*2014), Heath y Bizer (2011), Hausenblas (2009), Hyland *et al.* (2014), Abella (2011), Zuiderwijk *et al.* (2012), Zaveri *et al.* (2012), Ruiz (2012), Montenegro *et al.* (2011), OMG (2019), OMG (2014a), Atenas y Havemann (2013) y Dietze *et al.* (2009).

- En la fase de *Diseño del Metamodelo*, se establecieron las reglas necesarias para definir la sintaxis y la semántica del modelo. Lo cual involucró: la construcción de una ontología para describir recursos educativos, la definición de los vocabularios controlados, y las estrategias de mapeo de datos utilizadas. Se aclara que la construcción de la ontología no se enfocó en el guion pedagógico del recurso o en sus estrategias didácticas, ni en su diseño instruccional, dado que estos elementos hacen parte de la planeación y diseño del REA, y no son partes de elementos necesarios para su descripción, búsqueda y recuperación. En esta fase se revisaron y construyeron componentes tales como:

    - Taxonomía Metadatos REA.
    - Metadatos REA a enriquecer.
    - Dimensiones de Calidad a aplicar.
    - Modelo de Datos RDF.
    - Definición de reglas para la Semántica del modelo.
    - Diseño del Modelo de Datos: Ontología REA, Vocabularios, Mapeo de Datos.

    La definición de estos componentes permitió diseñar un modelo de datos que genera instancias de metadatos de recursos educativos enlazados, además de mapear sus metadatos a RDF Graph.

    La taxonomía de los recursos educativos, además de identificar los metadatos usados, definió cuáles de ellos son obligatorios en el proceso de descripción de REA. Adicionalmente, las dimensiones de calidad de datos identificaron las características de calidad a considerar dentro de la abstracción y modelado de los recursos educativos. La conjugación de estos elementos generó un modelo de datos para describir REA expresado en una ontología, proporcionando mejoras en la disponibilidad de los REA y en la confianza provista de cara a las necesidades de los usuarios.

Seguido a la construcción del modelo de datos, esta fase planteó sugerencias con relación a la aplicación del modelo LOD a la descripción de los REA, además de identificar los metadatos que eran susceptibles de enriquecimiento a través de enlace de datos.

Los elementos descritos enmarcaron los requerimientos que se deben satisfacer al momento de generar una instancia de datos, producto de la aplicación del metamodelo. Esta fase, como resultado adicional, presenta el diseño de estrategias de mapeo de los metadatos que describen los REA, tanto para las aplicaciones que los consumen como para mejorar la reutilización a través de su articulación con el mundo de los datos enlazados.

- En la fase de *Despliegue del Metamodelo*, se establecieron tanto el ambiente tecnológico como las plataformas a utilizar para implementar el metamodelo y así poder describir y enriquecer los REA. En este proceso, se parametrizaron los componentes tecnológicos, ajustándolos a los mapeos de aplicación y de dataset provistos por el metamodelo diseñado. En esta fase se revisaron y construyeron componentes tales como:

    - Entorno tecnológico de enlace de datos.
    - Diseño de Arquitectura Tecnológica.
    - Implementación del Metamodelo.

- Finalmente, en la fase de *Validación*, se realizaron cuatro enfoques de validación al modelo para la descripción y el enriquecimiento de REA. En primera instancia, se validó el funcionamiento del metamodelo tanto en la generación de los mapeos (para aplicación y para dataset), como en la gestión de propiedades de los REA.

  En segunda instancia, se ejecutaron pruebas que validaron el funcionamiento adecuado del despliegue del modelo de datos en un ambiente tecnológico. La inserción de datos enlazados permitió verificar la adecuada construcción de los RDF Graph y la obtención de respuestas ante las consultas SPARQL ejecutadas.

  En tercera instancia, se validó la carga de diferentes tipos de REA en la estructura de propiedades dispuesta por el modelo de datos. En última instancia, y considerando la naturaleza y abstracción del tema trabajado, esta investigación no se abordó explícitamente desde una etapa experimental. Sin embargo, se abordó el método Goal–Question-Metric (GQM), utilizado por Yaghouti *et al.* (2015), Behkamal *et al.* (2014), Behkamal (2014), Li *et al.* (2019) y Zahedi y Fathian (2016) en temas de evaluación de calidad de los datos en ambientes LOD, con el propósito de validar el modelo de datos a través de la observación de un conjunto de métricas definidas para tal fin.

  Por último, se planteó una revisión de los diferentes aportes proporcionados por el metamodelo diseñado, con relación a la calidad de los datos, a la disponibilidad y a la reutilización de los recursos educativos.

A continuación se detallan cada uno de los hitos y procesos de las fases abordadas para el diseño e implementación del metamodelo propuesto.

CAPÍTULO V DISEÑO DEL METAMODELO PARA LA DESCRIPCIÓN Y ENRIQUECIMIENTO DE LOS REA.

En primer lugar, es necesario tener en cuenta que el diseño de la estructura del metamodelo para la descripción y enriquecimiento de REA, se soportó a partir de recomendaciones y estudios realizados por diversos autores tales como Evans *et al.* (2003), Kleppe *et al.* (2003), Quintero y Anaya (2007), OMG (2014b), Pérez (2008), Montenegro *et al.* (2011), OMG (2019), OMG (2014a) y Haslhofer y Klas (2010). Con base en estas consideraciones, y teniendo en cuenta los hallazgos identificados en el capítulo anterior, el diseño del metamodelo para la descripción y el enriquecimiento de REA se enmarcó en los siguientes contextos (Figura 10):

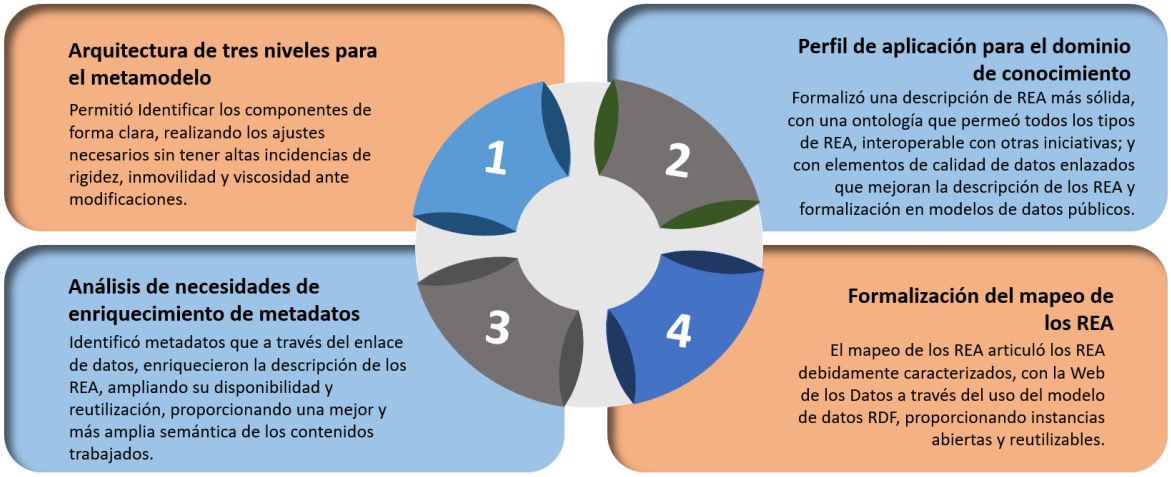


Figura 10. Contextos que definieron el metamodelo.

Por último, la implementación del metamodelo provee un modelo de datos para describir REA, más sólido y enriquecido, que además permea modelos de datos existentes, usados para describir recursos de acuerdo a las necesidades particulares de las organizaciones. A continuación se detalla cada uno de los pasos ejecutados para el diseño del metamodelo.

5.1. Construcción teórica del Metamodelo.

Después de explorar conceptos como los datos enlazados y el marco general de calidad de datos enlazados, junto con un barrido del estado de la cuestión en cuanto a procesos de apertura y enriquecimiento de datos; a continuación se presenta la construcción teórica que enmarcó los elementos del metamodelo.

En primera instancia, como lo definió González -citado por Rodríguez (2017)-, el meta-metamodelo (Capa 3) define las características del metamodelo (Capa 2), es decir, el tener un meta-metamodelo (Capa M3), permite que exista un metamodelo para cada dominio del conocimiento a tratar, mientras se tiene un meta-metamodelo común a todos estos para así poder realizar operaciones sobre ellos, como pueden ser transformaciones automáticas, validaciones y búsquedas. Un ejemplo de esta definición se encuentra en el concepto de “clase” de UML, concepto que puede verse como una instancia del correspondiente elemento del nivel M3, en donde se define de forma precisa ese concepto, así como sus características y las relaciones con otros elementos (por ejemplo, una clase es un clasificador, y por tanto puede tener asociado un comportamiento, y además dispone de un conjunto de atributos y de operaciones).

Dado que para esta investigación se trabajó con los REA como único dominio de conocimiento, no se requirió de ninguna operación entre metamodelos ni instanciación de las clases del meta-

metamodelo, razón por la cual no se definió un nivel superior M3 del metamodelo. Sin embargo, para definir la estructura y la semántica a las que debe conformarse el modelo que describe los REA, se consideraron las abstracciones, las conceptualizaciones básicas y las relaciones del metamodelo, necesarias para la descripción y enriquecimiento de los REA mediante especificaciones LOD, con base en elementos de calidad de datos. Dentro de las abstracciones y conceptos que se tuvieron en cuenta, se abordaron los siguientes elementos: (i) Recurso Educativo, (ii) Metadatos, (iii) Calidad de Datos, y (iv) Datos Enlazados. Estas abstracciones permitieron, además de tener un conjunto de conceptos comunes, definir un conjunto de restricciones y reglas a tener en cuenta en la construcción de los modelos de datos.

En segunda instancia, se identificó que MDA propone la definición y uso de modelos a diferentes niveles de abstracción, así como la posibilidad de generar código automáticamente a partir de los modelos definidos y de las reglas de transformación entre dichos modelos (Mora *et al.*, 2006; Quintero y Anaya, 2007). Para esta investigación es prudente aclarar que el diseño de este metamodelo no conllevó la generación automática de código a partir del modelo y de las reglas definidas (Mora *et al.*, 2006), teniendo en cuenta que el artefacto principal obtenido corresponde a un modelo abstracto de datos para describir los REA, compuesto de una serie de propiedades y parametrizaciones mejoradas a través de la aplicación de elementos de calidad de datos y enriquecimiento de datos. Por ende, dentro de las perspectivas de MDA no se contó con un punto de vista de plataforma específica. Por el contrario, y considerando el objeto principal de esta investigación, el punto de vista independiente de la computación y el punto de vista independiente de la plataforma, se presentan a continuación manteniendo las debidas proporciones acorde al enfoque y alcance del tema de investigación.

Para la construcción del metamodelo, en primera instancia se detallan las condiciones iniciales que brindan un mayor soporte al diseño del mismo.

5.2. Nivel M2 - El Metamodelo: Condiciones Iniciales.

Dadas las abstracciones provistas anteriormente, y considerando que el punto de vista independiente de la computación se centra en el entorno del sistema y los requerimientos para el mismo (Truyen, 2006; Mora *et al.*, 2006), la Figura 11 detalla las condiciones iniciales que enmarcaron los elementos del metamodelo, y definieron tanto la estructura como las restricciones del esquema de metadatos diseñado.

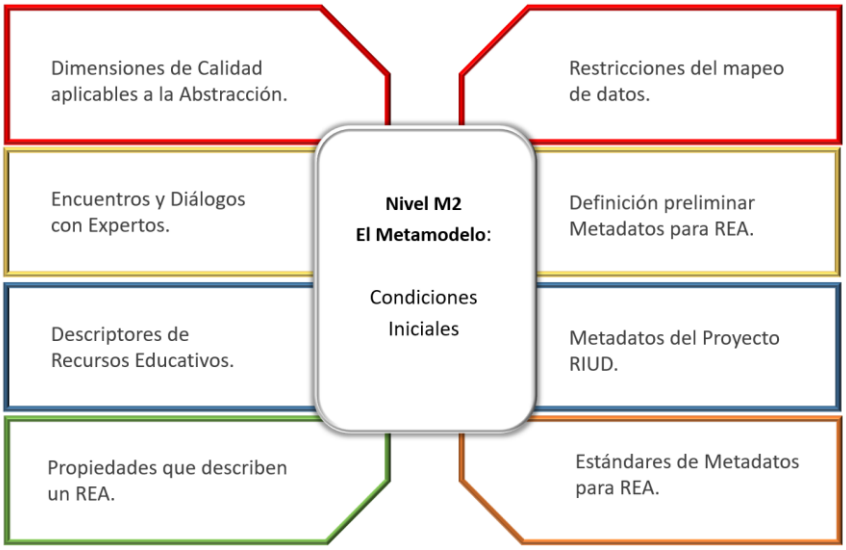


Figura 11. Contextos que definieron el metamodelo.

A continuación se describe y profundiza cada una de las condiciones iniciales consideradas en el diseño del metamodelo propuesto. Como primera condición inicial, y con el fin de iniciar el proceso de descripción de propiedades que permitan la búsqueda y recuperación de un REA, se planteó la caracterización básica de un REA.

5.2.1. Propiedades que describen un REA.

Posterior a la definición de recurso educativo (véase Sección 2.1), con el propósito de profundizar sobre los elementos requeridos por el metamodelo para la construcción de un modelo de datos que describa un REA, se caracterizaron las propiedades básicas que describen un recurso educativo. Acorde al alcance de esta investigación, la identificación de propiedades no se enfocó en aspectos tales como por ejemplo:

- El guion pedagógico del recurso o en sus estrategias didácticas.
- La secuencia para desarrollo de contenidos (mapa de contenidos).
- Propuesta de navegación (interface).
- Diseño de contenidos.
- Unidad gráfica para los espacios de interacción.
- Diseño instruccional.
- Técnicas de diseño utilizadas.
- La estructura de actividades resultantes.
- Presuposiciones del diseño.
- La rúbrica de evaluación.

Dado que estos elementos hacen parte de la planeación y diseño del material educativo, y de las formas o modos en cómo hacer que cierto conocimiento, habilidad o destreza sea aprendida (Mortera, 2010); y no hacen parte de la información que permita su descripción, atribución o reconocimiento como material educativo.

Por el contrario, se identificaron propiedades básicas de los REA necesarias para ejecutar la identificación, organización, búsquedas y recuperación de los recursos. Para esta identificación preliminar de propiedades se tomó como base a la Unesco - Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, dado que como agencia especializada de las Naciones Unidas, trata de establecer la paz mediante la cooperación internacional en materia de educación, ciencia y cultura. Adicionalmente, el sector de educación de la Unesco se centra en el seguimiento y análisis de los progresos mundiales en la adopción de los REA, así como en el apoyo a la elaboración de políticas nacionales relativas a los REA, mediante la celebración de talleres nacionales y regionales (Unesco, 2019).

La Figura 12 presenta un bosquejo de propiedades identificadas por la Unesco en investigaciones realizadas por Butcher *et al.* (2015) y Unesco (2015); que deben ser consideradas al momento de trabajar con un REA.



Figura 12. Propiedades básicas de un REA. Fuente: Butcher *et al.* (2015) y Unesco (2015).

Estas características, que posteriormente se transforman en metadatos, determinaron un conjunto inicial de propiedades provistas por un REA al momento de ser descrito con metadatos. En este conjunto de propiedades se observaron aspectos de caracterización del recurso (título, licencia, tipo, etc.), descripciones pedagógicos (objetivo, competencias, interactividad, demografía, área, etc.), ciclo de vida (versión, fecha, etc.) y soporte tecnológico (software de soporte). Estos elementos son esenciales al momento de transformar un material digital a un recurso educativo.

Para formalizar los metadatos identificados, como esquema de datos inicial, se identificaron esquemas de metadatos ampliamente reconocidos que representaron la abstracción de metadatos realizada.

5.2.2. Estándares de metadatos para REA.

Después de revisar los diferentes estándares de metadatos, se lograron identificar aspectos como:

a) Cada perfil de aplicación tiene una cantidad de propiedades o características en primer nivel. Adicionalmente, tiene calificadores que refinan las características, los cuales se encuentran ubicados en un segundo o hasta tercer nivel de jerarquía.
b) Una misma característica se puede o no incluir en un perfil de aplicación particular, según el tipo de estándar y el enfoque utilizado para construir el citado perfil.
c) Los perfiles de aplicación argumentan que como no todas las características definidas por los estándares son usadas, es prudente definir cuáles de estas características son implementadas según el dominio de conocimiento gestionado. Adicionalmente, se señala que una gran mayoría de perfiles de aplicación se definen sobre el esquema LOM.

Por lo tanto, considerando que la descripción de los recursos se hace de manera parcial y particularizada, aunado a la cantidad existente de estándares de metadatos, de perfiles de aplicación e iniciativas de trabajo; como resultado de un proceso de revisión de literatura

realizada, se identificaron aplicaciones de los siguientes estándares (Temesio, 2015; Hilera y Hoya, 2010; Pons *et al.*, 2011; OER-C, 2014):

- Dublin Core namespace – dcterms (DCMI, 2020).
- Dublin Core namespace legado – dc (Rühle *et al.*, 2017).
- Learning Object Metadata - LOM (IEEE, 2002).
- ISO/IEC 19788-5 Metadata Learning Resource – MLR (Pons et al., 2011).
- Learning Resource Metadata Initiative – LRMI (LRMI, 2013).
- Dublin Core Education Application Profile (Coyle, 2009).

Siendo Dublin Core y LOM los estándares más utilizados para la descripción de los recursos educativos (Martínez *et al.*, 2019; Yamada y Morimoto, 2010). Por otro lado, Dublin Core corresponde a uno de los estándares de metadatos que representa: i) mayor aceptación en los procesos de cosecha de datos originados desde diferentes fuentes, como por ejemplo las bibliotecas universitarias; y ii) mayor aceptación por la comunidad que aborda la publicación y apertura de datos (LAR, 2019).

Teniendo en cuenta que los esquemas de metadatos poseen una amplia gama de calificadores, algunos de los cuales no son considerados al momento de establecer la descripción de una entidad, se exploraron varios perfiles de aplicación para la descripción de recursos educativos con el fin de documentar qué otros descriptores han sido usados.

Dentro de los resultados de la revisión documental, se identificaron perfiles de aplicación internacionales, usados para la descripción de recursos. La Tabla 18 presenta algunos perfiles de aplicación internacionales más representativos, identificados por Marzal *et al.* (2006), Canabal *et al.* (2008) y Koutsomitropoulos *et al.* (2010).

Tabla 18. Ejemplos de Perfiles de Aplicación Internacionales.

Perfil de Aplicación	Basado en
The Gateway to Educational Materials (GEM) Metadata Element Set	Dublin Core
DC-ED	LOM
UK Learning Object Metadata Core	LOM
Sharable Content Object Reference Model (SCORM)	LOM
Canadian Core (CanCore) Learning Resource Metadata Application Profile	LOM
LOM-ES	LOM
UK LOM Core	LOM

Considerando el contexto colombiano, se identificaron iniciativas nacionales de perfiles de aplicación. La Tabla 19 presenta ejemplos de perfiles de aplicación nacionales identificados (MEN, 2013; Sena, 2017; MEN, 2012).

Tabla 19. Ejemplos de Perfiles de Aplicación Colombianos.

Perfil de Aplicación	Basado en
Metadatos IMS SENA	LOM

LOM-CO	LOM
Metadata - CEM Colombian Educational	LOM

Las variaciones detectadas en los perfiles de aplicación, tanto nacionales como internacionales, corresponden a dejar un conjunto base de características como obligatorias, un porcentaje adicional como recomendadas, y un restante de características como opcionales.

Para complementar la información de estándares de metadatos para describir REA, se examinaron propuestas de estándares de metadatos usadas en repositorios de recursos educativos. En esta exploración se identificó a IEEE LOM en primera instancia, y posteriormente a Dublin Core, como estándares de metadatos usados para describir los recursos gestionados. Dentro de los repositorios consultados se encuentran:

- Colombia Aprende – La Red de Conocimiento (Colombia Aprende, 2020).
- Red de Recursos Educativos en Abierto (Procomún, 2020).
- Learning Resource Exchange for Schools (LRES, 2020).
- OER Commons – Open Educational Resources (OER-C, 2020).
- Curriki - Free Education Materials for Teachers (Curriki, 2020).

En resumen, como resultado de la revisión de estándares y repositorios se identificó a IEEE LOM y Dublin Core como los esquemas de metadatos de mayor reconocimiento y uso. Por ende, como base para la construcción de la ontología que describe los REA, artefacto principal de esta investigación, se seleccionó el modelo de datos IEEE LOM. Adicionalmente, y dado que IEEE LOM tiene en común con Dublin Core un gran conjunto de elementos, se seleccionó éste último como el vocabulario sobre el cual se ejecuta el mapeo de los metadatos del perfil de aplicación PA-REA.

Posterior a esta fase, se continuó con la construcción y adecuación del conjunto de propiedades que permitieran describir los REA, complementando este conjunto inicial con propiedades recomendadas por diferentes autores.

5.2.3. Descriptores de recursos educativos.

Después de la identificación de un conjunto de propiedades básicas para describir los REA, se complementó esta descripción con características o propiedades aportadas por la literatura examinada. Investigaciones como las realizadas por MEN (2015), Martínez *et al.* (2019), LAR (2019), Atenas y Havemann (2013) y Castro *et al.* (2009), definieron un conjunto de descriptores básicos para los REA. Dentro de estas investigaciones, la Red de Repositorios de Acceso Abierto a la Ciencia, que efectúa procesos de recolección de metadatos en ocho (8) diferentes países de América Latina (LAR, 2019), definió como base para la descripción de REA los metadatos presentados en la Tabla 20.

Tabla 20. Metadatos definidos por LAR. Fuente: LAR (2019).

Restricción	Metadatos
Obligatorios	Título, Creador, Derechos / Nivel de accesibilidad, Fecha de publicación, Tipo / Tipo de documento, Identificador (corresponde a la URL que da acceso al objeto digital).
Obligatorio (si aplica)	Materia, Descripción, Editor, Idioma
Recomendados	Derechos / Licencias de uso-Derechos de autor, Colaborador (Obligatorio en caso de Director de Tesis), Tipo / Versión del documento, Formato, Fuente

Opcionales	Relación, Cobertura geo temporal, Audiencia

Los metadatos aportados por la exploración bibliográfica complementaron la descripción de los recursos educativos propuestos por la Unesco (2015), adicionando nuevas propiedades o en su defecto, refinando algunas previamente identificadas.

Seguidamente, y con el fin de contextualizar el esquema de datos para describir REA a un caso de estudio más específico, se exploró la propuesta de metadatos diseñada por el proyecto RIUD - Repositorio Institucional de la Universidad Distrital Francisco José de Caldas, el cual se describe a continuación.

5.2.4. Metadatos del proyecto RIUD – Universidad Distrital (Caso de Estudio).

El RIUD (RIUD, 2020), administrado por el Sistema de Bibliotecas de la Universidad Distrital Francisco José de Caldas, gestiona recursos tales como tesis, recursos de enseñanza, libros y revistas. Para describir estos recursos hace uso del estándar Dublin Core. La Tabla 21 muestra el esquema de metadatos utilizado por el proyecto RIUD.

Tabla 21. Esquema de metadatos usados por RIUD. Fuente: RIUD (2020).

Tipo	Metadatos
Tesis	Modalidad de grado, Título, Autor(es), Director de tesis, Email personal, Resumen, Abstract, Fecha de aprobación, Palabra(s) clave, Keywords, Patrocinador(es), Formato, Materias, Reconocimiento, Confidencialidad
OVA	Título, Otros títulos, Autor(es), Institución, Fecha de publicación, Resumen, Abstract, Palabra(s) clave, Keywords, Materias
Libros	Título, Otros títulos, Resumen, Abstract, Autor(es), Tipo de publicación, Edición, ISBN, Fecha de publicación, Palabra(s) clave, Keywords, Materias, Patrocinador(es), Formato, Parte de / capítulo de, Institución, Identificación de la editorial, Citación bibliográfica del libro, Lugar, Confidencialidad

Aunque el esquema de metadatos del RIUD se diseñó para documentar recursos de forma genérica, éste esquema aportó algunos elementos a tener en cuenta en la construcción del modelo de datos para describir los REA, producto de esta investigación, tales como el tratamiento de diferentes tipos de identificadores de los recursos, la consideración de diferentes tipos de contribuyentes que participan en el diseño o implementación de los recursos, entre otros.

De igual forma, el esquema de metadatos RIUD (RIUD, 2016) evidenció que la estructura de metadatos Dublin Core es susceptible de modificarse (IDEE, 2005) acorde a los requerimientos de metadatos del modelo propuesto (Tabla 22).

Tabla 22. Ejemplo de metadatos Dublin Core usados por RIUD. Fuente: RIUD (2016).

Propiedad	Metadatos RIUD
Director del trabajo de grado	dc.contributor
Autor	dc.creator
Resumen	dc.description
Abstract	dc.description.abstract
Palabras Clave	dc.subject
Keywords	dc.subject.keywords
Título	dc.title
Materias	dc.subject.lemb

Finalmente, y con el propósito de fortalecer la propuesta de metadatos, se llevó a cabo una serie de revisiones y entrevistas con expertos en el área de documentación y publicación de recursos de diversas universidades y entidades, tales como la Universidad Autónoma de Bucaramanga, el Centro de Recursos para el Aprendizaje y la Investigación de la Universidad Javeriana, la sección de análisis de información de la Biblioteca General Alfonso Borrero Cabal de la Universidad Javeriana, la dirección de la biblioteca de la Universidad Distrital Francisco José de Caldas, entre otras entidades. Estas entrevistas permitieron tener un enfoque más amplio, aterrizado y práctico de la propuesta.

5.2.5. Encuentros y entrevistas con expertos.

Adicional a la exploración de metadatos, estándares y recursos a utilizar, se realizaron entrevistas con personas expertas en el campo de la catalogación y gestión de recursos (Anexo 3). Como resultados generales de estos encuentros, se lograron identificar los siguientes aspectos:

- Aunque se han presentado varios proyectos en Colombia como el Banco Nacional de Recursos, éstos han declinado o fracasado por asuntos de interoperabilidad, licenciamientos, apertura de datos, entre otros factores.
- En Colombia, las organizaciones del sector público y educativo, cuentan con repositorios privados para gestionar sus recurso tales como: informes de gestión, tesis de grado, clases grabadas, indicadores, estadísticas, históricos, entre otros. Pero no se cuenta con una disponibilidad de modelos de datos abiertos que permitan la interoperabilidad.
- Dentro de los metadatos recomendados para describir recursos se encuentran: Título, contexto (para quien va dirigido, descripción, idioma, fecha de creación), contenidos o temáticas abordadas con el recurso, periodicidad (con que periodicidad se actualiza), tipos de interactividad o formatos (sonoro, interactivos, etc.), a quien va dirigido (tipo de población o nivel de formación, rango de edad), tipos de evaluación que maneja el recurso (práctica, de conocimiento, etc.), procedencia (autor, institución, versión), tipo de licenciamiento que maneja el recurso.
- No se cuenta con un esquema o perfil de aplicación para describir recursos educativos abiertos, ampliamente definido y compartido.
- La estructura de metadatos debe ser útil y no saturarse de información desde lo técnico, dado que restringe su uso al momento de describir el recurso. Por lo cual se recomienda tomar el estándar y construir un perfil de aplicación, acorde a las necesidades de la organización.
- LOM es uno de los esquemas más usados para describir recursos educativos abiertos.
- Dublin Core es uno de los estándares más usados para describir recursos, dada su flexibilidad para adaptar o adicionar nuevas características. Además, es uno de los estándares más usados para compartir datos entre organizaciones. Proyectos latinoamericanos como “LA Referencia” usan Dublin Core para recolectar datos.
- Repositorios institucionales pertenecientes a la Universidad Javeriana y la Universidad Distrital, por ejemplo, usan Dublin Core y plataformas como Dspace para gestionar los metadatos de los recursos almacenados.

Como resumen de esta fase, se identificó que i) tanto los perfiles de aplicación como los repositorios y los lineamientos definidos para recursos educativos, caracterizan las diferentes tipologías de recursos bajo un mismo esquema de propiedades, y ii) el esquema con mayor aplicación para la representación de recursos educativos corresponden a LOM. Adicionalmente, el estándar más utilizado para el trabajo de generación de metadatos de los recursos digitales corresponde a Dublin Core.

Con base en estas consideraciones y en la revisión de literatura realizada, se propuso el uso de LOM para la construcción del esquema de características que describen los REA. Por último, se

evidenció la necesidad de diseñar una ontología para describir todos los tipos de REA, bajo un mismo esquema de descriptores. El diseño de esta ontología y su correspondiente mapeo, deben considerar la reutilización de ontologías y vocabularios reconocidos (Fermoso *et al.*, 2008; Noy y McGuinness, 2001), o en su defecto optar por la implementación de un vocabulario que supla el mapeo de los descriptores utilizados (Guzmán *et al.*, 2012; Romero, 2018).

Para llevar a cabo el diseño de la ontología propuesta, considerando todos los aportes recopilados sobre metadatos para describir REA, se construyó una definición preliminar de metadatos para describir REA.

5.2.6. Definición preliminar de metadatos para recursos educativos.

Después de revisar las condiciones iniciales, el primer escenario abordado consistió en identificar la estructura de metadatos base para describir los recursos educativos. Para la construcción de esta estructura de metadatos se tuvieron en cuenta dos insumos: i) el análisis de condiciones iniciales, y ii) una comparación de estructuras de metadatos, realizada entre 7 diferentes iniciativas de metadatos (Anexo 4). Con base en estos insumos se definió el siguiente esquema de metadatos para la descripción de REA (Tabla 23):

Tabla 23. Metadatos seleccionados basado en bibliografía.

Categoría	Metadatos
General	Identificador, titulo, idioma, descripción (abstract), palabras claves
Ciclo de vida	Versión, rol, entidad, fecha
Técnica	Formato, tamaño, ubicación (URI), tipo, versión mínima
Educativo	Tipo de Interactividad, tipo de recurso, población objeto, nivel educativo, contexto educativo, rango de edad, descripción, idioma
Derechos	Licenciamiento, derechos del licenciamiento
Anotación	Persona, fecha, descripción.
Clasificación	Taxón

Teniendo en cuenta la variedad de metadatos usados para describir los recursos, se evidenció que las organizaciones construyen sus modelos de datos a partir de sus necesidades y proyectos particulares, limitando la expresividad del modelo que describe los REA.

Por lo tanto, como un resultado parcial de esta revisión y teniendo en cuenta lo descrito por Wason (2000), se consideró incluir:

a. Dentro de la categoría “Clasificación” los siguientes metadatos:
   - ***Área en la que se circunscribe el recurso***: Taxón o subdominio dentro del cual se enmarca el REA.
   - ***Asignatura en la que se circunscribe el recurso***: curso o materia para la cual inicialmente se desarrolló el REA.

b. Dentro de la categoría “Ciclo de Vida” en su subcategoría “Contribuyentes”:
   - ***Nombre de la institución***: Nombre de la organización principal que provee el REA.
   - ***Nombre del programa***: Nombre del programa, reconocido por una organización, que provee el REA

Anotando que estos metadatos son aplicables a aquellos contextos que así lo permitan, en el Anexo 4 se detalla la clasificación realizada. La inclusión de estos descriptores hace énfasis a que en contextos como el colombiano, se presentan situaciones cotidianas como:

- Los recursos educativos son desarrollados por los docentes para sus clases. Pero, tanto la disposición de tiempo como la actividad requerida, no se encuentran presentes en sus planes de trabajo.
- No se cuenta con procesos contractuales claros que definan las dedicaciones de tiempo para el diseño e implementación de REA, así como la claridad acerca de los derechos sobre los recursos producidos.
- No se contemplan políticas de educación en el ámbito nacional que permitan definir áreas de diseño curricular exclusivos sobre entidades educativas para el diseño e implementación de REA.

Por ende, no se ha registrado adecuadamente la procedencia de los recursos más allá del autor, siendo su contenido un valor literal. Por el contrario, los descriptores sugeridos proporcionan información de procedencia más completa, que además de fortalecer el modelo de datos, proveen información asociada a indicadores de calidad y desempeño de programas, instituciones, acreditación voluntaria de programas e instituciones, etc.

Por último, y como estrategia para mejorar la definición de metadatos para la descripción de REA, se continuó con la aplicación de las dimensiones calidad de datos a la abstracción de propiedades de los REA, proceso que se amplía a continuación.

5.2.7. Dimensiones de calidad aplicables al proceso de abstracción de datos.

Considerando las reglas necesarias para definir la semántica de los modelos, como parte de las condiciones iniciales, se analizaron las dimensiones de calidad de datos enlazados aplicables al esquema de metadatos preliminar. Estas dimensiones fueron definidas como parte del alcance esta investigación.

Dado que en gran parte de las investigaciones examinadas, las dimensiones de calidad se evaluaron sobre las instancias de datos, se consideró como aporte valioso para la construcción de este metamodelo, la identificación de propiedades que brinden parámetros de calidad desde la abstracción del mundo real, como primera etapa del ciclo de vida del enlace de datos.

Para llevar a cabo esta tarea, se revisó el alcance de cada métrica de las dimensiones de calidad. Esta revisión permitió identificar elementos de metadatos que debería contener la abstracción del REA con relación a cada métrica. Al terminar la revisión de las dimensiones de calidad y sus métricas, y con el fin de incluir estos elementos al perfil de aplicación, se buscó relacionar los elementos identificados con las categorías del esquema IEEE LOM, para luego ser mapeados a Dublin Core. Estos procesos se realizaron con el fin de brindar elementos de metadatos que proporcionaran criterios explícitos de calidad en los modelos instanciados. A continuación se presentan los criterios de calidad de datos contextual y de confianza que se utilizaron para complementar la abstracción de los REA. Es importante resaltar que para efectos propios de la propuesta, solo se contemplaron aquellas métricas cuya información requerida podía ser representada en la abstracción de propiedades de los REA.

a) **Dimensión Contextual:** La Tabla 24 muestra la información contextual que se identificó como requerida por los modelos de datos.

Tabla 24. Información Contextual requerida por los modelos de datos.

Dimensión	Métricas	Información Requerida
Relevancia	Uso de atributos de meta información.	Aparición de términos relevantes dentro de los metadatos tales como: palabras claves, título alternativo, etc.
Completitud	Grado en que las clases y propiedades no faltan.	Representación de las clases y propiedades de una ontología.
	Grado en que los valores de una propiedad no faltan.	Representación de todos los valores para una propiedad.
	Grado en que los objetos del mundo real están completos.	Representación de todos los objetos del mundo real.
	Grado en el que no faltan enlaces.	Identificación de propiedades a enriquecer con enlace de datos.

Las dimensiones contextuales (Zaveri *et al.*, 2012) establecieron la necesidad de complementar la descripción de los REA, aportando elementos de la siguiente manera:

- Completitud del esquema: el grado en que las clases y propiedades de una ontología están representadas, por lo tanto se puede llamar "completitud de ontología". Por ende, dado que un REA va más allá de la descripción de un libro, se requirió que el modelo diseñado incluyera:

  - ***Esquemas de identificación***: International Standard Book Number - ISBN para los libros, International Standard Serial Number - ISSN para las revistas, Digital Object Identifier - DOI para las tesis, libros, otros; International Standard Name Identifier - ISNI para la institución, etc. Esta propiedad hace parte de la descripción general del recurso.
  - ***Citación***: mención de la fuente del recurso. Principalmente está compuesta de nombre del autor, año, nombre del recurso, tipo del recurso, y datos de procedencia que se consideren pertinentes. Esta propiedad hace parte de la descripción general del recurso.
  - ***Descripción educativa del recurso:*** Dentro de esta descripción se consideran elementos tales como competencias desarrolladas por el recurso, contexto educativo en el cual se desarrolla el recurso, nivel educativo o entorno dentro del cual se da el aprendizaje, interactividad o nivel de comunicación entre el recurso y el usuario. Esta propiedad hace parte de la descripción educativa del recurso.

- Completitud de la Propiedad: para procurar que no falten valores para una propiedad específica, se definieron aspectos como los vocabularios controlados que parametrizan los respectivos valores para las propiedades.
- Completitud de la Población: todos los objetos del mundo real de un tipo particular, deben estar representados en los datasets. Para tal fin, en el diccionario de datos se especificaron aspectos como:

  - Tamaño del formato: Tamaño digital del recurso en i) bytes para recursos como imágenes, textos, etc., o ii) en tiempo para recursos como audios, por ejemplo. Esta propiedad hace parte de la descripción técnica del recurso.

- Completitud de la Interconexión: interconexión de las instancias del dataset. Para ello se realizó un análisis de los campos que son sujeto de enriquecimiento a través de enlace de datos.
- Uso de atributos de meta información relevante, que provea información importante para la consulta de los usuarios. Para tal fin se consideró necesario definir:

- Palabras claves: lista clasificada de términos en un campo particular, para su uso en indexación y recuperación de información. Esta propiedad hace parte de la descripción general del recurso.
- Título alternativo: segundo título que identifica un recurso. Esta propiedad hace parte de la descripción general del recurso.
- Determinar cuáles campos o atributos deben ser considerados como obligatorios y cuales como opcionales.

b) **Dimensión de Confianza:** La Tabla 25 muestra la información de Confianza que se identificó como requerida por los modelos de datos.

Tabla 25. Información de Confianza requerida por los modelos de datos.

Dimensión	Métricas	Información Requerida
Verificabilidad	Verificar la información del publicador.	Declaración del autor y sus colaboradores, el editor de los datos y su fuente.
	Verificar la autenticidad del dataset.	Uso de un vocabulario de procedencia.
Credibilidad	Meta información sobre la identidad del proveedor de información.	Declaración del proveedor / y contribuyentes principales.
Procedencia	Metadatos sobre un dataset.	Declaración de título, contenido y URI del dataset.
	Cálculo de recomendaciones de confianza personalizadas.	Declaración de información de procedencia.
	Calcular la confiabilidad de las declaraciones RDF.	Declaración de información de procedencia.
	Detectar la fiabilidad y credibilidad de una persona (editor).	Declaración de información de procedencia.
	Adquirir confianza de contenido de los usuarios.	Identificación de propiedades sujetas de enriquecimiento, que transferirán confianza a los usuarios.
	Detección de confiabilidad y credibilidad de una fuente de datos.	Anotaciones de confianza hechas para derivar una evaluación de la confiabilidad y credibilidad de las fuentes.
	Asignar valores de confianza a datos / fuentes / reglas.	Uso de ontologías de confianza, que permitan la interoperabilidad entre dominios.
Licenciamiento	Indicación de una licencia legible por las máquinas.	Declaración de una licencia en el dataset.
	Indicación de una licencia legible por persona.	Declaración de una licencia en la documentación del dataset o su fuente.
	Permisos para usar el dataset.	Declaración de una licencia que indique si está permitida la reproducción, distribución, modificación o redistribución.
	Indicación de atribuciones.	Declaración de si el uso de los datos se atribuye de la misma manera especificada por el autor o licenciante.
	Indicación de Copyleft or ShareAlike	Declaración de Copyleft o ShareAlike.
Reputación	Reputación del Publicador.	Declaración del editor en el dataset.

Estas dimensiones establecieron la necesidad de complementar la descripción del REA, adicionando elementos que:

- Describan el ciclo de vida del recurso, generando mayor confianza en el recurso a través de la verificación de sus actualizaciones (Fecha de creación, fecha de actualización, etc.). Esta propiedad hace parte del ciclo de vida del recurso.
- Complementen su procedencia, a través de incluir información que verifique la procedencia y genere confianza sobre el recurso consultado. Dentro de estos elementos de procedencia se determinaron:

    - Autor.
    - Editor.
    - Otros colaboradores (Proveedor de contenido, diseñador gráfico, licenciante, experto, implementador, validador, etc.).
    - Anotaciones realizadas al recurso.
    - Personas o entidades que realizan las anotaciones sobre el recurso.

    Estas propiedades hacen parte de los contribuyentes y anotaciones del recurso.

- Aporten en la verificabilidad y credibilidad, a través de la inclusión de elementos de procedencia como el lugar, grupo y departamento bajo el cual se construyó el REA (Propiedades que hacen parte de los contribuyentes del recurso). Lo anterior no solamente genera credibilidad en el autor, sino que apoya la credibilidad y generación de goodwill de la institución y grupo de investigación. Adicionalmente, diferentes recursos se pueden construir como obras por encargo, ante lo cual se debe dar crédito tanto al autor como a la organización que realiza el encargo.
- Complementen la confianza en los contenidos, a través de la identificación de propiedades que son susceptibles de enriquecimiento por medio de datos enlazados, permitiendo de esta forma mejorar y enriquecer la información de la instancia de datos.
- Complementen la procedencia a través de la inclusión de propiedades que aporten a la interoperabilidad entre dominios de datos.
- Complementen su licenciamiento, describiendo el tipo de licenciamiento y los derechos adquiridos. Propiedades que hacen parte de los derechos del recurso.
- Realicen una clasificación educativa del recurso.

c) **Dimensión de Accesibilidad:** En cuanto a la categoría de accesibilidad, sus dimensiones (disponibilidad, rendimiento, seguridad y tiempo de respuesta) se orientan a aspectos relacionados con la forma en que se pueden acceder y recuperar los datos instanciados. Adicionalmente, su valoración se orienta a la implementación y evaluación de aspectos tecnológicos independientes del modelo de datos. Razones por las cuales, después de ser revisadas, se decidió excluir estas dimensiones de la construcción del esquema de metadatos para la descripción y enriquecimiento de los REA.

5.2.8. Restricciones del mapeo de datos.

Para mapear el esquema de propiedades diseñado a un esquema de metadatos reconocido, como lo describe Calderón (2006), se consideró que estándares como Dublin Core poseen elementos opcionales y repetibles, y además permiten el uso de calificadores opcionales para cada elemento. Los calificadores aumentan la especificidad y precisión de los metadatos. El DCMI (Dublin Core Metadata Initiative) reconoce dos clases de calificadores (Calderón, 2006):

a. Refinación de elementos: Estos calificadores hacen que el significado de un elemento sea más específico. Un elemento refinado comparte el significado del elemento no calificado, pero con un alcance más restrictivo. Las definiciones de términos para refinamiento de elementos deben estar públicamente disponibles.
b. Esquema de codificación: Estos calificadores identifican esquemas que ayudan en la interpretación del valor de un elemento. Estos esquemas incluyen vocabularios controlados y notaciones formales o reglas de parseo. Un valor expresado usando un esquema de codificación, será entonces un símbolo (token) escogido de un vocabulario controlado (por ejemplo, un término de un sistema de clasificación) o un String con formato de acuerdo con una notación formal (por ejemplo, "2002-01-01" como la expresión estándar de una fecha).

Un ejemplo de la aplicación de estos calificadores se encuentra en expresiones como las definidas por el modelo presentado en esta investigación:

dc.contributor.author = “Pedro Pérez”
dc.contributor.uri = “http://www.miservidor.com/pedro”

En estas definiciones, una persona contribuye a un recurso bajo el rol de autor. Esta persona tiene nombre y URI para su identificación.

Por otro lado, el modelo de datos RDF define relaciones binarias o tripletas, que constan de una relación denominada predicado o propiedad, dirigida entre el sujeto y el objeto (Yu, 2014). Al intentar serializar los elementos descritos anteriormente, se conforma una tripleta cuyo predicado y objeto se configurarían como se muestra a continuación:

dc:contributor “Pedro Pérez”, “http://www.miservidor.com/pedro” ;

Según Rector y Noy (2006), en los lenguajes de la Web Semántica, como RDF y OWL, una propiedad es una relación binaria: se utiliza para vincular dos individuos o un individuo y un valor. Sin embargo, en algunos casos, la forma natural y conveniente de representar ciertos conceptos es usar relaciones para vincular a un individuo con más de un individuo o valor. Los casos particulares se presentan de la siguiente manera:

- Atributos adicionales que describen una relación: Se crea un individuo que representa la instancia de relación en sí, con enlaces desde el sujeto de la relación a esta instancia, y con enlaces desde esta instancia a todos los participantes que representan información adicional sobre dicha instancia.
- Diferentes aspectos de la misma relación: En la mayoría de las interpretaciones, una instancia de una relación no puede verse como una instancia de una relación binaria con atributos adicionales adjuntos. Más bien, es una instancia de relación que enlaza al sujeto y el objeto complejo, representando diferentes hechos sobre su relación.

La Figura 13 presenta una alternativa de solución al manejo de tripletas RDF n-arias.

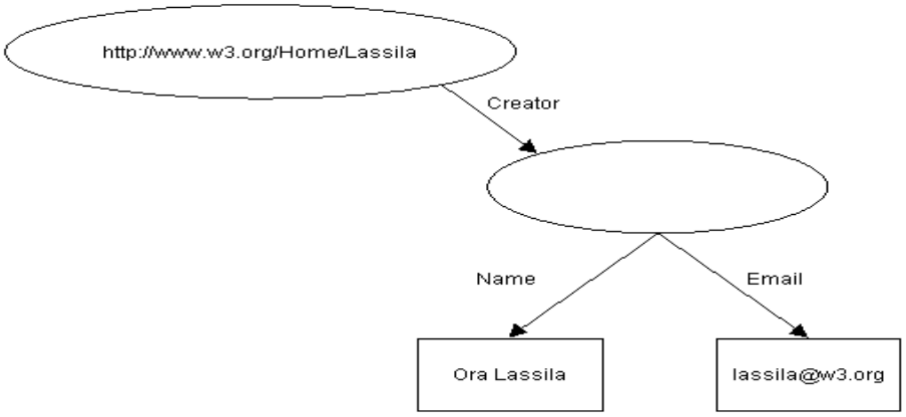


Figura 13. Grafo para una tripleta n-aria. Fuente: Rodríguez (2009).

Teniendo en cuenta que la definición del modelo producto de esta investigación incluye una serie de calificadores para refinar propiedades, se operaron restricciones de mapeo RDF para tripletas n-arias. Por lo tanto se consideró la opción de incluir instancias adicionales al modelo para simplificar dichas tripletas a tripletas binarias. Sin embargo, y con el fin de simplificar el modelado, se usaron estándares reconocidos (Fermoso *et al.*, 2008) que disminuyeron la complejidad de representación, al contar con los mismos calificadores en niveles superiores de profundidad.

En términos generales, la Figura 14 resume los principales componentes definidos en el nivel M2 del metamodelo, los cuales además determinan la sintaxis y la semántica del esquema o modelo de datos en la siguiente capa.

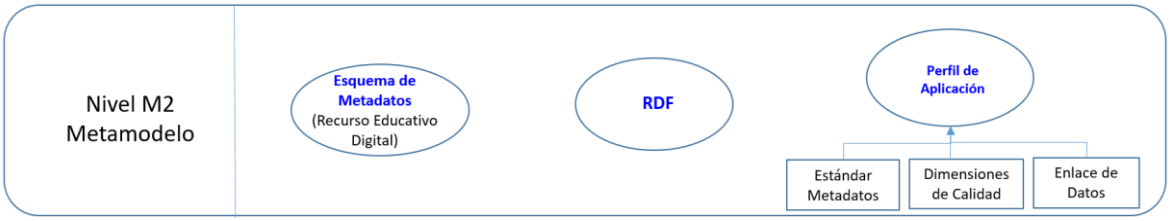


Figura 14. Nivel M2 del Metamodelo diseñado.

En este nivel M2 se especificaron las entidades tenidas en cuenta en la construcción de un modelo de datos para describir y enriquecer los REA:

a. El esquema de metadatos: Entidad que parametriza la caracterización de propiedades y valores, realizada sobre los recursos educativos.
b. Perfil de aplicación: Entidad que parametriza la descripción de REA de acuerdo a tres componentes. El primer componente corresponde al estándar de metadatos usado para representar el esquema de propiedades que describe un REA. Dicho estándar define un conjunto de restricciones para la representación de los datos. El segundo componente corresponde a las dimensiones de calidad de datos aplicadas, que además de complementar la descripción de los recursos a través de la inclusión de metadatos, parametrizan las necesidades de propiedades que debe proveer el modelo de datos con el fin de proporcionar insumos para evaluar la respectiva dimensión. Por último, el modelo de enlace de datos ofrece una serie de restricciones que debe cumplir el modelo de datos para describir y enriquecer un REA.
c. RDF como modelo de datos, además de proveer la sintaxis para la construcción de los modelos, establece un conjunto de restricciones aplicables al momento de construir las tripletas, como por ejemplo el uso de tripletas binarias.

Estos tres elementos del metamodelo definieron las reglas de representación y la forma como debían ser creados los modelos de datos que representan la descripción de los REA.

Como resultado de la revisión de condiciones iniciales, que contextualizan los tres componentes de la capa M2 del metamodelo, se identificaron descriptores adicionales los cuales fueron tenidos en cuenta al momento de establecer la taxonomía de los REA. Estos descriptores, además de proveer una descripción más detallada y completa del recurso, aportaron elementos para mejorar los índices de calidad de los dataset consultados. Por otro lado, estos descriptores adicionales agregaron información a los recursos más allá de requisitos particulares de una organización, haciéndolos extensibles a las características que requiera cualquier organización que ingrese al mundo del modelado y enriquecimiento de recursos educativos; además de proveer información que responda a principios de datos abiertos como la oportunidad.

Después de haber documentado el punto de vista independiente de la computación, a continuación se describen los artefactos tenidos en cuenta desde el punto de vista independiente de la plataforma.

5.3. Nivel M1: Perfil de Aplicación – Recursos Educativos Abiertos (PA-REA).

Considerando que esta investigación no genera una herramienta de software, para el punto de vista independiente de la plataforma no se describieron operaciones o funcionalidades en particular. Por el contrario, se hizo uso de lenguajes de modelado de propósito general para visualizar los artefactos que se produjeron en las fases descritas a continuación. Aclarado lo anterior, acorde a la revisión de literatura realizada y las condiciones iniciales definidas, se determinó que los diferentes tipos de REA son descritos con un mismo conjunto de características. Con relación a las particularidades de cada recurso, éstas son descritas por características calificadoras del recurso, cuyo contenido corresponde a elementos de vocabularios definidos para estas particularidades. Considerando lo anteriormente descrito, la Figura 15 presenta las actividades que se desarrollaron para llevar a cabo la construcción del PA-REA.

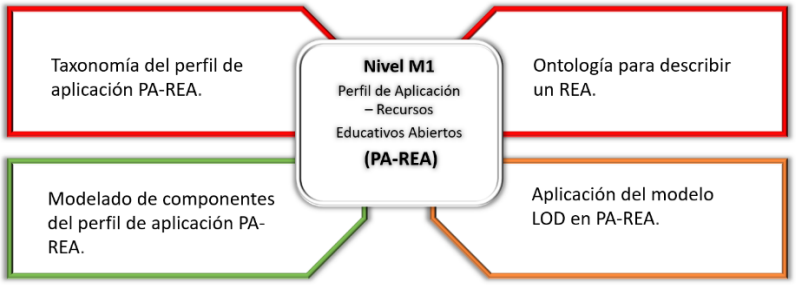


Figura 15. Nivel M1 del Metamodelo: PA-REA.

Como primer paso para diseñar el perfil de aplicación, se identificaron los términos (en particular los nombres de los conceptos y de las relaciones) utilizados para describir los REA a publicar. Este conjunto de términos constituyen la taxonomía de un REA.

5.3.1. Taxonomía del perfil de aplicación PA-REA.

Para la construcción de esta taxonomía, se usó el diagrama de clases como método para el refinamiento de atributos. El eje principal de la taxonomía para el perfil de aplicación PA-REA, se basó en la entidad fundamental REA. Esta entidad fundamental provee una serie de términos organizados en un esquema jerárquico y por tanto, en categorías y subcategorías (Codina y Pedraza, 2017), que permiten su identificación y reutilización. En términos generales, y

considerando los enfoques de disponibilidad y capacidad de reutilización (Abella *et al.*, 2018) de los REA:

a. Pueden existir varias versiones de un recurso educativo (por ejemplo, una versión en inglés y una versión en francés). Cada versión está potencialmente disponible en diferentes formatos (por ejemplo, en especificaciones IMS (Global Learning Consortium, Inc.) o como paquete SCORM - Shareable Content Object Reference Model. Finalmente, puede haber varias copias de una versión de recurso educativo dada en un formato dado (Massart y Shulman, 2011).
b. Por otro lado, un recurso educativo debe proveer una descripción detallada de los derechos provistos para su uso.
c. De igual forma, debe detallar las características educacionales que aborda, junto con una clasificación en el área y programa trabajado.
d. Adicionalmente, y considerando las dimensiones de calidad de datos vinculados, debe proveer información necesaria para el manejo de su procedencia.

Por ende, la descripción de los REA es fundamental para el descubrimiento y el intercambio de éstos, dado que proporcionan un medio para distinguir entre diferentes aspectos de los recursos. Para la construcción de esta taxonomía se tuvieron en cuenta:

- Los metadatos comunes identificados en la revisión de literatura,
- Los metadatos sugeridos como adicionales en la revisión de literatura,
- Los metadatos identificados como necesarios según las dimensiones de calidad.

Con relación a la descripción anterior, la Figura 16 muestra un diagrama de clases que detalla el refinamiento de los atributos realizado para describir los REA, que constituye la taxonomía propuesta en esta investigación.

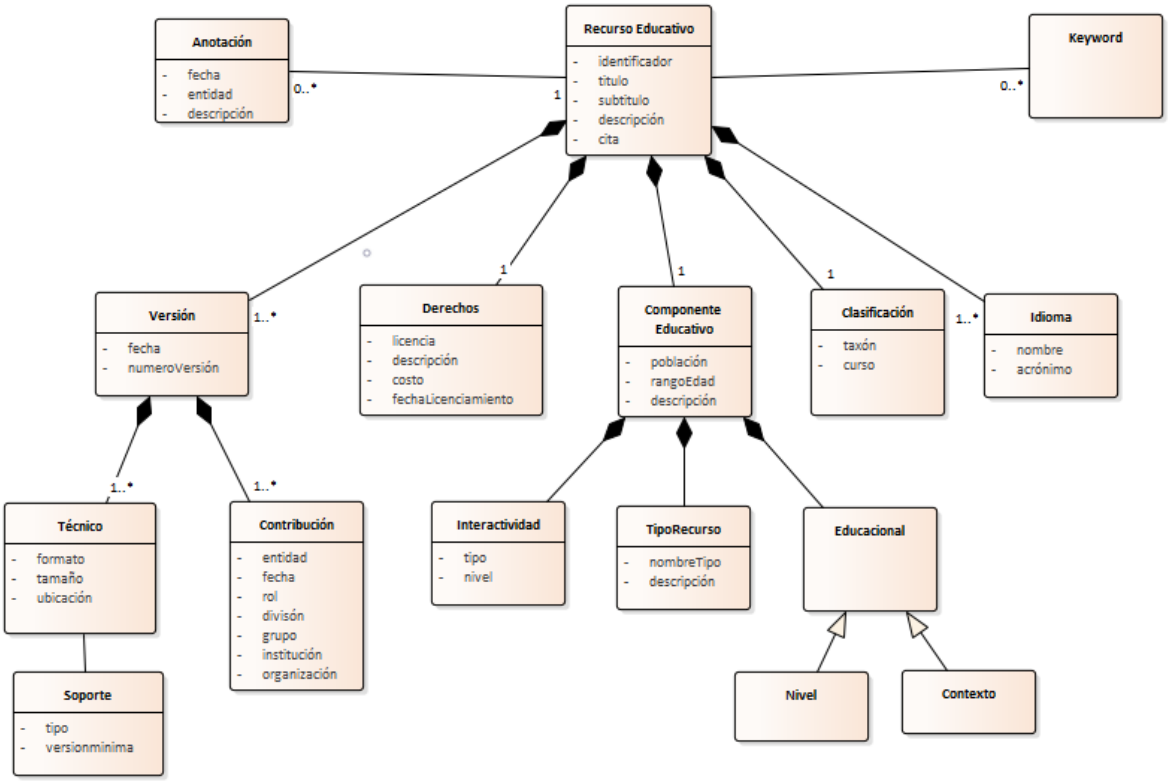


Figura 16. Taxonomía PA-REA.

Esta taxonomía se desarrolló con base en la construcción teórica realizada, y adicionalmente se tomó como base los elementos del esquema IEEE LOM (Wikipedia, 2011), con el fin de construir y adaptar la ontología que describe y representa el dominio de conocimiento, permitiendo de esta manera formalizar la descripción de conceptos y relaciones inherentes al perfil de aplicación PA-REA.

5.3.2. Ontología para describir un REA.

Para el desarrollo de la ontología se realizó un estudio previo de los esquemas utilizados para definir los REA (Anexo 4). En cuanto a los términos utilizados para describir la información, las buenas prácticas LOD recomiendan el uso de vocabularios existentes tanto en el ámbito del dominio de conocimiento como en la Web Semántica. De igual forma, y con el fin de mejorar la interoperabilidad y la compatibilidad semántica de la información, se recomienda el uso de esquemas de metadatos ampliamente reconocidos.

Por lo tanto, para la definición de esta ontología, se utilizaron términos extraídos de IEEE LOM. Ahora bien, dado que el perfil de aplicación PA-REA tiene como base el esquema IEEE LOM (IEEE, 2002) – como modelo de datos para la definición del esquema de metadatos-; PA-REA está formado por categorías acorde a sus funcionalidades, donde cada una de ellas está compuesta de elementos de metadatos que describen los REA. Bajo estos criterios, el contenido de PA-REA corresponde al refinamiento de siete componentes de metadatos, de acuerdo a la exploración de literatura realizada sobre esquemas de metadatos y de perfiles de aplicación, cruzada con los requerimientos de calidad de datos enlazados trabajados en esta investigación. Estos componentes son descritos en la Figura 17.



Figura 17. Perfil de Aplicación para REA.

Para particularizar la información provista por cada REA, las categorías PA-REA poseen calificadores que describen de forma más refinada los recursos. Estas categorías de PA-REA,

además de organizar los calificadores, permiten al metamodelo clasificar un nuevo calificador en el caso de plantearse la necesidad de ser incluido. La Figura 18 detalla los elementos calificadores de cada componente del perfil de aplicación PA-REA.

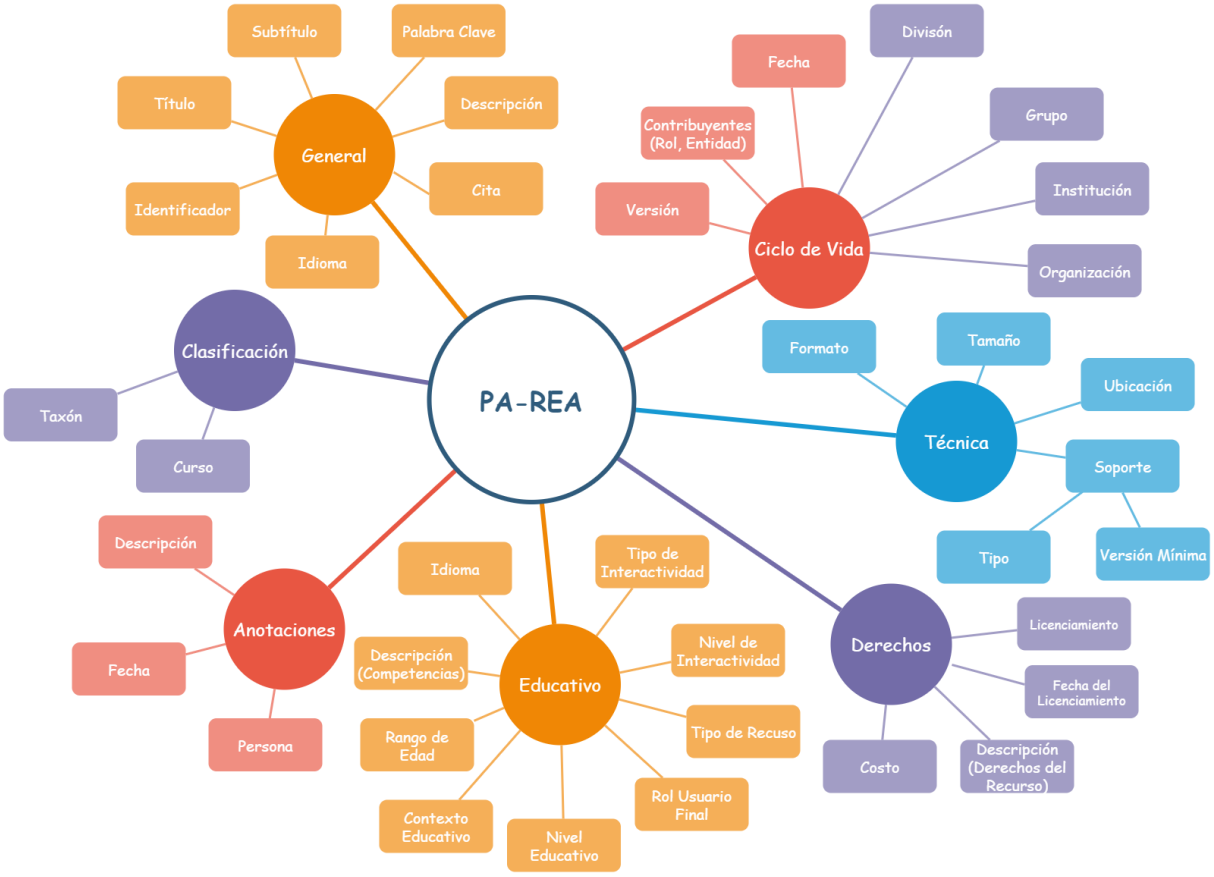


Figura 18. Elementos y Calificadores PA-REA.

Acorde a lo descrito por los autores Fermoso *et al.* (2008), el ámbito trabajado corresponde al modelo para describir y enriquecer REA, parametrizado por el estándar IEEE LOM, donde los términos definidos por esta ontología coinciden con las categorías de metadatos definidas por el estándar IEEE LOM.

5.3.3. Modelado de componentes del perfil de aplicación PA-REA.

Con base en la ontología que describe los REA, propuesta por el perfil de aplicación PA-REA, la Tabla 26 presenta cada uno de los elementos y calificadores de PA-REA, junto con su criterio de obligatoriedad y alineación con IEEE LOM.

Tabla 26. Detalle del diccionario de Datos de PA-REA.

Elemento	Descripción	Criterio
Identificador	GENERAL.IDENTIFIER - LOM 1.1	Ob
Título	GENERAL.TITLE - LOM 1.2	Ob
Subtítulo	GENERAL.DESCRIPTION.SUBTITLE - LOM 1.4	Op
Palabras Claves	GENERAL.KEYWORD - LOM 1.5	Ob
Descripción (Abstract)	GENERAL.DESCRIPTION - LOM 1.4	Ob
Cita	GENERAL.DESCRIPTION.CITE - LOM 1.4	Op
Idioma General	GENERAL.LANGUAGE - LOM 1.3	Ob
Versión	LIFE CYCLE.VERSION – LOM 2.1	Ob
Fecha Versión inicial	LIFE CYCLE.CONTRIBUTE.DATE – LOM 2.3.3	Ob
Contribución	LIFE CYCLE.CONTRIBUTE – LOM 2.3	Re

Elemento	Descripción	Criterio
Rol	LIFE CYCLE.CONTRIBUTE.ROLE – LOM 2.3.1	Ob
Entidad	LIFE CYCLE.CONTRIBUTE.ENTITY – LOM 2.3.2	Ob
Fecha del Ciclo de Vida	LIFE CYCLE.CONTRIBUTE.DATE – LOM 2.3.3	Ob
Formato Técnico	TECHNICAL.FORMAT – LOM 4.1	Ob
Tamaño del Formato	TECHNICAL.SIZE – LOM 4.2	Re
Ubicación Técnica	TECHNICAL.LOCATION – LOM 4.3	Ob
Tipo Técnico	TECHNICAL.REQUIREMENT.ORCOMPOSITE.TYPE – LOM 4.4.1.1.	Re
Versión Mínima	TECHNICAL.REQUIREMENT.ORCOMPOSITE.MINIMUNVERSION – LOM 4.4.1.3.	Re
Tipo de interactividad	EDUCATIONAL.INTERACTIVITYTYPE – LOM 5.1	Re
Nivel de interactividad	EDUCATIONAL.INTERACTIVITYLEVEL – LOM 5.3	Op
Tipo de Recurso	EDUCATIONAL.LEARNING RESOURCE TYPE – LOM 5.2	Ob
Rol Usuario Final	EDUCATIONAL.INTENDED END USER ROLE - LOM 5.5	Ob
Nivel Educativo	EDUCATIONAL.CONTEXT - LOM 5.6	Re
Contexto Educativo	CLASSIFICATION.DESCRIPTION – LOM 9.2	Re
Rango de Edad	EDUCATIONAL.TYPICAL AGE RANGE - LOM 5.7	Re
Descripción Educativa (Competencias)	EDUCATIONAL.DESCRIPTION – LOM 5.10	Ob
Idioma Educativo	EDUCATIONAL.LANGUAGE – LOM 5.11	Ob
Licenciamiento y Otras Restricciones	RIGHTS.COPYRIGHT AND OTHER RESTRICTIONS - LOM 6.2	Ob
Descripción de los Derechos	RIGHTS.DESCRIPTION - LOM 6.3	Ob
Costo de los Derechos	RIGHTS.COST - LOM 6.1	Op
Fecha del licenciamiento	RIGHTS.DATE - LOM 6	Ob
Persona	ANNOTATION.ENTITY – LOM 8.1	Op
Fecha	ANNOTATION.DATE – LOM 8.2	Op
Descripción	ANNOTATION.DESCRIPTION – LOM 8.3	Op
División	LIFECYCLE.CONTRIBUTOR.DIVISION – LOM 8.3	Op
Taxón	CLASSIFICATION.TAXON PATH.TAXON – LOM 9.2.2	Ob
Curso	CLASSIFICATION.PURPOSE – LOM 9.1	Op
Institución	LIFECYCLE.CONTRIBUTOR.INSTITUTION – LOM 8.3	Ob
Organización	LIFECYCLE.CONTRIBUTOR.ORGANIZATION – LOM 8.3	Ob
Grupo	LIFECYCLE.CONTRIBUTOR.GROUP – LOM 8.3	Op

Ob: Obligatorios. Re: Recomendados. Op: Opcionales.

Es importante anotar que atendiendo a lo descrito por autores como MEN (2012), Sánchez-Alonso *et al.* (2007), Smothers y Giordano (2015), Wason (2000), Friesen (2004), Koutsomitropoulos *et al.* (2010) y Bustos et al. (2016), para algunos elementos de metadatos se tomaron las siguientes decisiones:

a. Para “Fecha de versión inicial”, “Curso” y “Contexto educativo”, dado que dentro de IEEE LOM no se identificaron elementos que se ajustaran al requerimiento de forma directa, se optó por alternativas dentro del esquema que permitieron su descripción, para lo cual se

utilizaron las categorías de Ciclo de Vida y Clasificación, dado que el propósito de esta última categoría es proporcionar un lugar para colocar información que describa el recurso de una manera que no se proporciona en la estructura de metadatos en otro lugar (Wason, 2000).

b. Para “Subtítulo”, “Cita”, “Fecha de Licenciamiento”, “División”, “Institución”, “Organización” y “Grupo” se adicionaron elementos al esquema IEEE LOM con el fin de complementar la descripción de estos contribuyentes y la descripción general del REA.

Para revisar a mayor detalle la definición de los calificadores, las pautas para su diligenciamiento y los vocabularios definidos para algunos de estos calificadores, el perfil de aplicación PA-REA y su correspondiente diccionario de datos se detallan en el Anexo 5.

Seguidamente a la construcción de la ontología, y con el fin de enriquecer la descripción de los REA, se planteó el proceso de aplicación del modelo LOD al perfil de aplicación PA-REA.

5.3.4. Aplicación del modelo LOD en PA-REA.

Para este apartado se tuvo en cuenta que el metamodelo diseñado se orienta a generar modelos de datos para el enriquecimiento de REA bajo especificaciones LOD. Por ende, el perfil de aplicación diseñado tiene como finalidad la publicación de REA atendiendo a parámetros del modelo de Linked Open Data. Para tal fin, se presenta cada uno de los 5 niveles del modelo LOD y su tratamiento en el perfil de aplicación PA-REA:

a. ***Nivel No. 1: Publicar los datos en la Web bajo una licencia abierta***: Para este nivel, el perfil de aplicación PA-REA recomienda que el productor o proveedor del REA publique el recurso en la Web, haciendo uso de un licenciamiento que garantice la apertura del recurso. Para ello se puede hacer uso de plataformas de gestión de recursos o repositorios de recursos.

   El diseño didáctico y pedagógico del REA, la definición y aplicación de los requerimientos contractuales, legales, licenciamiento, etc., necesarios para abrir el recurso educativo no corresponden al alcance de ésta investigación. Sin embargo, para una adecuada aplicación del modelo LOD, tanto el recurso como el modelo de datos que lo describe, deben contar con condiciones de apertura tales como metadatos debidamente diligenciados y actualizados, el REA debe estar publicado bajo una licencia abierta, tanto el REA como los metadatos deben estar disponibles en línea, la estructura del modelo de metadatos debe ser pública y accesible, entre otras condiciones necesarias para el adecuado trabajo.

b. ***Nivel No. 2: Publicar los metadatos en la Web bajo un formato estructurado***: Para este nivel, el perfil de aplicación PA-REA recomienda que el productor o proveedor del REA proporcione los metadatos, ajustados al modelo diseñado, en un archivo, por ejemplo, con formato Excel. Este conjunto de metadatos puede desplegarse como información embebida en una página HTML, permitiendo al consumidor tener acceso a ellos y explorarlos.

c. ***Nivel No. 3: Publicar los metadatos en la Web bajo un formato no propietario***: Para este nivel, el perfil de aplicación PA-REA recomienda que el productor o proveedor del REA proporcione los metadatos, ajustados al modelo diseñado, en un archivo no propietario, por ejemplo CSV.

Estos tres niveles contribuyen a mejorar la disponibilidad del recurso, al contar con un esquema de metadatos mejor descrito y apoyado en elementos de calidad de datos enlazados. Adicionalmente, las recomendaciones de metadatos propuestas por PA-REA contribuyen a mejorar la confianza de los usuarios al proporcionar elementos relevantes y oportunos para las consultas y la toma de decisiones de los usuarios consumidores. Por último, es importante tener

en cuenta que este primer mapeo corresponde a una estructura de metadatos mapeada en Dublin Core, denominado mapeo PA-REA para Aplicación.

En cuanto a la reutilización de los REA, y atendiendo a los constructos definidos para el metamodelo, con el fin de avanzar en los dos siguientes niveles del modelo LOD, el perfil de aplicación PA-REA debe ser mapeado al modelo de datos RDF atendiendo los estándares sugeridos, generando como resultado un archivo denominado dataset. Este proceso configura un mapeo PA-REA para Dataset. Las consideraciones tenidas en cuenta son:

d. ***Nivel No. 4: Use URIs que identifiquen los recursos***: Para este nivel, el perfil de aplicación PA-REA requiere que el productor o proveedor identifique en la Web cada REA con una URI. Esta URI debe ser alimentada adicionalmente en la estructura de metadatos del recurso. De igual forma, los datasets publicados en la Web deben ser identificables, con el fin de poder ser enlazados.

e. ***Nivel No. 5: Enlace los recursos con otros similares para proporcionar contexto facilitando la reutilización.*** Para tal fin, al construirse el mapeo de los metadatos al modelo de datos RDF, los objetos de algunas tripletas serán definidos como URI, con el fin de enlazar información que enriquece y proporciona contexto a la descripción del recurso en mención. Atendiendo a la completitud de interconexión (dimensión de calidad contextual) y la dimensión de relevancia, se identificaron elementos de metadatos susceptibles de ser enriquecidos a través del enlace de datos, con base en el quinto nivel del modelo de Linked Data. Para realizar esta tarea, se planteó:

- La generación de ejemplos de diligenciamiento de los metadatos, con el fin de detectar campos susceptibles de ser enriquecidos con otros datos que les provean contexto.
- La exploración de estos contextos en diferentes fuentes y repositorios como Wikipedia, Europeana, Datahub, entre otros. Para tal fin se tuvo en cuenta el uso de fuentes que aporten credibilidad y confianza a los usuarios sobre la información que se está enlazando para proporcionar contexto. Por ende, se descartaron fuentes como por ejemplo perfiles de Facebook para complementar y enriquecer la información de los autores.

La Tabla 27 ilustra ejemplos de diligenciamiento de los metadatos que describe un REA.

Tabla 27. Ejemplos de diligenciamiento de metadatos PA-REA.

Elemento	L	V	U	Ejemplo
Identificador	X			ISBN 978-3-16-148410-0
Titulo	X			Introducción a las Vocales
Subtitulo	X			Explorando las Vocales
Palabras Claves		X		Vocales, vocablos, letras.
Descripción (Abstract)	X			Recurso dirigido a niños en edad escolar con interés en adquirir los conceptos básicos de identificación y pronunciación de las vocales.
Cita	X			ELIASH, Humberto y TELLEZ, Andrés. Estaciones de servicio: combustibles, comestibles, arquitecturales. Diseño, 7(46): 102-107, Noviembre 1997.
Idioma General		X		ESP
Fecha versión inicial	X			11/10/1998
Versión	X			Versión 1.2
Contribución – Rol		X		Autor
Contribución - Entidad			X	https://www.wikidata.org/wiki/Q5878
Contribución – Fecha	X			11/10/2000
Formato Técnico		X		Animación/multimedia. Flash

Elemento	L	V	U	Ejemplo
Tamaño del Formato	X			225 MB
Ubicación Técnica			X	https://www.curriki.org/oer/Suma-de-los-%C3%A1ngulos-de-un-tri%C3%A1ngulo-642182
Tipo Técnico	X			Sistema operativo Linux
Versión Mínima	X			Sistema operativo Windows 10
Tipo de interactividad		X		Descriptivo
Nivel de interactividad		X		Media
Tipo de Recurso		X		Juego educativo
Rol Usuario Final		X		Estudiante
Nivel Educativo		X		Básica Primaria
Contexto Educativo		X		Educación Formal
Rango de Edad	X			U-12
Descripción Educativa	X			Este recurso permite a los estudiantes identificar y analizar las principales funciones de edición de texto provistas por el editor de texto.
Idioma Educativo				ESP
Licenciamiento y Otras Restricciones		X		Si
Descripción de los Derechos			X	Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). https://creativecommons.org/licenses/by-nc-sa/4.0/
Costo de los Derechos		X		No
Fecha de Licenciamiento	X			11/10/2000
Persona			X	https://www.udistrital.edu.co/
Fecha	X			11/11/1976
Descripción				Este instrumento evalúa e aprendizaje llevado a cabo en las lecciones anteriores
Institución				Universidad
Organización				https://www.udistrital.edu.co/
Grupo				Programa
División				Ingeniería de Sistemas https://ingenieria.udistrital.edu.co/mod/page/view.php?id=233
Taxón				Programación
Curso				Programación Avanzada

L: Literal. V: Vocabulario. U: URI

De igual forma, y atendiendo a la estructura provista por el modelo de datos RDF, el comportamiento del dataset como sujeto u objeto de enlace debe ser registrado al interior de la estructura del dataset.

En resumen, la Tabla 28 define cuales elementos de los metadatos se plantearon como literales y cuales como URI. Aquellos metadatos que se identificaron como URI, cumpliendo con lo expuesto en el nivel 5 del modelo LOD, se enlazan a otros recursos para proporcionar mayor contexto al recurso en cuestión. Adicionalmente, al ser identificado con una URI, el REA puede ser enlazado por metadatos de otros recursos, lo cual contribuye a incrementar la reutilización y la visibilidad del mismo.

Tabla 28. Calificadores con contenido Literales y URI.

Categoría	Metadatos Literales	Metadatos URI
General	Identificador, Titulo, Subtitulo, Palabras Claves, Descripción, Idioma, Cita	

Categoría	Metadatos Literales	Metadatos URI
Ciclo de vida	Versión, Fecha, Fecha de versión inicial	Contribuyentes (Entidad)
Técnica	Formato, Tamaño, Tipo, Versión mínima	Ubicación técnica(URI)
Educativo	Tipo de Interactividad, Nivel de Interactividad, Tipo de recurso, Población o usuario final, Nivel Educativo, Contexto Educativo, Rango de edad, Descripción, Idioma	
Derechos	Derechos provistos por el Licenciamiento, Costo, Fecha del licenciamiento	Licenciamiento
Anotación	Fecha, Descripción.	Persona
Clasificación	Institución, Grupo, Taxón, Curso	Organización, División

Esta diferenciación aporta mayor enriquecimiento semántico a los REA publicados en la Web, al enlazar información que no solamente complementa al recurso, sino que brinda información oportuna y de interés al usuario consumidor del dataset. Por otro lado, el mapeo al modelo de datos RDF permite que el dataset pueda ser enlazado a The Linked Open Data Cloud (McCrae, 2020), cumpliendo con las restricciones que para ello se han definido, considerando que como requisito base se deben proveer http: // (o https: //) URI resolubles y los datos RDF deben proveerse en uno de los formatos RDF populares (RDFa, RDF / XML, Turtle, N-Triples).

Como lo describe la tabla anterior, en cuanto a los metadatos a enriquecer con URIs, se especificaron:

a. El autor principal: El autor principal se identifica con un nombre, sin embargo, dicho autor maneja una URI que le permite ser identificado en la Web. Ejemplos de esta identificación en la Web pueden ser: perfiles de CvLAC - Currículum Vitae de Latinoamérica y el Caribe, LinkedIn, entre otros. Por esta razón, el autor como contribuyente del recurso, aporta dos propiedades al modelo de datos. Los demás contribuyentes (publicador, revisor, sponsor y otros contribuyentes) se describen a través de su rol y un literal con su nombre.
b. El recurso puede estar almacenado localmente (referenciado con un handle o identificador) o en la Web (referenciado con una URI). Por tal razón, la ubicación técnica se maneja con una URI que referencia la ubicación del recurso.
c. Dentro de las dimensiones de calidad, además de proveer una descripción del tipo de licenciamiento, se requirió que el licenciamiento tuviera una URI que enlace una ampliación de información acerca de los derechos definidos para el tipo de licenciamiento otorgado.
d. En cuanto a las anotaciones realizadas a un recurso, además de proveer el tipo de anotación realizada y la fecha de su ejecución, se requirió complementar dicha anotación con la identificación de la persona o entidad que realizó la anotación. Para tal fin, y dado que las anotaciones proporcionan confianza acerca del tratamiento realizado al recurso, se decidió identificar a la persona que hace la anotación, a través de la URI de su perfil.
e. Por último, se consideró que la identificación institucional de los proveedores del recurso no solamente se basan en un nombre sino también en un tipo. Por lo tanto se adicionaron institución (tipo de entidad: Universidad, corporación, colegio, etc.), organización (nombre de la entidad), grupo (tipo de unidad dentro de la institución: programa, grupo de investigación, departamento, etc.) y división (nombre de la unidad dentro de la organización). Por otro lado, las organizaciones, grupos, etc., ahora cuentan con URIs que

los identifican en la Web, por lo cual se optó por trabajar estas propiedades como URIs y no como literales.

Después de aplicar el modelo LOD al esquema de metadatos que describen los REA, esquema definido por el perfil de aplicación PA-REA, se procedió a formalizar la ontología a través del mapeo de los metadatos a los modelos de datos codificados que permiten su procesamiento por otras aplicaciones.

5.4. Nivel M1: Mapeo del perfil de aplicación PA-REA.

Atendiendo a los enfoques de disponibilidad y reutilización de los REA, que guiaron esta investigación, se consideró que el perfil de aplicación PA-REA debía ser procesado por las aplicaciones que gestionan los metadatos, y adicionalmente ser exportado a grafos RDF, con el fin de ser enlazado a otros datasets, además de ser consumido a través de consultas SPARQL. Con base en ello, el proceso de mapeo del perfil de aplicación PA-REA se construyó para ser usado en dos momentos (Figura 19):

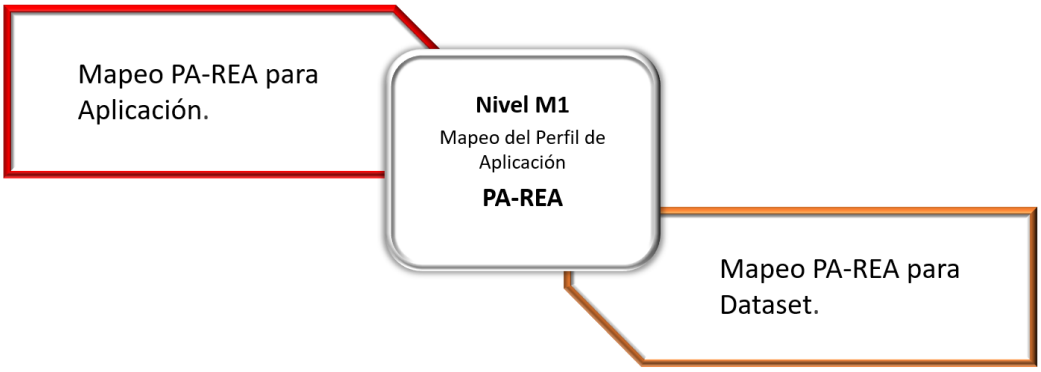


Figura 19. Mapeos del Perfil de Aplicación PA-REA.

A continuación se describen cada uno de estos mapeos, considerando la reutilización de ontologías ya existentes, que para el caso del mapeo para dataset se hizo necesario dado las restricciones de mapeo al modelo de datos RDF (véase Sección 5.2.9).

5.4.1. Mapeo PA-REA para aplicación.

Este mapeo se construyó para ser implementado en las plataformas que gestionan los REA. Estas plataformas gestionan un conjunto de propiedades de los REA, en un esquema definido, para capturar y almacenar su información. Dada la expresividad de los vocabularios seleccionados (correspondientes a Dublin Core Metadata Initiative), el perfil de aplicación PA-REA identificó alternativas de mapeo en diferentes propiedades y calificadores, para los elementos del esquema de metadatos. Considerando la ambientación realizada en el Anexo 6, la Tabla 29 muestra el ***Mapeo PA-REA para Aplicación***, expresado en Dublin Core, y definido para la descripción de los recursos educativos que alimenta las plataformas de gestión.

Tabla 29. Esquema de Mapeo PA-REA para Aplicación.

Elemento	Elemento DC
Identificador	dc.identifier.doi, dc.identifier.uri, dc.identifier.isbn, dc.identifier.issn
Título	dc.title
Subtítulo	dc.title.alternative
Palabras Claves	dc:subject
Descripción (Abstract)	dc.description.abstract
Cita	dc.identifier.citation

Elemento	Elemento DC
Idioma General	dc.language.iso
Fecha de Versión Inicial	dc.date.created
Versión	dc.description.version
Contribución Esta subdividida en: Rol, Entidad	dc.contributor, dc.contributor.author, dc.creator, dc.contributor.advisor, dc.description. sponsorship dc.publisher
Fecha del Ciclo de Vida	dc.date.available, dc.date.updated, dc.date.accessioned
Formato Técnico	dc:format
Tamaño del Formato	dc:format:extent
Ubicación Técnica	dc.description.uri
Tipo Técnico	dc.relation.requires
Versión Mínima	dc.relation.minimunversion
Tipo de interactividad	dc.description. interactivityType
Nivel de interactividad	dc.description.interactivitylevel
Tipo de Recurso	dc:type
Rol Usuario Final	dc.description.audience
Nivel Educativo	dc.description.educationlevel
Contexto Educativo	dc.description. KnowledgeGrouping
Rango de Edad	dc.description. typicalAgeRange
Descripción Educativa	dc:description
Licenciamiento y Otras Restricciones	dc:rights
Descripción de los Derechos	dc.rights.uri
Costo de los Derechos	dc. rights.cost
Fecha del Licenciamiento	dc.date.copyright
Persona (anotación)	dc. relation. annotationentity
Fecha (anotación)	dc. relation. annotationdate
Descripción (anotación)	dc.relation. annotationdesc
Institución	dc.relation.institution
Organización	dc.relation. organization
Grupo	dc.relation. group
División	dc.relation.division
Taxón	dc.relation. EducationalSubject
Curso	dc.relation.course

Es importante anotar que el ***Mapeo PA-REA para Aplicación*** proporciona un conjunto de aportes, dentro de las cuales se resaltan:

- Este mapeo es susceptible de ser usado por plataformas o repositorios, en la fase de ingesta de datos, dado que Dublin Core corresponde a uno de los vocabularios con mayor uso.
- Este mapeo se alinea con versiones reducidas e incompletas de mapeos realizados por otras instituciones, lo cual permite migrar sus datos hacia este modelo de datos que aporta elementos para incrementar la calidad de sus modelos.
- Además del uso de un estándar reconocido, este mapeo plantea una estandarización en la definición y categorización de calificadores, necesarios para describir los REA.
- Este mapeo también aporta la estandarización tanto en pautas de diligenciamiento como en vocabularios controlados, elementos necesarios para el adecuado diligenciamiento de las instancias de datos.

5.4.2. Mapeo PA-REA para Dataset.

El proceso de Metadata Crosswalk involucra el mapeo de los elementos, la semántica y sintaxis de un esquema de metadatos a otro, facilitando el intercambio de metadatos (Note, 2011). En términos generales, el mapeo de metadatos, como lo manifiesta Haslhofer (2008), consiste en:

> "*Desde la perspectiva del modelo, un mapeo de metadatos define las relaciones estructurales y semánticas entre los elementos del modelo a nivel de esquema y los valores de contenido a nivel de instancia. Para representar tales relaciones, cualquier mecanismo de mapeo requiere un conjunto de elementos de mapeo con una sintaxis y una semántica bien definidas*."

El proceso de Crosswalk cobró vital importancia al momento de enlazar y compartir recursos como datasets en la Web de los Datos. Por ello, este metamodelo aporta esta fase con el fin de que las descripciones de los REA, gestionadas por una plataforma local, sean enriquecidas y compartidas en la Web de los Datos, proporcionando sustento a los dos grandes enfoques de los datos enlazados: la disponibilidad y la reutilización (Abella *et al.*, 2018).

El Crosswalk diseñado en esta investigación, transforma o mapea los metadatos adquiridos en el proceso de ingesta de los REA (usando el Mapeo PA-REA para Aplicación), a un esquema de tripletas provistas por el modelo de datos RDF. Como resultado, este mapeo denominado ***Mapeo PA-REA para Dataset***, genera un dataset que es gestionado por un servidor SPARQL, y es consumido a través de consultas.

Dadas las restricciones planteadas por el modelo de datos RDF, los calificadores a un segundo y tercer nivel no son mapeados de forma simple en tripletas, por lo cual se ubicaron otros vocabularios, cuyos esquemas de descripción a primer nivel cumplen con estas restricciones. Para tal fin, el mapeo de los metadatos que describen los REA, definidos en el perfil de aplicación PA-REA, se basó en el uso de los vocabularios y ontologías como DC, DCTERMS, BIBO, FOAF, y LRMI. La ampliación de la construcción del Crosswalk PA-REA se encuentra descrito en el Anexo 6.

En cuanto a los prefijos definidos para el mapeo, éstos corresponden a las siguientes sentencias:

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix parea: <http://parea.com/parea#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix lrmi: <http://purl.org/dcx/lrmi-terms/> .
@prefix aiiso: <http://vocab.org/aiiso/schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ja: <http://jena.hpl.hp.com/2005/11/Assembler#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

Es importante resaltar que, derivado de la conjugación de las dimensiones de calidad de datos enlazados con la abstracción de metadatos realizada, se requirió complementar la información de procedencia. Este complemento se enfocó en las propiedades de institución, programa, área y materia, propiedades bajo las cuales se diseñó y construyó el recurso. Después de revisar diferentes ontologías de metadatos, se identificó a AIISO - Academic Institution Internal Structure Ontology, como una ontología que proporciona clases y propiedades para describir la estructura organizacional interna de una institución académica.

Por otro lado, después de llevar a cabo búsquedas en el catálogo de vocabularios disponibles para reutilizar denominado Linked Open Vocabularies – LOV (LOV, 2020), y dada la imposibilidad

de mapear algunos términos desde los vocabularios seleccionados, se usó un namespace adicional, tal como se presenta en la siguiente sentencia:

@prefix parea: < http://parea.com/parea#> .

La descripción de los términos del vocabulario PA-REA se encuentra en el Anexo 7. La Figura 20 describe el Mapeo PA-REA para Dataset, producto del proceso del Metadata Crosswalk PA-REA ejecutado sobre el modelo de datos PA-REA para aplicación. En este modelo se especificaron los respectivos prefijos y los esquemas utilizados, y de igual forma se especificaron los componentes definidos: recursos, literales, vocabularios y enlaces.

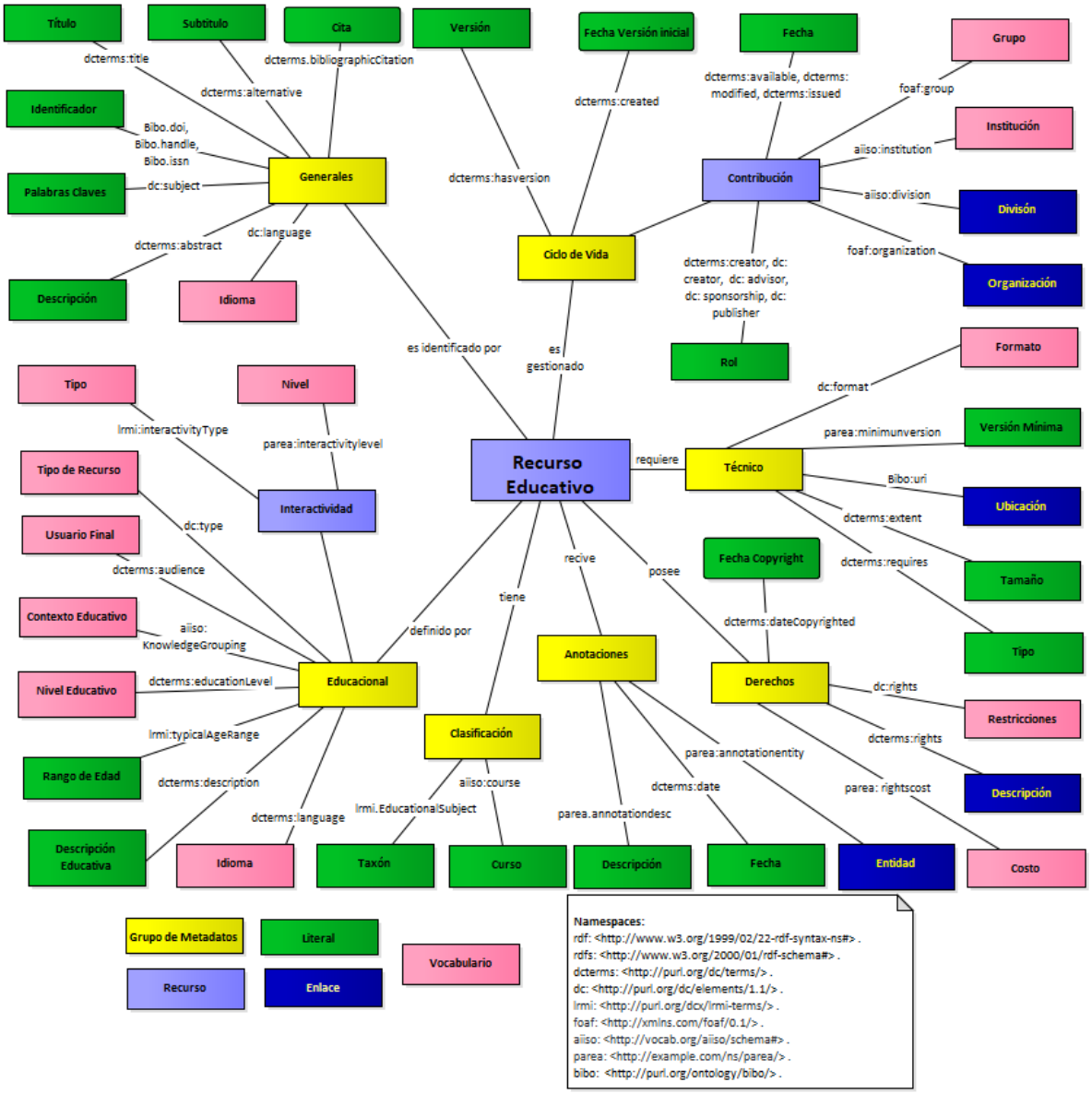


Figura 20. Mapeo PA-REA para Dataset.

La Figura 21 resume los principales procesos definidos en el nivel M1 del metamodelo diseñado, los cuales enmarcaron el modelo de representación de REA. Es importante resaltar que esta capa adoptó LOM como lenguaje de representación de metadatos, y Turtle como formato para expresar y serializar datos en RDF. Este último componente se alinea con los requerimientos de la nube LOD (McCrae, 2020) para el enlace de datos.

Acorde a la figura anterior, y dado que los elementos del nivel M1 del metamodelo son instancias del nivel M2, esta capa definió:

a. La estructura del perfil de aplicación que describe los REA, atendiendo a las directrices y restricciones de los estándares de metadatos, las dimensiones de calidad de datos trabajadas y los procesos de enlace de datos.
b. La estrategia de construcción de tripletas RDF que modela las propiedades y relaciones del perfil de aplicación.
c. La formalización de un modelo de representación de REA, el cual puede ser instanciado como modelo de datos para gestionar REA en plataformas o para gestionar datasets en un ambiente de datos enlazados.

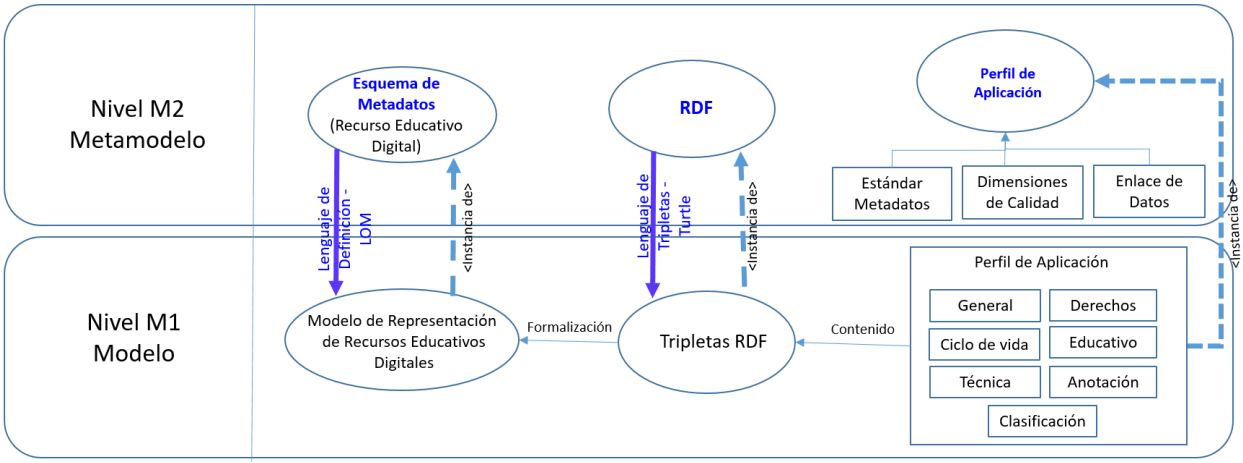


Figura 21. Nivel M2 – Modelo de Metadatos.

En resumen, el aporte de esta capa del metamodelo se centró en:

- Aportar un modelo de datos abierto, enriquecido y más completo, para la ingesta y gestión de los metadatos de los REA.
- Aportar un modelo de datos mapeado a Dublin Core, para mejorar la ingesta y gestión de los REA en las plataformas de gestión de recursos.
- Aportar un modelo de datos mapeado al modelo de datos RDF, para mejorar la disponibilidad y reutilización de los REA a través de la Web de los Datos.

La Figura 22 resume el proceso realizado y los artefactos obtenidos:

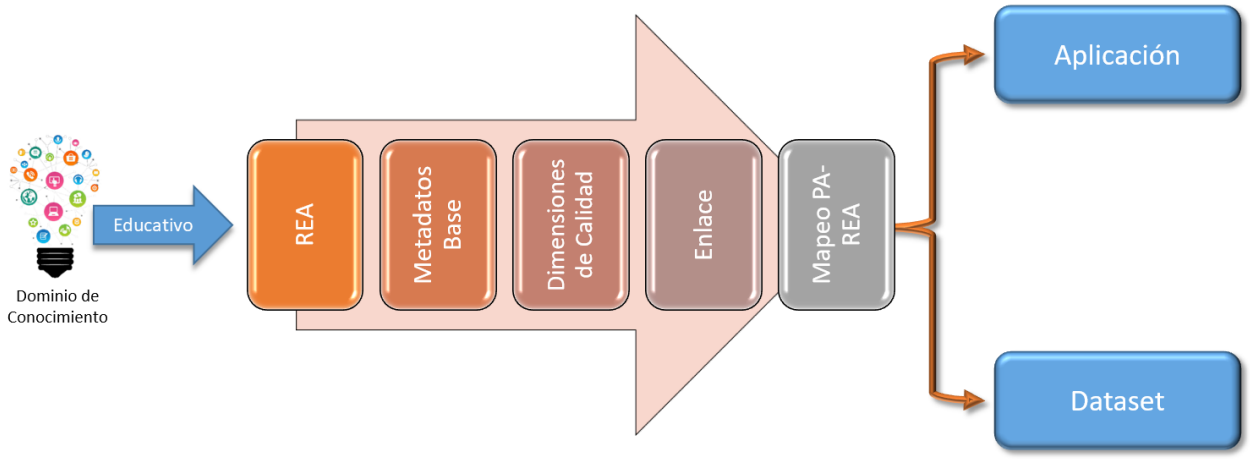


Figura 22. Mapeos del Perfil de Aplicación PA-REA.

Considerando los componentes y restricciones definidos en las capas M2 y M1 del metamodelo, a continuación de presenta la formalización de los mapeos provisto para el perfil de aplicación.

5.5. Nivel M1: Formalización del Metamodelo.

A continuación se expone la formalización del metamodelo a través de la aplicación de las diferentes directrices y restricciones, definidas por las capas superiores del metamodelo, para construir mapeos para aplicación y para dataset.

5.5.1. Formalización del mapeo PA—REA para aplicación.

Para formalizar la aplicación del modelo de datos, cuyas instancias generan REA descritos por un esquema de metadatos, se tuvieron en cuenta tres aspectos esenciales que son encapsulados por el modelo de representación:

- El esquema: Que para este caso corresponde a Dublin Core.
- La propiedad: Que corresponde a las características definidas por el perfil de Aplicación PA-REA.
- El calificador: Que corresponden a sub propiedades que especifican o refinan la información que contiene una propiedad, y que se encuentran definidas bajo el esquema Dublin Core.

Estos elementos son provistos en el perfil de aplicación PA-REA para aplicación, y se reflejan en instancias de REA, almacenados en las plataformas de gestión de recursos (Figura 23).

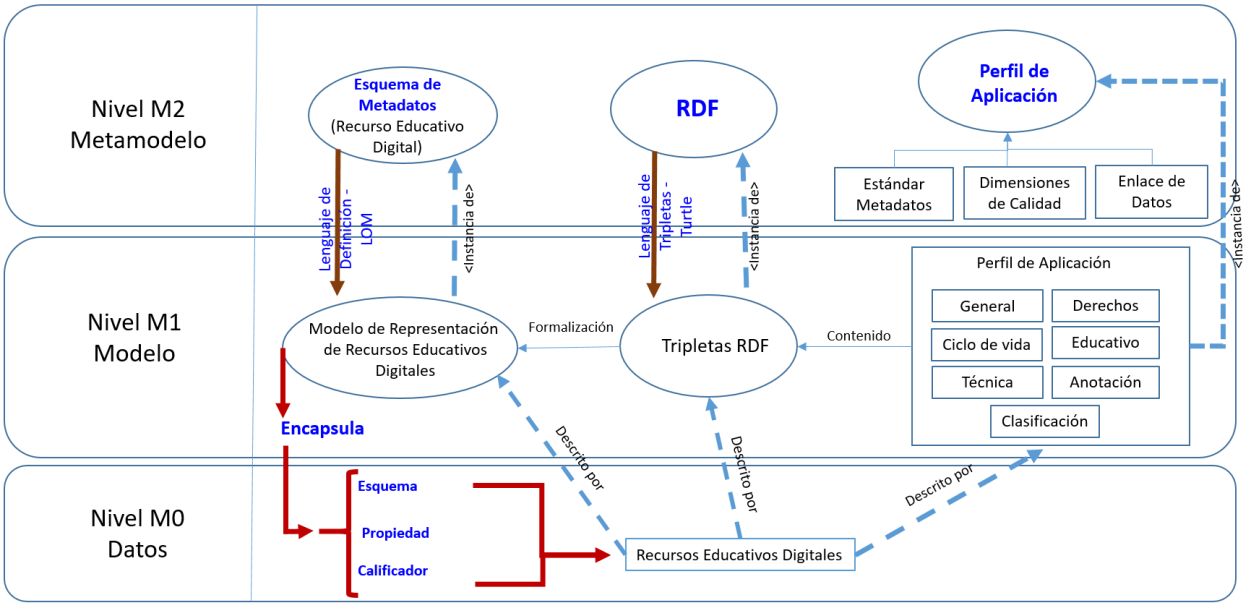
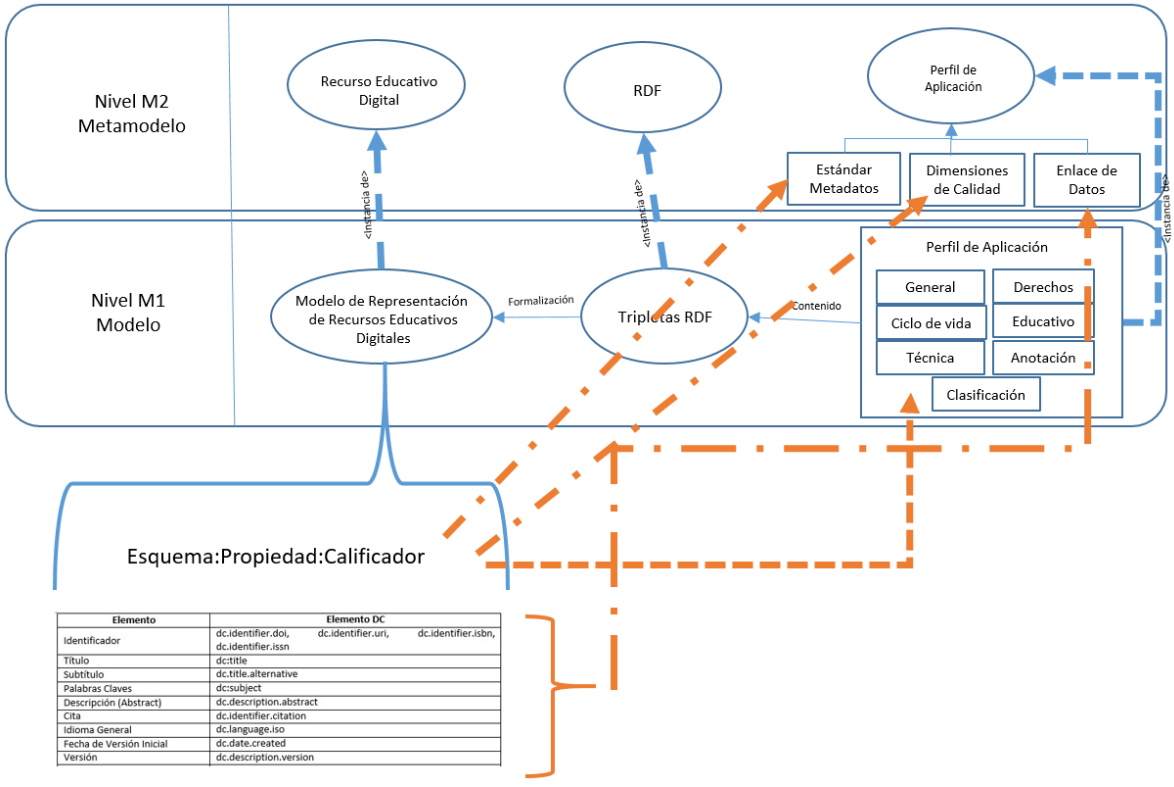


Figura 23. Nivel M0: Instancia de Datos para Aplicación.

Acorde a lo anterior, la Figura 24 muestra la estructura del esquema que deben seguir todos los tags de metadatos susceptibles de ser mapeados en Dublin Core.



Elemento	Elemento DC
Identificador	dc.identifier.doi, dc.identifier.uri, dc.identifier.isbn, dc.identifier.issn
Título	dc:title
Subtítulo	dc.title.alternative
Palabras Claves	dc:subject
Descripción (Abstract)	dc.description.abstract
Cita	dc.identifier.citation
Idioma General	dc.language.iso
Fecha de Versión Inicial	dc.date.created
Versión	dc.description.version

Figura 24. Esquema de Mapeo DC para Aplicación.

La figura anterior evidencia que los aspectos: esquema, propiedad y calificador, son parametrizados por las directrices y restricciones del estándar de metadatos y las dimensiones de calidad (Nivel M2). Ahora bien, las propiedades y los calificadores son definidos por el Perfil de Aplicación (Nivel M1). Por último, la identificación y tratamiento de aquellos datos susceptibles

de enriquecerse son parametrizados por el modelo de enlace de datos (Nivel M2). Lo anterior evidencia que la formalización expresada en el nivel M1 corresponde a instancias del nivel M2.

5.5.2. Formalización del Mapeo PA-REA para Dataset.

Para formalizar el mapeo del modelo de datos para Dataset, cuyas instancias generan metadatos mapeados en RDF, se tuvieron en cuenta dos aspectos esenciales:

- La sintaxis, vista como los elementos, las reglas y valores para describir el modelo de datos.
- La semántica, vista como los respectivos significados de las tripletas construidas, junto con el proceso de mapeo realizado.

Estos elementos son provistos en el perfil de aplicación PA-REA para dataset, y se reflejan en las tripletas formalizadas por RDF como metamodelo que parametriza el proceso de mapeo (Figura 25).

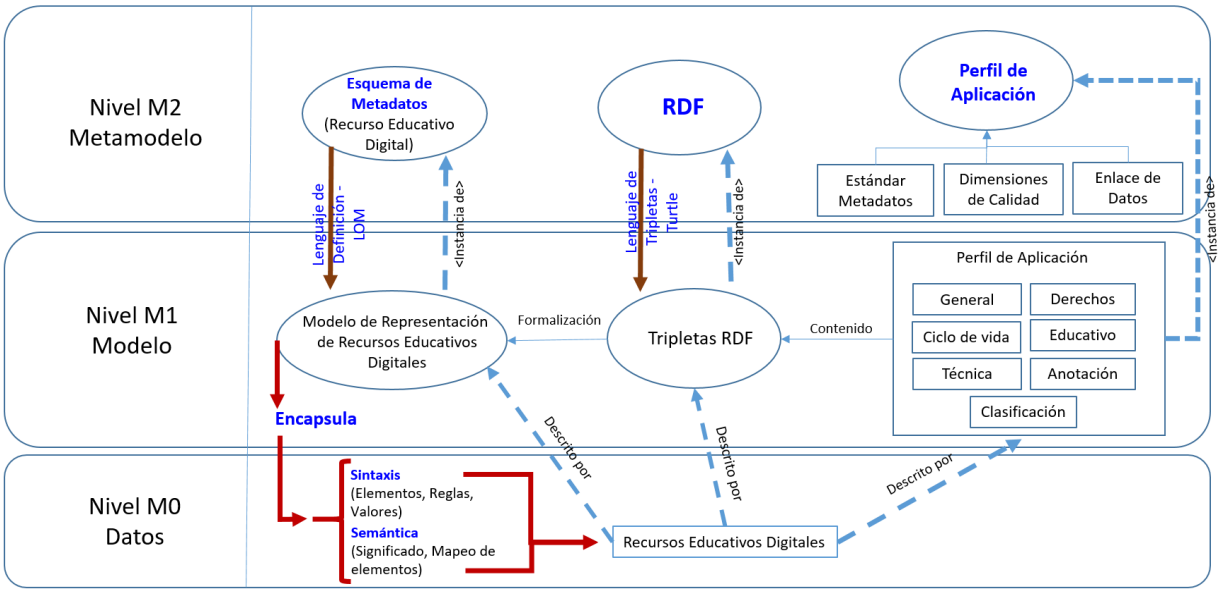


Figura 25. Nivel M0 – Instancias de Datos para Dataset.

Acorde a lo anterior, la Figura 26 muestra el protocolo que deben seguir todos los tags de metadatos susceptibles de ser mapeados en RDF.

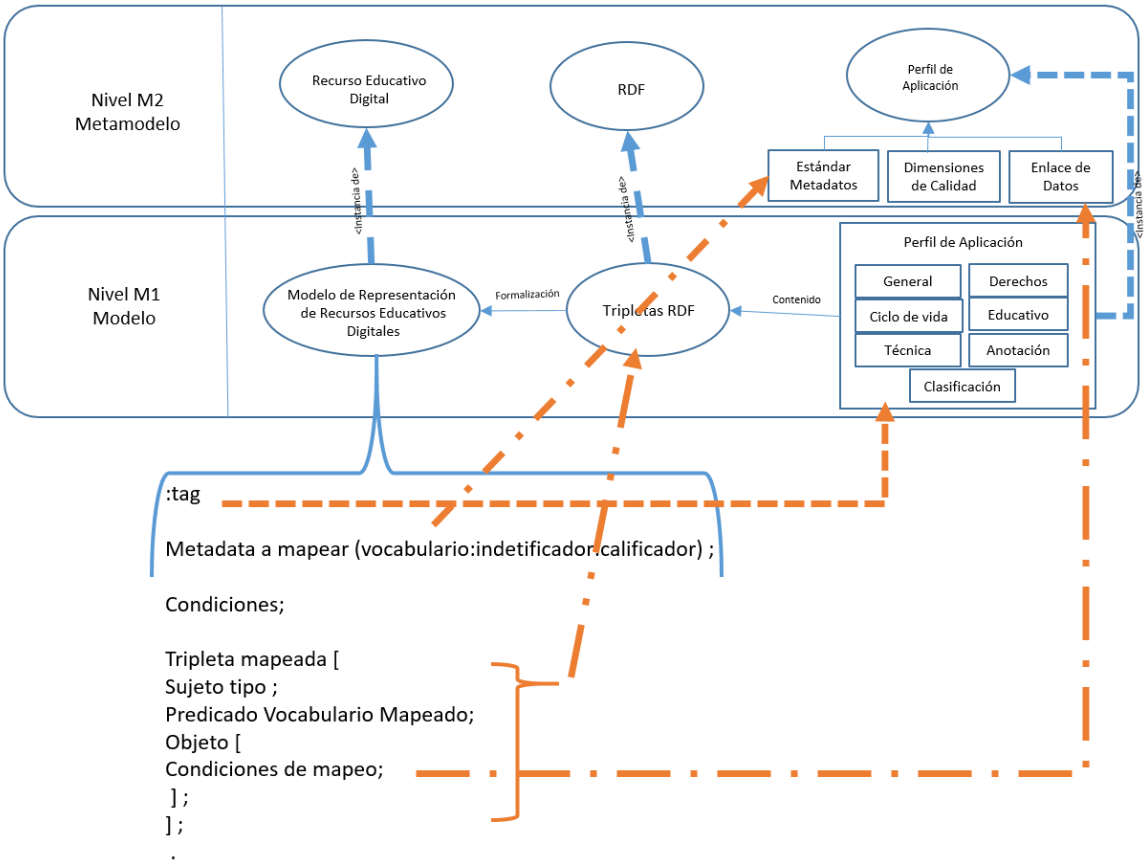


Figura 26. Esquema de mapeo RDF para PA-REA.

Dado que los REA son almacenados en plataformas que gestionan sus metadatos, el esquema de mapeo de la Figura 26 muestra la siguiente estructura de formalización:

a. Un :tag, que indica y diferencia el metadato a mapear.
b. Una construcción Vocabulario:identificador:calificador, que indica el esquema usado por la plataforma de gestión de recursos, junto con el identificador y calificador – característica o metadato, que se va a mapear. Esta construcción está definida por el perfil de aplicación PA-REA, como componente que debe ser parametrizado en la plataforma de gestión de REA para realizar el proceso de ingesta.
c. Unas Condiciones que debe cumplir dicha construcción, para iniciar su mapeo a RDF.
d. Un bloque de construcción de tripleta, en donde se definen el sujeto, el predicado y el objeto. En esta sección es importante aclarar que: i) el predicado corresponde a la relación provista por el vocabulario seleccionado; ii) el objeto puede ser parametrizado por condiciones que definan su construcción.

Este protocolo demuestra el uso y cumplimiento de las directrices provistas por las entidades definidas en el nivel M2 del metamodelo. Al seguir dichas directrices y restricciones, se generan modelos para la representación de REA en RDF.

5.6. Conclusiones del Capítulo.

El metamodelo diseñado en esta investigación consta de las siguientes capas (Figura 27):

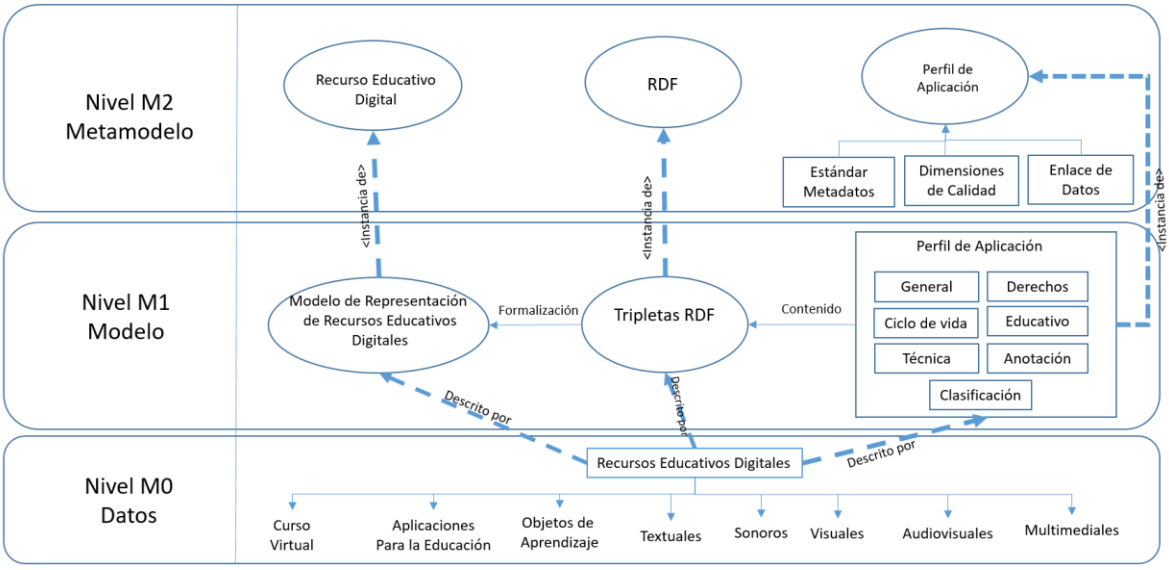


Figura 27. Metamodelo para la descripción y enriquecimiento de REA.

Estas capas cumplen las siguientes funciones:

i. ***El nivel M0*** consta de datos del mundo real, que corresponde a REA tales como curso virtual, aplicaciones para la educación, objetos de aprendizaje, textuales, sonoros, visuales, audiovisuales, multimediales. Estos recursos están expresados en instancias de metadatos, las cuales son instancias del modelo de representación de REA (Capa M1). Este modelo de representación refleja los elementos del esquema de metadatos PA-REA.

ii. ***El nivel M1*** consta de modelos, que se consideran instancias válidas del metamodelo. Este nivel incluye:

- La instancia del perfil de aplicación de metadatos, soportada en un lenguaje de definición de metadatos – LOM, y esquematizada en Dublin Core para formalizar la representación del recurso para aplicaciones. Este modelo de metadatos encapsula los elementos de:

  - El esquema: que para este caso corresponde a Dublin Core.
  - La propiedad: Que corresponde a las características definidas por el perfil de Aplicación PA-REA
  - El calificador: Que corresponden a sub propiedades que especifican o refinan la información que contiene una propiedad, y que se encuentran definidas bajo el esquema Dublin Core.

- Dicha instancia esta expresada en un modelo de datos RDF, que formaliza el mapeo a Dataset y la representación de un REA. Este modelo de metadatos encapsula los elementos de:

  a. Semántica (significado) formal, la cual se da a través del mapeo de los elementos del modelo a las entidades correspondientes en un determinado dominio de aplicación.
  b. Sintaxis (los elementos legales y las reglas sobre cómo y qué valores se pueden asignar a los elementos), para serializar las descripciones de metadatos. Esta sintaxis se define utilizando una cierta notación, que puede ser simbólica, textual o gráfica.

iii. ***El nivel M2*** consta de metamodelos, expresados en modelos de representación de REA, RDF y descripción de los recursos (metadatos). Cada uno de estos componentes tiene un

metamodelo, que determina no solo la estructura sino también las relaciones entre los tipos de entidad en un modelo (M2 en la figura).

En cuanto al Modelo de Representación de Recursos Educativos Abiertos, artefacto principal de este metamodelo; el modelo propuesto provee una mayor cantidad de metadatos a diligenciar, lo que conllevaría erróneamente a pensar que la inclusión de más descriptores a la abstracción de un REA se convierte en una tarea engorrosa de diligenciamiento, llegando al punto de contar con instancias de datos vacías o mal diligenciadas. Esta forma de ver la inclusión de nuevos metadatos, unida al tratamiento particularizado de los datos, son factores que han contribuido en detrimento no solo del enlace de los datos sino de la calidad de los datos abiertos en sí mismo.

Por el contrario, la estructura del modelo propuesto responde a requerimientos de calidad de datos enlazados, y a un diseño desprovisto de individualismos o particularidades adicionadas por las organizaciones. Esta investigación provee un modelo de metadatos flexible, con baja viscosidad, inmovilidad y rigidez gracias a la estructura de metamodelo utilizada para su construcción; con unos componentes claramente definidos que aportan valor no solamente a la apertura de datos, sino también al enriquecimiento y enlace de REA a la Web de los Datos.

CAPÍTULO VI IMPLEMENTACIÓN DEL MODELO DE DATOS.

En primera instancia, esta investigación presenta un metamodelo como resultado principal, componente abstracto que define las construcciones teóricas y reglas (Nivel M2) sobre las cuales se genera el modelo de datos (Nivel M1) para describir y enriquecer los REA. Previo a la validación del metamodelo fue necesario llevar a cabo la implementación de un ambiente tecnológico con el fin de validar el despliegue de uno de sus niveles: el M0, que corresponde a las instancias generadas de metadatos que describen los REA, considerando los mapeos propuestos: para aplicación y para dataset (Figura 28).

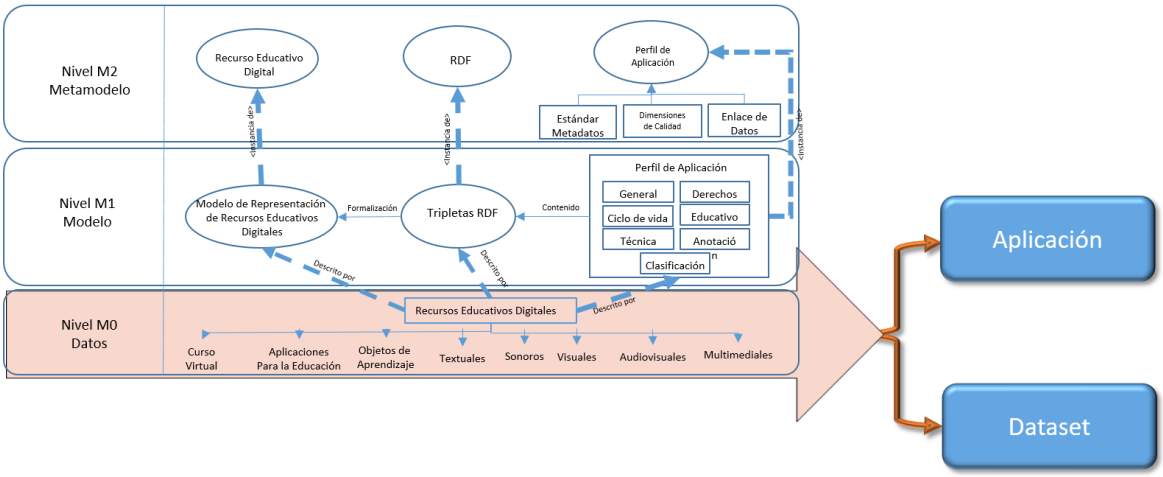


Figura 28. Nivel M0 para despliegue.

Por ende, este capítulo presenta el proceso de despliegue realizado para implementar el modelo de datos. Para tal fin, el capítulo se encuentra dividido en dos secciones principales: la primera sección corresponde al diseño y despliegue del ambiente tecnológico que soporta el modelo de datos. La segunda sección corresponde al despliegue del modelo de datos en la infraestructura configurada.

6.1. Despliegue del ambiente tecnológico.

Para el despliegue del ambiente tecnológico que permitió implementar el modelo de datos, se consideraron los siguientes componentes:

6.1.1. Plataformas seleccionadas.

Para desplegar la infraestructura que soporta el modelo de datos, se analizaron consideraciones expuestas por autores como Da Silva (2019), Fernández (2014), Jis (2020), Bodero *et al.* (2019) y Bankier y Gleason (2014). Estos autores examinaron diferentes herramientas teniendo en cuenta referentes de apertura, accesibilidad, soporte, madurez e interoperabilidad. De acuerdo al análisis realizado se utilizaron las siguientes plataformas:

i. Plataforma para la gestión de recursos. Estas plataformas recopilan, organizan y difunden la información generada, además de gestionar el acceso a los contenidos. Dado el enfoque de datos enlazados trabajado en esta investigación, uno de los criterios base para la selección de la plataforma consistió en que ésta debía permitir la interacción con datos enlazados a través del modelo de datos RDF.

Como plataforma para la gestión de recursos se seleccionó Dspace (Lyrasis, 2020a), dados los buenos puntajes obtenidos en cuanto a factores como infraestructura, gestión de contenido,

interfaces de usuario, administración de recursos, preservación, interfaces RDF, interoperabilidad e integración (Bankier y Gleason, 2014; Da Silva, 2019). Dspace es un software de código abierto, que recolecta, preserva y distribuye material digital. Así mismo, según el Directorio Global de Repositorios de Acceso Abierto - OpenDOAR (Jis, 2020), Dspace es la plataforma de gestión de recursos más utilizada, con un 39% del mercado de instalaciones en todo el mundo, instalaciones que representan una comunidad de usuarios activa y en continuo crecimiento (Figura 29).

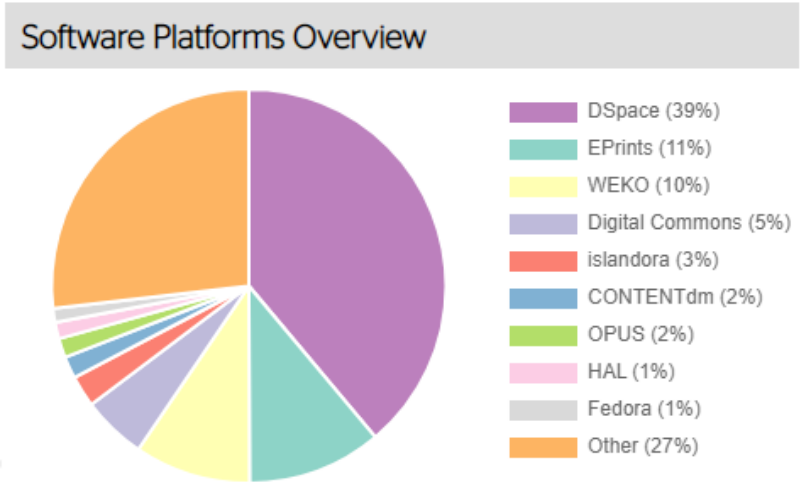


Figura 29. Plataformas de Gestión de Recursos 2020. Fuente: Jis (2020).

Esta plataforma para la gestión de recursos fue desarrollado por MIT Libraries (MIT, 2020) y Hewlett Packard Labs (HP Labs, 2020). Desde sus inicios cuenta con diferentes comunidades de desarrolladores, administradores de repositorios y otros interesados, que contribuyen al desarrollo de módulos y resolución de problemas. Dspace cuenta con una arquitectura que maneja comunidades, colecciones e ítems. Estas agrupaciones gestionan los diferentes recursos, además de ejecutar un proceso de ingesta y curación de datos (Figura 30).

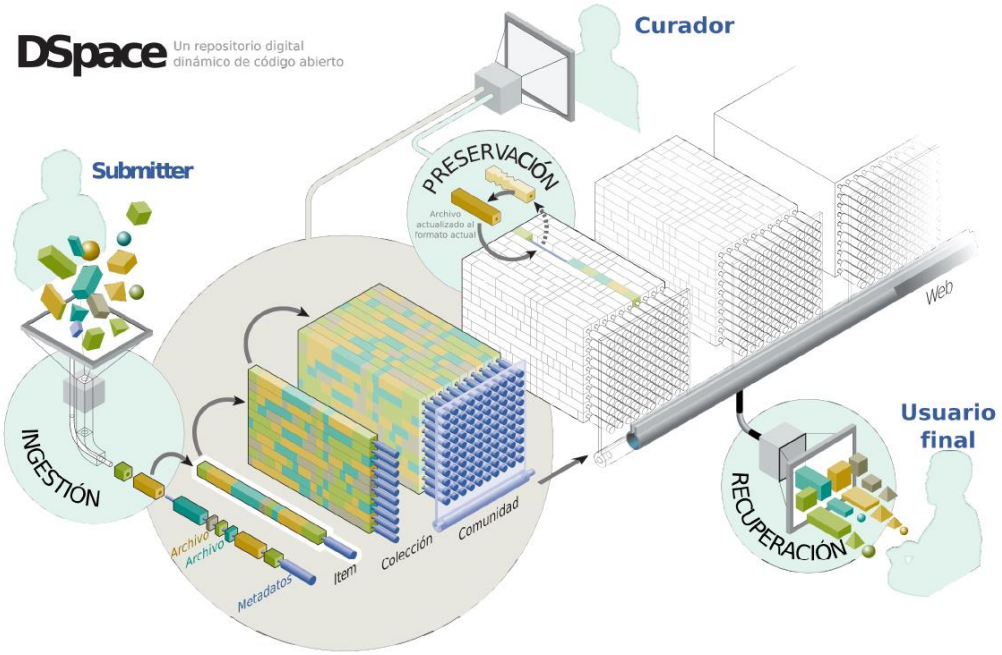


Figura 30. Arquitectura Dspace. Fuente: Donohue (2018).

De conformidad con alcances del proyecto y desde el contexto colombiano, Dspace es una plataforma de uso común en bibliotecas académicas y repositorios institucionales, como por ejemplo:

1. Repositorio Institucional de la Universidad del Norte (Colombia).
2. Repositorio Institucional de la Universidad de Cuenca (Ecuador).
3. Repositorio Institucional de la Universidad Distrital Francisco José de Caldas (Colombia).
4. Repositorio Institucional de la Pontificia Universidad Javeriana (Colombia).
5. Banco de Recursos Digitales del MEN (Colombia).
6. Difusión de los metadatos en los formatos ESE o EDM al proveedor OAI-PMH de Dspace.
7. Repositorio digital en la Universidad Central “Marta Abreu” de Las Villas (Cuba).

Dentro de las principales características de Dspace permite manejar múltiples tipos de archivos, reconociendo automáticamente los formatos más comunes (DOC, PDF, XLS, PPT, JPEG, MPEG, TIFF). Por otro lado, maneja Qualified Dublin Core – QDC (Lyrasis, 2020b) como metadatos predeterminados, esquema que se puede ampliar agregando esquemas personalizados similares a QDC. Por último, Dspace ha incorporado un proceso de serialización denominado RDFizer, el cual exporta su contenido a grafos RDF o RDF Graph.

ii. ***Servidor SPARQL.*** Para llevar a cabo la selección del servidor se examinaron estudios realizados por autores como Le Ngoc *et al.* (2016), DB-Engines (2020), Bellini y Nesi (2018) y Poyato (2014), quienes consideraron indicadores de licenciamiento, compatibilidad con estándares, estabilidad en el mercado, apertura y gestión RDF. Como resultado de este análisis se seleccionó Apache Jena Fuseki (Apache Jena, 2020), considerando que aunque Fuseki tiene un desempeño promedio en cuanto a ejecución de consultas, su compatibilidad con los procesos de serialización de Dspace es muy alta. Fuseki es una estructura Java de código abierto, usada para la implementación de aplicaciones para Web Semántica y datos enlazados. Su API permite: gestionar datos de grafos RDF, realizar consultas vía SPARQL, y soportar ontologías especificadas en OWL, con diversos razonadores internos. Algunas de las especificaciones de Fuseki se detallan en la Tabla 30.

Tabla 30. Algunas especificaciones de Fuseki. Fuente: Bonduel (2019).

Name triplestore	GeoSPARQL querying	Multiple repositories per DB instance	Named graphs	Supported reasoning schemes	Licensing options	Bulk loading
Apache Jena Fuseki	No*	Yes	Yes	RDFS/OWL rule engines**	Open source (Apache License v2.0)	Yes
Stardog	Yes	Yes	Yes	RDFS, OWL DL/QL/RL, SL (includes SWRL), SHACL	Proprietary: Enterprise licenses (60 day trial), free academic license	Yes
Ontotext GraphDB	No***	Yes	Yes	RDFS, OWL QL/RL, custom rulesets	Proprietary: Enterprise license (60 day trial), standard license (60 day trial), free license	Yes
BBN Parliament	Yes	No	Yes	Selection of RDFS and OWL rules (config), SWRL	Open source (BSD)	Yes

* Limited custom geospatial queries possible over WGS84 and GeoSPARQL WKT data (newest versions might include GeoSPARQL querying)
** Jena modules can be connected for extensions: http://jena.apache.org/documentation/inference/
*** Limited custom geospatial queries possible over WGS84 data (GraphDB functions)

Es importante resaltar que la API de Fuseki tiene la capacidad de consumir RDF Graph exportados por plataformas como Dspace, lo cual permite una compatibilidad alta entre las aplicaciones para el manejo de datos en RDF, y por ende facilita el proceso de despliegue del mismo.

6.1.2. Despliegue de las plataformas seleccionadas.

Como se muestra en la Figura 31, el despliegue del ambiente tecnológico se realizó en dos escenarios:

a. La instalación de las APIs de Dspace y Fuseki. Para instalar la API de Dspace, previamente se instaló un sistema de gestión de bases de datos, como por ejemplo PostgreSql (PostgreSQL, 2020), el cual gestiona el almacenamiento de los componentes Dspace.
b. Se ejecutó el respectivo despliegue de Dspace y Fuseki en Apache Tomcat (ASF, 2020), como contenedor de servlets usado para compilar y ejecutar aplicaciones Web realizadas en Java. Estos despliegues permiten a los usuarios tener acceso Web a Dspace y Fuseki.

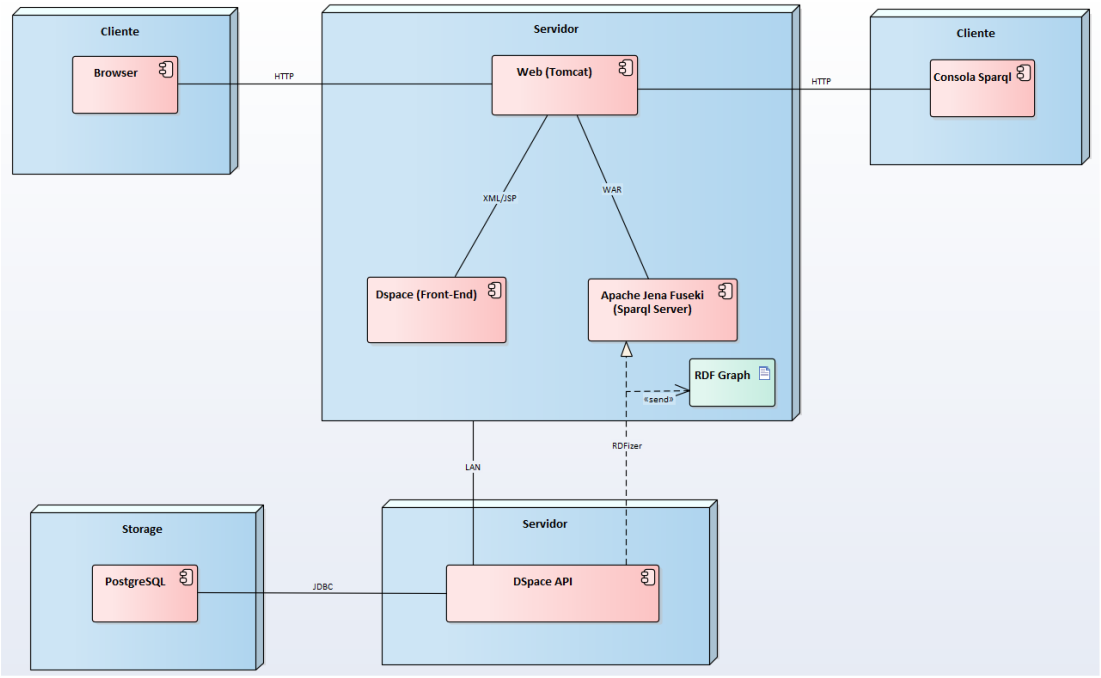


Figura 31. Despliegue del Ambiente Tecnológico.

A través del proceso de serialización denominado RDFizer, realizado sobre Dspace, se provee un artefacto RDF Graph el cual es almacenado en el TripleStore (espacio de memoria o caché para almacenar los datos convertidos) para luego ser consumido por Fuseki.

Seguidamente al despliegue del ambiente tecnológico, se procedió a realizar el despliegue del modelo de datos en los componentes que así lo requerían. A continuación de describe el despliegue realizado.

6.2. Despliegue del Modelo de Datos.

Posterior a la configuración de las herramientas, se desplegó el modelo de datos a través de la parametrización de diferentes componentes dentro de las herramientas seleccionadas. Considerando que el metamodelo diseñado describe los recursos educativos abiertos en dos ambientes: Mapeo PA-REA para Aplicación y Mapeo PA-REA para Dataset, es importante señalar que ésta parametrización se realizó en tres escenarios diferentes, los cuales se describen a continuación.

6.2.1. Parametrización de la ingesta y almacenamiento.

De acuerdo a lo expuesto en el Anexo 6, como preparación para la ingesta de datos, el registro de metadatos para Dspace se parametrizó acorde al mapeo PA-REA definido para la aplicación. Para tal fin, al registro Dublin Core existente en el "Metadata Registry" de Dspace, se adicionaron y modificaron los tags necesarios para completar el proceso de Metadata Crosswalk definido en esta investigación, proceso que permite mapear los metadatos de un esquema a otro (Figura 32).

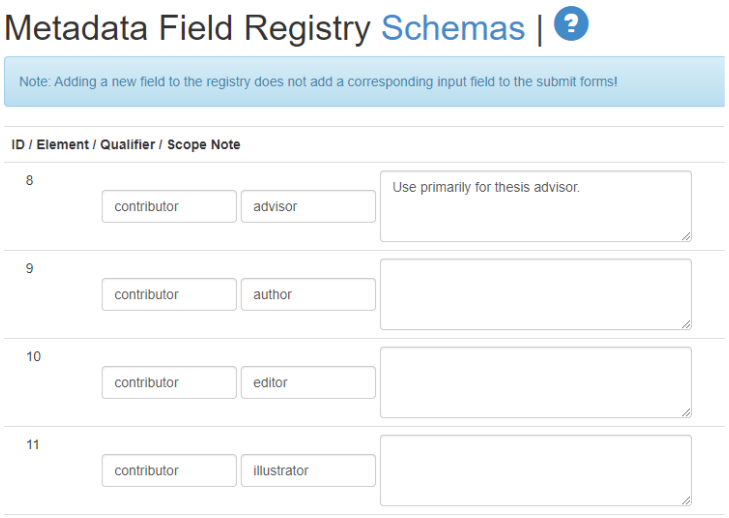


Figura 32. Modificación del Metadata Registry.

De igual forma, y con el fin de fortalecer la completitud de propiedad (dimensión de calidad contextual), se prepararon los vocabularios controlados, que proporcionaron el contenido adecuado para los campos que así lo requerían. La Figura 33 muestra un ejemplo de vocabulario controlado ingresado a la plataforma Dspace.

```
<node id="NivelEducativoCategories" label="NEducativo Categories">
    <isComposedBy>
        <node id="SCB11" label="Preescolar">
        </node>
        <node id="SCB12" label="Basica Primaria">
        </node>
        <node id="SCB13" label="Media">
        </node>
        <node id="SCB14" label="Media Profesional">
        </node>
        <node id="SCB15" label="Superior Tecnico">
        </node>
        <node id="SCB16" label="Superior Tecnologo">
        </node>
        <node id="SCB17" label="Superior Profesional">
        </node>
        <node id="SCB18" label="Superior Postgrado">
        </node>
        <node id="SCB19" label="Otro">
        </node>
    </isComposedBy>
</node>
```

Figura 33. Ejemplo de Vocabulario Controlado.

Bajo este contexto, se llevó a cabo un análisis para el diseño de vocabularios controlados, identificando aquellos metadatos cuyo contenido requerían de la estandarización de los valores de ingesta. A continuación se identifican las propiedades o metadatos para las cuales se diseñaron vocabularios controlados, y cuyos contenidos parametrizados se encuentran en detalle en el Anexo 5:

- Audiencia
- Contexto Educativo

- Contribuyentes
- Editores
- Formatos
- Grupos
- Idiomas
- Instituciones
- Nivel Educativo
- Nivel y tipo de interactividad
- Tipo de Recursos

Por último, para la captura de datos, las interfaces de ingesta se modificaron acorde a la estructura de metadatos diseñada en el modelo de datos propuesto. La Figura 34 muestra un ejemplo de la estructura modificada de una interfaz de ingesta.

```
<form name="traditional">
  <page number="1">

    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>title</dc-element>
      <dc-qualifier></dc-qualifier>
      <repeatable>false</repeatable>
      <label>Titulo del Recurso</label>
      <input-type>onebox</input-type>
      <hint>Ingrese el titulo principal del recurso.</hint>
      <required>Debe ingresar el titulo principal del recurso.</required>
    </field>
    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>title</dc-element>
      <dc-qualifier>alternative</dc-qualifier>
      <repeatable>true</repeatable>
      <label>Titulo Alternativo</label>
      <input-type>onebox</input-type>
      <hint>Si el recurso tiene un titulo alternativo, por favor ingreselo aca.</hint>
      <required></required>
    </field>
```

Figura 34. Ejemplo de Interfaz de Ingesta Dspace.

La estructura expuesta en la figura anterior presenta metadatos correspondientes al “Metadata Registry” los cuáles son gestionados por Dspace, y que de manera previa fue modificado de conformidad con el modelo de metadatos provisto.

Después de haber parametrizado este proceso de ingesta y almacenamiento sobre Dspace haciendo uso del modelo de datos diseñado, se procedió a parametrizar el proceso de mapeo RDF, el cual se describe a continuación.

6.2.2. Parametrización del mapeo RDF de REA.

De acuerdo a lo expuesto en el Anexo 6, para mapear los metadatos REA almacenados en la plataforma de gestión al modelo de datos RDF, se configuró un archivo de mapeo que recibe los metadatos Dublin Core, provenientes mapeo PA-REA para aplicación, y genera tripletas RDF. Esta transformación se lleva a cabo de acuerdo a los parámetros sintácticos y semánticos definidos en el mapeo PAR-REA de los datasets. Para este proceso, se consideraron los siguientes aspectos:

a. Ajustes de mapeo. Las estrategias de mapeo RDF gestionadas por Dspace, se deben ajustar a la estructura de mapeo que el Metadata Crosswalk definió para el dominio de conocimiento.
b. Verificación de Sintaxis. Los Metadata Names que reciben cada atributo Dublin Core, deben corresponder exactamente (mayúsculas, minúsculas, símbolos, etc.) a los definidos tanto en el Metadata Registry como en las interfaces de Ingesta de Dspace.

c. Aplicación de Condiciones. La inclusión de las “condiciones” de mapeo debe obedecer a las reglas provistas por el metamodelo RDF.
d. Parametrización del enlace de datos. Dado que el perfil de aplicación PA-REA definió algunos componentes como susceptibles de enlazar datos, es importante identificar cómo se realiza este mapeo.
e. Estrategias de Mapeo. El diseño de las estrategias de mapeo de literales y URIs se ajustan acorde a la plataforma de gestión de recursos que se esté utilizando.

Atendiendo a estas consideraciones, a continuación se detalla el proceso de mapeo, de forma general:

a. Mapeo RDF de elementos con contenidos literales:

:creator	Tag que identifica el elemento a mapear.
dm:metadataName "dc.creator" ;	Metadato a mapear, de acuerdo al Metadata Crosswalk para la plataforma de gestión.
dm:creates [	Inicializa un nodo en blanco describiendo la tripleta RDF a crear.
dm:subject dm:DSpaceObjectIRI ;	Usa DSpaceObject‘s URI como sujeto.
dm:predicate dc:creator ;	Usa la relación dc:creator como predicado.
dm:object dm:DSpaceValue ;] ; .	Usa el contenido literal del elemento como objeto (literal).

b. Mapeo RDF de elementos con contenidos URI:

:creator	Tag que identifica el elemento a mapear.
dm:metadataName "dc.creator" ;	Metadato a mapear, de acuerdo al Metadata Crosswalk para la plataforma de gestión.
dm:condition "^((?!http://dx.doi.org/)(?!http://hdl.handle.net/)(?!http://localhost:8080/jspui/handle/)).*" ;	Se usa esta regla o condición para ignorar cadenas que empiecen dx.doi, hdl.handle o localhost, dado que son URLs propias del sistema.
dm:creates [	Inicializa un nodo en blanco describiendo la tripleta RDF a crear.
dm:subject dm:DSpaceObjectIRI ;	Usa DSpaceObject‘s URI como sujeto.
dm:predicate dc:creator ;	Usa la relación dc:creator como predicado.
dm:object [a dm:ResourceGenerator ; dm:pattern "$DSpaceValue" ;] ;] ;	Crea una URI como objeto. Para ejecutar esta tarea, toma el literal URL introducido

.	al recurso, y lo transforma a una URL a través del generador.

Por último, se definieron unas recomendaciones para la construcción de las consultas SPARQL y visualizaciones de datos producto de la ejecución de las mencionadas consultas.

6.2.3. Parametrización de la visualización y consulta.

Para parametrizar las visualizaciones y consultas realizadas desde Fuseki, éstas deben construirse atendiendo a los parámetros de mapeo definidos en el Mapeo PA-REA para dataset. Por otro lado, dado que Fuseki es un servidor SPARQL 1.1 que proporciona los protocolos de consulta, actualización y almacenamiento de grafos SPARQL (Apache Jena, 2020; MVNRepository, 2014), las parametrizaciones de las visualizaciones y consultas deben construirse atendiendo a esta recomendación (W3C, 2013). Un ejemplo básico de consulta sobre un RDF Graph se presenta en la Figura 35.

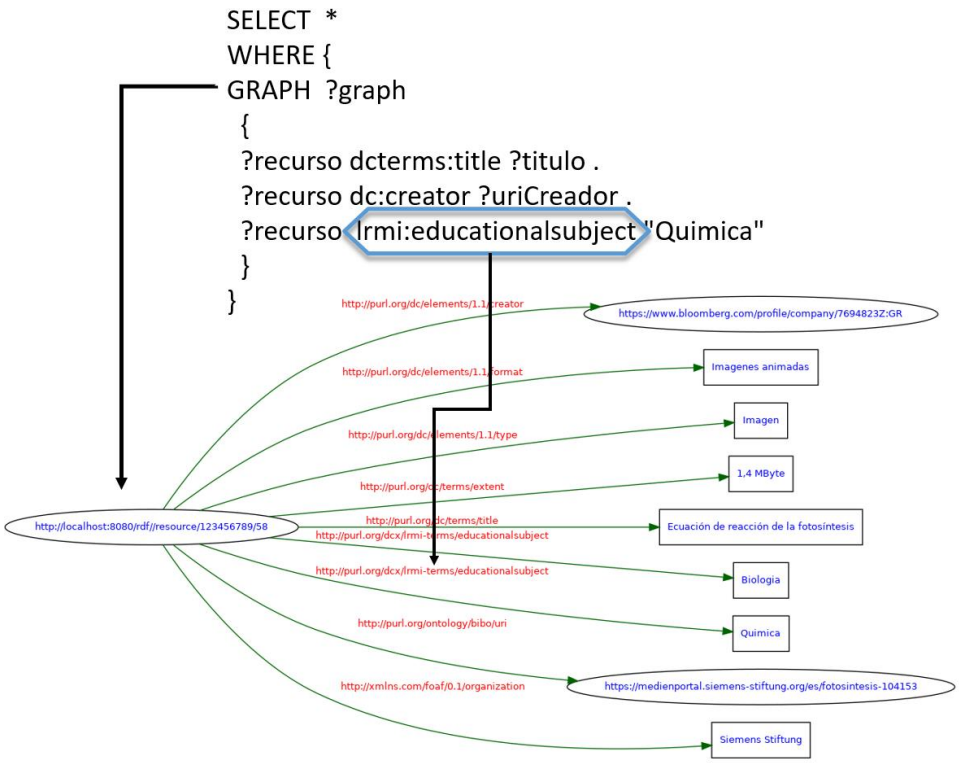


Figura 35. Ejemplo de construcción de consultas.

6.3. Conclusiones del Capítulo.

En primer lugar, como se presenta en la Figura 36, los mapeos definidos por el metamodelo pueden tener diferentes destinos, como lo son:

a. Almacenamiento: La descripción de los recursos se almacena en la base de datos que soporta el funcionamiento de la herramienta Dspace.
b. Interfaz: los metadatos de los recursos pueden ser gestionados a través de las interfaces provistas por la herramienta Dspace.
c. Nube LOD: Al contar con metadatos enriquecidos, y adicionalmente ser exportados a grafos RDF que se alinean con los requisitos de publicación de la nube LOD, los grafos generados pueden estar enlazados a grafos de conocimiento a través de las URI utilizadas.

d. Consultas SPARQL: Los grafos generados pueden ser consultados a través de puntos de acceso SPARQL que ofrece la plataforma SPARQL provista por Fuseki.

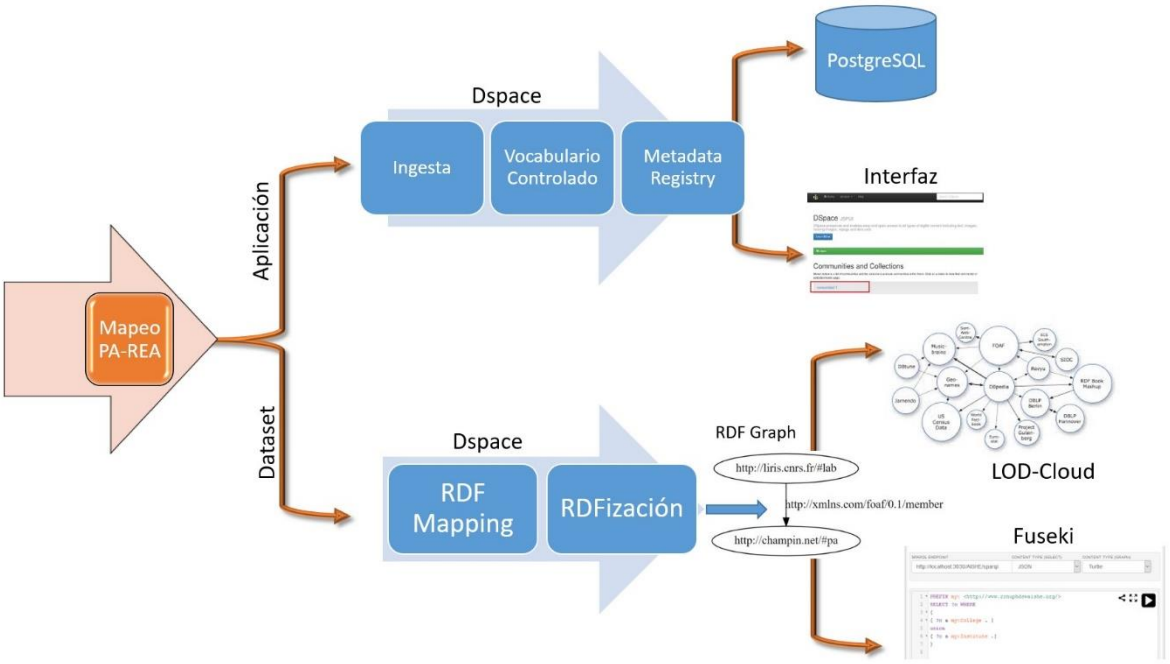


Figura 36. Procesos de Mapeos PA-REA para Aplicación y para Dataset.

Basado en el proceso descrito anteriormente, el despliegue del perfil de aplicación PA-REA generado en esta investigación, ofrece beneficios tanto de accesibilidad como de reutilización. Así mismo, al ser publicado de forma abierta, dispone del uso de metadatos que se han complementado y enriquecido, con el propósito de aumentar su visibilidad y accesibilidad a través de enlaces sobre la Web de los Datos.

En segundo lugar, en la implementación realizada se parametrizaron todos los componentes involucrados en el despliegue del modelo de datos, tales como interfaces, registro de metadatos, estructuras de mapeos y vocabularios controlados. Esta parametrización se construyó acorde a las restricciones del modelo de datos (Nivel M1) definido por el metamodelo, con el fin de capturar y gestionar las instancias de metadatos REA (Nivel M0 del metamodelo). La implementación realizada se usó para llevar a cabo procesos de validación y conformidad con el modelo de datos, descritos en el siguiente capítulo.

CAPÍTULO VII VALIDACIÓN DEL METAMODELO.

En términos generales, el proceso de validación se diseña con el propósito de evidenciar que un sistema cumpla con las expectativas del cliente (Sommerville, 2011). Con relación a esta investigación, el cual se orientada al diseño de un metamodelo para describir y enriquecer REA basados en especificaciones LOD, se consideraron cinco escenarios importantes para llevar a cabo la evaluación y el cumplimiento de la propuesta planteada, escenarios que se describen en Figura 37.

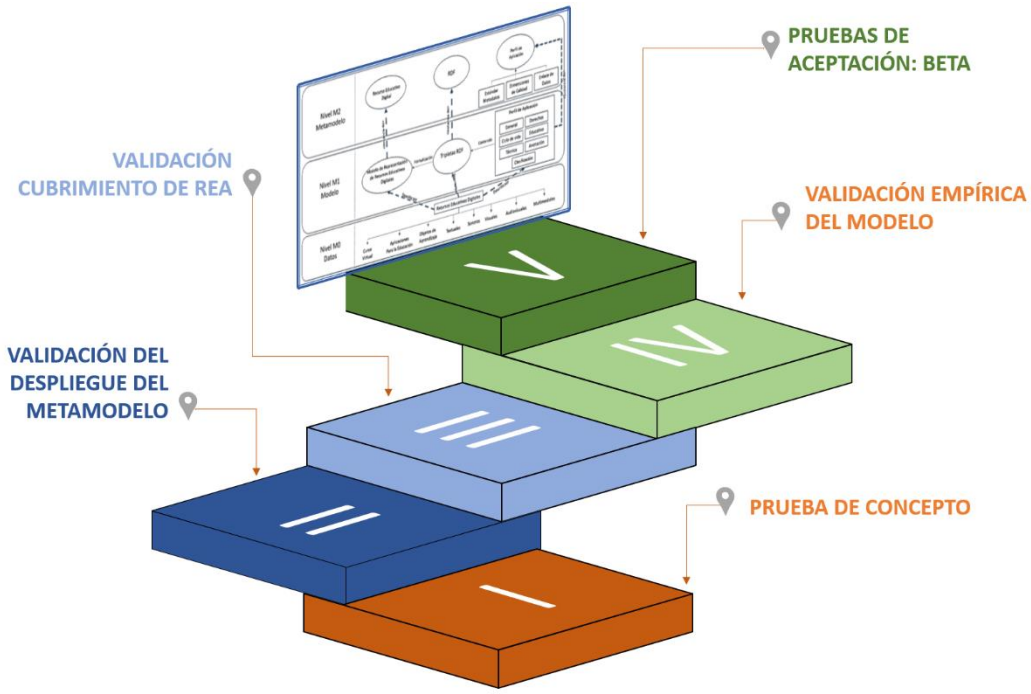


Figura 37. Estadios de validación del metamodelo propuesto.

Con relación a las validaciones definidas, y acorde a lo descrito por Sommerville (2011) y Pressman (2010), cada una de ellas se enfocó en:

- Prueba De Concepto: Este tipo de pruebas permitieron validar el funcionamiento del metamodelo (Belloir *et al.,* 2019). Su enfoque correspondió a validar que el metamodelo diseñado dispusiera de los elementos estructurales requeridos para la generación del modelo de datos que describe y enriquece los REA.
- Validación del Despliegue del Modelo: Después de desplegar el metamodelo, y atendiendo a dos componentes específicos: a) el factor de Relevancia - Dimensión de calidad contextual (Zaveri *et al.*, 2012), y b) las fases de prueba en el proceso de desarrollo de software (Torres, 2008; Sommerville, 2011), considerando las respectivas salvedades acerca de lo abstracto del producto generado (Belloir *et al.,* 2019); se realizaron pruebas unitarios, de integración y de sistema, tanto del modelo de datos como del ambiente tecnológico de despliegue, con el fin de verificar su adecuado funcionamiento. Tanto el despliegue como las pruebas se realizaron sobre la infraestructura tecnológica proporcionada por el CECAD - Centro de Computación de Alto Desempeño, perteneciente a la Universidad Distrital Francisco José de Caldas.
- Validación de Cubrimiento de REA: Su enfoque consistió en confrontar las descripciones de los diferentes tipos de REA versus el modelo de datos diseñado. Para llevar a cabo esta prueba, se realizó una alimentación manual de los diferentes tipos de REA definidos para esta investigación (MEN, 2012).

- Validación Empírica del modelo: Esta validación se orientó en evaluar la calidad inherente del modelo de datos y su interoperabilidad frente a otros modelos de datos que describen recursos. Para llevar a cabo este propósito, se utilizó el método Goal–Question-Metric (GQM), método utilizado por autores tales como Yaghouti *et al.* (2015), Behkamal *et al.* (2014), Behkamal (2014), Li *et al. (*2019) y Zahedi y Fathian (2016), para temas de evaluación de calidad de los datos en ambientes LOD.
- Pruebas de Aceptación de Usuario: Considerando lo expuesto por Galin (2004) y Tian (2005), estas pruebas se enfocan en verificar si el sistema está listo y en condiciones para su uso. Para este caso, considerando que el componente construido corresponde a un metamodelo el cual no involucra el desarrollo de software, este tipo de pruebas se orientaron a explorar la opinión de los usuarios (pruebas beta), con relación al modelo de datos obtenido como instancia del metamodelo. Para esta prueba se trabajó con dos grupos de población (Expertos y estudiantes) seleccionados al azar, buscando incluir personas que tuvieran contacto con el RIUD, dada su aplicación como caso de estudio.

A continuación se describen los procesos que se realizaron para evaluar y validar cada uno de los escenarios descritos.

7.1. Prueba de concepto del metamodelo diseñado.

En primera instancia, como lo describe Pressman (2010), la prueba de concepto muestra si el producto de software o su función separada son adecuados para resolver un problema comercial en particular. En segunda instancia, como plantea Villalobos (2020), un metamodelo es una estructura de entidades, asociaciones y restricciones que permite representar modelos conceptuales. Cada entidad tiene un nombre que es único, un conjunto de atributos, un conjunto de asociaciones con otras entidades y un conjunto de restricciones. De acuerdo a los elementos provistos por la anterior definición, para validar la conformidad del metamodelo, se realizaran las siguientes pruebas de concepto:

- Validación del mapeo para aplicación.
- Validación del mapeo para dataset.
- Validación de la gestión de propiedades

Con base en lo expuesto por Belloir *et al.* (2019), y dado que se opera sobre un producto abstracto como lo es un metamodelo, las dos primeras pruebas de concepto se orientaron en establecer si los componentes del metamodelo interactúan de la forma esperada para producir los mapeos requeridos. Como criterios de éxito se determinaron que los mapeos generados (nivel Mo del Metamodelo) fueran conformes a las restricciones y condiciones que se definieron en los niveles M1 y M2 del metamodelo. Para la última prueba de concepto se orientó a determinar la funcionalidad de la gestión de propiedades y su articulación con cada uno de los mapeos definidos (aplicación y dataset). Como criterio de éxito de esta prueba de concepto se definió que la gestión (ingreso/eliminación, modificación) de variables en el metamodelo cumpliera con los requisitos del modelo (Capa M1) y permitiera generar instancias de datos (nivel M0) conformes al modelo actualizado.

A continuación se presenta la formalización de cada una de las pruebas de concepto para el metamodelo, en cada uno de sus mapeos: de aplicación y de dataset, junto con la prueba de gestión de propiedades.

7.1.1. Validación del mapeo para aplicación.

La Figura 38 muestra un ejemplo de la formalización del modelo de representación de REA mapeado para aplicación, para una parte de un recurso en especial.

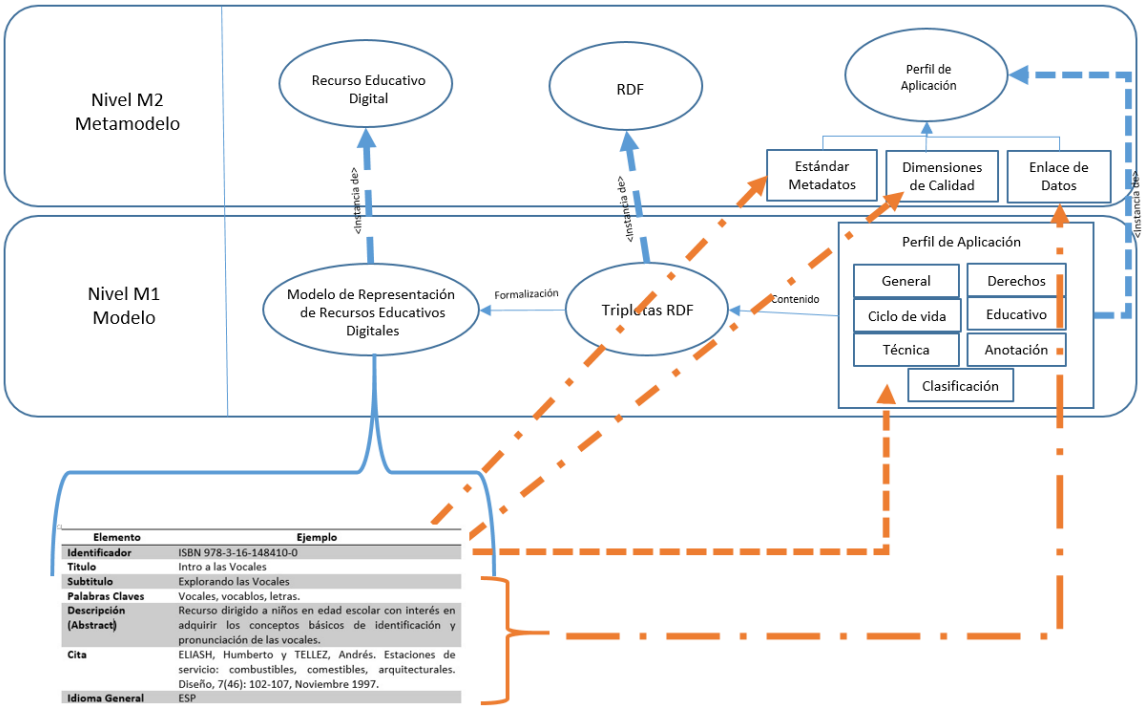


Figura 38. Ejemplo de Esquema de Mapeo PA-REA para Aplicación.

En el mapeo de aplicación se observó:

- La instancia del recurso (nivel M0) corresponde a la estructura del modelo de representación de REA (nivel M1), que a su vez responde a una entidad REA definida en el metamodelo (nivel M2).
- La instancia del recurso (nivel M0) tiene un conjunto de propiedades mapeadas en Dublin Core, que corresponde a una de las restricciones provistas por la entidad del nivel M2. Estas propiedades están enmarcadas en un perfil de aplicación, y sus valores están parametrizados de acuerdo a un diccionario de datos que refina el perfil de aplicación. Este perfil de aplicación (nivel M1) es una instancia parametrizada adicionalmente por restricciones definidas por las entidades “Dimensiones de Calidad” y “Enlace de Datos” (Nivel M2).
- La instancia contiene un valor para cada uno de las propiedades definidas en el metamodelo, y este valor está descrito de acuerdo al tipo de dato del atributo definido en el respectivo diccionario de datos.

En resumen, para el mapeo de aplicación, cada entidad, asociación y propiedad está claramente definida dentro de la estructura del metamodelo, lo cual evidencia que la instancia generada del REA (en su mapeo para aplicación) cumple con las reglas que se definieron en el metamodelo, lo cual implica que el modelo es conforme con las reglas definidas en el metamodelo. Como criterio de aceptación, se observa que se construye de forma adecuada el mapeo de propiedades de un REA bajo Dublin Core, atendiendo a las características del modelo de representación de REA.

7.1.2. Validación del mapeo para dataset.

La Figura 39 presenta un ejemplo de la formalización del modelo de representación de REA mapeado para dataset, para un tag en especial.

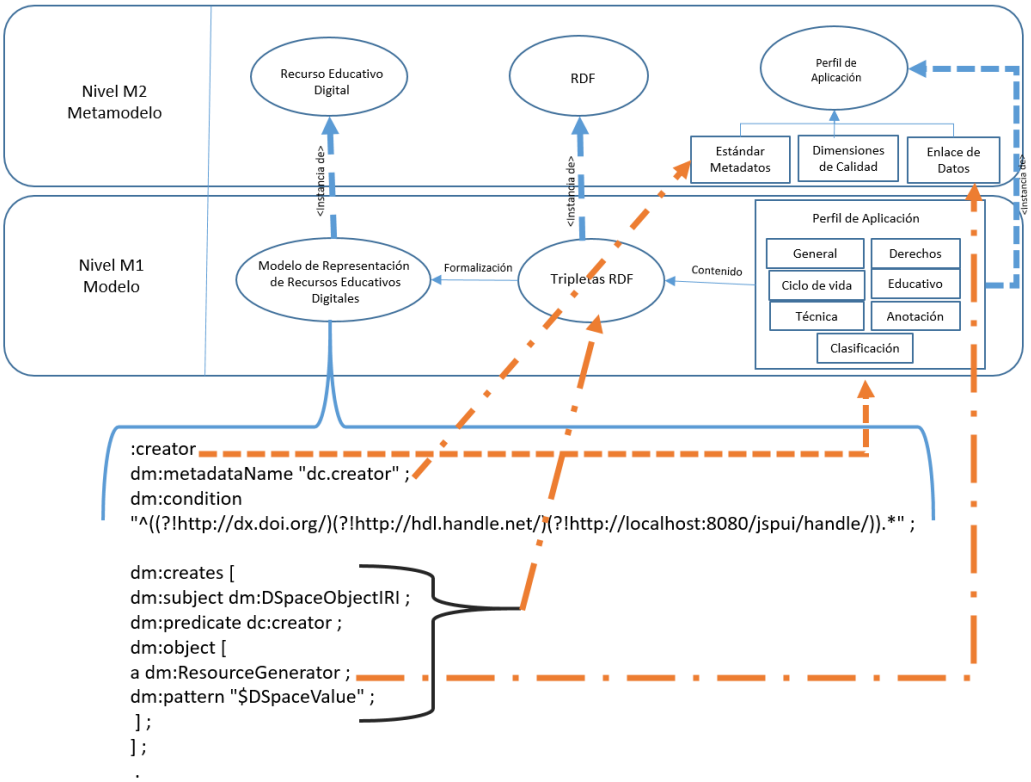


Figura 39. Ejemplo del esquema de mapeo RDF para PA-REA.

En la figura anterior se observa el mapeo de la propiedad “Creador” de un REA. Este elemento, además de estar definido dentro del modelo de datos PA-REA; está alineado con los requerimientos aportados por las dimensiones de calidad de datos enlazados, y provisto como elemento susceptible de enlace. En el ejemplo se tiene:

1. El metadato a mapear corresponde a “dc.creator”, identificador definido en el vocabulario Dublin Core (dc), usado por la plataforma de gestión de recursos para describir el REA.
2. Dado que en la construcción del modelo de datos PA-REA se definió que éste atributo gestiona la URI del creador, la “condición” que debe cumplir éste identificador es que no corresponda a una URI identificadora del recurso.
3. La tripleta creada tiene como “predicado” la relación “dc.creator”, lo que identifica que el vocabulario de entrada al mapeo corresponde al mismo vocabulario de salida. Adicionalmente, el contenido de este atributo mapeado, correspondiente al “Objeto”, es una URI cuyo patrón corresponderá a la URI ingresada en la plataforma de gestión.

En resumen, la Figura 40 muestra un ejemplo de cómo los elementos del nivel M1 corresponden a instancias del nivel M2 del metamodelo. Además, evidencia el cumplimiento de cada una de las directrices y restricciones definidas por cada entidad para el diseño y la formalización del metamodelo, acorde al perfil de aplicación PA-REA.

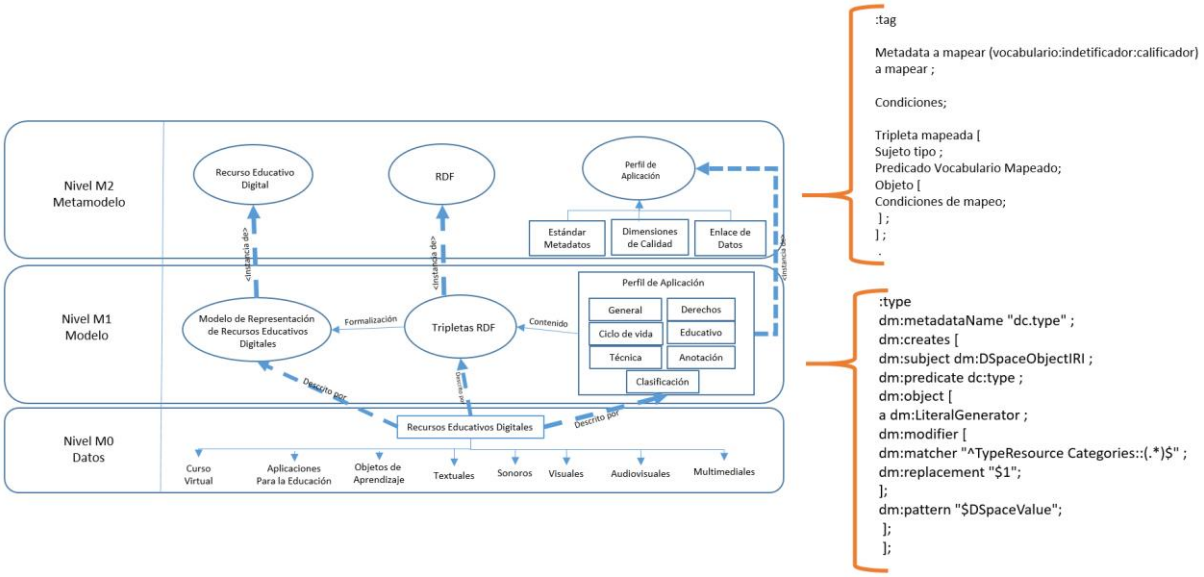


Figura 40. Formalización del Metamodelo al Modelo de Datos.

El proceso de formalización descrito anteriormente genera datasets de metadatos REA, que corresponden al nivel M0 del metamodelo diseñado. Para el mapeo de dataset se observó:

- Los datasets contienen URIs, según los metadatos enriquecidos, a recursos bien sean internos o externos, contribuyendo a incrementar los grafos de conocimiento existentes.
- A través del uso de consultas formuladas por intermedio de puntos de acceso SPARQL, la información de los datasets es accesible y reutilizable por usuarios consumidores externos.
- La instancia del recurso (nivel M0) responde a la estructura del modelo de representación de REA (nivel M1), que a su vez responde a una entidad REA definida en el metamodelo (nivel M2).
- La instancia del recurso (nivel M0) tiene un conjunto de propiedades mapeadas en RDF. Este mapeo de tripletas (Nivel M1) responde a un conjunto de reglas definidas por el Modelo de Datos RDF (Nivel M2), reglas que permiten el mapeo de las propiedades del REA que están enmarcadas en un perfil de aplicación. Dicho perfil de aplicación (nivel M1) es una instancia parametrizada del nivel M2.
- La instancia del recurso mapeado para dataset contiene tripletas claramente definidos que conservan la estructura Sujeto – Predicado - Objeto.

En resumen, para el mapeo de dataset, cada entidad, asociación y propiedad está definida dentro de la estructura del metamodelo, lo cual evidencia que la instancia generada del REA (en su mapeo para dataset) cumple con las reglas que se definieron en el metamodelo, lo cual implica que el modelo corresponde con las reglas definidas en el metamodelo. Como criterio de aceptación, se observa que se construye de forma adecuada el mapeo de propiedades de un REA en tripletas RDF, atendiendo a las características del modelo de representación de REA y el modelo de datos RDF.

Después de validar los mapeos definidos por el metamodelo diseñado en esta investigación, se procedió a validar la gestión de propiedades haciendo uso de las estructura de capas del metamodelo construido como producto de esta investigación.

7.1.3. Validación de la gestión de propiedades para los mapeos de datos.

Considerando lo expuesto por Villalobos (2020) y Belloir *et al.* (2019), con esta prueba de concepto se buscó validar que la gestión de propiedades que ingresan, se modifican o salen del

modelo de datos para describir los REA, cumple con las restricciones y reglas del metamodelo, además de ser conformes con cada uno de sus niveles. Para tal fin, el nivel M1 constituye el eje clave para la modificación del modelo de datos, bien sea para mapeo de aplicación o de dataset. Al tener el requerimiento de introducir/modificar/eliminar una propiedad que describa un REA, este nivel conserva los constructos teóricos del nivel M2, pero tiene la capacidad de alterar los modelos usados en el nivel M0.

Para el caso de ingresar una nueva propiedad al mapeo PA-REA para aplicación, la Figura 41 describe el proceso realizado:

1. La nueva propiedad, derivada del esquema de metadatos del REA, es clasificada en el perfil de aplicación:

    - La propiedad “XXX” es analizada en el diccionario de datos, con el fin de determinar aspectos como su tipo de datos, las pautas para su diligenciamiento, si se debe alimentar con algún tipo de vocabulario controlado, etc.
    - La propiedad “XXX” es ubicada en una de las categorías del perfil de aplicación. En caso de no encajar con alguna categoría, dada las restricciones del nivel M2, se puede insertar una nueva categoría.
    - Acorde a las restricciones de la entidad “Enlace de Datos”, se analiza si la propiedad “XXX” es susceptible de enlazar datos.
    - Acorde a las restricciones de la entidad “Estándar de Metadatos”, se mapea la propiedad “XXX” a Dublin Core.

2. La propiedad “XXX” es representada en el Modelo de Representación de REA para Aplicación, el cual incluye la propiedad en el modelo de datos.

3. La propiedad “XXX” es encapsulada en el esquema de representación de Dublin Core, en el formato esquema:propiedad:calificador.

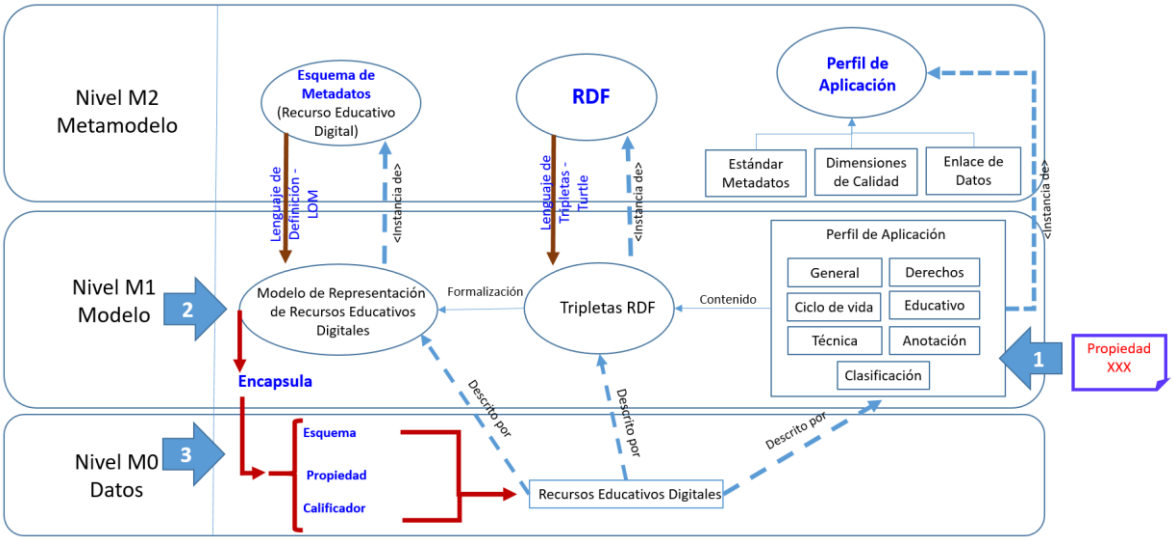


Figura 41. Inserción de una Propiedad al Mapeo PA-REA para Aplicación.

Al realizar el proceso anterior, la propiedad “XXX” queda incluida dentro del modelo de datos que se instancia en el nivel M0. De igual forma, se evidencia que el proceso de gestión de propiedades (ingreso, modificación y eliminación) del modelo de datos que describe REA se adapta a las restricciones definidas por el metamodelo.

Para el caso de ingresar una nueva propiedad al Mapeo PA-REA para Dataset, la Figura 42 describe el proceso realizado:

1. La nueva propiedad, derivada del esquema de metadatos del REA, es clasificada en el perfil de aplicación.

2. La propiedad es ajustada a las reglas definidas por el Modelo de Datos RDF.

3. Esta nueva propiedad es representada en el Modelo de Representación de REA para Dataset, donde se incluye la propiedad en el modelo de datos.

4. Esta propiedad es encapsulada en el esquema de representación de tripletas RDF, en el formato: Sujeto – Predicado – Objeto.

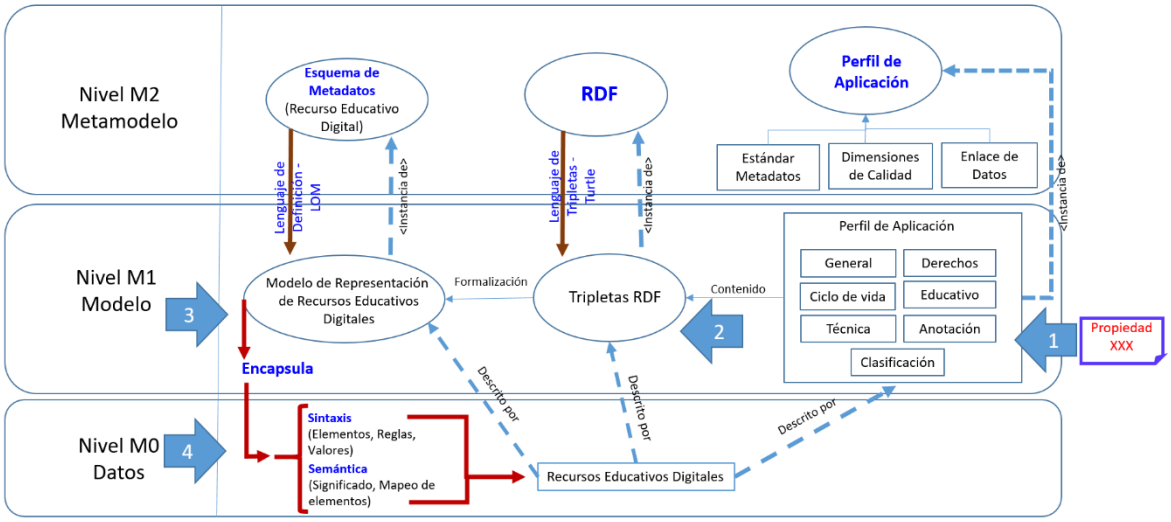


Figura 42. Inserción de una Propiedad al Mapeo PA-REA para Dataset.

Lo anterior evidencia, como criterio de éxito de esta prueba de concepto, que el metamodelo diseñado esgrime las facilidades necesarias para un adecuado y eficiente manejo del modelo de datos para describir REA. Además, identifica de forma clara cada uno de los componentes afectados ante la gestión de una propiedad, y por lo tanto, se ajusta a las restricciones y reglas definidas por cada componente del nivel M2.

Después de realizar la respectiva validación del metamodelo, se procedió a validar el despliegue de su artefacto principal: el modelo de datos para describir y enriquecer REA.

7.2. Validación del Despliegue del Modelo.

Seguidamente a la prueba de concepto, y dado que se hace uso de herramientas de software, se procedió a la validación del despliegue del modelo de datos para describir REA, artefacto que se materializó a través de la modificación y prueba de diferentes componentes del ambiente tecnológico provisto para el mencionado despliegue. Atendiendo a las modificaciones realizadas, se consideró necesario llevar a cabo las siguientes pruebas (Torres, 2008; Sommerville, 2011; Belloir *et al.,* 2019; Galin, 2004; Tian 2005):

- Unitarias: Prueban el diseño y el comportamiento de cada componente del sistema una vez construido. Dado que las herramientas de software provistas en el despliegue ya se

encontraban desarrolladas, estas pruebas se orientaron a verificar el correcto funcionamiento de cada uno de los componentes modificados con el modelo de datos para describir REA, como por ejemplo el despliegue del modelo de datos y el despliegue de los mapeos construidos.

- Integración: Prueban la correcta relación entre los componentes del sistema a través de sus interfaces y si ellas cumplen con la funcionalidad establecida. Para este tipo de pruebas se validó la correcta relación de los mapeos a lo largo de su tránsito por las interfaces de las diferentes plataformas: la ingesta de elementos de metadatos, el mapeo de los elementos de metadatos a Dublin Core, el mapeo de metadatos a tripletas RDF.
- Sistema: Prueban el sistema comprobando su funcionalidad y atributos de calidad. El sistema es probado en un ambiente lo más parecido posible al ambiente operacional. Para tal fin se definieron pruebas que validaran el tránsito de los elementos de metadatos desde su ingesta hasta su mapeo respectivo (para aplicación y para dataset), haciendo uso de las herramientas definidas para el ambiente operacional propiamente dicho: Dspace y Fuseki.

Con base en la tipología de pruebas descritas anteriormente y en el ciclo de vida de los datos enlazados, descrito en el Marco Teórico (véase Sección 2.4.5), y en el esquema LOD de 5 niveles, se diseñó una secuencia de pruebas (Figura 43) para verificar, como criterio de éxito, el adecuado funcionamiento del modelo de datos generado para cada una de las parametrizaciones dispuestas en la implementación. Adicionalmente, y considerando lo expuesto por Abella *et al.* (2018), este conjunto de pruebas se dividió en dos secciones:

- Disponibilidad: Se procuró evaluar que los elementos de metadatos que describen un REA fueran ingresados, mantenidos y disponibles en una aplicación para la gestión de recursos. Adicionalmente, en este proceso se buscó verificar el adecuado uso del mapeo del perfil de aplicación PA-REA para aplicación, y el almacenamiento de los metadatos en formatos abiertos que se ajustaran con los requisitos del modelo de datos enlazados.
- Reutilización: Se determinó evaluar que los elementos de metadatos que describen un REA, además de ser mapeados para dataset, ofrecieran los elementos necesarios para abordar los niveles 4 y 5 del modelo de datos enlazados.

Por último, es prudente aclarar que para las pruebas realizadas no se manejó un porcentaje de éxito y de fracasos de la prueba, dado que la funcionalidad del software en sí misma ya se encontraba definida y desarrollada. Por ende, los resultados que presentan las siguientes pruebas se orientaron a mostrar el adecuado despliegue y funcionamiento de los componentes tras la modificación surtida por el modelo de datos utilizado y los respectivos mapeos construidos a través del uso del metamodelo propuesto.

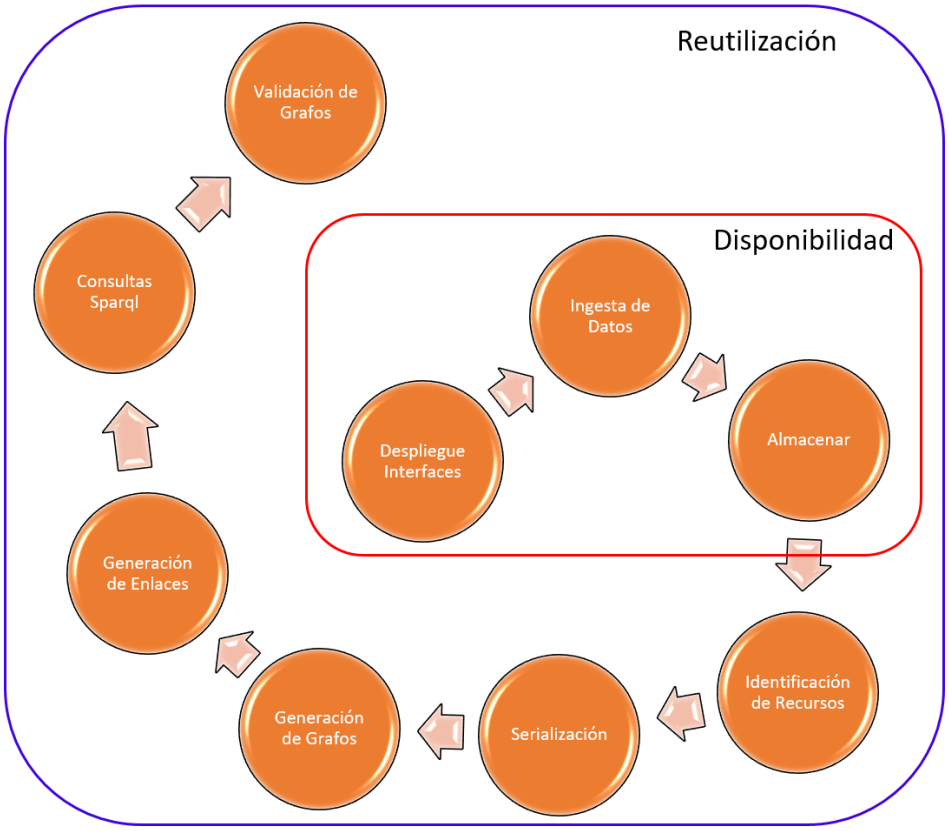


Figura 43. Pruebas del Despliegue del Modelo de Datos.

A continuación se presenta el enfoque y los resultados de cada una de las pruebas realizadas al despliegue del modelo de datos, bajo las herramientas previamente definidas y configuradas.

7.2.1. Pruebas de despliegue de interfaces.

Dado que el modelo de datos es parametrizado tanto en el Metadata Registry como en las interfaces de ingesta, el objeto de esta prueba unitaria fue verificar que las interfaces desplegaran adecuadamente las propiedades del modelo. Para tal fin, se procedió a cargar las interfaces de Dspace, con el fin de verificar su adecuado funcionamiento. La Figura 44 presenta el despliegue de una interfaz, donde se evidencia:

a. Propiedades definidas por el modelo.
b. Propiedades definidas como literales: como por ejemplo el título y el título alternativo.
c. Propiedades definidas como URI: como por ejemplo la URI del recurso.
d. Obligatoriedad de la propiedad: propiedades señaladas con asteriscos.
e. Propiedades cuyo contenido corresponde a vocabularios controlados: propiedades que despliegan categorías, como el caso de tipo de recurso.

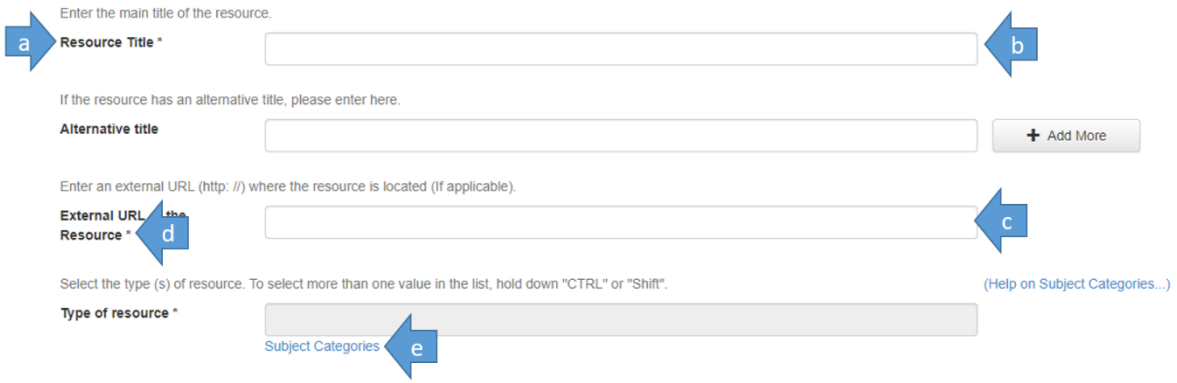


Figura 44. Despliegue de Interfaz de Ingesta de metadatos.

7.2.2. Prueba de Ingesta de Datos.

El objeto de esta prueba unitaria fue corroborar que los datos se ingresen de acuerdo a las características y requisitos de cada propiedad. Para la ejecución de esta prueba se procedió a verificar las restricciones de obligatoriedad, los vocabularios controlados, la cantidad y formato de los datos se desplegaran adecuadamente. La Figura 45 evidencia que la ingesta de datos se realiza acorde al modelo de datos diseñado y a las restricciones definidas.

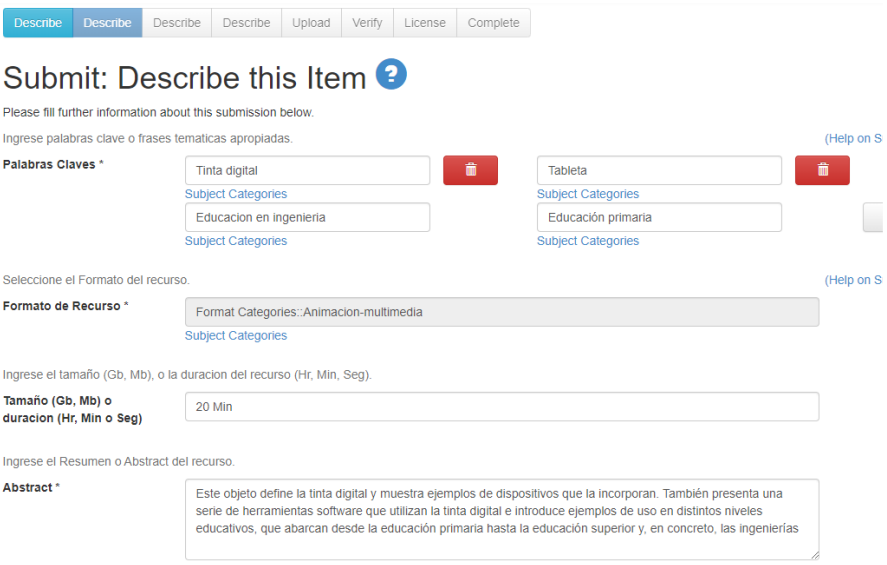


Figura 45. Prueba de Ingesta de metadatos.

7.2.3. Prueba de almacenamiento de recursos digitales.

Esta prueba de integración se realizó en tres momentos:

a. El almacenamiento de los metadatos del recurso: Para esta prueba se almacenó el conjunto de metadatos de un REA y se procedió a verificar el despliegue de la información almacenada. Esta prueba verificó que las restricciones del modelo de datos parametrizaron la ingesta. La Figura 46 evidencia que los metadatos son almacenados por la plataforma de gestión de recursos, y que los respectivos contenidos son almacenados acorde a las restricciones definidas, como por ejemplo: la construcción de las URI, la carga de los vocabularios controlados, etc.



Figura 46. Prueba de Almacenamiento de Metadatos.

b. El uso adecuado del mapeo del perfil de aplicación PA-REA para aplicación: Esta prueba verificó que el mapeo PA-REA para aplicación se ejecutó correctamente. Para esta prueba se desplegó todo el contenido de mapeo de aplicación definido por el perfil de aplicación PA-REA. La Figura 47 evidencia que el despliegue realizado de los metadatos es acorde al modelo de datos mapeado para aplicación, diseñado para ser mantenido en la plataforma de gestión de recursos.



Figura 47. Despliegue del Metadata Registry de un Recurso.

c. El almacenamiento de los metadatos en archivos de formato abierto: Esta prueba verificó que los metadatos del recurso, además de estar disponibles para su consulta en la plataforma de gestión, se exportan y almacenan en un formato de datos que se alinea a los requerimientos del modelo LOD. Para esta prueba, después de haber realizado la ingesta y almacenamiento de los metadatos de un REA, se procedió a exportar su contenido en un archivo cuyo formato se alinee con las restricciones del modelo LOD. La Figura 48 evidencia que los metadatos son almacenados en un formato CSV.

```
id,collection,dc.contributor.advisor,dc.contributor.author,dc.co
ntributor[en_US],dc.creator[en_US],dc.date.accessioned,dc.date.a
vailable,dc.date.copyright,dc.date.created,dc.date.updated,dc.de
scription.KnowledgeGrouping[en_US],dc.description.abstract[en_US
],dc.description.audience[en_US],dc.description.educationlevel[e
n_US],dc.description.interactivityType[en_US],dc.description.int
eractivitylevel[en_US],dc.description.provenance[en],dc.descript
ion.typicalAgeRange[en_US],dc.description.uri[en_US],dc.descript
ion[en_US],dc.format[en_US],dc.identifier.uri,dc.language.iso[en
_US],dc.publisher[en_US],dc.relation.EducationalSubject[en_US],d
c.relation.course[en_US],dc.relation.group[en_US],dc.relation.ha
sversion[en_US],dc.relation.institution[en_US],dc.relation.organ
ization[en_US],dc.rights.uri[en_US],dc.rights[en_US],dc.subject[
en_US],dc.title[en_US],dc.type[en_US]
"bc0974b5-84e0-4436-b698-380a516d4b13","123456789/3","RodrÃguez,
Cristina||GonzÃ¡lez, JosÃ©","Zamit, Virginia||RodrÃguez.,
Mary","Contributor Categories::Contenido::Experto::Caterine
Lagos||Contributor Categories::Diseno::Diseno
Instruccional::Roberto
Arcila","https://www.facebook.com/layumi1990","2020-07-19T23:30:
33Z","2020-07-19T23:30:33Z","2015","2015","2015","CEducativo
```


Figura 48. Metadatos almacenados en formato CSV**.**

Las pruebas descritas anteriormente evidenciaron la aplicación y cumplimiento de las respectivas reglas y restricciones definidas por el metamodelo para el mapeo de aplicación. De igual forma, las instancias de metadatos generadas corresponden a instancias del modelo de datos expuesto en el nivel M1 del metamodelo.

7.2.4. Prueba de identificación de recursos.

El objeto de esta prueba de sistema fue corroborar la identificación de los metadatos del recurso a través de la asignación de una URI (local o en la Web). La Figura 49 evidencia que cada descripción de un REA es identificada de forma particular e irrepetible.

Resource title:	Concepto de Punteros en C
Alternative title:	Concepto de Punteros en C
Author:	Carlos Casallas
Author's URI:	https://open.spotify.com/album/2f8psgefNAeQztCpBnARL4
Key Words:	C;Punteros
Upload date:	7-Oct-1996
Editor:	Carlos Marin
Serie:	12456;
Local Identifier:	http://localhost:8080/jspui/handle/123456789/13
ISSN:	1563-2598
Appears in Collections:	Programación

Figura 49. Identificación de la descripción del REA.

Esta URI identifica la descripción del recurso, con lo cual además de hacerla diferente a otras descripciones, permite que sea enlazada por un dataset.

7.2.5. Prueba de serialización de metadatos.

El objeto de esta prueba de integración y sistema fue corroborar el proceso de serialización realizado como artefacto resultante de la gestión de los metadatos. Esta serialización utilizó el mapeo del perfil de aplicación PA-REA para dataset. En la Figura 50 se observa que cada modificación realizada en la estructura de metadatos de cada recurso, es serializada a los respectivos RDF Graph. Se aclara que para cada componente, Dspace genera un RDF Graph, el cual es exportado al TripleStore de Fuseki.

```
[2020-03-23 22:29:00] Admin      INFO  [1] GET http://localhost:8080/fuseki/$/server
[2020-03-23 22:29:00] Admin      INFO  [1] 200 OK (25 ms)
[2020-03-23 22:41:33] Fuseki     INFO  [2] PUT http://localhost:8080/fuseki/dspace/data?graph=http%3A%2F%2
080%2Frdf%2F%2Fresource%2F123456789%2F1
[2020-03-23 22:41:33] Fuseki     INFO  [2] Body: Content-Length=-1, Content-Type=application/rdf+xml, Char
f/XML : Count=9 Triples=9 Quads=0
[2020-03-23 22:41:34] Fuseki     INFO  [2] 200 OK (1,324 s)
[2020-03-23 22:41:34] Fuseki     INFO  [3] DELETE http://localhost:8080/fuseki/dspace/data?graph=http%3A%2
3A8080%2Frdf%2F%2Fresource%2F123456789%2F77
[2020-03-23 22:41:35] Fuseki     INFO  [3] 404 No such graph: http://localhost:8080/rdf//resource/12345678
[2020-03-23 22:42:10] Fuseki     INFO  [4] PUT http://localhost:8080/fuseki/dspace/data?graph=http%3A%2F%2
080%2Frdf%2F%2Fresource%2F123456789%2F77
[2020-03-23 22:42:10] Fuseki     INFO  [4] Body: Content-Length=-1, Content-Type=application/rdf+xml, Char
f/XML : Count=7 Triples=7 Quads=0
[2020-03-23 22:42:11] Fuseki     INFO  [4] 201 Created (1,069 s)
[2020-03-23 22:42:11] Fuseki     INFO  [5] PUT http://localhost:8080/fuseki/dspace/data?graph=http%3A%2F%2
080%2Frdf%2F%2Fresource%2F123456789%2F78
[2020-03-23 22:42:11] Fuseki     INFO  [5] Body: Content-Length=-1, Content-Type=application/rdf+xml, Char
f/XML : Count=7 Triples=7 Quads=0
[2020-03-23 22:42:12] Fuseki     INFO  [5] 201 Created (979 ms)
[2020-03-23 22:42:22] Fuseki     INFO  [6] PUT http://localhost:8080/fuseki/dspace/data?graph=http%3A%2F%2
080%2Frdf%2F%2Fresource%2F123456789%2F78
[2020-03-23 22:42:22] Fuseki     INFO  [6] Body: Content-Length=-1, Content-Type=application/rdf+xml, Char
f/XML : Count=7 Triples=7 Quads=0
[2020-03-23 22:42:23] Fuseki     INFO  [6] 200 OK (812 ms)
```

Figura 50. Despliegue de la RDFizacion de un Recurso.

7.2.6. Prueba de generación de RDF Graph.

El objeto de esta prueba de integración y sistema fue corroborar la generación correcta de los grafos RDF, al ser serializados en el TripleStore. Como resultado, en la Figura 51 se observa la carga de las instancias de metadatos de cada recurso, expresadas en RDF Graph bajo un formato Turtle. Estos grafos son gestionados por Fuseki, a través de cargas automáticas.

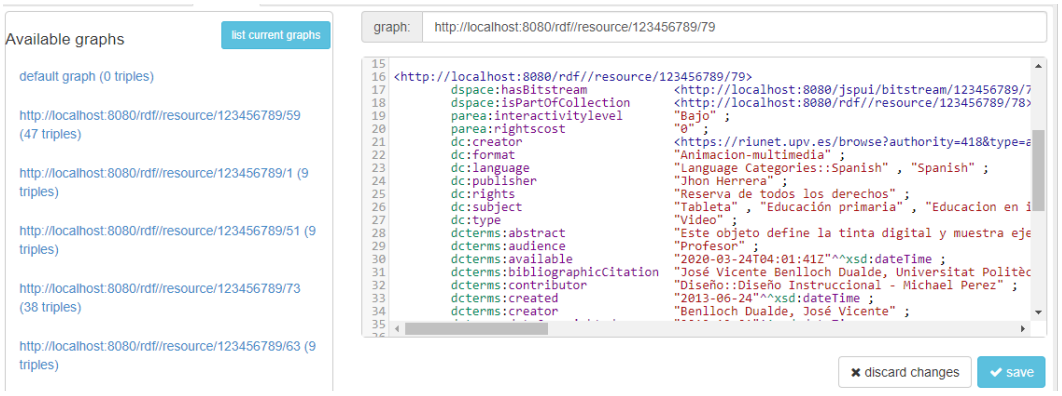


Figura 51. RDF Graph Cargado en Fuseki.

A continuación se muestra un ejemplo de un RDF Graph serializado bajo las restricciones definidas por la mapeo PA-REA para dataset:

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix parea: <http://parea.com/parea#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix fuseki: <http://jena.apache.org/fuseki#> .
@prefix dspace: <http://digital-repositories.org/ontologies/dspace/0.1.0#> .
@prefix lrmi: <http://purl.org/dcx/lrmi-terms/> .
@prefix tdb: <http://jena.hpl.hp.com/2008/tdb#> .
@prefix aiiso: <http://vocab.org/aiiso/schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ja: <http://jena.hpl.hp.com/2005/11/Assembler#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://localhost:8080/rdf//resource/123456789/57>
dspace:hasBitstream                <http://localhost:8080/jspui/bitstream/123456789/57/1/Proteina.txt> ;
dspace:isPartOfCollection          <http://localhost:8080/rdf//resource/123456789/56> ;
parea:annotationdesc               "Usado en grados: 1.º a 4.º grado; 5.º a 6.º grado; 7.º a 9.º grado;
parea:annotationentity             "Autor" ;
dc:creator                         <https://www.bloomberg.com/profile/company/7694823Z:GR> ;
dc:format                          "Imágenes" ;
dc:language                        "Language Categories::Spanish" , "Spanish" , "French" , "Language
                                   Categories::French", "Language Categories::English", "English";
dc:rights                          "licencia CC BY-SA 4.0 internacional." ;
dc:subject                         "Luz" , "Energía renovable" , "Energía" , "Ecología" , "Fotosíntesis" ,
                                   "Medio ambiente (general)", "Energía solar";
dc:type                            "Imagen" ;
dcterms:abstract                   "Información e ideas: Un ejemplo de la conversión de energía radiante
                                   en energía química."@en-US ;
dcterms:audience                   "Aprendiz (Alumno)" ;
dcterms:available                  "2020-02-15T18:02:15Z"^^xsd:dateTime ;
dcterms:bibliographicCitation      "Arbol. MediaHouse GmbH. 2019.";
dcterms:created                    "2016-03"^^xsd:dateTime ;
dcterms:creator                    "MediaHouse GmbH" ;
dcterms:date                       "2019"^^xsd:dateTime ;
dcterms:dateCopyrighted            "2017"^^xsd:dateTime ;
dcterms:description                "Identificación de fuentes de fotosíntesis"@en-US ;
dcterms:educationlevel             "Media" , "Básica Primaria" ;
dcterms:extent                     "319,6 kByte" ;
dcterms:hasPart                    <http://localhost:8080/jspui/bitstream/123456789/57/1/Proteina.txt> ;
dcterms:hasVersion                 "1.0" ;
dcterms:isPartOf                   <http://localhost:8080/rdf//resource/123456789/56> ;
dcterms:issued                     "2020-02-15T18:02:15Z" ;
dcterms:modified                   "2019-08-22"^^xsd:dateTime ;
dcterms:rights                     <https://www.siemens-stiftung.org/> ;
dcterms:title                      "Árbol" ;
lrmi:educationalsubject            "Biología", "Conocimiento del medio natural, social y cultural" ;
bibo:uri                           <https://medienportal.siemens-stiftung.org/es/fotosintesis-104153> ;
void:sparqlEndpoint                <http://localhost:8080/fuseki/dspace/sparql> ;
aiiso:institution                  "Otro" ;
aiiso:knowledgegrouping            "Educación Formal" ;
foaf:homepage                      <http://localhost:8080/jspui> ;
```

Al verificar tanto la estructura como la generación del RDF Graph, el mapeo realizado cumple con los requerimientos sintácticos establecidos por el modelo de datos RDF y verificados por un validador (Figura 52).

226	<http://localhost:8080/rdf//resource/123456789/60>	dc:subject	"Generación de energía eléctrica"
227	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:created	"2016"^^xsd:dateTime
228	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:dateCopyrighted	"2017"^^xsd:dateTime
229	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:hasVersion	"1.0"
230	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:description	"Sin descripcion"@en-US
231	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:rights	<https://www.siemens-stiftung.org/>
232	<http://localhost:8080/rdf//resource/123456789/60>	dcterms:modified	"2019-08-22"^^xsd:dateTime
233	<http://localhost:8080/rdf//resource/123456789/60>	dc:format	"Documento electronico"

Figura 52. Resultados del Validador RDF Turtle.

La Figura 52 también prueba y verifica tanto la adecuada construcción de tripletas, como la generación de metadatos enriquecidos a través del enlace de recursos.

7.2.7. Prueba del enlace de datos.

La verificación del manejo de datos enlazados, acorde al nivel 5 del modelo LOD, como prueba de sistema se realizó en dos momentos:

- Verificación en la plataforma de gestión de recursos: Para corroborar el uso y adecuado manejo de datos enlazados que enriquecen la descripción de los REA, la Figura 53 muestra el despliegue de URIs en campos que fueron definidos para tal fin, de acuerdo a la construcción de modelo de datos para describir REA.

Resource title:	Cálculo en el área de Albañilería
Author:	Zamit, Virginia Rodríguez., Mary
Creator's URI:	https://www.facebook.com/layumi1990
Key Words:	Matematicas;Area
Editor:	Editors Categories::Jhon Herrera
Abstract:	entender el trabajo de Albañil, cómo calcular el material y finalmente comprender porque es necesario contratar uno, para la construcción.
Educative description:	Aprender sobre los conocimientos Matemáticos en la Albañilería, y en particular sobre el cálculo de Área.
Local Identifier:	http://localhost:8080/jspui/handle/123456789/17
Appears in Collections:	Programación

Figura 53. Enlace de datos desde la plataforma de gestión de recursos.

- Verificación del enlace de datos en los grafos RDF: Para corroborar el enlace de datos sobre los grafos, se realizaron consultas SPARQL con el fin de recuperar información de los RDF Graph:

```
prefix void: <http://rdfs.org/ns/void#>
prefix parea: <http://parea.com/parea#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix fuseki: <http://jena.apache.org/fuseki#>
prefix dspace: <http://digital-
repositories.org/ontologies/dspace/0.1.0#>
prefix lrmi: <http://purl.org/dcx/lrmi-terms/>
prefix tdb: <http://jena.hpl.hp.com/2008/tdb#>
prefix aiiso: <http://vocab.org/aiiso/schema#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix ja: <http://jena.hpl.hp.com/2005/11/Assembler#>
prefix dcterms: <http://purl.org/dc/terms/>
prefix bibo: <http://purl.org/ontology/bibo/>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix dc: <http://purl.org/dc/elements/1.1/>

SELECT ?titulo ?autor ?url_autor
WHERE {
        GRAPH ?graph{
                ?recurso dcterms:title ?titulo .
                ?creador dcterms:creator ?autor .
                ?url_creador dc:creator ?url_autor
         }
}
LIMIT 25
```

La Figura 54 muestra el resultado de la consulta SPARQL, donde se detallan el título del recurso, el nombre del autor y la URI del autor.

6	"El blog, espacio interactivo para la clase de Matematicas"	"Acosta, Diana"	<http://www.cumbrecp.com/congresosec/conferencistas/item/472-diana-acosta-colombia.html>
7	"La Edad Media: ¿Qué es el Feudalismo?"	"Herrera, Catalina"	<https://co.linkedin.com/in/catalina-herrera-807ba326>
8	"Introductory Biology"	"Ray, Diviya"	<https://biology.mit.edu/profile/angelika-amon/>
9	"Introductory Biology"	"Amon, Angelika"	<https://biology.mit.edu/profile/angelika-amon/>
10	"Introductory Biology"	"Sive, Hazel"	<https://biology.mit.edu/profile/angelika-amon/>
11	"Questions at Issue"	"Gosse, Edmund"	<https://es.wikipedia.org/wiki/Edmund_Gosse>
12	"¿Cómo determinan el clima del planeta los rayos del sol?"	"Cortez, Diego"	<https://www.colombiaaprende.edu.co/es/usuarios/diegocortes>
13	"Sistema Internacional de Unidades (SI)"	"Martinez, Misael"	<https://co.linkedin.com/in/martinezotero>

Figura 54. Resultados de Mapeo.

El campo “creador” aparece como una URI activa, la cual al ser seleccionada despliega el contenido enlazado a este tag de metadatos en una nueva ventana del browser. Esta consulta permitió verificar:

- La adecuada carga de metadatos enriquecidos a través del enlace de datos.
- El adecuado despliegue de las URI que enlazan recursos externos al dataset.
- El enriquecimiento semántico del REA provisto por la información recuperada al acceder a las URI obtenidas de la consulta.

Las pruebas descritas anteriormente evidenciaron el funcionamiento de la aplicación y cumplimiento de las respectivas reglas y restricciones definidas por el metamodelo para el mapeo de dataset. De igual forma, las instancias de metadatos generadas corresponden a instancias del modelo de datos expuesto en el nivel M1 del metamodelo, que adicionalmente se encuentran conformes al componente de enlace de datos del nivel M2 del metamodelo.

7.2.8. Prueba de consultas.

Para confrontar los resultados de la implementación de los artefactos producidos por el metamodelo (mapeo para aplicación y mapeo para dataset), a continuación se presenta un ejemplo de consulta de datos en Dspace y en Fuseki, representativa como prueba de sistema. La Figura 55 muestra los resultados obtenidos al consultar el título y autor de los recursos cuya palabra clave es “Ecología”.

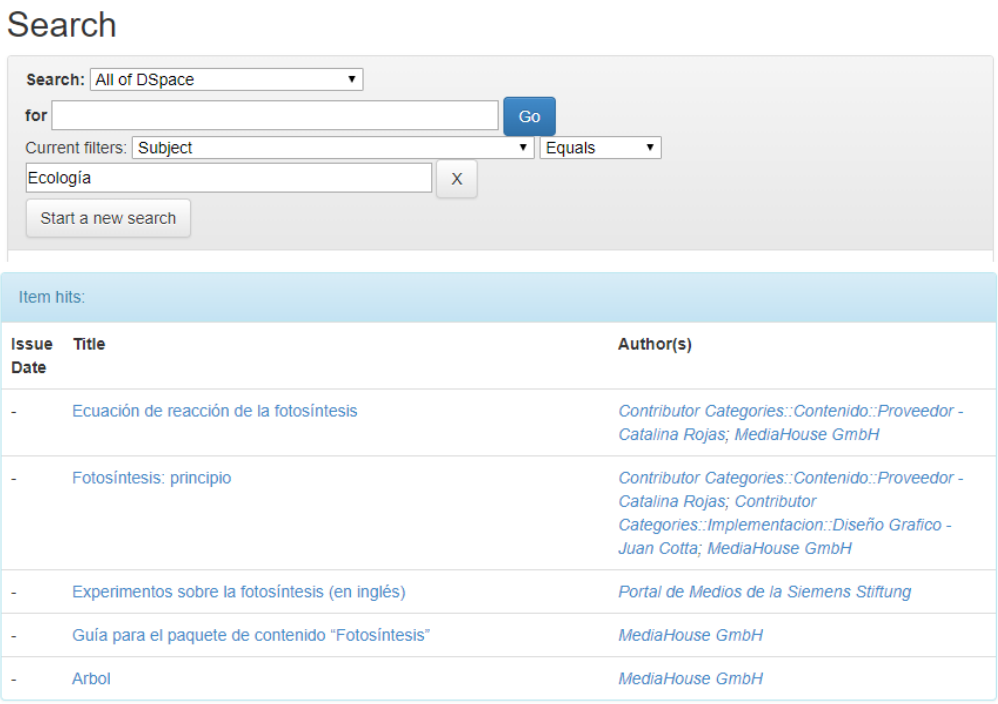
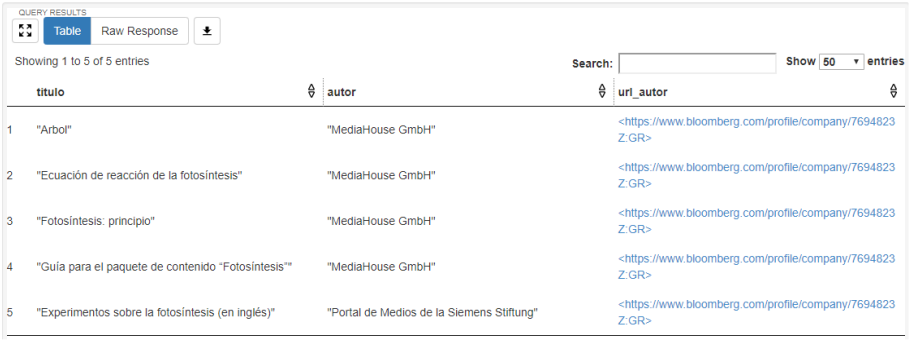


Figura 55. Consulta realizada en Dspace.

En cuanto al mapeo para dataset, la consulta realizada al RDF Graph, usando los mismos parámetros, validó la generación de los mismos resultados (Figura 56).

```
SELECT ?titulo ?autor ?url_autor
WHERE {
        GRAPH ?graph{
                ?recurso dcterms:title ?titulo .
                ?creador dcterms:creator ?autor .
                ?url_creador dc:creator ?url_autor .
                ?recurso dc:subject "Ecología"
         }
}
LIMIT 25
```



QUERY RESULTS

Table | Raw Response

Showing 1 to 5 of 5 entries

Search: Show 50 entries

	titulo	autor	url_autor
1	"Arbol"	"MediaHouse GmbH"	<https://www.bloomberg.com/profile/company/7694823Z:GR>
2	"Ecuación de reacción de la fotosíntesis"	"MediaHouse GmbH"	<https://www.bloomberg.com/profile/company/7694823Z:GR>
3	"Fotosíntesis: principio"	"MediaHouse GmbH"	<https://www.bloomberg.com/profile/company/7694823Z:GR>
4	"Guía para el paquete de contenido "Fotosíntesis""	"MediaHouse GmbH"	<https://www.bloomberg.com/profile/company/7694823Z:GR>
5	"Experimentos sobre la fotosíntesis (en inglés)"	"Portal de Medios de la Siemens Stiftung"	<https://www.bloomberg.com/profile/company/7694823Z:GR>

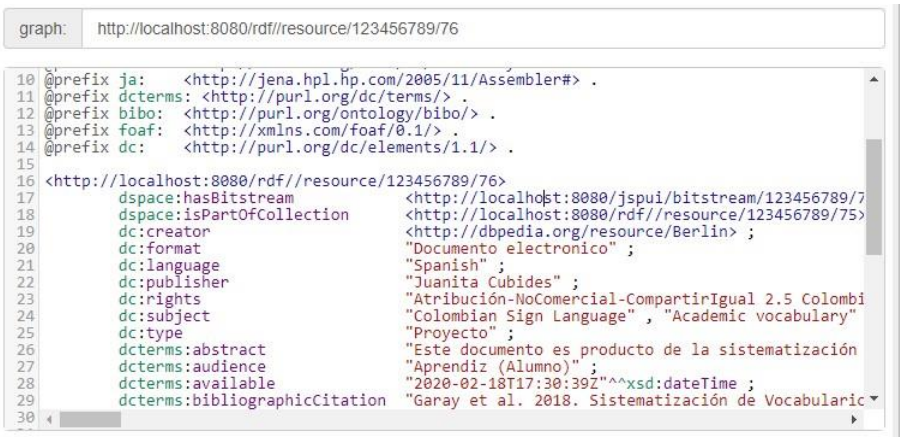
Figura 56. Consulta realizada en Fuseki.

Esta comparación evidenció que las instancias de metadatos (nivel M0 del metamodelo), bien sea mapeadas para aplicación o para dataset, cumplen con las propiedades y restricciones definidas por el modelo de datos (nivel M1 del metamodelo).

7.2.9. Validación de grafos.

Para ejecutar la verificación de la generación de grafos RDF, en esta prueba de integración y sistema se realizaron los siguientes pasos:

a. Al realizar el proceso de serialización, Dspace genera un RDF Graph en formato Turtle, de cada elemento contenido (Figura 57).



graph: http://localhost:8080/rdf//resource/123456789/76

```
@prefix ja:    <http://jena.hpl.hp.com/2005/11/Assembler#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo:  <http://purl.org/ontology/bibo/> .
@prefix foaf:  <http://xmlns.com/foaf/0.1/> .
@prefix dc:    <http://purl.org/dc/elements/1.1/> .

<http://localhost:8080/rdf//resource/123456789/76>
        dspace:hasBitstream             <http://localhost:8080/jspui/bitstream/123456789/7
        dspace:isPartOfCollection       <http://localhost:8080/rdf//resource/123456789/75>
        dc:creator                      <http://dbpedia.org/resource/Berlin> ;
        dc:format                       "Documento electronico" ;
        dc:language                     "Spanish" ;
        dc:publisher                    "Juanita Cubides" ;
        dc:rights                       "Atribución-NoComercial-CompartirIgual 2.5 Colombi
        dc:subject                      "Colombian Sign Language" , "Academic vocabulary"
        dc:type                         "Proyecto" ;
        dcterms:abstract                "Este documento es producto de la sistematización
        dcterms:audience                "Aprendiz (Alumno)" ;
        dcterms:available               "2020-02-18T17:30:39Z"^^xsd:dateTime ;
        dcterms:bibliographicCitation   "Garay et al. 2018. Sistematización de Vocabularic
```

Figura 57. Ejemplo de RDF Graph Turtle Exportado.

En esta figura se observa que el predicado dc:creator tiene una URI externa: http://dbpedia.org/resource/Berlin.

b. A través del uso de validadores RDF como por ejemplo IDLab Turtle Validator (IDLab, 2020), se verificó la correcta sintaxis del grafo (Figura 58).

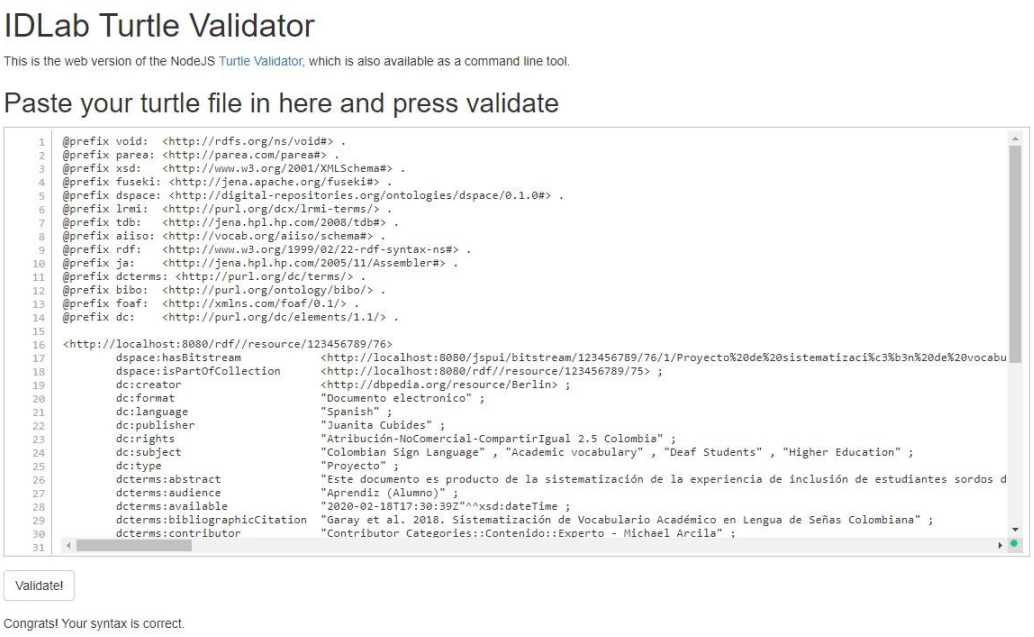
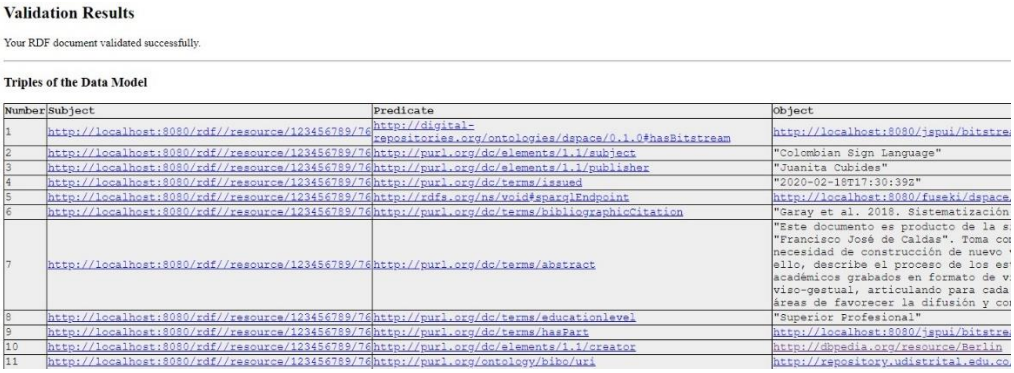


Figura 58. Validación del RDF Graph Turtle.

c. Por medio de un Validation Service (W3C, 2006), se hizo el parseo del grafo y su exportación a tripletas e imagen de grafo.



Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://localhost:8080/rdf//resource/123456789/76	http://digital-repositories.org/ontologies/dspace/0.1.0#hasBitstream	http://localhost:8080/jspui/bitstre
2	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/elements/1.1/subject	"Colombian Sign Language"
3	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/elements/1.1/publisher	"Juanita Cubides"
4	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/terms/issued	"2020-02-18T17:30:39Z"
5	http://localhost:8080/rdf//resource/123456789/76	http://rdfs.org/ns/void#sparqlEndpoint	http://localhost:8080/fuseki/dspace
6	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/terms/bibliographicCitation	"Garay et al. 2018. Sistematización
7	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/terms/abstract	"Este documento es producto de la s "Francisco José de Caldas". Toma co necesidad de construcción de nuevo ello, describe el proceso de los es académicos grabados en formato de v viso-gestual, articulando para cada áreas de favorecer la difusión y co
8	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/terms/educationlevel	"Superior Profesional"
9	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/terms/hasPart	http://localhost:8080/jspui/bitstre
10	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/dc/elements/1.1/creator	http://dbpedia.org/resource/Berlin
11	http://localhost:8080/rdf//resource/123456789/76	http://purl.org/ontology/bibo/uri	http://repository.udistrital.edu.co

Figura 59. Tripletas RDF del Grafo exportado.

En estas tripletas (Figura 59) se verifica en la línea 10 la aparición del enlace externo a DBpedia. Al dar clic sobre la URI respectiva, es cargado automáticamente el enlace destino. En cuanto a la generación del grafo, la Figura 60 visualiza la carga de la URI de DBpedia como un enlace.

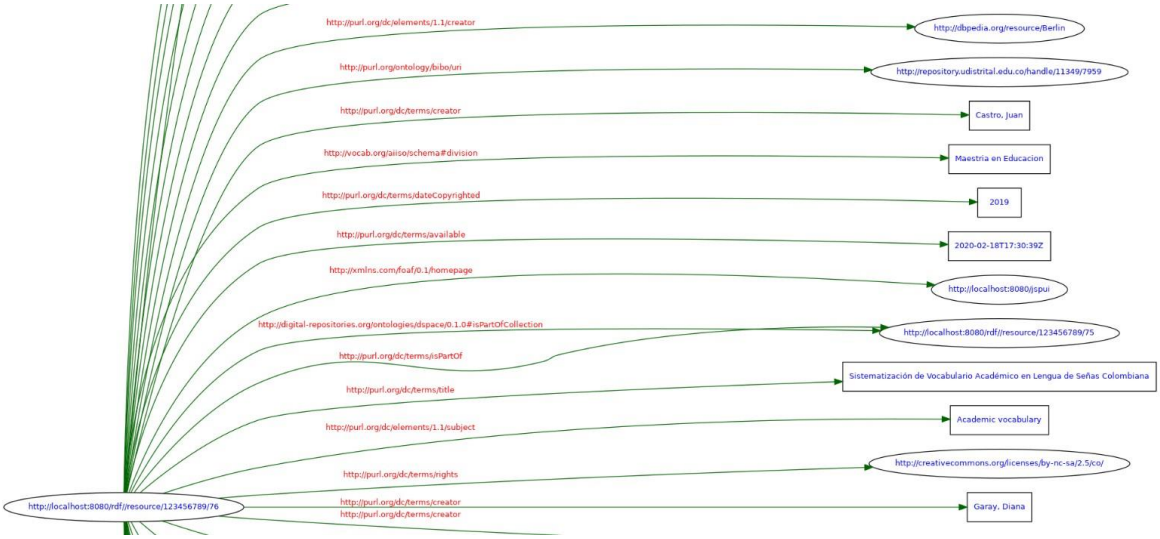


Figura 60. Visualización del Grafo RDF.

Tanto en las tripletas como en el grafo se puede observar la presencia de enlaces internos y externos. Estos últimos enlazan información que enriquecen las propiedades de los recursos.

7.2.10. Requisitos de conexión a la nube LOD.

Como lo expresa The Linked Open Data Cloud (McCrae, 2020), existe un procedimiento para contribuir con datasets a la nube LOD. Para enviar un dataset, se debe publicar datos de acuerdo con los principios de datos enlazados:

- Debe haber http: // (o https: //) URI resolubles.
- Deben resolver, con o sin negociación de contenido, los datos RDF en uno de los formatos RDF populares (RDFa, RDF / XML, Turtle, N-Triples).
- El dataset debe contener al menos 1000 tripletas.
- El dataset debe estar conectado a través de enlaces RDF a un dataset que ya está en la nube LOD. Esto significa que su dataset debe usar URI del otro dataset o viceversa. Se requieren al menos 50 enlaces.
- El acceso a todo el dataset debe ser posible mediante RDF Crawling, RDF dump o punto de acceso SPARQL.

Estos requisitos son manejables bajo el modelo de datos generado, dado que el mapeo de dataset, provisto por el perfil de aplicación PA-REA, genera grafos en formatos RDF Turtle, y cada grafo cuenta con al menos 5 propiedades que son enriquecidas a través de enlace de datos a recursos externos, propiedades que pueden ser enriquecidas al enlazar URIs de datasets que estén en la nube LOD. En cuanto al punto de acceso SPARQL, Fuseki tiene la capacidad de generar un punto de acceso para la consulta de datos.

Por último, dado que las instancias de metadatos de REA están almacenadas en un repositorio o plataforma, y cada una cuenta con un identificador irrepetible, fácilmente se puede superar la cantidad mínima de enlaces requeridos, al proveer recursos enlazables desde la nube LOD. En cuanto a los datos requeridos para subir el dataset a la nube LOD, son muy pocos y no presentan un alto grado de dificultad para su diligenciamiento (Figura 61).

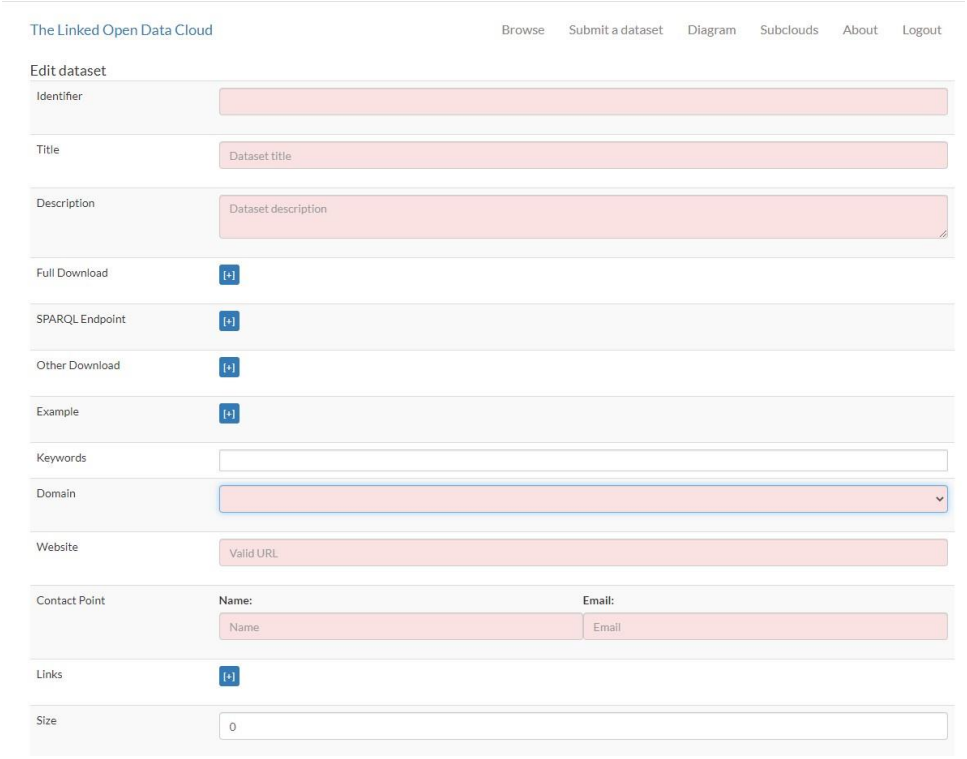


Figura 61. Datos requeridos por LOD Cloud.

El despliegue realizado del modelo de datos, superó las pruebas técnicas ejecutadas sobre el workflow por el cual transitan los metadatos de los REA, desde la ingesta de los metadatos hasta los respectivos mapeos, extrayendo la información requerida a través de las interfaces de consulta de Dspace y Fuseki.

El Crosswalk realizado, cumplió con todas las restricciones provistas por el modelo de datos RDF y definidas en el nivel M2 del metamodelo, tales como enlace de datos, estándares de metadatos, dimensiones de calidad, modelo de datos RDF y propiedades de los REA. Lo cual permite la obtención de un RDF Graph que representa fielmente un REA mapeado en Dublin Core. Es decir, las instancias obtenidas en el nivel M0, tanto en DC como en RDF son equivalentes en cuanto a propiedades y contenidos.

Posterior a la validación del despliegue del modelo, se procedió a validar el cubrimiento de propiedades provisto por el modelo de datos, de acuerdo a los diferentes tipos de REA clasificados para esta investigación (véase sección 2.1.1).

7.3. Validación del cubrimiento de REA.

Teniendo en cuenta que los REA son clasificados en diferentes categorías (MEN, 2012), este tipo de prueba se diseñó con el fin de validar que los elementos de metadatos del modelo de datos, producto del metamodelo propuesto, cubren los diferentes tipos de REA atendiendo a sus particularidades. Para llevar a cabo esta prueba, alimentó de forma manual la interfaz de ingesta de Dspace y se verificó la carga de las propiedades, según sus particularidades.

De acuerdo con la tipología de REA (véase Sección 2.1.1), las Tablas 31, 32 y 33 evidencian que la descripción de cada tipo de REA es cubierto por las propiedades del modelo de datos propuesto, considerando que como conclusión de la revisión del estado de la cuestión, se diseñó un solo modelo de datos para describir los diferentes tipos de recursos. Se aclara que los

ejemplos presentados corresponden a recursos educativos reales, por lo que la falta de completitud de algunos campos del modelo corresponde a la falta de descripción de los recursos por parte de sus proveedores, lo cual dificulta su clasificación, consulta y recuperación como recursos educativos.

Tabla 31. Ejemplos de Descripción de REA -I.

Elemento	Curso Virtual	Apps para Educación	Objetos de Aprendizaje
Identificador	-	-	-
Título	UNAB Bilingüe - Curso virtual de inglés para docentes y administrativos	Khan Academy	Classical Genetics Simulator
Subtítulo	-	-	-
Palabras Claves	Idiomas, Inglés, Extensión	Apps, Educación	genetics, biology, computer-based, drosophila, plant, simulation, online, inheritance, interactive, lab
Descripción (Abstract)	Crear una comunidad educativa global donde el inglés se constituye en pilar indispensable en la internacionalización y por ende en la creación de una cultura acorde a las exigencias del mundo actual.	El lugar perfecto para repasar aquel contenido que no les ha quedado claro a tus hijos/as, consultar aquella duda que quedo pendiente de resolver y encontrar la solución a ese problema que se les resiste.	A customizable teaching tool that simulates genetic inheritance including linkage over generations. An instructor can set broad or specific parameters for the virtual populations that their students examine. Model organisms such as Arabidopsis, Drosophila, or mouse can be used.
Cita	-	-	-
Idioma General	EN	ES	EN
Fecha de Versión Inicial	Jul 22, 2016	Sep, 2010	agosto 29, 2008
Versión	1.0	-	2
Contribución	Autor: Departamento de Lenguas	Autor: Tiching	Author: Ben Adamczyk
Fecha del Ciclo de Vida	Actualización: Jul 4, 2018	-	Updated: julio 26, 2020
Formato Técnico	Sitio Web	Animación/multimedia	Sitio Web, Flash
Tamaño del Formato	30Mb	25 Mb	45 Mb
Ubicación Técnica	https://www.unab.edu.co/publicaciones/curso-virtual-ingl%C3%A9s-docentes-y-administrativos	http://co.tiching.com/khan-academy/	http://cgslab.com/
Tipo Técnico	Windows	Android	Android
Versión Mínima	Windows 10 o superior	7.8	10
Tipo de interactividad	Descriptiva	-	-

Elemento	Curso Virtual	Apps para Educación	Objetos de Aprendizaje
Nivel de interactividad	Alta	-	-
Tipo de Recurso	Módulo	Aplicación	-
Rol Usuario Final	Aprendiz, Docente	-	-
Nivel Educativo	Superior - Profesional	-	-
Contexto Educativo	-	-	-
Rango de Edad	-	-	-
Descripción Educativa	Fortalecer el inglés como segunda lengua en la formación docente	Fortalecer las competencias básicas de matemáticas	A web-based genetics lab, allowing students to apply lessons in Mendelian genetics to real-world scenarios
Licenciamiento y Otras Restricciones	Módulo de acceso restringido bajo contraseña y usuario provistos por la Unab.	App de acceso abierto a través de plataformas móviles	Unknown
Descripción de los Derechos	https://creativecommons.org/	https://creativecommons.org/	Unknown
Costo de los Derechos	0	0	0
Fecha del Licenciamiento	Jul 4, 2018	Sep, 2010	agosto 29, 2008
Persona (anotación)	-	-	-
Fecha (anotación)	-	-	-
Descripción (anotación)	-	-	-
Institución	Universidad Autónoma de Bucaramanga	Corporación Tiching	Universidad de Wisconsin
Organización			
Grupo	Dpto. de Lenguas	-	-
División		-	-
Taxón	Idioma Extranjero	-	Science and Technology / Agriculture and Environmental Sciences
Curso	Inglés I	-	-

Tabla 32. Ejemplos de Descripción de REA –II.

Elemento	Textual	Sonoro	Visual
Identificador	-	-	-
Título	Taller de Multimedia	Cómo hacer un uso adecuado del tablero en clase	New Science
Subtítulo	-	-	-
Palabras Claves	-	-	lightning strike, cloud to cloud, static electricity discharge, powerlines, thunderstorm, mesa, Arizona
Descripción (Abstract)	-	3 pensamientos renovados para	-

Elemento	Textual	Sonoro	Visual
		aprovechar el pizarrón al máximo.	
Cita	Prendes Espinosa, Mª Paz e Isabel Solano Fernández (2001). "Taller de Multimedia", Universidad de Murcia, presentación en el Congreso de Oviedo. Disponible en: <bibliotecadigital.tamaulipas.gob.mx/>.	-	-
Idioma General	ES	ES	EN
Fecha de Versión Inicial	2001	Oct 3. 2020	-
Versión			
Contribución	Autor: Prendes Espinosa, Mª Paz Autor: Solano Fernández, Isabel	-	Autor: Vann, W.
Fecha del Ciclo de Vida	-	-	-
Formato Técnico	Documento Electrónico	Audio	Imagen
Tamaño del Formato	500 Kb	9 Min	100 Kb
Ubicación Técnica	https://www.goconqr.com/es/mindmap/3255702/taller-de-multimedia	https://open.spotify.com/episode/2XYQyh8EOJNSv2axoPO1nC?go=1&utm_source=embed_v3&t=0&nd=1	http://www.edupic.net/Images/Weather/lightning5428.jpg
Tipo Técnico	Procesador de texto	Reproductor de Audio	Visor de Imágenes
Versión Mínima	2010	-	-
Tipo de interactividad	-	-	-
Nivel de interactividad	-	-	-
Tipo de Recurso	-	Broadcast	Imagen
Rol Usuario Final	-	-	-
Nivel Educativo	-	-	-
Contexto Educativo	-	-	-
Rango de Edad	-	-	-
Descripción Educativa	Fortalecer la abstracción teórica de los elementos multimediales	Fortalecer el uso de mediaciones como el tablero	Abstraer los elementos de los rayos
Licenciamiento y Otras Restricciones	-	-	-
Descripción de los Derechos	-	-	-
Costo de los Derechos	-	-	-
Fecha del Licenciamiento	-	-	-
Persona (anotación)	-	-	Vann, W.
Fecha (anotación)	-	-	2020

Elemento	Textual	Sonoro	Visual
Descripción (anotación)	-	-	Imagen satelital Nasa-5
Institución	Universidad de Murcia	Profes Digitales	-
Organización			-
Grupo	-	-	-
División	-	-	-
Taxón	-	-	-
Curso	-	-	-

Tabla 33. Ejemplos de Descripción de REA –III.

Elemento	Audiovisual	Multimedia
Identificador		
Título	Introducción a los tipos textuales	Las nubes
Subtítulo		
Palabras Claves	Lingüística	Vocales
Descripción (Abstract)	Este recurso en formato video muestra los tipos de mensaje de texto, y cómo dependen de la intención del emisor al producir su mensaje. Los contenidos de este recurso están dirigidos a docentes y estudiantes de Educación Básica Secundaria y Educación Media para el área de Lenguaje.	Pulsa sobre las nubes y aprende las vocales.
Cita		-
Idioma General	ES	ES
Fecha de Versión Inicial	2014	-
Versión	1.0	-
Contribución	Autor: López Castro, Juan Carlos	Autor Pipoclub
Fecha del Ciclo de Vida	-	-
Formato Técnico	Video	Animación/multimedia
Tamaño del Formato	15 Min	10 Min
Ubicación Técnica	https://www.youtube.com/watch?v=7IDI89Hp4UI	https://www.pipoclub.com/juegos-para-ninos-gratis/juego-nubes-2.html
Tipo Técnico	Chrome	Chrome
Versión Mínima	5	5
Tipo de interactividad	-	-
Nivel de interactividad	-	-
Tipo de Recurso	-	Juego educativo
Rol Usuario Final	-	-
Nivel Educativo	-	-
Contexto Educativo	-	-
Rango de Edad	-	0-3
Descripción Educativa	Fortalecer los tipos textuales en el desarrollo de la lingüística	Fortalecer la identificación y asociación de las vocales
Licenciamiento y Otras Restricciones	-	-
Descripción de los Derechos	-	-
Costo de los Derechos	-	-

Elemento	Audiovisual	Multimedia
Fecha del Licenciamiento	-	-
Persona (anotación)	-	-
Fecha (anotación)	-	-
Descripción (anotación)	-	-
Institución	Corporación Educatina	Pipo Club
Organización	-	-
Grupo	-	-
División	-	-
Taxón	Lingüística	-
Curso	-	-

En las tablas anteriores se evidenciaron aspectos tales como:

- El tamaño del formato varía según el tipo de recurso (peso en bytes o tiempo en minutos). Sin embargo, la parametrización de la propiedad permite que estas variaciones puedan ser alimentadas en la propiedad.
- Los vocabularios controlados definidos para la propiedad “tipo de recurso” permiten el manejo de los diferentes tipos de REA. Lo cual evidencia que el metamodelo presentado es flexible tanto en la inserción de una nueva propiedad para los recursos existentes, como para incluir un nuevo tipo de recurso educativo.
- Para que un material digital adquiera la connotación de recurso educativo abierto, debe proveer un conjunto de propiedades que permitan su descripción educativa, procedencia, clasificación en el dominio de conocimiento, entre otros.

En resumen, el modelo de datos para describir REA producto de esta investigación, mejorado a través de la inclusión de elementos de calidad de datos y enriquecido a través del uso del modelo LOD, como criterio de éxito cubre todas las clasificaciones de REA, mejorando su disponibilidad y capacidad de reutilización (Figura 62).

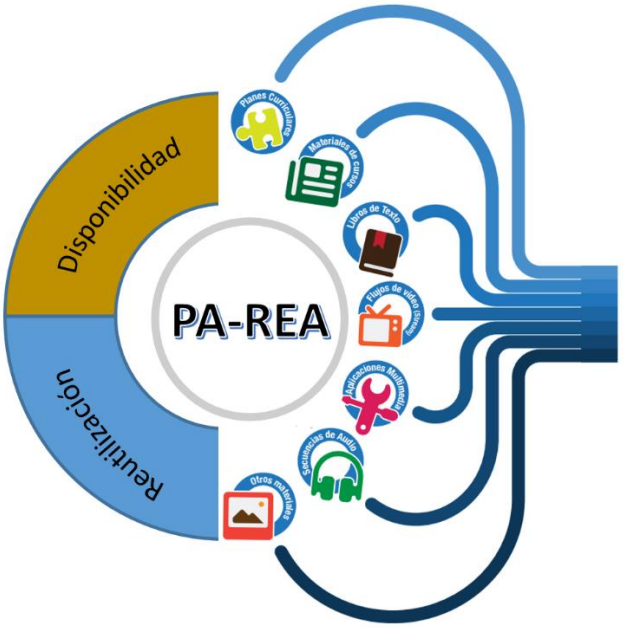


Figura 62. Cubrimiento de REAs por PA-REA. Adaptado de MEN (2012).

Como último escenario de validación del modelo para la descripción y enriquecimiento de REA, se procedió a realizar una validación empírica del modelo haciendo uso del enfoque GQM.

7.4. Validación Empírica del Modelo Diseñado.

Para validar empíricamente el modelo de datos REA, inicialmente se consideraron dos aspectos:

- Los modelos y métricas para evaluar la calidad de los datos no se pueden aplicar directamente sobre el modelo LOD, dado a aspectos como la calidad de la representación de datos o la coherencia con respecto a la información presentada en los otros datasets publicados, adicionando que los mecanismos de inferencia de conocimiento sobre la datos enlazados generalmente siguen una suposición de mundo abierto (Behkamal *et al.*, 2014; Baliyan y Kumar, 2018).
- La calidad de los datos se relaciona con dos perspectivas diferentes: (i) una perspectiva interna, que implica, por un lado, los datos en sí y, por otro lado, las ontologías que describen los datos; y (ii) una perspectiva externa que está impulsada por problemas y contenidos típicos relacionados con un dominio específico de los datos (Bonduel *et al.*, 2019).

Por tales razones, el proceso de validación del modelo se enfocó desde una perspectiva interna, que responde a cómo la descripción de metadatos realizada responde a criterios de calidad de datos enlazados (Zaveri *et al.*, 2012), en especial a las dimensiones de calidad trabajadas: Contextual y Confianza.

Considerando la dinámica de representación de los datos enlazados, para validar el modelo diseñado en esta investigación se usó el enfoque Goal-Question-Metric (GQM), enfoque en el cual los objetivos se refinan gradualmente en varias preguntas y cada pregunta se refina en métricas. Además, una métrica se puede utilizar para responder varias preguntas (Yaghouti *et al.*, 2015). Aunque este enfoque fue propuesto inicialmente sobre Ingeniería de Software, se ha venido aplicando en diferentes dominios de conocimiento, convirtiéndose en una metodología aceptada para la definición de métricas en el contexto de la evaluación de la calidad de los datos (Sarcia, 2010). Este enfoque también es utilizado para evaluar la calidad de los datos en ambientes LOD, como lo evidencian las investigaciones realizadas por Yaghouti *et al.* (2015), Behkamal *et al.* (2014), Behkamal (2014), Li *et al.* (2019) y Zahedi y Fathian (2016). A continuación se detalla tanto el enfoque como el proceso realizado.

7.4.1. Enfoque de validación para evaluar el modelo de datos: Modelo GQM.

El enfoque de GQM se basa en la suposición de que para que una organización mida de manera intencionada, primero debe especificar los objetivos para sí misma y sus proyectos. Luego debe rastrear esos objetivos a los datos que definen esos objetivos operacionalmente, y finalmente proporcionar un marco para interpretar los datos con respecto a los objetivos establecidos (Van Solingen y Berghout, 1998; Bassili *et al.*, 1994).

Como lo describen Behkamal *et al.* (2014), citando a Bassili *et al.* (1994), el modelo de medición tiene tres niveles:

- Nivel conceptual: se definen los objetivos específicos para un objeto de medición por una variedad de razones, desde diferentes puntos de vista en relación con un entorno particular.
- Nivel operativo: se desarrolla un conjunto de preguntas para caracterizar el camino hacia el logro de un objetivo específico ya definido en el nivel conceptual.
- Nivel cuantitativo: se define un conjunto de métricas para responder a las preguntas que se definieron en el nivel operativo, que se abordan de forma cuantitativa.

La Figura 63 muestra la operacionalización de estos tres niveles GQM, planteados para esta investigación.

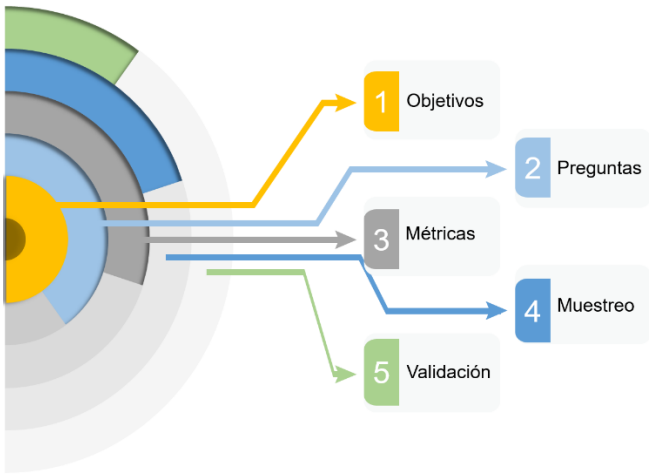


Figura 63. Aplicación de los niveles GQM.

Atendiendo a los procesos adelantados por Behkamal *et al.* (2014), Zahedi y Fathian (2016) y Behkamal (2014), a continuación se describe la operacionalización de cada uno de estos niveles aplicados a la validación del modelo de datos para REA.

7.4.2. Definición del Objetivo.

El objetivo motivador para el proceso de validación corresponde a:

Evaluar de la calidad inherente del Modelo de Datos PA-REA desde el punto de vista de las dimensiones de calidad de datos enlazados, en el contexto de LOD.

En este objetivo se especifica:

- Un propósito: evaluar.
- Un objeto de medición: Modelo de Datos PA-REA.
- Un punto de vista: dimensiones de calidad.
- Un problema: calidad inherente.
- Un entorno: LOD.

Como se estableció anteriormente, como constructos teóricos del metamodelo, se consideraron dos dimensiones para la calidad inherente al modelo LOD, por lo tanto, el objetivo principal se descompone en dos sub metas:

- Evaluación de la calidad contextual del modelo de datos desde la perspectiva de calidad de datos enlazados en el contexto de LOD;
- Evaluación de la calidad de confianza del modelo de datos desde la perspectiva de calidad de datos enlazados en el contexto de LOD;

Para trabajar con estas dimensiones, es importante recordar que Zaveri *et al.* (2012) definió un conjunto de sub dimensiones para cada uno de las dimensiones trabajadas, las cuales fueron analizadas en la construcción del modelo de datos, razón por la cual influyen en la construcción de las preguntas y las métricas que a continuación se plantean.

7.4.3. Formulación de preguntas.

A continuación se presentan las preguntas definidas para abordar los objetivos secundarios propuestos.

Calidad Contextual:
Las dimensiones contextuales dependen en gran medida del contexto de la tarea en cuestión, así como de las preferencias subjetivas del consumidor de datos. Hay tres dimensiones contextuales (Zaveri *et al.*, 2012):

- Completitud: se refiere al grado en que toda la información requerida está presente en un dataset en particular. En términos de datos enlazados, se clasifica de la siguiente manera: (a) Completitud del esquema, el grado en que se representan las clases y propiedades de una ontología, por lo que se puede llamar "integridad de la ontología", (b) Completitud de la propiedad, medida de los valores faltantes para una propiedad específica, (c) La Completitud de la población es el porcentaje de todos los objetos del mundo real de un tipo particular que están representados en los datasets y (d) La Completitud de la interconexión, debe considerarse especialmente en los datos enlazados y se refiere a grado en que las instancias del dataset están interconectadas.
- Cantidad de datos: se refiere a la cantidad y el volumen de datos que es apropiado para una tarea particular.
- Relevancia: se refiere a la provisión de información que está de acuerdo con la tarea en cuestión e importante para la consulta de los usuarios.

Con base en estos conceptos, las preguntas planteadas son:

- ¿Se abstraen y representan las entidades y propiedades del mundo real?
- ¿Se representan los valores posibles para una propiedad?
- ¿Se identifican las propiedades a enriquecer con el enlace de datos?

Calidad de Confianza:
Las dimensiones de confianza son aquellas que se centran en la confiabilidad del dataset. Hay cinco dimensiones de confianza (Zaveri *et al.*, 2012):

- Procedencia: se refiere a los metadatos contextuales que se centran en cómo representar, administrar y usar información sobre el origen de la fuente.
- Verificabilidad: se refiere al grado en que un consumidor de datos puede evaluar la exactitud de un dataset y, en consecuencia, su confiabilidad.
- Credibilidad: se define como el grado en que la información se acepta como correcta, verdadera, real y creíble.
- Reputación: es un juicio realizado por un usuario para determinar la integridad de una fuente. Se asocia principalmente con un editor de datos, una persona, organización, grupo de personas o comunidad de práctica en lugar de ser una característica de un dataset. El editor de datos debe ser identificable para un determinado dataset.
- Licencias: se define como la concesión de un permiso para que un consumidor reutilice un dataset en condiciones definidas.

Con base en estos conceptos, las preguntas planteadas son:

- ¿Se representan la información del editor y demás colaboradores?
- ¿Se representan el ciclo de vida del dataset?
- ¿Se representa la información del título, contenido y URI del recurso?
- ¿Se representa la información de anotaciones del recurso?
- ¿Se establecen todos los datos necesarios de procedencia para el recurso?
- ¿Se usan ontologías y vocabularios de confianza para la representación del recurso?
- ¿Los recursos cuentan con una declaración de licenciamiento para su uso?

- ¿Los recursos cuentan con una descripción de los derechos que circunscribe el licenciamiento otorgado?
- ¿Los recursos cuentan con Declaración de Copyleft o ShareAlike?
- ¿Los recursos cuentan con una adecuada descripción educativa?

7.4.4. Definición de métricas.

En el nivel cuantitativo se definieron las métricas que abordan las preguntas construidas para las dos dimensiones de calidad definidas. Es importante anotar tres cosas con respecto al diseño de las métricas: (i) las métricas definidas son susceptibles de ser complementadas en investigaciones futuras; (ii) la aplicación de las métricas puede variar en la evaluación de otros modelos, dado que existen repositorios enfocados en un solo tipo de recurso; (iii) Algunas métricas fueron tomadas de autores como Behkamal *et al.* (2014), Zahedi y Fathian (2016) y Behkamal (2014), dada la afinidad de su enfoque de medición.

A continuación, la Tabla 34 detalla las métricas definidas para el primer sub tema.

Tabla 34. Métricas definidas para la Dimensión de Calidad Contextual.

Evaluación de la calidad contextual del modelo de datos desde la perspectiva de calidad de datos enlazados en el contexto de LOD		
Pregunta	**Métrica**	**Definición**
¿Se abstraen y representan las entidades y propiedades del mundo real?	M1. Ratio de propiedades faltantes de las entidades del mundo real.	El número de propiedades que contiene la abstracción de entidades con respecto a las propiedades definidas en el modelo diseñado dividido por el número total de propiedades que contiene el modelo diseñado.
¿Se representan los valores posibles para una propiedad?	M2. Ratio de asignación de valores para una propiedad.	El número de propiedades cuyo valores son definidos con respecto a los vocabularios definidos en el modelo diseñado dividido por el número total de propiedades que contienen vocabularios controlados el modelo diseñado.
	M3. Ratio de propiedades con valores fuera de rango.	El número de propiedades contienen valores atípicos con respecto a los rangos de valores aceptables definidos en el modelo de datos dividido por el número total de propiedades con vocabularios controlados.
¿Se identifican las propiedades a enriquecer con el enlace de datos?	M4. Ratio de propiedades enriquecidas con vinculación de datos.	El número de propiedades contienen valores enriquecidos con datos enlazados respecto a las propiedades definidas en el modelo de datos dividido por el número total de propiedades enriquecidas con datos enlazados definidas en el modelo.

La tabla anterior presenta la definición de cuatro métricas para la calidad contextual. Las cuatro métricas se plantearon para responder las preguntas sobre Completitud, exceptuando la Completitud de la Población dado que dentro de las instancias consultadas no se cuenta con un repositorio que gestione todos los tipos de recursos identificados y gestionados por el metamodelo. Adicionalmente, la métrica 4 aporta elementos para la dimensión de relevancia, la cual se complementa con métricas del segundo subtema. Por el contrario, la cantidad de datos no se evaluó dado que su aplicación se orienta a las instancias de datos.

A continuación, la Tabla 35 detalla las métricas definidas para el segundo sub tema.

Tabla 35. Métricas definidas para la Dimensión de Calidad de Confianza.

Evaluación de la calidad de confianza del modelo de datos desde la perspectiva de calidad de datos enlazados en el contexto de LOD		
Pregunta	**Métrica**	**Definición**
¿Se representan la información del editor y demás colaboradores?	M5. Ratio de propiedades que describen colaboradores.	El número de propiedades contienen valores que describen los colaboradores respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades definidas para colaboradores definidas en el modelo.
¿Se representan el ciclo de vida del dataset?	M6. Ratio de propiedades que describen el ciclo de vida.	El número de propiedades contienen valores que describen el ciclo de vida respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades definidas para el ciclo de vida, definidas en el modelo.
¿Se representa la información del título, contenido y URI del recurso?	M7. Ratio de propiedades de identificación del recurso (título, contenido, URI).	El número de propiedades de identificación del recurso (título, contenido, URI) respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades definidas para la identificación del recurso, definidas en el modelo.
¿Se representa la información de anotaciones del recurso?	M8. Ratio de propiedades que contienen información de anotaciones del recurso.	El número de propiedades que contienen información de anotaciones del recurso respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades que contienen información de anotaciones del recurso, definidas en el modelo.
¿Se establecen todos los datos necesarios de procedencia para el recurso?	M9. Ratio de propiedades que contienen información de procedencia del recurso.	El número propiedades que contienen información de procedencia del recurso respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades que contienen información de procedencia del recurso, definidas en el modelo.
¿Se usan ontologías y vocabularios de confianza para la representación del recurso?	M10. Ratio de ontologías de confianza usadas para describir los recursos.	El número ontologías de confianza usadas para describir los recursos respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de ontologías de confianza usadas para describir los recursos, definidas en el modelo.
¿Los recursos cuentan con una declaración de licenciamiento para su uso?	M11. Ratio de licenciamiento los recursos.	El número propiedades que definen el licenciamiento de los recursos respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades que definen el licenciamiento de los recursos, definidas en el modelo.
¿Los recursos cuentan con una descripción de los derechos que circunscribe el licenciamiento otorgado?	M12. Ratio de descripción de los derechos otorgados a los recursos.	El número propiedades que describen los derechos otorgados a los recursos respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades que describen el licenciamiento de los recursos, definidas en el modelo.

Evaluación de la calidad de confianza del modelo de datos desde la perspectiva de calidad de datos enlazados en el contexto de LOD		
Pregunta	**Métrica**	**Definición**
¿Los recursos cuentan con Declaración de Copyleft o ShareAlike?		
¿Los recursos cuentan con una adecuada descripción educativa?	M13. Ratio de descripción educativa de los recursos.	El número propiedades que contienen la descripción educativa de los recursos respecto a las propiedades definidas para tal fin en el modelo de datos dividido por el número total de propiedades que contiene la descripción educativa de los recursos definidas en el modelo.

La tabla anterior presenta la definición de nueve métricas para la calidad de Confianza. Las dos primeras métricas plantean responder elementos de verificabilidad, credibilidad y reputación. Las cuatro siguientes métricas plantean responder elementos de procedencia. Las tres últimas métricas plantean responder elementos de licenciamiento.

En resumen, y como lo plantean Behkamal *et al.* (2014), las métricas propuestas se contextualizan en el marco de medición basado en propiedades (Briand *et al.*, 1996), el cual proporciona cinco tipos de métricas que incluyen tamaño, longitud, complejidad, acoplamiento y cohesión. Las métricas se definieron en escala de Relación, y se enmarcan dentro del tipo de Complejidad, dado que cumplen las propiedades de no-negativo, no nulo, simetría y monotonicidad.

7.4.5. Definición del componente muestral.

Para ejecutar la validación, y dado que: (i) los recursos educativos se describen bajo un mismo esquema de datos, el cual es apropiado según las necesidades particulares, y (ii) el análisis realizado no se enfocó en la segmentación de los recursos; en primer paso, para el diseño experimental se planteó un muestreo aleatorio simple (Castro, 2008), el cual da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado. Considerando la proliferación de REA, es extremadamente complejo determinar un tamaño poblacional de REA, por lo que se planteó la Ecuación (1) para la obtención del tamaño muestral a trabajar:

$$n = \frac{Z^2 * p * q}{e^2} \qquad \textbf{(1)}$$

Ecuación 1. Fórmula para obtener tamaño poblacional

Para llevar a cabo esta estimación se aceptó un margen de error del 4% con un nivel de confianza del 95%. Considerando que cada REA tiene la misma probabilidad de éxito o fracaso, el resultado de la muestra estimada es presentado en la Ecuación (2):

$$n = \frac{1.96^2 * 0.5 * 0.5}{0.04^2} = 600 \qquad \textbf{(2)}$$

Ecuación 2. Cálculo de Tamaño poblacional

Como resultado de esta estimación, para el diseño experimental se planteó una muestra de 600 REA. En primer paso para la obtención de esta muestra, se identificaron instancias de repositorios de datos que proveen diferentes tipos de recursos digitales. La Tabla 36 presenta las instancias de datos consultadas.

Tabla 36. Instancias de Datos Consultadas.

Tipos de Recursos	#	Nombre de la Instancia
Varios	1	El Portal de Medios de la Siemens Stiftung
Varios	2	Eduteka Icesi
Varios	3	MIT Open Course Ware
Libros	4	Proyecto Gutenberg
Varios	5	Colombia aprende
Libros	6	iBiblio
Varios	7	OER Commons
Varios	8	Smithsonian Institute
PDF, Videos	9	Discovery Channel: Impact(Ed) International
Imágenes, Animaciones	10	Visible Earth (A catalog of Nasa images)
Varios	11	Internet Archive (Open Educational Resources)
Varios	12	Open Learning Initiative (Carnegie Mellon University)
Cursos	13	Sofia - Sharing of Free intellectual Assets
Varios	14	Procomún
Varios	15	Curriki
Varios	16	Merlot
Cursos	17	Connexions - Open Stax CNX
Cursos	18	CEDEC - Centro Nacional de Desarrollo Curricular en Sistemas No Propietarios
Cursos	19	Open Michigan
Documentos	20	Repositorio Universidad de Antioquia
Varios	21	Universidad Politécnica de Valencia
Varios	22	Universidad Distrital Francisco José de Caldas

En segundo paso, para los denominadores de las métricas, se obtuvieron indicadores propios del modelo de datos diseñado. Los indicadores obtenidos corresponden a los siguientes:

- M1: 37 propiedades con 46 elementos.
- M2: 11 propiedades usan vocabularios controlados.
- M4: 6 propiedades con valores enriquecidos.
- M5: 1 propiedad con 6 elementos para Colaboradores.
- M6: 3 propiedades con 6 elementos para el Ciclo de Vida.
- M7: 3 propiedades para identificar título, contenido y URI.
- M8: 3 propiedades para las Anotaciones.
- M9: 6 propiedades para la Procedencia.
- M10: 6 ontologías usadas para construir el modelo de datos.
- M11: 2 propiedades para el Licenciamiento: URL y fecha de Licenciamiento.
- M12: 2 propiedades para el Licenciamiento: Descripción y Costo del Licenciamiento.
- M13: 7 propiedades para la Descripción Educativa del Recurso.

Seguidamente se obtuvieron indicadores generales de cada una de las instancias de recursos consultadas. La Tabla 37 muestra el total de recursos y propiedades que expone cada instancia.

Tabla 37. Cantidad de Recursos y Propiedades de las Instancias Consultadas.

#	Nombre de la Instancia	Tipos de Recursos	Cantidad de Recursos Totales	Numero de Propiedades	Recursos de Muestra
1	El Portal de Medios de la Siemens Stiftung	Varios	4000	16	22
2	Eduteka Icesi	Varios	1286	6	22
3	MIT Open Course Ware	Varios	2400	13	22
4	Proyecto Gutenberg	Libros	1249	10	22
5	Colombia aprende	Varios	100000	19	22
6	iBiblio	Libros	2500 - P	8	22
7	OER Commons	Varios	52034	16	23
8	Smithsonian Institute	Varios	39.309.900	17	23
9	Discovery Channel: Impact(Ed) International	PDF, Videos	NI	5	23
10	Visible Earth (A catalog of Nasa images)	Imágenes, Animaciones	NI	10	23
11	Internet Archive (Open Educational Resources)	Varios	2287	19	23
12	Open Learning Initiative (Carnegie Mellon University)	Varios	NI	17	23
13	Sofia - Sharing of Free intellectual Assets	Cursos	NI	8	23
14	Procomún	Varios	94874	25	23
15	Curriki	Varios	100105	18	23
16	Merlot	Varios	NI	23	23
17	Connexions - Open Stax CNX	Cursos	NI	15	23
18	CEDEC - Centro Nacional de Desarrollo Curricular en Sistemas No Propietarios	Cursos	NI	12	23
19	Open Michigan	Cursos	NI	16	23
20	Repositorio Universidad de Antioquia	Documentos	11598	17	23
21	Universidad Politécnica de Valencia	Varios	96524	20	23
22	Universidad Distrital Francisco José de Caldas	Varios	17750	9	23

P: Proyectos, NI: No Informa

Como tercer paso, se identificaron el total de recursos y propiedades particulares que expone cada instancia de repositorios REA consultada. En la recolección de esta información se tuvo en cuenta los siguientes aspectos:

- En la recolección de información de recursos, se tuvieron en cuenta tanto repositorios con fines educativos (63.6%) como otros con fines generales (36.4%). Sin embargo, los recursos provistos por repositorios generales se pueden usar con fines didácticos y educativos, razón por la cual se hace necesario una mejor descripción de sus propiedades.

- Para la cantidad de propiedades, se contaron solo las propiedades distinguibles, y no aquellas que estaban contenidas en otras propiedades.
- Para la identificación de vocabularios controlados, los repositorios no reportaron ni permitieron verificar para qué campos usaban vocabularios controlados. La verificación se realizó comparando de forma manual el contenido de los metadatos presentados.

La Tabla 38 presenta los resultados obtenidos de la recolección de indicadores particulares.

Tabla 38. Indicadores de Instancias Consultadas.

#	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11
1	5	1	0	1	1	3	0	0	1	1	3
2	0	0	2	1	1	3	0	1	0	0	1
3	1	0	0	1	1	3	0	1	0	0	1
4	0	0	0	1	0	2	0	0	0	0	0
5	4	0	1	0	0	3	0	0	0	0	5
6	1	0	0	0	0	3	0	0	0	0	1
7	5	1	0	2	2	3	0	1	2	0	2
8	6	1	0	2	2	3	0	1	1	1	2
9	0	0	0	0	0	3	0	0	0	0	0
10	0	0	0	0	1	3	0	1	0	0	0
11	5	0	0	1	1	3	3	1	2	0	0
12	0	0	0	2	1	3	0	1	2	0	0
13	1	0	0	1	0	3	0	0	0	0	0
14	6	0	0	4	3	3	0	2	2	1	7
15	5	0	0	1	0	3	0	0	2	0	5
16	8	0	0	2	2	3	0	3	2	1	4
17	2	0	0	1	2	3	3	1	2	0	3
18	0	0	0	0	0	3	0	0	0	0	0
19	2	0	0	1	2	3	0	0	2	0	3
20	0	0	0	2	3	3	0	2	2	0	0
21	8	0	0	1	1	3	0	1	2	0	7
22	0	0	0	2	3	3	0	1	2	0	1

I1 – Vocabularios Controlados. I2 – Metadatos Enriquecidos.
I 3 – Valores Fuera de Rango. I 4 – Metadatos de Colaboradores.
I 5 – Metadatos de Ciclo de Vida. I 6 – Metadatos de Título, Descripción y URL.
I 7 – Metadatos de Anotaciones. I 8 – Metadatos de Procedencia.
I 9 – Metadatos de Declaración de Licencia. I10 – Derechos y costo de Licenciamiento.
I11 – Metadatos de Descripción Educativa.

Como resultado de esta revisión inicial se encontraron hallazgos como:

1. Limitado uso de metadatos para la descripción de recursos, dejando por fuera metadatos relevantes para la descripción de los mismos, o cargando otros con información adicional que no les corresponde. Por ejemplo, los datos de procedencia y contribución agrupaban más de un tipo de contenido, los colaboradores solo se referían a los Autores, el ciclo de vida solo relacionaba la fecha de creación o la fecha de actualización.
2. Problemas de completitud y de consistencia en la información diligenciada en estos metadatos. Por ejemplo, la descripción educativa no era uniforme en todos los recursos, no

se detallaban las competencias, y el manejo de las edades de la población objeto presentaba errores de información fuera de rango.
3. Uso de diferentes metadatos para describir los mismos tipos de REA.
4. Uso de literales para el diligenciamiento de los metadatos, dejando entrever un bajo nivel de enriquecimiento de sus metadatos.
5. La información requerida para describir los REA se presentaba como contenido HTML y no dentro de una estructura de metadatos, por ejemplo descripción educativa, colaboradores.

Estas problemáticas se originan en abstracciones limitadas y particulares de los recursos publicados, lo que disminuye la posibilidad de reutilizarlos y enriquecerlos. Por otro lado, aunque en gran proporción se identificó al lenguaje de descripción de recursos educativos – LOM (IEEE, 2002) como esquema de datos para describir REA, existe una disparidad de metadatos usados para describirlos (Sena, 2017; Marzal *et al.*, 2006; Canabal *et al.*, 2008; Koutsomitropoulos *et al.*, 2010; MEN, 2013; MEN, 2012). Por último, las organizaciones construyen los modelos de datos para que respondan a sus necesidades y proyectos particulares, lo cual limita la expresividad del modelo que describe los recursos.

7.4.6. Validación del Modelo de Datos.

Como resultados generales de la aplicación de las métricas, se identificaron los valores presentados en la Figura 64.

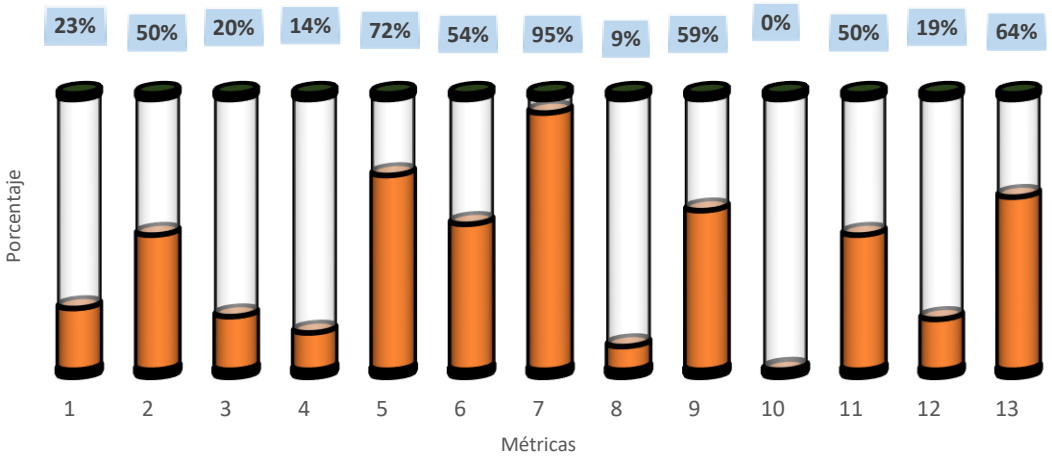


Figura 64. Resultados Generales de Aplicación de Métricas.

En la figura se observan dos comportamientos específicos: (i) Para la métrica 10, los repositorios no reportan el tipo de ontología o vocabulario usado; (ii) Para la métrica 7, el 95% de las instancias maneja los tres elementos básicos requeridos: Título, Descripción y URI del recurso, los cuales se definieron como obligatorios en el modelo diseñado.

Con relación a las demás métricas, se observa un nivel muy inferior con relación a las características del modelo diseñado. A continuación se detallan los hallazgos particulares identificados (Tablas 39 y 40).

Tabla 39. Observación de Métricas – Instancias 1 – 11.

	Instancias de Metadatos Consultadas										
MÉTRICA	1	2	3	4	5	6	7	8	9	10	11
M1	0.42	0.15	0.34	0.26	0.5	0.21	0.42	0.44	0.13	0.26	0.5
M2	0.45	0	0.09	0	0.36	0.09	0.45	0.54	0	0	0.45

	Instancias de Metadatos Consultadas										
M3	0	0.18	0	0	0.09	0	0	0	0	0	0
M4	0.16	0	0	0	0	0	0.16	0.16	0	0	0
M5	0.16	0.16	0.16	0.16	0	0	0.33	0.33	0	0	0.16
M6	0.16	0.16	0.16	0	0	0	0.33	0.33	0	0.16	0.16
M7	1	1	1	0.66	1	1	1	1	1	1	1
M8	0	0	0	0	0	0	0	0	0	0	1
M9	0	0.16	0.16	0	0	0	0.16	0.16	0	0.16	0.16
M10	-	-	-	-	-	-	-	-	-	-	-
M11	0.5	0	0	0	0	0	1	0.5	0	0	1
M12	0.5	0	0	0	0	0	0	0.5	0	0	0
M13	0.42	0.14	0.14	0	0.71	0.14	0.28	0.28	0	0	0

Tabla 40. Observación de Métricas – Instancias 12 – 22.

	Instancias de Metadatos Consultadas										
MÉTRICA	12	13	14	15	16	17	18	19	20	21	22
M1	0.44	0.21	0.65	0.47	0.6	0.39	0.31	0.42	0.44	0.52	0.23
M2	0	0.09	0.54	0.45	0.72	0.18	0	0.18	0	0.72	0
M3	0	0	0	0	0	0	0	0	0	0	0
M4	0	0	0	0	0	0	0	0	0	0	0
M5	0.33	0.16	0.66	0.16	0.33	0.16	0	0.16	0.33	0.16	0.33
M6	0.16	0	0.5	0	0.33	0.33	0	0.33	0.5	0.16	0.5
M7	1	1	1	1	1	1	1	1	1	1	1
M8	0	0	0	0	0	1	0	0	0	0	0
M9	0.16	0	0.33	0	0.5	0.16	0	0	0.33	0.16	0.16
M10	-	-	-	-	-	-	-	-	-	-	-
M11	1	0	1	1	1	1	0	1	1	1	1
M12	0	0	0.5	0	0.5	0	0	0	0	0	0
M13	0	0	1	0.71	0.57	0.42	0	0.42	0	1	0.14

Según los resultados mostrados en las tablas anteriores, y considerando que los recursos pertenecientes a la misma instancia conservan un esquema igual de descripción, se realizaron las siguientes apreciaciones:

- En cuanto a la métrica 1, sólo cinco instancias de recursos alcanzaron un 50% o más de la cantidad de propiedades descriptoras definidas por el modelo diseñado (Figura 65), evidenciando una mayor completitud en propiedades por parte de este modelo, además de proporcionar más elementos de calidad de datos que aportan en la mejora de la calidad de las instancias generadas.

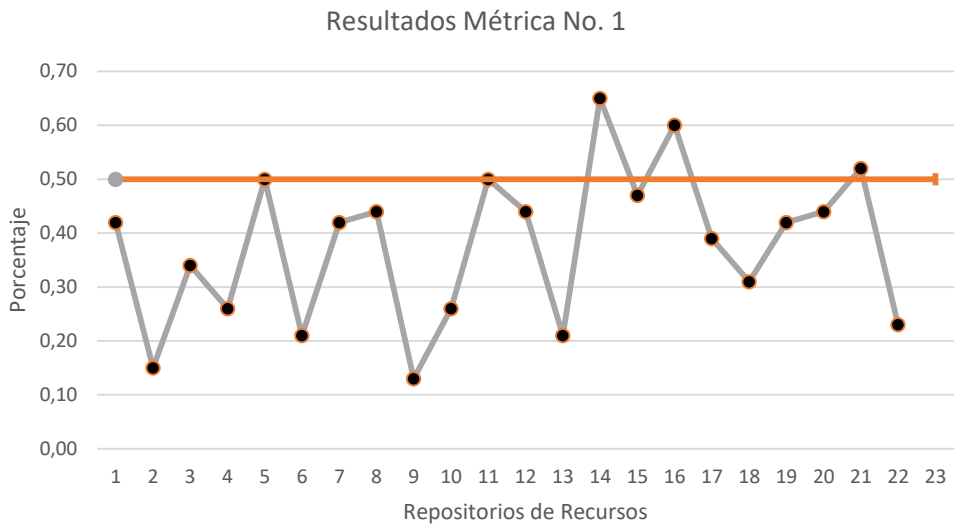


Figura 65. Resultados Métrica 1

- Para la métrica 2, en comparación a los vocabularios definidos en el modelo diseñado, el 59% de las instancias usaban entre 0% y 10% de vocabularios controlados para definir el contenido de las propiedades de sus recursos (Figura 66), lo que incrementa el riesgo de tener una alta disparidad de valores o en su defecto valores inapropiados para las propiedades. Situación que se evidenció en la métrica 3, donde el 10% de las instancias presentaban recursos con valores fuera de rango. Por su parte, solo un 9% de las instancias presentaban un alto uso de vocabularios controlados en la descripción de sus recursos, con relación a los dispuestos en el modelo diseñado. Esta situación puede deberse al limitado uso de propiedades que restringe la expresividad de la descripción de los recursos.

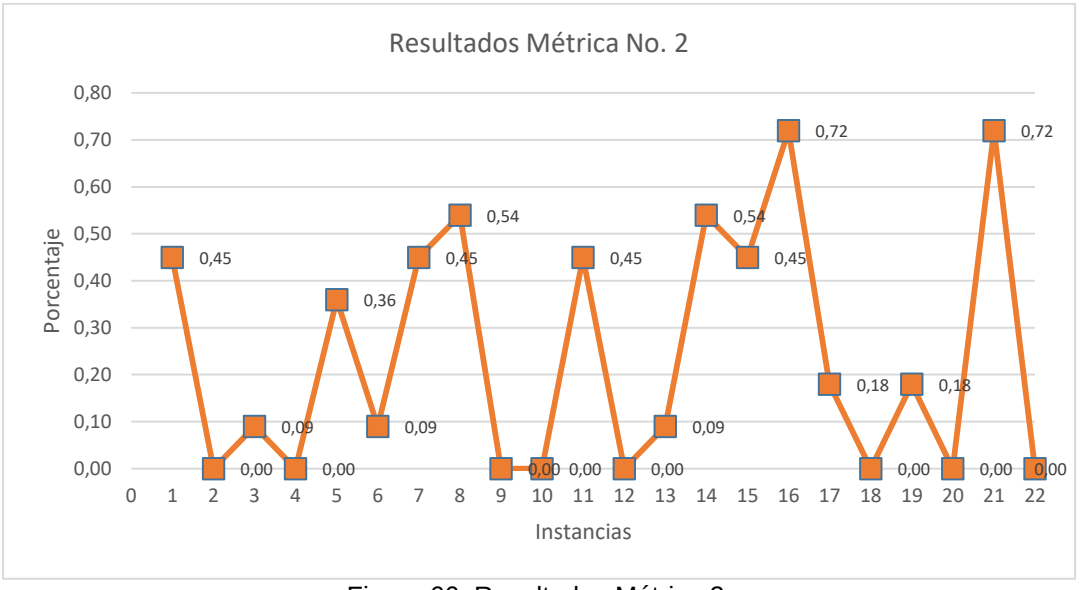


Figura 66. Resultados Métrica 2.

- Con relación a la métrica 4, solo el 13.5% de las instancias manejaban valores enriquecidos para la descripción de sus recursos (este porcentaje solo representa una propiedad del recursos para cada una de las tres instancias identificadas). Lo anterior dejó entrever una muy baja aplicación del enriquecimiento de los metadatos versus el modelo diseñado en esta investigación.

- Para la métrica 5, la propiedad “colaboradores” de sus recursos correspondía solamente al autor en el 72% de las instancias. En contraposición, el modelo diseñado plantea 6 propiedades para el diligenciamiento de la información de los colaboradores.
- Para la métrica 6, los recursos del 54% de las instancias usaban la fecha de creación o de actualización como dato del ciclo de vida. Por su parte, el modelo diseñado maneja datos de ciclo de vida como creación, carga, actualización del recurso.
- Para la métrica 8, los recursos de solo 2 repositorios manejaban datos de anotaciones. En uno de estos repositorios, las anotaciones se usan para completar la descripción educativa del recurso. Elementos que están claramente definidos en el modelo diseñado.
- Para la métrica 9, los recursos del 40.9% de las instancias no manejaban información de procedencia. Los recursos de otras instancias manejaban una cantidad reducida de propiedades relacionadas con la procedencia.
- Para la métrica 11, referida a la identificación del licenciamiento, los recursos del 36% de las instancias no manejaban un licenciamiento explícito. Mientras que el 50% de las instancias presentaban una URI de identificación del tipo de licencia utilizada para sus recursos.
- Para la métrica 12, los recursos del 81% de las instancias no presentaban una descripción de los derechos adquiridos a través de la licencia manejada o la falta de especificación de la licencia utilizada.
- Para la métrica 13, los recursos del 36% de las instancias no presentaban una descripción educativa de los mismos.

En cuanto a los recursos de repositorios con fines educativos trabajados en la muestra, se identificó que aunque Procomún (Procomún, 2020) y Merlot poseían mejores descripciones de sus recursos (Figura 67), esta descripción es permeada e interoperable por el modelo de datos diseñado. Adicionalmente, el modelo de datos diseñado maneja una mayor cantidad de propiedades descriptoras que profundizan y complementan la descripción de los REA.

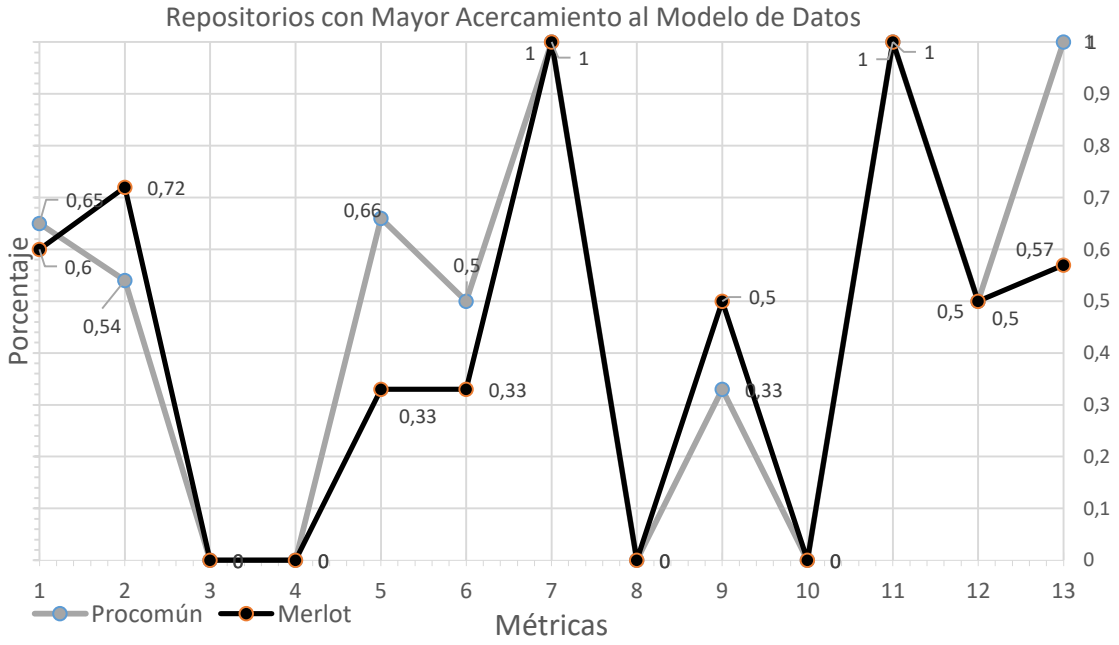


Figura 67. Comparación Propiedades Procomún y Merlot.

En resumen, con relación al objetivo motivador del proceso de validación empírica, desde el punto de vista de las dimensiones de calidad contextual y de confianza, a través de los resultados de las métricas se observó que el modelo de datos diseñado proporciona elementos de calidad que, además de mejorar la descripción de los REA, permite la interoperabilidad con descripciones

provistas por otras organizaciones. Adicionalmente, el modelo promueve la generación de confianza en los usuarios, al proveer metadatos que permiten evaluar y comparar el recurso de una forma más integral, desde diferentes aristas tales como su contexto educativo, su procedencia, sus características particulares, entre otros aspectos.

Por ende, el modelo de datos para describir y enriquecer REA propuesto en esta investigación, proporciona elementos de calidad inherentes, que aportan tanto a mejores instancias de metadatos como resultados de consultas más oportunos para los usuarios.

7.5. Prueba de Aceptación de Usuario.

Las pruebas de aceptación se enfocan en verificar si el sistema está listo y en condiciones para su uso (Galin, 2004; Tian, 2005). Considerando que el metamodelo construido, como componente abstracto, no involucra el desarrollo de software, la prueba de aceptación realizada se orientó a explorar la opinión de los usuarios sobre los beneficios provistos por el metamodelo (pruebas de aceptación categoría beta), con relación al modelo de datos obtenido como instancia del metamodelo (Serrano, 2013; Sánchez, 2016).

Para llevar a cabo esta prueba beta, se contó con la participación dos grupos de personas:

- Expertos cuyo quehacer involucraba bien sea el diseño, el uso o el tratamiento de REA. Dentro de estos participantes se invitaron expertos de diferentes universidades (áreas de archivística y gestión documental, diseño recursos educativos, educación e ingeniería), junto con expertos que administran plataformas como el RIUD (caso de estudio). Para esta prueba se contó con la participación de 15 expertos.
- Estudiantes de la Universidad Distrital, quienes han tenido contacto con: a) Recursos educativos, dada el uso del enfoque virtual promovido por el Gobierno Nacional, como estrategia para mitigar los contagios por Covid-19; y b) con el RIUD, como plataforma de gestión de recursos institucional de la Universidad Distrital Francisco José de Caldas, además de configurarse como caso de estudio en esta investigación. Para esta prueba se contó con la participación de 25 estudiantes.

A cada una de estas poblaciones, seleccionados de forma aleatoria, se les aplicó una encuesta con temáticas referentes al modelo de datos diseñado en esta investigación.

7.5.1. Resultados del grupo poblacional de expertos.

En primera instancia, los resultados muestran que la población encuestada se alineó con la estructura de elementos de metadatos provisto por el perfil de aplicación PA-REA, proporcionando niveles de obligatoriedad símiles (Tabla 41) a los definidos en el Crosswalk realizado.

Tabla 41. Nivel de Obligatoriedad de elementos de metadatos según Expertos.

Elemento	Obligatorio	Recomendado
Identificador	X	
Título	X	
Subtítulo		X
Palabras Claves	X	
Descripción (Abstract)	X	
Cita	X	

Elemento	Obligatorio	Recomendado
Idioma General	X	
Fecha de Versión Inicial		X
Versión	X	
Contribución (Rol, entidad)		X
Fecha del Ciclo de Vida		X
Formato Técnico	X	
Tamaño del Formato		X
Ubicación Técnica	X	
Tipo Técnico	X	
Versión Mínima	X	
Tipo de interactividad		X
Nivel de interactividad		X
Tipo de Recurso	X	
Rol Usuario Final		X
Nivel Educativo	X	
Contexto Educativo	X	
Rango de Edad		X
Descripción Educativa		X
Licenciamiento y Otras Restricciones	X	
Descripción de los Derechos	X	
Costo de los Derechos	X	
Fecha del Licenciamiento	X	
Persona (anotación)		X
Fecha (anotación)		X
Descripción (anotación)		X
Institución	X	
Organización	X	
Grupo		X
División		X
Taxón	X	
Curso	X	

Por otro lado, se obtuvieron resultados tales como:

- El 86.7% de los encuestados consideró que los elementos de metadatos que describen un REA deberían enriquecerse.
- El 93,3% de los encuestados consideró que es necesario el uso de vocabularios controlados para estandarizar la alimentación de los elementos de metadatos.
- El 93,3% de los encuestados consideró que los REA deberían ser publicados bajo licencias abiertas.
- El 73.3% de los encuestados consideró que un REA se debía describir con suficiente información que permita identificarlo, determinar si es lo que se está buscando y poderlo recuperar. Por el contrario, un 26.7% de los encuestados consideró que una descripción básica del REA (título, descripción, palabras claves) es suficiente.
- Para los expertos prima la descripción educativa, clasificación y descripción técnica del recurso.

Estos resultados permiten evidenciar la necesidad de describir de mejor forma los REA, además de enriquecer elementos de metadatos estratégicos a través del enlace de datos. En respuesta a esta necesidad, el 86.7% de los encuestados consideró que el modelo propuesto en esta investigación maneja y proporciona los elementos suficientes para caracterizar las propiedades de un REA, evidenciando una amplia aceptación del modelo de datos para describir REA, producto del metamodelo trabajado.

7.5.2. Resultados del grupo poblacional de estudiantes.

La población encuestada evidencia un alto interés por una descripción del REA, una caracterización técnica y un licenciamiento que les permite poder identificar el REA, saber si lo pueden usar según el tipo de licencia y el tipo de requerimientos técnicos necesarios para su uso (Tabla 42), aspectos que son suplidos por el perfil de aplicación PA-REA, provisto por esta investigación.

Tabla 42. Grado de interés de elementos de metadatos según estudiantes.

Elemento	Alto interés (%)	Bajo interés (%)
Identificador		
Título	91,30%	
Subtítulo		13%
Palabras Claves	78,30%	
Descripción (Abstract)	87%	
Cita		39,10%
Idioma General		47,80%
Fecha de Versión Inicial	56,50%	
Versión		30,40%
Contribución (Rol, entidad)	60,90%	
Fecha del Ciclo de Vida	58%	
Formato Técnico	95,50%	
Tamaño del Formato	59,10%	

Elemento	Alto interés (%)	Bajo interés (%)
Ubicación Técnica	86,40%	
Tipo Técnico	86,40%	
Versión Mínima	86,40%	
Tipo de interactividad	52,20%	
Nivel de interactividad		43,50%
Tipo de Recurso	91,30%	
Rol Usuario Final	52,20%	
Nivel Educativo	60,90%	
Contexto Educativo	56,50%	
Rango de Edad		26,10%
Descripción Educativa		39,10%
Licenciamiento y Otras Restricciones	87%	
Descripción de los Derechos		47,80%
Costo de los Derechos	52,20%	
Fecha del Licenciamiento		30,40%
Persona (anotación)		21,70%
Fecha (anotación)	52,50%	
Descripción (anotación)	87%	
Institución	65,20%	
Organización	65,20%	
Grupo		30,40%
División		30,40%
Taxón	90,90%	
Curso		45,50%

Otros resultados obtenidos corresponden a:

- El 91.7% de los encuestados consideró que los elementos de metadatos que describen un REA deberían enriquecerse.
- El 100% de los encuestados consideró que los REA deberían ser publicados bajo licencias abiertas.
- Las anotaciones realizadas a los recursos no presentan un marcado interés para los estudiantes.

Resultados que, aunque se diferencian un poco de los requeridos por los expertos, dadas sus necesidades y enfoques; son cubiertos por el perfil de aplicación diseñado en esta investigación.

En términos generales, las pruebas de aceptación de usuario convergen en que el modelo para describir REA cubre con las necesidades particulares de cada rol, constituyéndolo en un modelo más completo, enriquecido y que responde a los requerimientos de disponibilidad y reutilización.

7.6. Conclusiones del Capítulo.

En cuanto al metamodelo, se observó que su estructura permite la gestión de propiedades de los recursos educativos, de una forma ágil y sencilla, disminuyendo el nivel de rigidez, inmovilidad y viscosidad al momento de manipular las propiedades y los mapeos de los recursos. Con relación a la disponibilidad y reutilización de los recursos, las pruebas realizadas permitieron evidenciar que la aplicación de las diferentes reglas y restricciones del metamodelo generan instancias de metadatos que se alinean a los requerimientos de los niveles del modelo LOD. Por ende, además de brindar una mejor descripción de los recursos, proporcionan elementos como el enriquecimiento y los mapeos RDF, que al articularse con los grafos de conocimiento, colaboran con el incremento de la visibilidad y reutilización de los recursos.

En general, como se observó en la validación realizada, el modelo diseñado ofrece una serie de bondades que además de proporcionar elementos de calidad, como modelo de datos, a las instancias que de él se deriven, permite su interoperabilidad con otros modelos, además de alinearse con proyectos de índole nacional e internacional, proporcionando elementos de enriquecimiento que insertan los recursos, publicados como RDF Graph, a la nube de datos enlazados.

Además del enriquecimiento e interoperabilidad proporcionada por el modelo de datos propuesto, las propiedades adicionales permiten generar una descripción más sólida y completa del REA. Esta descripción proporciona elementos como por ejemplo autor, editor, anotaciones, organizaciones, departamentos, entre otros, cuyos valores pueden ser contrastados con rankings, indicadores, etc., con el fin de proporcionar evaluaciones de calidad para los usuarios consumidores. Estos elementos de calidad permiten fortalecer la confianza de los usuarios sobre los datos publicados. Adicionalmente, el modelo de datos proporciona elementos necesarios para el proceso de compartir los datos, definidos dentro de los principios FAIR - Findable, Accessible, Interoperable, Reusable (Sixto *et al.*, 2019), elementos que son descritos citando a Wilkinson *et al*. (2016):

a. Encontrables:
   - Los datos y metadatos están debidamente identificados.
   - Los datos están descritos con suficientes metadatos.
   - Los metadatos claramente identifican claramente los datos que describen.
   - Los datos y metadatos están registrados o indexados en un recurso que sea fácil de localizar.

b. Accesibles:
   - Los datos y metadatos pueden ser recuperables con su identificador, utilizando un protocolo estandarizado de comunicación.
   - Los metadatos son accesibles.

c. Interoperables:
   - Los datos y metadatos utilizan un lenguaje formal, accesible y ampliamente aplicable para la representación del conocimiento.
   - Los datos y metadatos utilizan vocabularios para su descripción.
   - Los datos y metadatos incluyen referencias a otros datos o metadatos.

d. Reutilizables:

- Los datos y metadatos son descritos de manera rica, con una pluralidad de atributos precisos y relevantes.
- Se recomienda la publicación de los datos y metadatos con una licencia que permita su uso y reutilización de forma clara y accesible.
- Los datos y metadatos asocian información sobre su procedencia detallada.

Por último, la estructura de metamodelo planteada, aporta una mayor flexibilidad para adicionar nuevas reglas o parámetros al constructo teórico que define la construcción del modelo de datos diseñado.

CAPÍTULO VIII APORTES Y DISCUSIONES DEL METAMODELO DISEÑADO.

Después de validar el metamodelo propuesto, a continuación se describen los diferentes aportes proporcionados tanto por el metamodelo como por el modelo de datos para describir y enriquecer REA, derivado del metamodelo diseñado en esta investigación.

8.1. Aportes del Metamodelo Diseñado

Para detallar los aportes esbozados por el modelo propuesto para describir y enriquecer REA, conforme a las directrices y restricciones definidas por el metamodelo diseñado; estos aportes se abordan desde las ópticas definidas en la Figura 68.

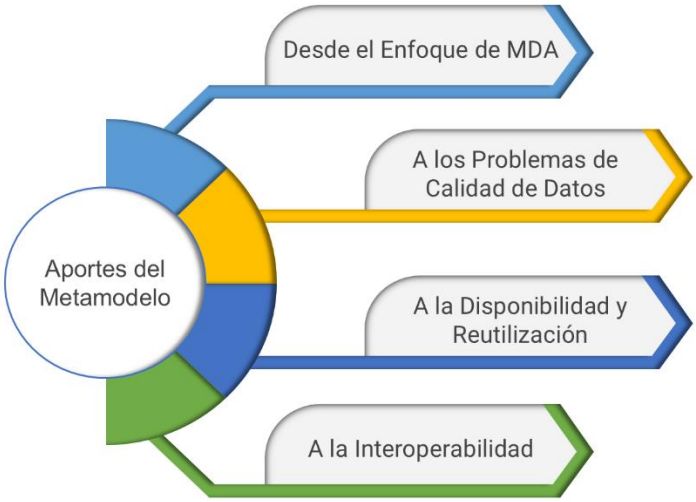


Figura 68. Resumen de Aportes del Metamodelo.

8.1.1. Aportes desde la promesa de MDA.

Como lo describen Truyen (2006) y Montenegro (2017), la promesa de MDA es facilitar la creación de modelos con el objetivo de flexibilidad a largo plazo. Aunque el metamodelo diseñado en esta investigación, dado el nivel de abstracción del tema trabajado, no conlleva desarrollo en el punto de vista de plataforma específica, los aportes desde el enfoque proporcionado por MDA a través del metamodelo diseñado se expresan en términos de:

- Obsolescencia de la tecnología: dado que el metamodelo diseñado genera un modelo de datos para describir y enriquecer REA, y que dicho modelo puede ser mapeado bien sea para aplicación o para datasets; la nueva infraestructura de implementación se puede integrar más fácilmente con el modelo datos haciendo las configuraciones y parametrizaciones requeridas por la nueva versión del ambiente tecnológico.
- Portabilidad: el modelo de datos obtenido de la implementación del metamodelo, se puede migrar a nuevos entornos y plataformas. En primera instancia, la ontología que describe los REA es independiente de la plataforma, permitiendo su adaptación a otros entornos y plataformas. En segunda instancia, los mapeos de los metadatos que describen los REA, tanto para aplicación como para dataset, se construyeron haciendo uso de vocabularios y modelos de datos ampliamente reconocidos y usados en el mercado, como lo son Dublin Core y el modelo de datos RDF respectivamente.
- Calidad: la separación formal de vistas y componentes, proporcionada por este enfoque, permite tener una visión más clara de los mismos, además se logra identificar

adecuadamente elementos, restricciones y reglas. Esto permite la creación de artefactos con mayor consistencia y confiabilidad.

- Mantenimiento: la disponibilidad de los modelos proporciona un acceso directo a la especificación de los componentes, simplificando las tareas de mantenimiento, además de flexibilizar los procesos de modificación o actualización de los mismos.
- Pruebas: el modelo de datos para describir REA se puede validar directamente contra los requisitos, así como compararlos y validarlos contra diversas propuestas de descripción de metadatos para REA.

Lo anterior evidencia que el diseño de un modelo de datos para describir REA basado en un enfoque de metamodelo, además de aportar la flexibilización del proceso de diseño y mantenimiento, aporta más claridad en cuanto a los componentes utilizados y su interacción. Además provee una forma ágil y sencilla de identificar y manipular elementos y reglas ante procesos de actualización o modificación de la estructura de propiedades que describen los REA.

8.1.2. Aporte de PA-REA a los problemas de calidad de datos.

Como se observa en la Tabla 43, cada uno de los elementos del modelo de datos mejorado y enriquecido, además de estar ajustados a los lineamientos del metamodelo, aportan a una o varias de las dimensiones de calidad trabajadas.

Tabla 43. Aporte del Metamodelo a las Dimensiones de Calidad.

Elemento	Mapeo PA-REA	Dimensión de Calidad
Identificador	Bibo.doi Bibo.handle Bibo.issn	Completitud de Esquema
Título	dcterms:title	Confianza → Verificabilidad Confianza → Credibilidad
Subtítulo	dcterms:alternative	Relevancia
Palabras Claves	dc:subject	Relevancia
Descripción (Abstract)	dcterms:abstract	
Cita	dcterms:bibliographicCitation	Completitud de Esquema
Idioma General	dc:language	
Fecha de Versión Inicial	dcterms:created	Confianza→ Credibilidad
Versión	dcterms:hasversion	
Contribución Esta subdividida en: Rol, Entidad	dcterms:creator dc:creator (URI del autor) dc: advisor dc: sponsorship dc:publisher	Confianza → Verificabilidad Confianza → Credibilidad
Fecha del Ciclo de Vida	dcterms:available dcterms:modified dcterms:issued	Confianza → Verificabilidad Confianza → Credibilidad
Formato Técnico	dc:format	
Tamaño del Formato	dcterms:extent	Completitud de población
Ubicación Técnica	Bibo:uri	Procedencia
Tipo Técnico	dcterms:requires	
Versión Mínima	parea:minimunversion	
Tipo de interactividad	lrmi:interactivityType	Completitud de Esquema
Nivel de interactividad	parea:interactivitylevel	Completitud de Esquema
Tipo de Recurso	dc:type	
Rol Usuario Final	dcterms:audience	
Nivel Educativo	dcterms:educationLevel	Completitud de Esquema
Contexto Educativo	Aiiso:KnowledgeGrouping	Completitud de Esquema
Rango de Edad	lrmi:typicalAgeRange	

Elemento	Mapeo PA-REA	Dimensión de Calidad
Descripción Educativa	dcterms:description	Completitud de Esquema
Licenciamiento y Otras Restricciones	dc:rights	Licenciamiento
Descripción de los Derechos	dcterms:rights	Licenciamiento
Costo de los Derechos	parea: rightscost	Licenciamiento
Fecha del Licenciamiento	dcterms:dateCopyrighted	Licenciamiento
Persona (anotación)	parea:annotationentity	Confianza → Credibilidad Procedencia
Fecha (anotación)	dcterms:date	Confianza → Credibilidad Procedencia
Descripción (anotación)	parea:annotationdesc	Confianza → Credibilidad Procedencia
Institución	Aiiso:Institution	Procedencia
Organización	Foaf:organization	Procedencia
Grupo	Foaf:group	Procedencia
División	Aiiso:division	Procedencia
Taxón	lrmi.EducationalSubject	Completitud de Esquema
Curso	Aiiso:course	Completitud de Esquema

Estos aportes se traducen en instancias de metadatos con elementos que contribuyen a mejorar los problemas de calidad de datos sufridos por los datasets. Algunos de estos aportes de mejora se presentan en:

- Mejorar la generación de confianza, dado que ofrecen un conjunto mayor y más detallado de propiedades que permiten evaluar la confianza de los datasets contra diferentes estrategias como los rankings de contribuyentes, instituciones, grupos que generan los recursos, autores, entre otros aspectos.
- Mejorar la información de procedencia del dataset, al incluir más elementos tanto de tipo como de denominación, que detallan formalmente la procedencia de los recursos.
- Mejorar la información de licenciamiento, al poder contar con información adicional a una URI de licencia, que contextualiza los derechos y restricciones con que cuenta un recurso.
- Mejorar la información de la descripción educativa, al proveer información que amplía el detalle de las características educativas de los recursos, y no se limita a proveer una descripción genérica que intrínsecamente lleva incluidas las competencias del recurso, por ejemplo.

8.1.3. Aporte de PA-REA a la Disponibilidad y Reutilización de los REA.

Considerando lo expuesto por Abella *et al.* (2018), quienes describen que el modelo de enlace de datos se enfatiza en dos aspectos relevantes: la disponibilidad (niveles 1, 2 y 3) y la capacidad de reutilización (niveles 4 y 5); el perfil de aplicación PA-REA aporta:

a. Bajo el aspecto de disponibilidad, se provee un modelo de datos para describir y enriquecer REA:

   - Susceptible de ser publicado en la Web.
   - Abierto, permitiendo de esta forma su conocimiento, uso, gestión y armonización.
   - Interoperable con otros modelos de datos que describen recursos.
   - Disponible en formatos que se alinean con los requerimientos del modelo LOD.
   - Que contiene una mejor descripción de los recursos educativos abiertos.
   - Que contiene elementos enriquecidos a través del enlace de datos.
   - Que aporta elementos de calidad de datos enlazados.
   - Que aporta elementos que mejoran la confianza generada en los usuarios consumidores.

Aportes que se alinean con propuestas como la Carta Internacional de Datos Abiertos, promovida en 2015 (Cepal, 2020; ODC, 2020).

b. Bajo la capacidad de reutilización, se provee un modelo de datos donde:

- Cada instancia de metadatos es identificable con una URI, permitiendo de esta manera poder ser enlazados por otros datasets.
- Cada instancia de metadatos tiene elementos que enlazan otros datasets, con el fin de enriquecer la descripción de la instancia.
- Las instancias de metadatos, mapeadas a RDF, pueden adquirir una mejor visibilidad, al enlazar los objetos de sus tripletas a diferentes grafos de conocimiento.
- Las instancias de metadatos mapeadas en RDF tiene un mayor nivel de accesibilidad, al poder ser consultadas a través de puntos de acceso SPARQL.

Aportes que evidencian cómo el metamodelo propuesto permite la generación de instancias de metadatos que alcanzan el quinto nivel del modelo de datos enlazados, proporcionando tanto un mayor contexto para los recursos, como elementos de enriquecimiento a otros recursos publicados en la Web.

8.1.4. Interoperabilidad del modelo de datos con otros proyectos.

En primera instancia, se planteó corroborar la interoperabilidad del modelo de datos diseñado contra las instancias de recursos consultadas para la validación del modelo. Para este propósito, se revisó la estructura de metadatos de cada una de las 22 instancias, con el fin de determinar el conjunto de metadatos que gestionaban. La Tabla 44 muestra los resultados obtenidos.

Tabla 44. Interoperabilidad de Metamodelo diseñado.

	Instancias de Metadatos Consultadas																					
Metadatos (Modelo diseñado)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Identificador			A	A	A	A		A												A	A	
Título	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Subtítulo					A															A		
Palabras Claves	A	A	A	A	A	A	A	A		A	A	A	A	A	A	A	A	A	A	A	A	A
Descripción (Abstract)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Cita						A																
Idioma General	A			A			A	A			A			A	A	A	A	A	A	A	A	
Fecha de Versión Inicial				A	A		A	A		A	A	A		A	A	A	A		A	A	A	
Versión																	A					
Contribución (Rol, Entidad)	A	A	A	A	A	A	A	A		A	A	A	A	A	A	A	A	A	A	A	A	A
Fecha del Ciclo de Vida	A						A			A	A	A				A	A		A	A	A	A
Formato Técnico	A		A	A	A		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Tamaño del Formato	A			A	A			A	A	A	A										A	
Ubicación Técnica	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		A	A	A	A	
Tipo Técnico	A		A		A		A			A	A	A		A	A	A		A			A	
Versión Mínima					A							A		A		A						

	Instancias de Metadatos Consultadas																					
Metadatos (Modelo diseñado)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Tipo de interactividad														A								
Nivel de interactividad														A							A	
Tipo de Recurso	A	A					A	A				A		A	A	A		A	A	A	A	
Rol Usuario Final												A		A		A					A	
Nivel Educativo	A		A		A		A	A				A	A	A		A		A	A		A	
Contexto Educativo	A						A				A			A	A	A		A	A		A	
Rango de Edad														A								
Descripción Educativa					A							A	A	A	A	A		A	A			
Idioma Educativo					A									A								
Licenciamiento y Restricciones	A		A		A		A	A			A	A		A	A	A	A		A	A	A	A
Descrip. de los Derechos	A		A		A		A	A			A	A		A	A	A	A		A	A		A
Costo de los Derechos																A						
Fecha del Licenciamiento					A																	
Persona (anotación)											A			A	A	A	A					
Fecha (anotación)											A			A	A	A	A					
Descripción (anotación)											A			A	A	A	A					
Institución								A														
Organización								A			A									A	A	
Grupo			A				A													A		
División																						
Taxón	A		A		A	A		A			A	A		A	A	A	A		A	A	A	A
Curso																						

A: Aplica

Como resultado de la comparación del modelo de datos generado por el metamodelo, versus los metadatos requeridos por cada una de las plataformas analizadas, se encuentra que el modelo de datos diseñado en esta investigación:

- Cubre los metadatos requeridos por cada instancia, al igual que los analizados en el Anexo 4, lo cual incrementa la interoperabilidad con los otros modelos de datos.
- Aporta elementos de calidad de datos enlazados, que contribuyen a subsanar problemas identificados en las instancias de datos.
- Maneja metadatos similares haciendo uso de vocabularios controlados, lo cual disminuye problemas como valores fuera de rango, o multiplicidad de denominaciones o etiquetados para una misma propiedad.
- Complementa la información requerida para describir los recursos educativos, lo cual además de verificar la aplicación de las dimensiones de calidad para datos enlazados, propuestas en el metamodelo; mejora y refina los criterios de búsqueda, además de cumplir con requerimientos de calidad de datos enlazados.

- Por último, en las instancias de datos revisadas, al igual que en modelo diseñado, se maneja un mismo set de metadatos para describir los diferentes tipos de REA.

Factores que evidencian que el modelo desarrollado en esta investigación, además de cubrir las necesidades expuestas de mejor descripción y enriquecimiento, provee un dialogo más cercano con otros perfiles de aplicación disponibles, lo que contribuye a un mejor nivel de interoperabilidad entre repositorios u organizaciones que aborden la tarea de abrir y describir sus recursos.

8.2. Otras aportes del metamodelo diseñado.

Además de proporcionar mecanismos flexibles para gestionar las propiedades de un REA y ser extensible a modelar la tipología de REA definida para esta investigación, la implementación de un metamodelo permitió que el modelo de datos generado fuera interoperable con otros modelos de datos, además de poder abstraerse a otros tipos de recursos (Figura 69).

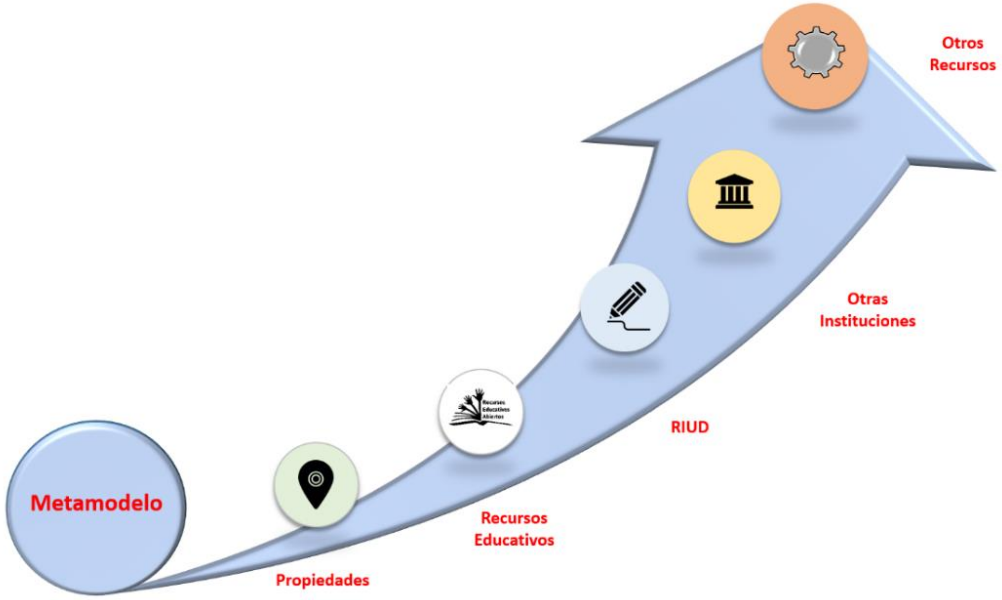


Figura 69. Otros Aportes del Metamodelo.

A continuación se describen brevemente los diferentes aportes proporcionados por la implementación del metamodelo.

8.2.1. Aplicación del modelo de datos en la Universidad Distrital.

La Universidad Distrital cuenta con el Repositorio Institucional de la Universidad Francisco José de Caldas – RIUD, herramienta donde se depositan los objetos digitales que corresponden a la producción intelectual de la Universidad. Con relación a la aplicación del modelo de datos para describir REA diseñado en esta investigación:

- El modelo datos producto del metamodelo propuesto, permite cubrir el modelo de datos establecido por el RIUD, por lo cual su aplicación mejoraría la descripción de los recursos gestionados por el RIUD.
- El modelo de datos del RIUD no aporte mayor información sobre aspectos como la procedencia de los recursos, la descripción educativa, etc., elementos claves al momento de abordar las problemáticas de calidad de datos. Por consiguiente, la implementación del modelo de datos propuesto, brindaría a la Universidad elementos de información, como por ejemplo datos de procedencia, tan requeridos para procesos como acreditación de programas y acreditación institucional.

- Ambas implementaciones hacen uso de Dspace como plataforma de gestión de datos, razón por la cual la modificación de la estructura de metadatos tendría un impacto más reducido, dado que se enfocaría en la implementación completa del mapeo para aplicación.
- El RIUD no cuenta aún con una estrategia para el enriquecimiento de los metadatos que describen sus recursos, por lo cual el modelo de datos diseñado, además de mejorar la descripción de los recursos almacenados, permite enlazar información que complemente dicha descripción.
- Las instancias de metadatos almacenadas en el RIUD no poseen un mapeo RDF a dataset. La implementación del mapeo de dataset del modelo diseñado contribuye a aumentar la visibilidad y reutilización de los recursos al ser enlazados a otros grafos de conocimiento.
- Con la incursión de la Universidad Distrital en la formación virtual, la generación de mediaciones se ha venido incrementando. Sin embargo, la descripción de estos recursos aún es muy limitada, y el modelo de datos del RIUD se queda corto para la descripción de estos recursos. Razón por la cual, el modelo diseñado contribuye a ampliar el alcance del RIUD en cuanto a catalogación y gestión de la información de recursos concierne.

Por ende, la aplicación del modelo propuesto al contexto de la Universidad Distrital aportaría no solo a la mejora en la descripción de sus recursos digitales, sino al enriquecimiento de los mismos. Lo cual se ve reflejado en el aumento de la disponibilidad y reutilización de los recursos producidos por la Universidad.

8.2.2. Aplicación en otras instituciones.

Considerando tres aspectos esenciales: i) El modelo de datos propuesto cubre diferentes perfiles de aplicación para describir recursos, tales como CEM (MEN, 2013), SENA (Sena, 2017), LOM-CO (MEN, 2012), RIUD (RIUD, 2016) ii) Dublin Core es uno de los estándares más usados para describir recursos, y iii) Dspace es una de las plataformas más usadas para gestionar recursos; por ende, la aplicación en otras instituciones del modelo de datos generado por el metamodelo diseñado es altamente posible.

Lo anterior se soporta en que diferentes instituciones han empezado a migrar sus estructuras de metadatos a proyectos como LA Referencia (LAR, 2019), cuyo modelo de datos es permeado por el modelo propuesto. Por otro lado, ante requerimientos como los generados por la pandemia del Coronavirus a lo largo de 2020, la apertura y la mejora en la descripción de los recursos educativos se han convertido en necesidades claves para afrontar las nuevas realidades generadas por esta pandemia (Bozkurt *et al.*, 2020).

Por ende, bien sea el uso total de este modelo de datos, la interoperabilidad de sus modelos con esta propuesta, o la armonización parcial de sus modelos de datos atendiendo a la estructura de metadatos provista en este modelo, se visualiza como una estrategia clave para mejorar la disponibilidad y reutilización de los recursos educativos abiertos.

8.2.3. Ampliación del metamodelo a otros tipos de recursos.

Ante requerimientos de ampliar la descripción y enriquecimiento a otros tipos de recursos, MDA y el trabajo con modelos permiten que el metamodelo diseñado en esta investigación sea modificado para cubrir este tipo de requerimientos.

Por ejemplo, la Universidad Distrital en los términos de la Ley 1712 de 2014 y el Decreto 2573 de 2014, ha venido trabajando en el tema de la apertura de sus datos, llegando a publicar sus datasets en plataformas como Datos Abiertos – Colombia. Sin embargo, sus datos abiertos, que describen diferentes tipos de recursos, se ajustan a diferentes parámetros y no cuentan con enriquecimiento provisto por la Web Semántica (Pérez *et al.*, 2015).

Ante esta problemática, al metamodelo propuesto en esta investigación se le puede adicionar una capa M3 (Meta-Metamodelo) que permita abstraer las reglas, entidades y propiedades comunes a los recursos abiertos en general que maneja la Universidad (Figura 70). De esta forma, y como lo definió González -citado por Rodríguez (2017)-, se contaría con un meta-metamodelo que define las características de los metamodelos, permitiendo que exista un metamodelo para cada dominio del conocimiento a tratar, mientras se tiene un meta-metamodelo común a todos estos para así poder realizar operaciones sobre ellos, como pueden ser transformaciones automáticas, validaciones y búsquedas.

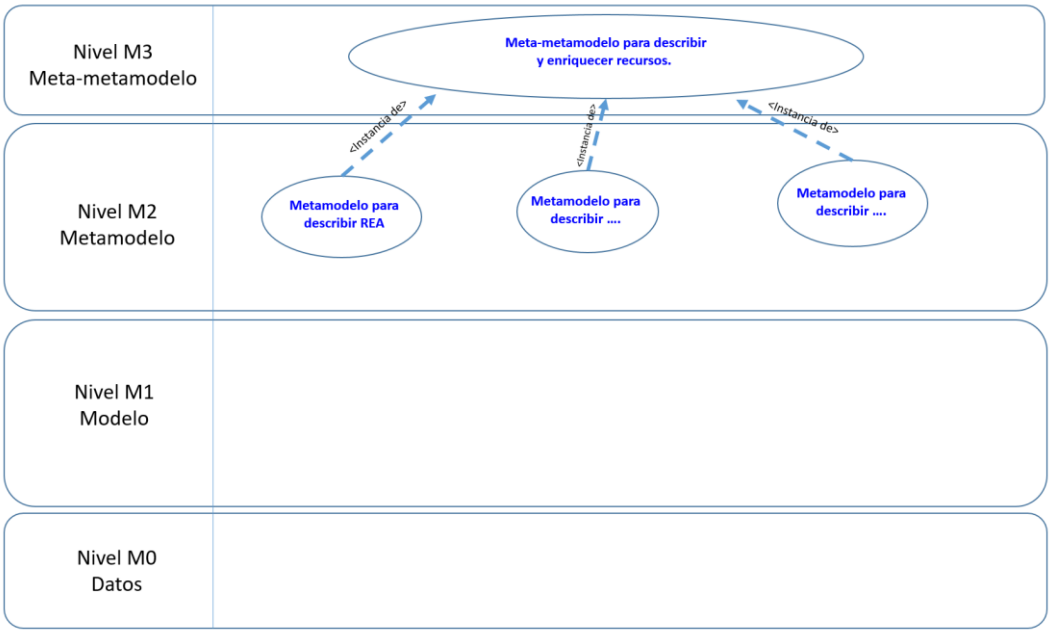


Figura 70. Adición de la capa M3 – Meta-metamodelo.

Esta capa de meta-metamodelo permitiría generar metamodelos para describir recursos educativos, procesos judiciales, contrataciones, quejas y reclamos, presupuestos, entre otros tipos de recursos que maneja la Universidad Distrital.

Un ejemplo básico de la construcción del nivel M3 del metamodelo propuesto, corresponde a la articulación con los modelos de datos trabajados por:

- PlanEsTIC-UD – Plan Estratégico de Tecnologías de Información y Comunicaciones – Universidad Distrital (Planestic, 2019; Planestic, 2020), quienes asesoran procesos de formación de integración didáctica de las TIC en las prácticas pedagógicas presenciales, y acompañan el desarrollo de programas con metodología virtual.
- La OAS – Oficina Asesora de Sistemas (OAS, 2020), quienes planean, proponen e implantan la sistematización de la información de actividades, procesos y tareas institucionales en las diferentes dependencias con miras de agilizar el funcionamiento administrativo, operativo y de planeación de la Universidad Distrital.

Al generar un nivel M3 para el metamodelo propuesto, se podrían articular modelos de datos de:

- Sistema de gestión para el concurso de méritos – JANO (JANO, 2020).
- Sistema de gestión de información docente – kyron (Kyron, 2017), que gestiona las hojas de vida de los docentes.

Al llevar a cabo esta articulación, se trabajarían con diferentes dominios de conocimiento por lo que se requeriría de operaciones entre metamodelos e instanciación de las clases del meta-metamodelo, rol que sería abordado por el nivel superior M3 del metamodelo (Figura 71).

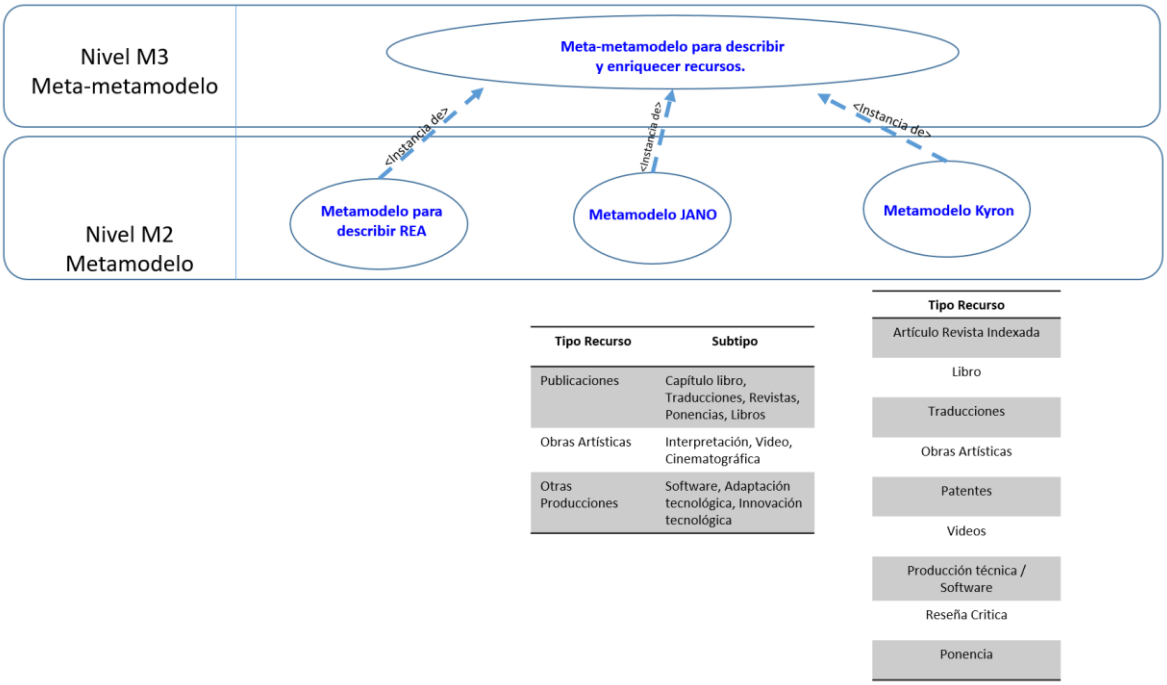


Figura 71. Ejemplo de la capa M3 – Meta-metamodelo.

Después de haber revisado los resultados obtenidos y los aportes esgrimidos por esta investigación, a continuación se exponen los argumentos que se discuten como mejoras y aportes al conocimiento, provistos por esta investigación

8.3. Discusión Planteada

La discusión planteada en esta investigación se orienta en la verificación de la hipótesis, la interoperabilidad y los aportes del metamodelo sobre los estudios previos.

En primera instancia, en cuanto a los elementos de la hipótesis planteada, se puede identificar que su conjugación aporta mejoras en la disponibilidad y reutilización de los recursos educativos abiertos. Lo anterior se verifica a través de:

- Como es argumentado por Wason (2000), MEN (2015), Martínez *et al.* (2019), LAR (2019), Atenas y Havemann (2013) y Castro *et al.* (2009), los diferentes tipos de recursos educativos se describen bajo un mismo conjunto de elementos de metadatos, lo que permite la generación de un único modelo para describir REA cuyas particularidades son manejadas por elementos calificadores, que especifican las características de los recursos. Por ende, el modelo propuesto definió un conjunto de propiedades aplicables a todos los tipos de REA. Por otro lado, considerando que: a) sistemas como la Universidad Distrital Francisco José de Caldas manejan diferentes tipos de recursos (entre ellos los recursos educativos), estos recursos poseen tanto características comunes como características particulares; y b) el desarrollo de modelos permite la generación de otros modelos que luego, al ser juntados, proveerán la solución a todo un sistema e independiza el desarrollo de las tecnologías empleadas (Montenegro *et al.,* 2011; OMG, 2019; OMG, 2014a); por ende, el trabajo con la arquitectura dirigida por modelos – MDA se convirtió en una elemento clave en esta

investigación dado que además de permitir la generación de un modelo para describir REAs, permite reconocer y trabajar la interoperabilidad entre los diferentes tipos de recursos gestionados por el sistema. Este aporte permite que, como trabajos futuros, se pueda ampliar el modelado a los diferentes tipos de recursos existentes atendiendo las directrices de los metamodelos.

- Autores como Navarrete y Luján (2015), Alyahibi y Alsharif (2019), Nahhas *et al.* (2018), Guy *et al.* (2014), Mouromtsev y d'Aquin (2016) y Pereira *et al.* (2018), han identificado que los recursos educativos abiertos presentan diferentes desafíos en cuanto a la apertura y enlace de datos tales como la interoperabilidad de los datos y recursos educativos, el acceso a los datos, enriquecimiento de material, entre otros. Como se observa en el estado de la cuestión presentado en esta investigación, se han construido diferentes iniciativas para abordar estas problemáticas, las cuales se han orientado a las últimas fases del ciclo de vida de datos abiertos y enlazados (McCrae, 2020; Hogan *et al.,* 2012; Herrera-Cubides *et al.,* 2016; Unión Europea, 2020; GODI, 2019; COL, 2015; Abella *et al.,* 2018; Smith-Yoshimura, 2018; Herrera-Cubides *et al.,* 2017; LOD2, 2019; Klímek *et al.,* 2018; Dietze *et al.,* 2012). Por ende, dos elementos cobraron especial importancia en esta investigación: la calidad de los datos y las primeras fases del ciclo de vida de los datos abiertos. Es por ello, que esta investigación abordó los elementos de calidad de datos desde la abstracción y modelado de las entidades – los REA, buscando definir aquellos elementos de metadatos que mejoraran la descripción de los REA con el fin de proveer instancias con mejores elementos para abordar los problemas actuales de calidad. Aunque en la hipótesis se propuso el trabajo de tres dimensiones de calidad de datos, dado el alcance de las métricas de la dimensión de Accesibilidad, ésta no se pudo incluir al no aportar elementos que impactaran de forma directa la abstracción y modelado de propiedades para la descripción de los REA. El abordar la calidad de los datos desde los procesos de abstracción y modelado se convierten en estrategias claves para superar diferentes problemáticas de calidad que se presentan en los datasets actuales tales como la falta de confianza ante elementos de metadatos pobres con relación a procedencia de los datasets, por ejemplo. Lo anterior dado que las dimensiones de calidad aportan diferentes componentes que permiten mejorar y complementar la descripción de los recursos desde diferentes ópticas, como lo muestra la Figura 72.



Figura 72. Aportes desde las Métricas de Calidad.

- En cuanto a la mejora en la disponibilidad y reutilización de los recursos (Abella *et al.*, 2018), el modelo propuesto en esta investigación mejora la descripción de los recursos incluyendo

más elementos de procedencia, licenciamiento, descripción educativa, entre otros elementos de metadatos, que además de mejorar la consulta y recuperación de los recursos, proporciona descripciones publicadas en la Web de forma abierta e identificables usando URIs. Factores que se convierten en componentes claves en el proceso de apertura de datos que viene adelantando diferentes países como Colombia, por ejemplo. Además, una mayor disponibilidad de los REA, procura nuevos avances en la labor de abrir y compartir el conocimiento, dando oportunidad de socializar y democratizar el mismo. En cuanto a la reutilización, el metamodelo propuesto articula el componente de enlace de datos, con el fin de proporcionar mayor contenido semántico a la descripción de los recursos. Aunque el enlace de datos ha experimentado un detrimento en su uso, el aporte en cuanto a calidad de datos y mejora en la descripción de los REA, proporcionado por esta investigación, brinda elementos para que el enlace de datos se convierta en elemento amplificador del conocimiento ligado a la descripción de los recursos. Por otro lado, la generación de grafos RDF haciendo uso de la descripción de los metadatos de los REA, se convierte en estrategia clave para la visibilidad de los recursos, además de poderse conectar a otros grafos de conocimiento, mejorando la accesibilidad y disponibilidad de los mismos.

Por último, aunque una cantidad mayor de elementos de metadatos a considerar podría ser una tarea engorrosa para algunas personas, el desarrollo de esta tarea engorrosa no se compara con los beneficios que recibe un individuo al poder consultar y recuperar un REA que se ajuste a sus necesidades, sin tener que consumir horas en procesos de búsqueda o acceder a recursos que no están dentro de su interés, afectando de esta manera su experiencia de usuario además consumir recursos adicionales como tiempo, dinero, energía, entre otros.

Es por ello que el metamodelo diseñado en esta investigación, no solamente contribuye a mejorar la calidad de las descripciones de los recursos, sino que además: a) ofrece la posibilidad de ser interoperable con otros tipos de recursos, b) procura una mejor disponibilidad de los recursos atendiendo criterios de calidad, b) mejora la visibilidad y reutilización de los recursos al interactuar con grafos de conocimiento que son posibles de enlazar a otros grafos para proporcionar mejor semántica a sus descripciones.

En segunda instancia, en cuanto a la interoperabilidad de las descripciones de los REA, la pesquisa permitió identificar que algunas de las instancias consultadas no distribuyen los valores de los metadatos de forma adecuada, acumulando información en metadatos que no les corresponden. En otros casos, los modelos de datos no refinan sus propiedades o características para no incluir sus valores en otras propiedades. Algunos ejemplos de estas situaciones son:

a. Resumen (Abstract) junto con descripción educativa (Competencias).
b. Describir las competencias como si fuera el resumen del recurso (Abstract).
c. Incluir el departamento y la institución, como datos del autor del recurso.
d. El formato da la imagen y el tamaño no se separan en formato técnico y tamaño del formato.
e. Documentar el nivel educativo como los grados para los cuales aplica el recurso.
f. Documentar el contexto educativo como el nivel educativo en el cual se aplica el recurso.
g. Denominar el tipo de recurso en formas diferentes: En Merlot un recurso se declara de tipo “Animación”, mientras que el mismo recurso en la Politécnica de Valencia se denomina “Polimedio”.

Estos hallazgos, identificados en las instancias consultadas, no generan búsquedas refinadas, no completan indicadores de calidad como por ejemplo la evaluación de la procedencia y la verificabilidad, además de que los resultados de las consultas no arrojan la información completa y oportuna.

Por ende, la aplicación del metamodelo diseñado complementa la información requerida para la caracterización con calidad de los recursos, y provee información para refinar las búsquedas a realizar.

Por otra parte, al comparar cualitativamente las estructuras de metadatos provistas para describir los recursos, se identifica que, además de basarse en el esquema IEEE LOM, el modelo diseñado en esta investigación tiene un cubrimiento de propiedades similares e interoperables con los modelos presentados por proyectos como:

- La Referencia (LAR, 2019): Red Federada de Repositorios institucionales de Publicaciones Científicas, proyecto gestado por Organismos de Ciencia y Tecnología de Latino América, al cual Colombia está suscrita a través de Colciencias. En este proyecto se definen un conjunto de 16 metadatos para la descripción de sus publicaciones (Figura 73), reduciendo el espectro de elementos de metadatos que podrían ser aplicables a la descripción de un REA, limitando de esta manera elementos que brindarían búsquedas y recuperaciones más asertivas y refinadas.

Metadato	Regla	Validación 2013-2014	Validación 2015	Comentario
TITLE	Contienen el campo título (dc:title)	Obligatorio	Obligatorio	Si el metadato se halla ausente, el registro es rechazado.
CREATOR	Contiene el campo autor (dc:creator)	Obligatorio	Obligatorio	Si el metadato se halla ausente, el registro es rechazado. En general se usa formato APA para nombres.
RIGHTS	Contiene campo del nivel de acceso (dc:rights).	Recomendado	Obligatorio	Ejemplo: openAccess. En caso de transformación se lleva bitácora por país. Ver sección vocabulario.info:eu-repo/semantics
	Contiene un campo de derechos de autor/condiciones de licencia (dc:rights).	Recomendado	Recomendado	Especificar la URL de la licencia, como en el caso de Creative Commons.
SUBJECT	Contiene el campo materia (dc:subject)	Recomendado	Obligatorio si es aplicable	Palabras clave, descriptores y/o códigos de clasificación.
DESCRIPTION	Contiene el campo descripción - resumen (dc:description)	Recomendado	Obligatorio si es aplicable	Insumo que permite aplicar técnicas de minería de textos. Se recomienda que se incluya una versión en español/portugués y en inglés para mayor visibilidad.
PUBLISHER	Contiene el campo editor (dc:publisher)	Sin validación	Obligatorio si es aplicable	Entidad responsable de hacer que el recurso esté disponible. Una persona, una organización o un servicio pueden ser un editor.
CONTRIBUTOR	Contiene el campo colaborador (dc:contributor)	Sin validación	Recomendado	Para el caso de las Tesis es Obligatorio si es aplicable colocar el nombre del Director (a) o Supervisor (a) de la Tesis.
DATE	Contiene el campo de fecha de publicación (dc:date)	Obligatorio	Obligatorio	Cumplimiento de la norma ISO-88601. Si el metadato se halla ausente, el registro es rechazado. Sólo en caso de transformación se obtendrá la menor de ellas y se asumirá como fecha de publicación.
TYPE	El campo Publication Type (dc:type) se encuentra conforme al vocabulario establecido.	Obligatorio	Obligatorio	Se recolecta: Artículos, Tesis de Magister, Tesis de Doctorado, Reportes. Si el metadato se halla ausente, el registro es rechazado. En caso de transformación se lleva bitácora por país. Ver sección vocabulario. info:eu-repo/semantics
	dc:type indica la versión del documento (Publication Version)	Recomendado	Recomendado	Ver sección vocabulario (accepted, published…). Ver sección vocabulario. info:eu-repo/semantics
FORMAT	Contienen un campo de formato (dc:format)	Recomendado	Recomendado	Ejemplo: pdf, xml, etc. http://www.iana.org/assignments/media-types/media-types.xhtml

IDENTIFIER	Contienen un campo de identificador (dc:identifier).	Obligatorio	Obligatorio	En caso que no exista una URL bien conformada, el registro será rechazado. Ejemplos: URI (Uniform Resource Identifier), URL, DOI (Digital Object Identifier) y Handle. Respecto al nombre de la institución y repositorio revise *Recomendaciones Específicas a Cosechadores Nacionales*.
SOURCE	Contienen un campo de fuente (dc:source)	Recomendado	Recomendado	Se utiliza para hacer la relación entre un objeto digital que es parte de un recurso mayor
LANGUAGE	Contienen un campo de idioma (dc:language).	Recomendado	Obligatorio si es aplicable	ISO 639-x, donde x puede ser 1,2 o 3. Mejor práctica: Usar ISO 639-3. En caso de transformación se lleva bitácora por país.
RELATION	Contiene un campo relación (dc:relation)	Sin validación	Opcional	Se utiliza para relacionar distintos registros de metadatos, que pertenecen a distintas versiones del mismo objeto digital (por ej. versión del autor y versión del editor; preimpresión, postimpresión, etc. El valor de dc:relation es el identificador del otro documento.
COVERAGE	Contiene un campo de cobertura (dc: coverage)	Sin validación	Opcional	Ver directrices Driver y OpenAIRE
AUDIENCE	Contiene un campo de audiencia (dc: audience)	Sin validación	Opcional	Ver directrices Driver y OpenAIRE

Figura 73. Metadatos Proyecto LA Referencia. Fuente: LAR (2019).

Varios repositorios institucionales de bibliotecas a nivel Colombia, están orientado la gestión de sus datos para el manejo interoperable de sus recursos bibliográficos bajo este proyecto. Esta orientación gestionaría todos los tipos de recursos, incluidos los educativos, bajo un mismo esquema de recursos genéricos.

- DRIVER - "*Digital Repository Infrastructure Vision for European Research*" (Visión de infraestructura de repositorios digitales para la investigación europea), es un proyecto fusionado con OpenAIRE infrastructure, que consiste en un recolector y buscador europeo que proporciona acceso a repositorios gratuitos de contenido académico con más de 2.500.000 de publicaciones científicas (artículos de revistas, tesis, libros, informes, etc.) que son recolectadas regularmente de más de 249 repositorios de 33 países (Driver, 2020). Aunque este proyecto tiene un conjunto de metadatos más amplio, se ajusta a descripciones de recursos genéricos, y por otro lado algunos de sus elementos no tienen relación con estándares como Dublin Core, lo cual sesga su interoperabilidad con estándares reconocidos a nivel de mercado (Ceriotto y Testa, 2011).

- Red Colombiana de Información Científica: proyecto conformado por el Ministerio de Ciencia, Tecnología e Innovación en el 2019, con el objetivo de articular los esfuerzos de los actores del Sistema Nacional de Ciencia, Tecnología e Innovación, para potenciar la visibilidad, circulación y gestión de la información científica colombiana (RedCol, 2020).

  Estas directrices se orientan a partir de los productos que son reconocidos como resultados de las actividades de grupo de investigación según el Modelo de Medición de Grupos de Investigación, Desarrollo Tecnológico o de Innovación y de Reconocimiento de Investigadores del Sistema Nacional de Ciencia, Tecnología e Innovación (RedCol, 2020).

- CEM-2: estándar definido por el Ministerio de Educación Nacional, para la implementación de la Estrategia Nacional de Recursos Educativos Digitales Abiertos (MEN, 2013).

Lo anterior evidencia que, a pesar de que algunos de estos proyectos se orientan a recursos académicos e investigativos, el metamodelo propuesto: a) cuenta con una estructura de metadatos interoperable con otras propuestas, b) el perfil de aplicación PA-REA permite la generación de RDF Graph que articulan y visibilizan los recursos descritos con otros grafos de conocimiento, c) los metadatos definidos también pueden ser susceptible de ser recolectado usando OAI-PMH, en el caso que la institución que lo implemente así lo requiera; y d) gracias a la estructura de metamodelo diseñada, se pueden construir modelos adicionales para la descripción de recursos en dominios particulares.

Por último, este tipo de alineación con otros proyectos y la atención a estándares y esquemas ampliamente reconocidos, procurados por el perfil de aplicación generado en esta investigación, son aspectos importantes para la descripción de los recursos, además de poder construir una interoperabilidad más sólida. Como lo declara la Doctora Vallejo: "*Es muy importante tener en cuenta los estándares establecidos para que no se incurra en errores que impidan facilitar procesos de intercambio de información, como es el caso del repositorio institucional de la Universidad Distrital - RIUD, que actualmente participa en la Red Colombiana de Información Científica con solo 7 recursos, porque la demás producción de la Universidad Distrital tiene problemas de metadatos para ser compartidas y por lo tanto no se pueden ver en dicho repositorio*" (R. Vallejo, comunicación personal, 4 de diciembre de 2020). Tal apreciación refuerza este modelo de datos, dado que atiende a estándares reconocidos, además de proveer una mejor descripción de los REA, lo cual amplía su espectro de interoperabilidad con otros modelos de datos para describir REA. Por último, en la revisión de literatura realizada, no se encontraron propuestas que articularan el enlace de datos de forma amplia a la descripción de recursos y la mejora semántica de su descripción. Estos referentes convierten a esta propuesta en una alternativa viable no solamente para la descripción de los recursos, sino para procurar mejoras en su disponibilidad, visibilidad y reutilización, además de propender por la apertura de datos y el aprovechamiento de las bondades del enlace de datos

8.4. Conclusiones del Capítulo.

Los aportes identificados evidencia que el diseño de este metamodelo, a través del cumplimiento de la promesa de MDA, genera un modelo de datos para la descripción y enriquecimiento de los REA, como se describe en la Figura 74, cuyas instancias aportan recursos mejor descritos, disponibles e interoperables con otros modelos que describen recursos educativos no solamente a nivel de los requerimientos de la Universidad Distrital Francisco José de Caldas sino a nivel nacional e internacional.

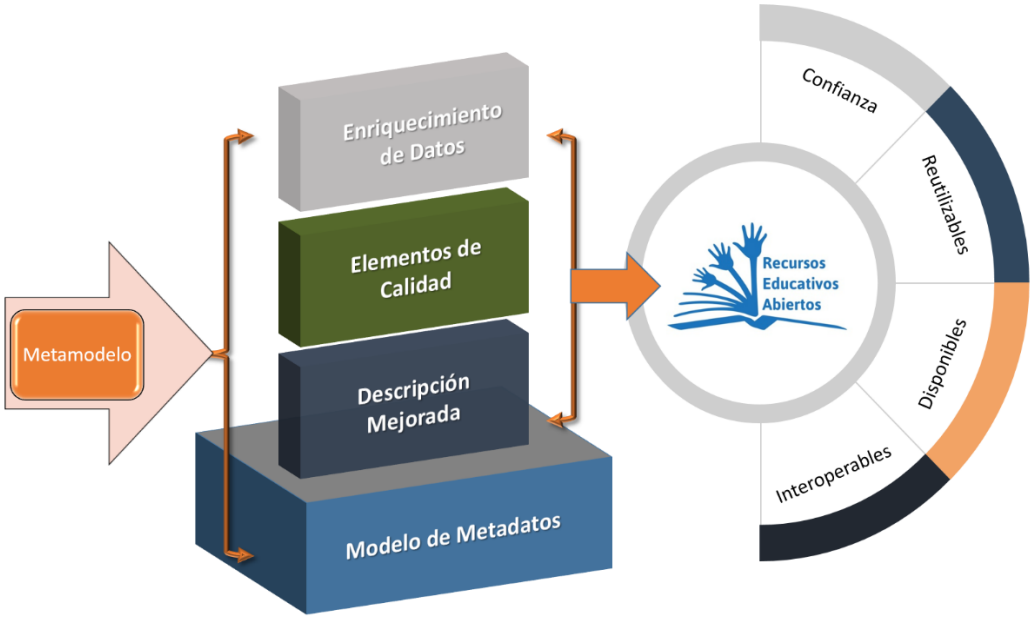


Figura 74. Alineación de la hipótesis con los resultados obtenidos.

Adicionalmente, a través de la inclusión de parámetros de calidad de datos y enriquecimiento de sus propiedades con base en el modelo de datos enlazados, este modelo proporciona metadatos que contribuyen tanto a mejorar la calidad de las instancias de metadatos, como a mejorar la confianza por parte de los usuarios dado que proporciona propiedades más refinadas que

además de proveer una descripción más completa y concisa, permite comparar y evaluar sus contenidos contra otros referentes.

De igual forma, el modelo de datos generado mejora la reutilización, visibilidad y accesibilidad de los recursos al describirlos a través del modelo de datos RDF, permitiendo de esta forma poder ser enriquecidos a través de su articulación con otros grafos de conocimiento, además de proporcionar la plataforma necesaria para enlazar sus instancias a la nube de datos LOD.

Las mejoras provistas en esta investigación en cuanto a la descripción y enriquecimiento de los REA, además de impactar de forma positiva la identificación, consulta y recuperación de los recursos educativos, proporciona un paso más en cuanto al acceso al conocimiento y mejora de la calidad de vida de las personas, dado que además de promulgar la apertura y socialización de los recursos bajo licencias abiertas, se busca poner a su disposición una mejor descripción del conocimiento, procurando mejoras en su visibilidad y reutilización. Este tipo de impactos procuran que a futuro las personas puedan acceder a un pool de recursos mejor descritos y complementados.

Por último, estos aportes confirman que los artefactos generados en esta investigación, propuestos en los objetivos planteados, se alinean con la hipótesis planteada bajo los criterios de interoperabilidad, disponibilidad, reutilización y confianza.

PARTE IV CONCLUSIONES E INVESTIGACIÓN FUTURA

CAPÍTULO IX CONCLUSIONES E INVESTIGACIÓN FUTURA.

9.1. Resumen de Objetivos y Resultados.

En esta sección, a modo de resumen, se presentan todo los resultados que se han generado en esta investigación. Para ello en la Tabla 45 se detallan los objetivos de esta investigación, enunciados en el Capítulo 1, junto con los resultados obtenidos a lo largo del proceso de investigación realizado.

Tabla 45. Resumen de Objetivos y Resultados.

Objetivo	Descripción	Resultado
O1	Llevar a cabo un diagnóstico de la Web de los Datos, con el fin de especificar los requerimientos funcionales y conceptuales sobre los procesos de enlace de datos y dimensiones de calidad de datos enlazados.	A
O2	Proponer, diseñar e implementar un metamodelo para la exposición y consumo de REA, basado en principios de calidad de datos enlazados, con el fin de contribuir a la generación de confianza en los consumidores de datos enlazados y a en las aplicaciones que hacen uso de ellos.	B
O3	Diseñar y construir un modelo instrumental que permita verificar las parametrizaciones planteadas, analizar y evaluar los resultados con el fin de establecer el grado de contribución que pueda prestar el modelo propuesto para el enlace de REA.	C

Los resultados obtenidos de este proceso de investigación se llevaron a cabo en las siguientes actividades:

A. Estudio que se abordó en dos capítulos de la memoria: i) Capítulo 2 asociado con el marco teórico, y ii) Capítulo 3 donde se expusieron los antecedentes de la apertura y el estado de la cuestión. Capítulos que permitieron esbozar los requerimientos claves para el diseño de la propuesta.
B. Actividad que se abordó a lo largo del Capítulo 5 de la memoria, mediante el diseño del metamodelo para la descripción y el enriquecimiento de los REA. Capítulo en el cual se establecieron las características de cada nivel del metamodelo, junto con los componentes que se conjugaron en su construcción.
C. Actividad que se desarrolló en tres capítulos: Capítulo 6, asociado a la implementación del modelo de datos, donde se llevó a cabo el despliegue de la infraestructura requerida y del modelo generado como artefacto principal del metamodelo; y el Capítulo 7, asociado a la validación del metamodelo, donde se llevó a cabo la prueba de concepto, la validación del despliegue, la validación del cubrimiento, la validación empírica, y la prueba de aceptación; y por último, el Capítulo 8, asociado a los aportes y discusiones del metamodelo diseñado.

9.2. Conclusiones.

La apertura y enriquecimiento de datos ha venido sufriendo diferentes problemas a lo largo de cada una de las etapas definidas en su ciclo de vida. Los REA, como recursos susceptibles de apertura y enriquecimiento no han sido ajenos a problemas. Estas problemáticas varían desde procesos de abstracción de propiedades deficientes o sujetos a necesidades particulares, hasta instancias de datos que no cuentan con la información suficiente, completa o adecuadamente diligencia. Para el caso de los REA, los problemas de calidad de datos en características como

la procedencia, el licenciamiento, características educativas, entre otros factores, contribuyen a sesgar y reducir la disponibilidad y reutilización de los mismos. Estas limitaciones en estas propiedades se reflejan en problemas para consultar, ubicar, recuperar, complementar y compartir la información que describen los REA publicados en la Web. Por otro lado, la conjugación de procesos como la apertura de recursos, la descripción y enriquecimiento de sus propiedades, y por último, el proceso de mapeo de los modelos de datos requeridos, generan un proceso de análisis y desarrollo de constructos teóricos, reglas y restricciones para llevar a cabo la implementación de cada uno de estos procesos. Por lo tanto, este proceso de análisis y desarrollo, además de conllevar la construcción de los esquemas requeridos para modelar los componentes de un sistema, debe buscar que estos esquemas provean lineamientos mínimo como: i) una mayor facilidad de cambiar este esquema, es decir, buscar asegurar una menor rigidez; ii) un menor grado de inmovilidad, en otras palabras, un menor grado de resistencia a ser reutilizado; y iii) mecanismos fáciles para realizar cambios, en otras palabras, una baja viscosidad.

Por último, este tipo de esquemas, con sus respectivas reglas y restricciones, debe ser extensible a parametrizar otros tipos de recursos, dado que: i) los REA conservan un mismo modelo de propiedades para su descripción; ii) adicional a los REA, los repositorios institucionales como las bibliotecas universitarias, por ejemplo, manejan otros tipos de recursos, para los cuales un modelo base de propiedades debe ser aplicable y extensible para su uso.

Es por ello que esta investigación propuso un metamodelo para la vinculación de REA en el contexto basado en LOD, a partir de principios de calidad como enfoques contextuales, accesibles y por último principios de confianza de los datos vinculados. Con base en estos criterios, entre otros expuestos a lo largo de la presente investigación, a continuación se presentan de manera desagregada las conclusiones para llevar a cabo el planteamiento del mismo.

9.2.1. Asociadas a la descripción, apertura y enriquecimiento de los REA.

En cuanto a la apertura de recursos, como se observa en las secciones 3.1., 3.2.1. y 3.2.2., se identificaron diferentes problemas asociados a la apertura tanto de los recursos en sí mismo como de sus metadatos. Problemas tales como qué datos abrir, percepción de posible pérdida de derechos sobre los recursos, problemas asociados a licenciamientos y formatos a usar para abrir de manera correcta los datos, entre otros factores. Por su parte, la falta de apertura de los conjuntos de metadatos que describen los recursos restringe la interoperabilidad de los mismos, llevando a procesos de negociar, consensuar y tratar de resolver problemas funcionales asociados a la interoperabilidad de los mismos (Ramesh *et al.*, 2015). Los REA, pertenecientes al dominio de los recursos, sufren de problemas de apertura tales como: i) una gran mayoría de REAs se encuentran publicados de forma propietaria en la Web; ii) a pesar de utilizar un mismo esquema de metadatos para describir los REA, los metadatos que describen estos REAs, no son públicos, o en su defecto tiene una cantidad reducida de elementos de metadatos, ajustados a necesidades particulares, para describir los recursos relacionados, denominados perfiles de aplicación. Es por ello que autores como Pereira *et al.* (2018) plantean retos como la integración de datos distribuidos entre repositorios heterogéneos, mediar y transformar datos que describen recursos y enriquecer e interconectar metadatos no estructurados.

En cuanto al enriquecimiento de datos en el dominio de los REA, como se observa en las secciones 3.2.7. y 3.2.8., aparte de los problemas allí descritos, se identificó el escaso uso de tecnologías de enriquecimiento semántico tales como datos enlazados, aplicables a este tipo de recursos. Aspecto que restringe la adición de contenido semántico que contribuya al enriquecimiento de la información que describe el REA; además de restringir a los usuarios consumidores de poder ampliar su proceso de aprendizaje, a través de la vinculación de más contenido de referencia que fortalezca su proceso de búsqueda y recuperación de información.

Por estos motivos, la mejora en la descripción de los REA, la apertura de los modelos de datos construidos y el enriquecimiento de los mismos, se convirtieron en puntos clave de esta investigación. Aunque la información siempre será heterogénea y los modelos de datos usados en cada conjunto de datos pueden ser distintos, ello no implica que se deje de lado aspectos que permitan mejorar la descripción de propiedades dentro de entidades del dominio de conocimiento. Por lo tanto, en primera instancia se identificó la necesidad de mejorar la disponibilidad y la reutilización de los REA mediante procesos de abstracción y modelamiento a través del uso de elementos de calidad de datos y elementos de enriquecimiento de datos. Como resultado de este proceso se consolidó una taxonomía de metadatos para describir REAs (Sección 5.3.). Esta taxonomía propone una descripción más amplia, que incluye diferentes aspectos como la procedencia, referentes educativos, entre otros elementos de metadatos; además de publicarla de forma abierta, permite que los usuarios puedan tener más y mejores elementos para buscar, identificar, consultar, recuperar y compartir los REA. En términos generales, los hace más disponibles para los usuarios. Por otro lado, el enriquecimiento de los elementos de metadatos permite fortalecer el conocimiento publicado al ser enlazado con otros grafos de conocimiento, proporcionando más contenido que complemente los procesos de aprendizaje, además de disminuir los tiempos de búsqueda y los esfuerzos consumidos para tal fin.

Dada estas características, se consideró importante el manejo de vocabularios controlados para el trabajo con los recursos educativos, con el fin de brindar elementos que, además de enriquecer los recursos publicados, éstos puedan ser disponibles y ser reutilizados. Como se presentó en los hallazgos identificados (Secciones 2.2., 3.1. y 3.2.), a pesar de existir diferentes propuestas de estándares como IEEE LOM, Scorm, LRMI, entre otras, los perfiles de aplicación identificados y analizados son de un alcance limitado, restringiendo particularidades de la caracterización de los REA que cobran especial énfasis al momento de verificar si un recursos responde a las necesidades del usuario (por ejemplo: temática, competencias a desarrollar, tipo de recurso, etc.); o por otro lado, permitir establecer indicadores requeridos para procesos de registro calificado y de acreditación de programas académicos, como se lleva a cabo en Colombia, donde elementos de metadatos como la institución, el tipo de institución, el grupo de investigación, entre otros datos acerca del REA, son de especial valor al momento de construir estos documentos maestros para ser remitidos a dichos procesos de aseguramiento de calidad en la educación.

Por ende, la apertura y el enriquecimiento de los REA no solo los vuelven más disponibles y reutilizables, adicionalmente mejoran los procesos de aprendizaje y la calidad de vida de los usuarios. Aunque un conjunto de elementos de metadatos mayor puede verse como una tarea difícil de cumplir, su completitud y consistencia aportan mucho más valor agregado a las personas que los usan, como se puede observar en la sección 7.5.

9.2.2. Asociadas a las dimensiones de calidad de datos.

Como se identificó en el estado de la cuestión (Secciones 3.1. y 3.2.), las investigaciones en cuanto a la revisión y evaluación de la calidad de los datos se han venido enfocando en superar problemas de calidad presentes en las instancias de datos existentes, focalizándose en atacar las últimas fases del ciclo de vida de los datos abiertos y enlazados (Navarrete y Luján, 2015; Alyahibi y Alsharif, 2019; Nahhas *et al.,* 2018; Guy *et al.,* 2014; Mouromtsev y d'Aquin, 2016). Trabajos que han venido dejando con poca participación en la mejora de la calidad de los metadatos a las etapas iniciales del ciclo de vida de apertura y enriquecimiento de datos: la abstracción de propiedades y el modelado de las entidades. Los resultados obtenidos en dichas etapas son susceptibles de ser mejoradas a través de la inclusión de elementos de calidad de datos proporcionados por dimensiones de calidad como las dimensiones contextual y de confianza. Por ende, uno de los aportes generados por esta investigación corresponde a que las mencionadas dimensiones de calidad permitieron identificar propiedades (Secciones 5.1. y 5.2.) que aportaron elementos de metadatos para describir los REA (Sección 5.3.), los cuales en diferentes perfiles de aplicación no se habían tenido en cuenta (Secciones 2.2., 3.1. y 3.2.).

El uso de estas dimensiones permitieron complementar la descripción de los REA adicionando elementos de metadatos tales como: esquemas de identificación, diferenciación de contribuyentes, ampliación de elementos de licenciamiento, entre otros elementos que permitieron superar falencias identificadas en la descripción de los REA (Secciones 3.1., 3.2., 5.2. y 5.3.). Estos elementos de metadatos adicionados proveen mayor información para posteriores evaluaciones de calidad en las instancias generadas.

Por otro lado, aunque en la investigación se planteó el trabajo de tres dimensiones de calidad: Contextual, Confianza y Accesibilidad, esta última no se aplicó al desarrollo de la abstracción de propiedades dado que sus métricas se enfocaron directamente a la evaluación de la infraestructura tecnológica que soporta la instancia de datos publicada, y que se encuentra ampliamente evaluadas en diferentes aportes realizados por autores como Bizer (2007), Hogan *et al.* (2010) y Hogan *et al.* (2012).

Por último, la inclusión de estas propiedades derivadas de la observación de las dimensiones de calidad de datos, permitieron ampliar el cubrimiento del modelo de datos propuesto (Sección 7.3.), además de superar la validación empírica y de aceptación (Secciones 7.4. y 7.5.) dado que permitieron ofrecer más elementos de metadatos que facilitaron interoperar con otros modelos, realizando un cubrimiento más amplio; y por otro lado, complementar las necesidades del usuario al momento de recuperar REAs en procesos de búsqueda.

Es por esto que ante la necesidad palpable de aportar en la mejora de la búsqueda, ubicación y recuperación de los REAs, esta investigación articuló diversos enfoques orientados a: (i) diferentes descripciones de REA con el fin de construir una descripción más sólida; (ii) integrar soluciones basadas en el movimiento de Datos Abiertos, (iii) evaluar dimensiones de calidad de datos que aportaron elementos de metadatos claves para posteriores valoraciones de calidad en las instancias de datos y, finalmente (iv) integración de procesos de enriquecimiento a través de tecnologías como Linked Data. Todo ello con el fin de proveer un modelo de descripción de REAs abierto, que opera con criterios de calidad, que usa un mismo esquema de metadatos para describir las propiedades y gestiona sus particularidades a través de calificadores.

En resumen, para aportar mejoras sustanciales a la disponibilidad y reutilización de los datasets que describen los REA, se requieren de condiciones mínimas como son: 1) tener una descripción de propiedades mejorada y enriquecida a través de enlace de datos, 2) atender dimensiones y criterios de calidad de datos desde su abstracción de propiedades, 3) estar publicada de forma abierta y estructurada en la Web, 4) en un formato no propietario bajo una licencia abierta, e 5) identificado por una URI; tal como se detalla en la sección 5.3. Estos criterios se consideran fundamentales al momento de ingresar los REA al mundo de los datos abiertos y enlazados (Sección 5.3.4.), dado que proporcionan descripciones más detalladas al momento de ser consultadas y recuperadas, y que se articulan con más conocimiento que les proporciona un mejor contexto al momento de usar y compartir estos REA en procesos de aprendizaje (Secciones 5.4., 7.2., 8.1. y 8.3.).

9.2.3. Asociadas al metamodelo como estrategia de abstracción y modelado.

Para llevar a cabo la articulación del perfil de aplicación (estándares de metadatos, dimensiones de calidad y enlace de datos), el modelo de datos RDF y el esquema de metadatos para describir los REA, se hizo uso de una estrategia que permitiera un modelamiento flexible e interoperable, que disminuyera la rigidez, inmovilidad y la viscosidad ante posibles cambios o escalamientos de las abstracciones realizadas.

Con base en estos requerimientos, se planteó la implementación de MDA como pilar de trabajo para el diseño de un metamodelo para describir y enriquecer REA. Los metamodelos, expresados

en modelos de representación de REA, RDF y descripción de los recursos (metadatos), determinaron no solamente la estructura sino también las relaciones entre los tipos de entidad en un modelo. De tal forma que el uso del metamodelo para flexibilizar y escalar la abstracción de los REAs, entidad principal del dominio de conocimiento, permitió no solamente generar instancias válidas y conformes a las diferentes reglas y restricciones, sino también proporcionar una estructura que permitiera extenderse a otros dominios de conocimiento gracias al trabajo de los modelos propuestos. Un ejemplo de ello se puede encontrar en los apartados 5.2. a 5.5.

El uso de metamodelos, además de permitir una gestión flexible y sencilla de los elementos que participaron en la construcción y gestión de elementos del modelo de datos para describir los REA (Sección 7.1.), permiten la extensión de esta propuesta hacia otros tipos de recursos y otras organizaciones, llevando a cabo parametrizaciones sencillas acorde a las reglas y restricciones adicionales (Secciones 8.2. y 8.3.).

Es por ello que el uso de metamodelos se convirtió en una estrategia clave en esta investigación, dado el manejo de sus niveles de responsabilidad, lo que permite un crecimiento hacia otros tipos de recursos y la generación de otros modelos conformes a las parametrizaciones de los niveles superiores.

9.2.4. Asociadas a la interoperabilidad con otros modelos.

En primera instancia, gracias a la ampliación de la base de propiedades para describir los REA, provista por las dimensiones de calidad trabajadas y el análisis de diferentes propuestas de perfiles de aplicación (Secciones 3.1., 3.2. y 5.2.), se construyó una taxonomía para describir los REAs (Sección 5.3.). Esta taxonomía, expresada en los mapeos propuestos (Secciones 5.4. y 5.5.), cubren los perfiles de aplicación de diferentes proyectos para la gestión de REA (Secciones 8.1. y 8.2.). Este nivel de cubrimiento proporciona un diálogo más fluido entre modelos, proveyendo mejores niveles de interoperabilidad entre los mismos. Tal es el caso del proyecto LA Referencia, proyecto para describir recursos a nivel de Latinoamérica, el cual es cubierto por la cantidad de propiedades provista por el modelo de datos producto de esta investigación. Por otro lado, un caso muy significativo corresponde al proyecto RIUD de la Universidad Distrital, proyecto que gestiona los recursos asociados a proyectos de grado generados en la Universidad. Para este caso, en el análisis provisto (Sección 5.2.4.) mostró que este proyecto caracterizaba de forma transversal todos los tipos de recursos, desde tesis hasta objetos virtuales. El modelo de datos provisto en esta investigación aporta al RIUD una base de propiedades más amplia (Sección 8.3.), que le permite poder categorizar de mejor forma los recursos, y por lo tanto, poder ofrecer mejores estrategias de búsqueda y recuperación de recursos.

En segunda instancia, gracias al trabajo con metamodelos, permitió aclarar el horizonte tanto en la construcción de una descripción transversal a todos los tipos de REA, como poder integrar otros tipos de recursos a través del diseño de nuevos modelos que cumplan con las restricciones y parámetros del metamodelo (Sección 8.3.3.). Situación que benefició a las organizaciones donde se opera con diferentes tipos de recursos, gracias al establecimiento de una serie de constructos teóricos, reglas y restricciones para el diseño de diferentes modelos para describir recursos. Un ejemplo cercano de los beneficios del uso de metamodelos, corresponde al RIUD, donde la aplicación de nuestra propuesta permitirá poder extender su esquema de gestión de recursos, a los diferentes tipos de recursos que demanda un repositorio institucional universitario, sin dejar de lado que este mismo metamodelo permitirá la generación de modelos adicionales para gestionar recursos de otras aplicaciones y unidades de la Universidad. Por lo tanto, la articulación de MDA en esta investigación, se planteó como una estrategia clave para el aprovechamiento y aporte significativo en la abstracción y modelado de diferentes tipos de recursos organizacionales, que no solamente son susceptibles de mejorar su descripción, sino que también son susceptibles de enriquecer semánticamente a través del enlace de información, la cual es provista por los grafos de conocimiento. En resumen, el uso de metamodelos permite

alivianar la creación de nuevos modelos dado que éstos se articulan con reglas, restricciones y abstracciones generadas, procurando un ambiente más interoperable y flexible, además de una gestión de las entidades y sus relaciones de forma más sencilla, al interior de la organización, como es el caso del estudio aplicado para la validación del mismo dentro de las instancias de la Universidad Distrital Francisco José de Caldas.

9.2.5. Asociadas al Modelo de Representación de REA.

En cuanto al Modelo de Representación de Recursos Educativos Abiertos, artefacto principal de este metamodelo; puede pensarse erróneamente que la inclusión de más descriptores a la abstracción de un REA se convierte en una tarea engorrosa de diligenciamiento, llegando al punto de contar con instancias de datos vacías o mal diligenciadas. Esta conceptualización se aleja de la realidad dado que la inclusión de nuevos metadatos que promuevan la completitud, la calidad, la consistencia y la confianza de la descripción de la entidad, unida al enriquecimiento de sus datos, son factores que contribuyen no solo del enlace de los datos, sino de la calidad de los datos abiertos en sí mismo. Adicionalmente, estos nuevos metadatos incluidos en la descripción de los REA, contribuyen en proporcionar elementos que permitan mejorar la calidad de vida de los usuarios, dado que proporciona información enriquecida que contribuyen a recuperar información que realmente se ajuste a sus procesos de formación, disminuyendo los tiempos, esfuerzos y energía consumidos al momento de llevar a cabo la consulta y recuperación de los recursos. Situación que fue verificada a través de las pruebas con expertos realizadas en esta investigación.

Por otro lado, esta investigación proporciona un modelo para describir y enriquecer REAs que responde a requerimientos de calidad de datos enlazados, y a un diseño más interoperable, desprovisto de limitaciones o restricciones de propiedades adicionadas por las organizaciones para modelar sus recursos de acuerdo a sus necesidades (Sección 3.1.). Adicionalmente, se configura como un modelo de metadatos flexible, con baja viscosidad y rigidez gracias a la estructura de metamodelo utilizada para su construcción; con unos componentes claramente definidos que aportan valor no solamente a la apertura de datos, sino también al enriquecimiento y enlace de REA a la Web de los Datos.

Por último, aportes adicionales de los artefactos construidos en esta investigación corresponden a: i) planteamiento de un metamodelo como el modelo para describir REA, resultantes de este proceso de investigación, pueden ser extrapolados para su uso a otros tipos de organizaciones educativas; ii) gracias al trabajo con metamodelos, la construcción de modelos para describir otros tipos de recursos, enriqueciendo semánticamente su contenido, puede ser tomada como una experiencia de trabajo para otros proyectos; iii) el modelo para describir REA obtenido en esta investigación es susceptible de ser interoperable con otros proyectos; iv) el modelo diseñado, en su mapeo de ingesta, puede llegar al punto de ser cosechado por otras aplicaciones, dado el trabajo realizado con estándares reconocidos para su construcción; aspecto que permite un diálogo más fluido con otros proyectos que busquen su articulación.

9.2.6. Asociadas a los mapeos generados.

Con relación a los perfiles de aplicación explorados, en su gran mayoría hacen uso de estándares reconocidos como IEEE LOM, pero se limitan a ser mapeados para la ingesta de los metadatos que describen los recursos (Secciones 3.1. y 3.2.). Autores como Nahhas *et al.* (2018), Alyahibi y Alsharif (2019) y Pereira *et al.* (2018), plantean como reto que la descripción de los recursos educativos se debe enriquecer semánticamente a través del uso de tecnologías como Linked Data pero, sin embargo, aún existen múltiples problemas de interoperabilidad, calidad de datos, entre otros factores que dificultan esta labor. Por lo tanto, esta investigación, además de complementar y enriquecer la descripción de los REA, visualizo como un gran aporte, el proporcionar un mapeo adicional que permitiera a los recursos poder ser reutilizados (Secciones

5.5., 6.2. y 7.1.) y potencialmente enlazados a los diferentes grafos de conocimiento publicados en la nube LOD (Sección 7.2.10.).

El conjunto de propiedades que constituyen el modelo diseñado en esta investigación se configura en los mapeos obtenidos al llevar a cabo la ingesta de metadatos. Los mapeos definidos por el metamodelo pueden tener diferentes destinos (Sección 5.4.) como lo son: (i) mapeo a un estándar reconocido y ampliamente usado (Dublin Core) que permite el almacenamiento local, la gestión de interfaces, la cosecha de datos, la interoperabilidad con otros modelos de datos trabajados por otros proyectos; (ii) mapeo a grafos de conocimiento RDF gracias al uso del modelo de datos RDF, permitiendo enlazar las descripciones de los REA con la nube LOD, la explotación a través de consultas SPARQL. Estas estrategias de mapeo, verificadas en las pruebas de verificación realizadas en esta investigación (Secciones 7.1. y 7.2.), además de visibilizar los REA de formas alternativas permite su ubicación y acceso a través de URIs y consultas SPARQL, y adicionalmente los alimenta de forma semántica con una descripción más sólida, proceso que contribuye al fortalecimiento de los procesos de aprendizaje dado que los actores del mismo pueden conectar y conceptualizar nuevos conocimientos con base en la información semántica proporcionada como enriquecimiento al REA trabajado.

Como se ha identificado y contrastado a lo largo del proceso de investigación, esta propuesta también se articula con la Carta Internacional de Datos Abiertos, al divulgar tanto la apertura de los recursos educativos como de sus descripciones o metadatos; se alinea con referentes de recursos educativos abiertos promovidos por la Unesco, donde se busca procurar un ambiente de conocimiento abierto, accesible a los diferentes tipos de población, con recursos de calidad, que además de optar por la democratización del conocimiento se orienta en la expansión del mismo hacia comunidades vulnerables, bajo el uso de licencias abiertas, buscando poner a su disposición una mejor descripción del conocimiento, procurando mejoras en su visibilidad y reutilización. Este tipo de impactos procuran que a futuro las personas puedan acceder a un pool de recursos mejor descritos y complementados.

9.2.7. Asociadas a la disponibilidad y reutilización de los REA.

Como es presentado por Abella *et al.* (2018), el proceso de apertura y enriquecimiento semántico de datos se enmarca dentro de dos componentes esenciales: la disponibilidad y la reutilización. Entendida la disponibilidad, no como el adecuado y permanente funcionamiento del ambiente tecnológico que soporta la publicación de los datasets, sino como el cumplimiento de los tres primeros niveles del enlace de datos.

Considerando los problemas de apertura y fuentes abiertas identificados en las secciones 3.1. y 3.2., los aportes provistos por el modelo de datos y los mapeos construidos en esta investigación (Secciones 5.2., 5.3. y 5.4.) contribuyen en términos de:

- Disponibilidad, al aportar una descripción de REAs más articulada con su dominio de conocimiento y con dimensiones de calidad de datos, descripción publicada de forma estructurada en formatos no propietario y bajo una licencia libre. Al plantear un conjunto de propiedades que incluye elementos de calidad de datos, aporta en la mejora de los procesos de identificación, consulta y recuperación de recursos. En otras palabras, los REA adquieren una mejor accesibilidad al proveer información más específica que permite refinar los criterios de búsqueda y recuperación (Secciones 6.2., 7.1. y 7.2.).
- Reutilización, al aportar un mapeo al modelo de datos RDF, proporcionando de esta forma la posibilidad de enriquecer sus metadatos a través del enlace de datos semánticos provenientes de diferentes grafos de conocimiento. Por otro lado, la mejora en la visibilidad de los recursos producidos se convierte en un aporte valioso, al poder ser enlazados y convertirse en complemento semántico de otros recursos (Secciones 7.1. y 7.2.).

En términos generales, como lo plantean Pereira *et al.* (2018): "*Si bien el contenido de los recursos educativos abiertos es, por definición, accesible y reutilizable, la publicación de sus metadatos datos enlazados, puede hacer que el contenido de diferentes repositorios sea más visible, accesible y conectable*". Por lo tanto, la articulación de la calidad de datos y el enriquecimiento de datos con la descripción de REAs, hace de esta investigación un aporte significativo al proceso de apertura y enriquecimiento que viene adelantándose en el dominio de conocimiento de los recursos educativos.

9.3. Recomendaciones.

En esta sección presentamos a modo resumen algunas de las recomendaciones por áreas de interés que se han recopilado del proceso de diseño e implementación del modelo de datos para describir REA, expuestos en los capítulos 3, 5 y 6.

9.3.1. Asociadas a la apertura de los REAs.

Aunque el proceso de apertura de datos se ha venido consolidando, como lo muestra el Open Data Barometer (WWWF, 2019), la sección 3.1 muestra que aún existen muchas barreras y problemáticas en cuanto a la apertura de datos. Problemáticas que afloran aún más para los REA, en contextos como el colombiano, por ejemplo, donde iniciativas de apertura de recursos educativos han fracasado. Con relación a este tema, a continuación se destacan las siguientes recomendaciones:

- *Establecimiento de políticas relacionadas con la apertura de los REA*. Con el fin de compartir, enriquecer y reutilizar los REA, se debe contar con un conjunto de políticas tanto nacionales como institucionales que definan clara y detalladamente los procesos de apertura, derechos de autor y licenciamiento abierto para los mismos. De igual forma, se definan un conjunto de estrategias, beneficios o exenciones que motiven a las instituciones, sobre todo las privadas, a abrir y compartir sus recursos, dado que la construcción de recursos es una actividad que demanda diferentes tipos de insumos. Por otro lado, se deben construir y proveer repositorios, con infraestructuras estables, que vayan más allá de referencias a recursos publicados en repositorios propietarios.
- *Fortalecimiento del cambio cultural*. La cultura de compartir no es algo muy fortalecido en diferentes comunidades, sobre todo cuando no se observa la retribución a un corto plazo. Por ende, se hace necesario un trabajo fuerte relacionado con la apertura de recursos educativos no solo a nivel intrainstitucional (muchas veces no conocemos que hay en nuestra propia casa), sino a nivel interinstitucional. Trabajo que debe estar respaldado por un conjunto de políticas e incentivos claros y alcanzables, acorde a la respectiva realidad nacional. Adicionalmente, se debe masificar el conocimiento del licenciamiento abierto, y de los beneficios y derechos que de ellos se derivan, con el fin de abordar barreras como la percepción de pérdida de derechos por parte del autor.
- *Reconocimiento a la labor de producción de los REA*. El diseño y producción de REA debe ser una labor reconocida claramente a nivel institucional. Para tal fin, se debe proveer de los insumos tangibles e intangibles necesarios para su implementación. Por otro lado, la labor de diseño e implementación de REA debe ser reconocida como parte de la labor docente, otorgándose los espacios y tiempos requeridos para tal fin, y no dar por entendido que su asignación de horas lectivas conlleva intrínsecamente la generación de estos recursos. Esta última situación es la que ha conllevado a que muchos docentes construyan y usen recursos educativos en sus clases, pero éstos no sean compartidos ni reutilizados.
- *Observar los requerimientos básicos para llevar a cabo la apertura de datos*. Para llevar a cabo esta labor, se deben considerar los principios básicos de apertura estipulados en la Carta Internacional de Datos Abiertos. Adicionalmente, y con la proyección de enriquecer los

datos, se deben considerar los principios de datos enlazados, con el fin de proveer una plataforma sólida para llevar a cabo esta labor.

9.3.2. Asociadas a la descripción de propiedades de los REA.

El contar con la información suficiente y accesible, publicada de forma abierta, sobre las propiedades pertinentes de los REA aumenta su nivel de disponibilidad y reutilización. Por ende se ha hace necesario contar con los datasets publicados en los repositorios, que permitan una clara y detallada descripción de los recursos. Con relación a este tema, a continuación se destacan las siguientes recomendaciones:

- *Apertura de los datasets que describen los REA*. Además de licenciar los recursos de forma abierta, es necesario proveer de forma abierta los conjuntos de propiedades que describen estos recursos. Ello con el fin de poder llevar a cabo procesos de búsqueda, identificación y recuperación de recursos acorde a las necesidades de los usuarios.
- *Descripción de propiedades.* Acorde a los principios de apertura, los REA deben estar disponibles para cualquier tipo de usuario, por ende sus descripciones deben desprenderse de requerimientos particulares. Por el contrario, estas descripciones deben proveer la suficiente información que permita a los usuarios poder realizar búsquedas más refinadas y recuperar recursos que se adecuen de mejor forma a sus necesidades.
- *Interoperabilidad de recursos.* Dado que los recursos distribuidos son susceptibles de ser integrados: a) los repositorios deben proveer servicios que permitan su explotación, b) la descripción de los Rea debe buscar proveer la suficiente información para su recuperación, acorde a requerimientos como los expuestos por la Unesco, c) se deben proveer interfaces de consulta amigables que provean servicios para la explotación de estos recursos distribuidos y heterogéneos, haciendo uso de metadatos mediadores y transformadores.

9.3.3. Asociadas a los estándares de metadatos.

Para asegurar una mejor interoperabilidad de la descripción de los recursos, la taxonomía de propiedades que configuran la descripción del REA, debe estar mapeada a estándares y esquemas ampliamente reconocidos. Este mapeo puede involucrar más de un estándar, lo cual involucra la construcción de un perfil de aplicación. Para la selección de un perfil de aplicación se debe considerar muy seriamente el alcance del mismo, pues una gran mayoría de estos están diseñados para las necesidades particulares de una organización con relación a lo que esperan de la descripción de sus recursos. Esta limitada descripción dificulta la interoperabilidad con otros repositorios, procesos de recolección de datos o alineamiento con propuestas nacionales o internacionales de descripción de recursos educativos.

9.3.4. Asociadas a la calidad de datos.

La calidad de la información provista por una instancia de datos depende de múltiples factores, dentro de los cuales se cuenta la abstracción de propiedades realizada para la construcción de los modelos de datos que la soportan. Con relación a este tema, a continuación se destacan las siguientes recomendaciones:

- *Identificar un conjunto de dimensiones y criterios de calidad a atender*. Al momento de abordar la apertura y enriquecimiento de datos, se debe ser consciente de la existencia de una serie de dimensiones y criterios de calidad, tales como la disponibilidad, la completitud, la confianza, etc.; presentes en diferentes etapas del ciclo de vida de apertura y enriquecimiento de datos. El abordar estas dimensiones desde etapas tempranas del proceso de apertura, proveer elementos para mejorar la calidad de las instancias de datos, y por ende, aportar en la mejora de los procesos: de búsqueda y recuperación de REA, de interoperabilidad con otros repositorios y modelos de datos, generación de confianza en los datos por parte de los usuarios.

- *Asegurar los insumos suficientes para observar la calidad de los datos*. Para llevar a cabo procesos de aseguramiento de calidad de datos, se deben proveer diferentes insumos a lo largo del ciclo de vida de apertura y enriquecimiento de datos, como por ejemplo proveer una buena curva de aprendizaje de las personas participantes del proceso, hacer uso de ontologías de domino reconocidas, tener un conocimiento amplio del dominio trabajado, entre otros factores que contribuyen a la inclusión de dimensiones y criterios de calidad, y por ende, unas mejores instancias de datos.

9.3.5. Asociadas al enriquecimiento semántico de los metadatos.

El enriquecimiento semántico de los metadatos cobra cada vez mayor importancia, al proporcionar un contexto más holístico del recurso, además de conectarlo con otros elementos, que pueden proporcionar más herramientas, a los usuarios que se encuentran desarrollando procesos de aprendizaje. Con relación a este tema, a continuación se destacan las siguientes recomendaciones:

- *Asegurar los insumos necesarios para el proceso de enriquecimiento*. Para llevar a cabo esta labor, se debe proveer de una serie de insumos para llevar a cabo esta tarea. Se deben considera aspectos como la formación del personal que realizará la labor, la infraestructura requerida para soportar este tipo de servicios, las transformaciones que se deben realizar al modelo de datos RDF, el uso de URIs para identificar los recursos, etc. Aunque para muchos manager o directivos, este alistamiento puede generar la idea de un valor monetario mayor, se deben considerar aspectos tales como: a) Existen múltiples herramientas libres que colaboran con este tipo de procesos. Esto disminuirá el esfuerzo para publicar los datos así como los costos incurridos; b) El valor agregado ofrecido al usuario puede superar la inversión realizada en llevar a cabo el proceso de enriquecimiento de datos.
- *Identificar los metadatos susceptibles de enriquecer semánticamente*. Para llevar a cabo el enriquecimiento de metadatos, se debe realizar un estudio concienzudo de aspectos como: a) que metadatos son susceptibles de enriquecer, b) Que fuentes de datos abiertos se pueden usar para enriquecer estos metadatos. Para tal fin, se puede hacer uso de los grafos de conocimiento publicados en la Web, de las fuentes abiertas que proveen diferentes tipos de datos, o de los repositorios que ofrecen este tipo de servicio, por ejemplo.
- *Proveer los servicios para la explotación de los datos enlazados.* Para tal fin se debe proveer de una infraestructura que provea un despliegue, mapeo y explotación de datos abiertos y enlazados. En el mercado existen diferentes herramientas, libres y propietarias, que se ajustan a diferentes requerimientos organizacionales, además de ofrecer diferentes estrategias de soporte. De igual forma, se pueden encontrar en la Web, diferentes estudios comparativos acerca de las bondades y la compatibilidad de cada una de estas herramientas.

9.3.6. Asociadas al uso de metamodelos.

A nivel organizacional, se pueden encontrar diferentes tipos de recursos trabajados, los cuales mantienen un conjunto de propiedades y restricciones en común. El uso de metamodelos para la gestión de los diferentes modelos de recursos, provee un mayor nivel de flexibilidad al momento de gestionar reglas o propiedades sobre sus descripciones. Por otro lado, proveen un menor nivel de inmovilidad, rigidez y viscosidad al momento de modificar reglas o propiedades de los recursos. Para tal fin, se debe realizar una caracterización detallada de sus propiedades y de los diferentes componentes que se conjugaran en la gestión de los recursos, con el fin de establecer una clara definición de responsabilidades de cada capa del metamodelo, y poder generar instancias y modelos de datos que respondan y sean conformes a las condiciones y restricciones de los niveles superiores.

9.4. Aportaciones Originales.

Como aportación central, esta investigación ofrece una alternativa para describir y enriquecer recursos educativos, basado en el uso de metamodelos y dimensiones de calidad de datos enlazados. Su propósito se orientó en proporcionar una taxonomía de propiedades más completa y enriquecida, observando elementos de calidad de datos, que ofreciera los correspondientes mapeos para aplicación y para dataset, de dichas propiedades. Esto proporcionará una mejora en la calidad y confianza de las instancias de datos generadas, además de proporcionar mejoras en la disponibilidad y reutilización de los recurso educativos abiertos publicados en la Web.

Entre las aportaciones complementarias que han surgido a partir de esta contribución principal, se pueden destacar las siguientes:

- Un análisis del estado de la cuestión de apertura y enriquecimiento de recursos educativos, donde se identificaron problemáticas y retos a abordar en el proceso de apertura y enriquecimiento de datos.
- El análisis, construcción y desarrollo de esquemas, reglas y restricciones, estructurados bajo los niveles de un metamodelo, utilizados para el modelado de recursos educativos abiertos.
- La articulación de las dimensiones de calidad de datos (de confianza y contextual), y del enriquecimiento semántico a la descripción de propiedades de los REA. Esta articulación generó una taxonomía de propiedades para describir y enriquecer REA.
- El diseño e implementación de dos tipos de mapeos de propiedades de los REA: a) Mapeo para aplicación, basado en Dublin Core, y b) Mapeo para Dataset, haciendo uso del modelo de datos RDF.
- La ejecución de un despliegue tecnológico que permitió la implementación del modelo de datos para describir los REA, y llevar a cabo su consecuente validación.
- La definición de un conjunto de recomendaciones para el proceso de apertura y enriquecimiento de propiedades que describen los REA.

9.5. Trabajos Futuros.

Esta tesis es el punto de partida para la construcción de un metamodelo para describir recursos más completos, que integre otros tipos de recursos como los manejados por proyectos como el Repositorio Institucional de la Universidad Distrital - RIUD, recursos de investigación, administrativos, entre otros. Situación que abre la puerta a múltiples trabajos futuros en áreas como por ejemplo:

- El fortalecimiento de los constructos teóricos que soportan el metamodelo.
- La parametrización de reglas y restricciones que permitan la inclusión de nuevos recursos.
- La inclusión de nuevos tipos de recursos, que conlleva por ende la generación de sus respectivos modelos de datos, conformes a las directrices del metamodelo.
- La modificación y parametrización de las aplicaciones que gestionan los modelos de datos que describen estos nuevos recursos.
- La articulación de estos nuevos modelos a la estructura del metamodelo.
- La generación de los nuevos esquemas de mapeo, a través dela implementación de procesos adicionales de Metadata Crosswalk, El despliegue y validación de los nuevos modelos.
- La implementación del modelo propuesto al sistema de gestión de recursos – RIUD.
- El diseño y fortalecimiento de políticas de apertura de datos, no solamente a nivel institucional de la Universidad Distrital, sino a diferentes niveles del sector público y privado.

- La articulación de los elementos de metadatos que describen los REA a procesos de acreditación de programas, como indicadores que aportan información valiosa al citado proceso de aseguramiento de calidad de programas académicos.

Este tipo de ampliaciones y generación de modelos fue verificado a través de las pruebas de concepto realizadas en esta investigación, lo que permite evidenciar la posibilidad de trabajar de forma vertical y horizontal sobre el aprovechamiento de las bondades ofrecidas por los metamodelos. De igual forma, se plantea como trabajo futuro la aplicación del enriquecimiento de datos en otros tipos de recursos que se articulen con el metamodelo propuesto, lo que permitiría enriquecer semánticamente su contenido.

9.6. Publicaciones derivadas de la investigación

A lo largo del desarrollo de esta investigación se han realizado una serie de publicaciones en diversos ámbitos de divulgación. Para empezar, se consideró importante señalar que el desarrollo de esta investigación generó un conjunto de hitos clasificados en ponencias, artículos, estancias de pasantía, proyecto de investigación financiado, proyectos de grado de maestría y de pregrado, y la ejecución de un curso taller desarrollado a través de un webinar (Figura 75).

Para el caso de los artículos y las ponencias, los hitos construidos aportaron en aspectos como:

- El desarrollo de la investigación, al proveer elementos que además de fortalecer la identificación de problemáticas en el dominio de conocimiento, proporcionaron argumentos más sólidos para mejorar la descripción y enriquecimiento de los REA, soportado sobre dimensiones de calidad de datos. De igual forma, y dados hallazgos presentados en estos hitos y en el análisis del dominio de conocimiento, se expusieron las bases acerca de la necesidad del diseño de un metamodelo que permitiera además de gestionar la descripción y enriquecimiento de los REA, contar con una estructura que permitiera modelar y gestionar otros tipos de recursos.
- La construcción y actualización de referentes para futuras investigaciones, acerca aspectos como el estado de la Web Semántica, la apertura y el enriquecimiento de datos. Referentes que además de sustentar el planteamiento expresado en esta investigación, también proporcionan sustento y argumentos para muchas oportunidades de investigación y mejora en temáticas como la apertura de datos, la apertura de REA, mejoramiento del modelado y descripción de diferentes tipos de recursos, el enriquecimiento de otros tipos de recursos, la mejora en la aplicación de la calidad de los datos, aplicación y uso de plataformas y herramientas para la apertura y enriquecimiento de datos.
- La aplicación particular de conceptos o estrategias sobre el dominio de conocimiento. Tal es el caso del consumo de servicios en plataformas de datos abiertos, la evaluación de la calidad de los datasets existentes a través de técnicas como lógica difusa, el diseño de herramientas que permitieran evaluar situaciones de calidad de datos tales como la actualización de los datasets, los formatos de publicación, la procedencia y los enlaces de que son sujetos. Este tipo de aplicaciones aportan elementos de trabajo para futuras investigaciones, además de poder evaluar aspectos de datasets existentes.

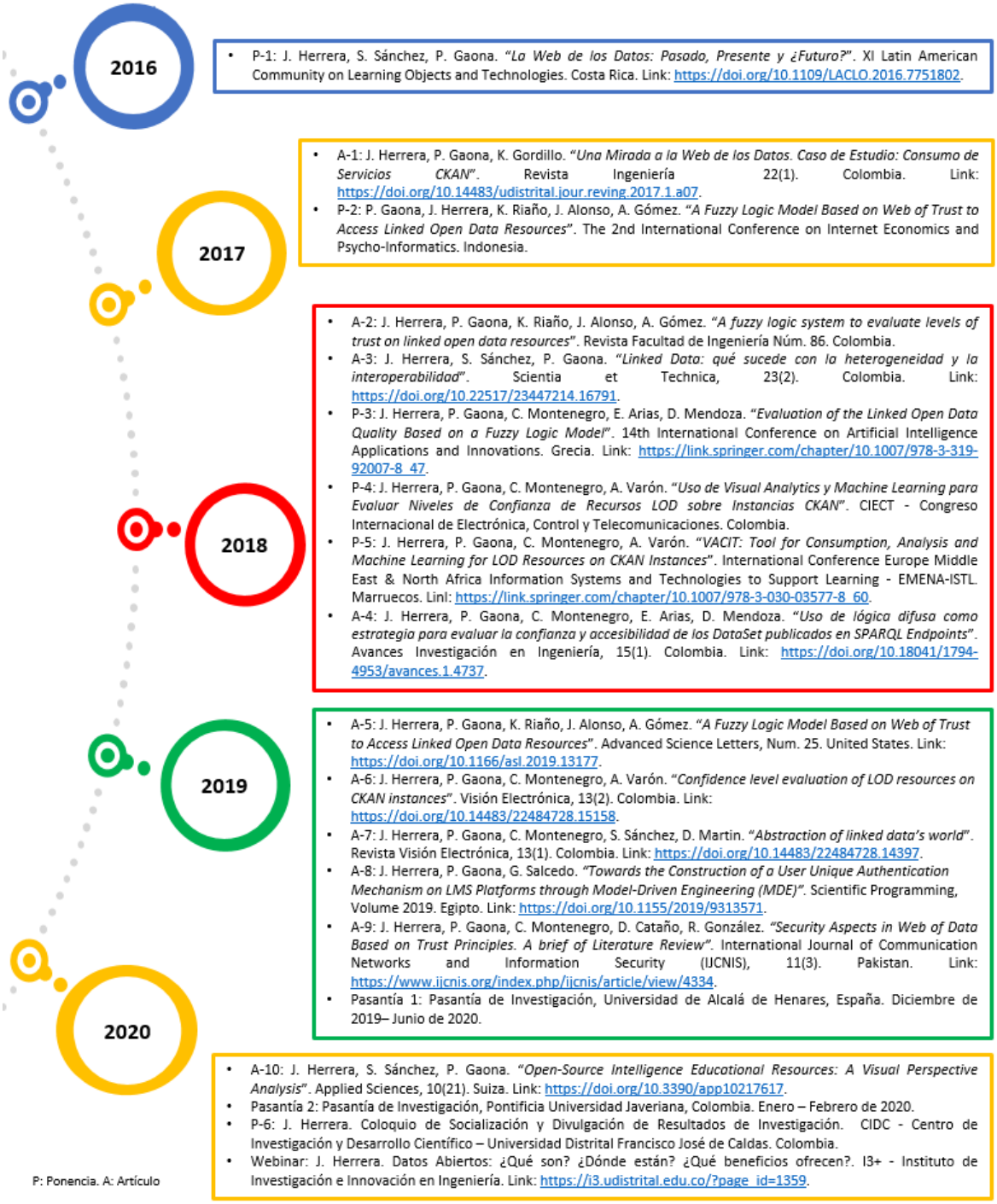


Figura 75. Productos generados en la Investigación.

- El desarrollo y fortalecimiento de los procesos de apertura y enriquecimiento de datos en Colombia, donde estos campos aun presentan una serie de barreras para su implementación.

Adicionalmente, los artículos y ponencias han aportado al fortalecimiento de las líneas de investigación de los grupos de investigación de la Universidad Distrital, además de contribuir en la mejora de indicadores de desarrollo y visibilidad de la investigación realizada institucionalmente.

En cuanto a las estancias de pasantía, realizadas en la Pontificia Universidad Javeriana (Colombia) y en la Universidad de Alcalá de Henares (España), además de contribuir en el fortalecimiento de los vínculos de colaboración interinstitucional, aportaron en el intercambio de conocimientos y experiencias, en el reconocimiento de realidades, en el fortalecimiento de esta investigación a través de la construcción de una alternativa para mejorar la descripción y enriquecimiento de los REA como factores fundamentales en la universalización y democratización del conocimiento.

Con relación al coloquio de socialización y el webinar, estas estrategias aportaron en la incursión de otros actores como estudiantes, docentes y administrativos en ambientes como los datos abiertos y la posibilidad no solamente de mejorar la descripción de recursos provista por proyectos como el RIUD, sino su aplicación a otras áreas donde se manejan otros tipos de recursos. Mejoras y orientaciones que fortalecieron el objeto de esta investigación: el trabajo con el concepto de metamodelo, permitiendo de esta forma poder aportar en el modelamiento y enriquecimiento de diferentes recursos trabajados en instituciones como la Universidad Distrital, como caso de estudio que puede ser extrapolado a otras instancias y dominios de conocimiento.

En resumen, como se muestra en la Figura 76, cada uno los hitos construidos (Ponencias y Artículos) además de sustentar cada uno de los objetivos perseguidos en esta investigación, evidenciando tanto la problemática identificada como la aplicación de la propuesta diseñada para atacar la mencionada problemática; aportaron diferentes elementos, reconocidos por la comunidad académica e investigativa, que además de proveer bases para esta investigación, proporcionan elementos para mejorar la apertura, disponibilidad, enriquecimiento y reutilización de los recursos educativos. Aspectos tales que se han vuelto cada vez más importantes en los procesos de compartir el conocimiento, en situaciones como las vividas en la actualidad gracias a las diferentes situaciones abordadas por la pandemia de la Covid-19.

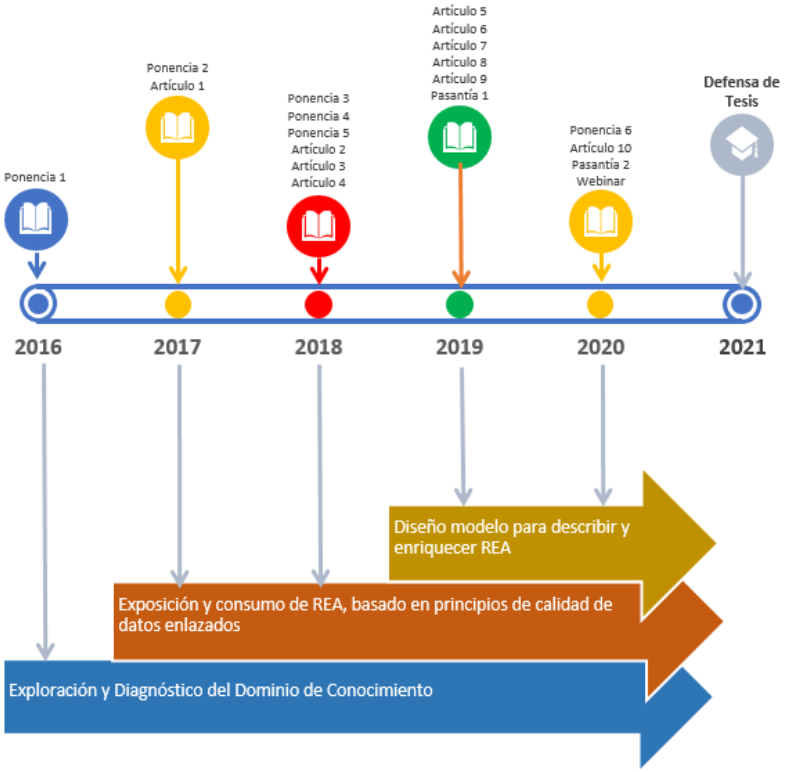


Figura 76. Aporte de los productos a los objetivos.

Con relación a la articulación de la investigación a nivel de pregrado y maestría, esta investigación aportó en el enlace y operacionalización de la apertura, el enriquecimiento de datos y los REA en temáticas específicas (Figura 77), permitiendo validar el interés y aplicación de la apertura y enriquecimiento de recursos en diferentes dominios de conocimiento. Adicionalmente, estas aplicaciones proporcionaron elementos y experiencias que fortalecieron el desarrollo de esta investigación, y que aportaron iniciativas de valor para la mejora de la calidad de vida de las personas como el caso de los ambientes inmersivos y los sistemas de monitoreo, basando su quehacer sobre la apertura de datos.

Por último, esta investigación también aporto en la búsqueda de la formalización del modelo de datos para describir y enriquecer REA, a través de un proyecto de investigación financiado por parte del Centro de investigación y Desarrollo Científico - CIDC, perteneciente a la Universidad Distrital Francisco José de Caldas. Bajo este proyecto, además de proponer una estructura de metadatos REA mejorada y enriquecida, se orientó en apoyar procesos como los adelantados por el RIUD, la Oficina Asesora de Sistemas y PlanesTic, orientados a fortalecer la formación virtual, la descripción de recursos y la apertura de datos.

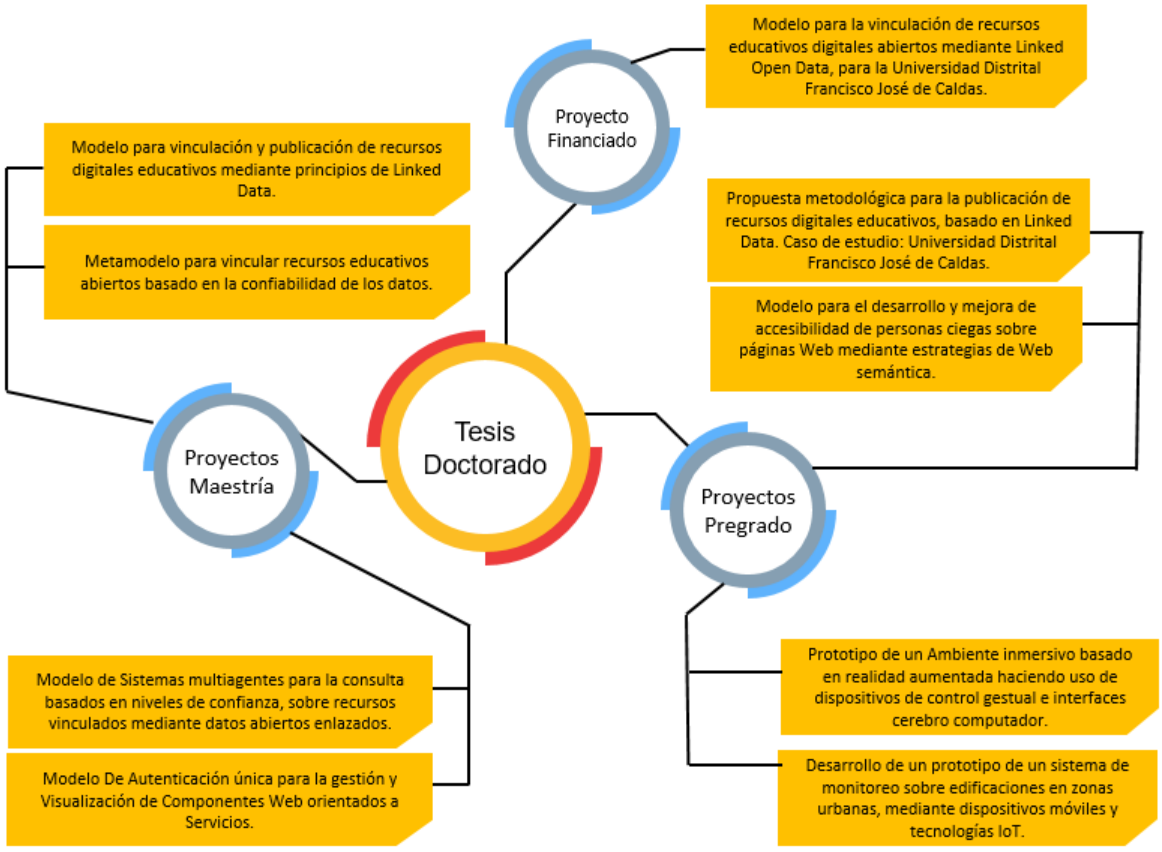


Figura 77. Aporte desde los proyectos de investigación.

A modo de resumen, la Tabla 46 categoriza los aportes y publicaciones derivados de esta investigación.

Tabla 46. Categorización de los aportes de la Investigación.

Tipo	Cantidad
Artículo Nacional Categoría A	0
Artículo Nacional Categoría B	2
Artículo Nacional Categoría C	3
Artículo Internacional Cuartil Q1	2

Tipo	Cantidad
Artículo Internacional Cuartil Q2	0
Artículo Internacional Cuartil Q3	4
Artículo Internacional Cuartil Q4	0
Ponencia Nacional	3
Ponencia Internacional	4
Proyectos Financiados	1
Proyectos de Maestría	4
Proyectos de Pregrado	4
Estancia Pasantía Nacional	1
Estancia Pasantía Internacional	1

REFERENCIAS BIBLIOGRÁFICAS

Abella, A. (2011). *Reutilización de información Pública y Privada en España*. Rooter. https://gobiernoabierto.navarra.es/sites/default/files/opendata/reutilizacion_informacion_publica_privada_od_rooter.pdf.

Abella, A., Ortiz-de-Urbina-Criado, M. y De-Pablos-Heredero, C. (2018). Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona. *El Profesional de la Información*, 27(2), pp. 375-382. https://recyt.fecyt.es/index.php/EPI/article/view/epi.2018.mar.16.

Al Fayez, Q. y Joy, M. (2014). A Framework for Linking Educational Medical Objects: Connecting Web2.0 and Traditional Education. *Web Information Systems Engineering – WISE 2014*. Lecture Notes in Computer Science, Volume 8787. https://doi.org/10.1007/978-3-319-11746-1_11.

Alyahibi, S. y Alsharif, Z. (2019). Using Linked Open Data (LOD) in the Educational Field. *International Journal of Science and Research (IJSR)*, 8(11), pp. 1064 – 1070. https://www.ijsr.net/search_index_results_paperid.php?id=ART20202727.

Antoniou, G. y Van Harmelen, F. (2008). *A Semantic Web Primer*. Information Systems. Massachusetts Institute of Technology. ISBN 978-0-262-01242-3. https://mitpress.mit.edu/books/semantic-web-primer-second-edition.

Apache Jena. (2020). *Apache Jena Fuseki*. https://jena.apache.org/documentation/fuseki2/.

Aranda, G. y Ruiz, F. (2005). Clasificación y ejemplos del uso de ontologías en Ingeniería del Software. *Workshop en Ingeniería del Software y Bases de Datos WISBD 2005.* https://pdfs.semanticscholar.org/88fa/9f42a63df74ac2dffabfe9da3ff70828a7a6.pdf.

Ariadne. (2017). *Ariadne Project EU*. https://www.ariadne-eu.org/.

Armas, M. (2012). *Model Driven Architecture (MDA)*. https://sg.com.mx/content/view/861.

ASF - The Apache Software Foundation. (2020). *Apache Tomcat*. http://tomcat.apache.org/.

Atenas, J. y Havemann, L. (2013). Quality Assurance in the Open: An Evaluation of OER Repositories. *InnoQual - The International Journal for Innovation and Quality in Learning*. https://core.ac.uk/download/pdf/141219719.pdf.

Atenas, J. (2014). *Estudio de calidad de los repositorios de recursos educativos abiertos en el marco de la educación universitaria*. Doctorado en Educación y Sociedad. Universidad de Barcelona. https://www.tesisenred.net/handle/10803/286031.

Auer, S., Buhmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P., Van Nuffelen, B., Stadler, C., Tramp, S. y Williams, H. (2012). Managing the Life-Cycle of Linked Data with the LOD2 Stack. En *11th International Semantic Web Conference*. http://iswc2012.semanticweb.org/sites/default/files/76500001.pdf.

Auer, S., Bryl, V. y Tramp, S. (2014). Linked Open Data – Creating Knowledge out of Interlinked Data. Result of the LOD2 Project. Lecture Notes in *Computer Science* (LNCS, volume 8661). Springer Open. https://link.springer.com/book/10.1007/978-3-319-09846-3.

Ávila, E. (2020). Los datos bibliográficos abiertos enlazados y su comportamiento en la recuperación de información. Investigación Bibliotecológica, 34(82), ISSN: 2448-8321, pp. 203-227. http://www.scielo.org.mx/pdf/ib/v34n82/2448-8321-ib-34-82-203.pdf.

Ayllón, J. (2012). *Linked Data en las Bibliotecas: El cao de la Biblioteca Nacional de España*. Trabajo de Fin de Master. Facultad de Comunicación y Documentación. Universidad de Granada. https://masteres.ugr.es/mic/pages/checklist1516/2012_tfmayllonmillanjuanmanuel/!.

Baliyan, N. y Kumar, S. (2018). SWAQ: a Semantic Web Application Quality Evaluation Framework. *Journal of Experimental & Theoretical Artificial Intelligence*, 30:6, 953-972. https://doi.org/10.1080/0952813X.2018.1509893.

Bankier, J. y Gleason, K. (2014). *Institutional repository software comparison*. UNESDOC - Biblioteca Digital. Unesco - Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000227115.

Baskauf, S. (2013). *Dublin Core in RDF*. TDWG RDF/OWL Best Practices Task Group. https://code.google.com/archive/p/tdwg-rdf/wikis/DublinCore.wiki.
Bassili, V., Caldiera, G. y Robach, H. (1994). *The goal question metric approach*. Encyclopedia of software engineering. https://www.cs.umd.edu/users/mvz/handouts/gqm.pdf.
Bauer, F. y Kaltenböck, M. (2012). *Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers*. Edition mono/monochrom, Vienna, Austria. ISBN: 978-3-902796-05-9. https://www.reeep.org/linked-open-data-essentials.
BBC. (2020). *The British Broadcasting Corporation*. https://www.bbc.com/.
Behkamal, B. (2014). Metrics-Driven Framework for LOD Quality Assessment. *The Semantic Web: Trends and Challenges. ESWC 2014*. Lecture Notes in Computer Science, vol 8465. https://doi.org/10.1007/978-3-319-07443-6_54.
Behkamal, B., Kahani, M., Bagheri, E. y Jeremic, Z. (2014). A Metrics-Driven Approach for Quality Assessment of Linked Open Data. *Journal of Theoretical and Applied Electronic Commerce Research,* 9(2), pp. 64-79. http://www.jtaer.com/may2014/Behkamal_Kahani_Bagheri_Jeremic_p5.pdf.
Bellini, P. y Nesi, P. (2018). Performance assessment of RDF graph databases for smart city services. *Journal of Visual Languages & Computing*, 45, pp. 24-38, ISSN 1045-926X. https://doi.org/10.1016/j.jvlc.2018.03.002.
Belloir, N., Buisson, J. y Bartheye, O. (2019). *Metamodeling NATO Operation Orders: a proofof-concept to deal with digitalization of the battlefield*. 14th Annual Conference System of Systems Engineering (SoSE). pp. 260-265. https://hal.archives-ouvertes.fr/hal-02181689/document.
Ben Ellefi, M. (2017). *Profile-based Dataset Recommendation for RDF Data Linking*. Tesis de Doctorado. Universit´e de Montpellier. https://www.researchgate.net/publication/331632808_Profile-based_Dataset_Recommendation_for_RDF_Data_Linking.
Berlanga, A., López, C., Morales, E. y García, F. (2005). *Consideraciones para Reforzar el Valor de los Metadatos en los Objetos de Aprendizaje*. Departamento de Informática y Automática, Universidad de Salamanca, España. http://avellano.fis.usal.es/~aberlanga/files/Pubs/SPDECE05_berlanga_lopezVF.pdf.
Berners-Lee, T. (2009). *The Next Web*. TED2009. TED.com. https://www.ted.com/talks/tim_berners_lee_the_next_web.
Berners-Lee, T. (2011). *Linked Data*. World Wide Web Consortium. http://www.w3.org/DesignIssues/LinkedData.html.
Bizer, C. (2007). *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*. PhD thesis, Freie Universität Berlin. http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/DisertationChrisBizer.pdf.
Bizer, C. (2009). *The Emerging Web of Linked Data*. IEEE Intelligent Systems. https://doi.org/10.1109/MIS.2009.102.
Bio2RDF. (2020). *Bio2RDF Project - Linked Data for the Life Sciences*. https://bio2rdf.org/.
Bl. (2020). La Biblioteca Británica. https://www.bl.uk/.
Blaney, J. (2018). Introducción a los Datos abiertos enlazados. The Programming Historian. https://doi.org/10.46430/phes0038.
BNB. (2020). The British National Bibliography. https://bnb.data.bl.uk/.
BNE. (2020). *La Biblioteca Nacional de España*. http://www.bne.es/es/Inicio/index.html.
BnF. (2020). *La Biblioteca Nacional de Francia*. https://www.bnf.fr/es.
Bodas, D., Fernández, P. y Soltero Domingo, F. (2004). Estándares en Tecnologías Educativas. *Enlaces - Revista del CES Felipe II*, Numero 1. ISSN 1695-8543. https://dialnet.unirioja.es/servlet/articulo?codigo=3692664.
Bodero, E., De Giusti, M., Radicelli, C. y Villacrés, E. (2019). Analyses of institutional Open Access digital repositories in Ecuador. *Revista Espacios*, 4(32). ISSN 0798-1015. http://www.revistaespacios.com/a19v40n32/19403215.html.

Bodoff, D., Hung, P. y Ben-Menachem, M. (2005). Web Metadata Standards: Observations and Prescriptions. *IEEE Software*, 22(1). https://doi.org/10.1109/MS.2005.25.

Bonduel, M. (2019). *RDF triplestores and SPARQL endpoints*. LDAC summer school 2019, Portugal. http://linkedbuildingdata.net/ldac2019/summerschool/files/07_Bonduel_triplestores_SPARQL_endpoints.pdf.

Bonduel, M., Dandan, R., Leone, V., Futia, G., Rosales-Mndez, H., Ozdowska, S., Palma, G. y Gangemi, A. (2019). *A Framework for LOD Validity using Data Quality Dimensions*. Chapter 4. 35-44. Linked Open Data Validity. A Technical Report. ISWS - International Semantic Web Research Summer School. https://arxiv.org/pdf/1903.12554.pdf.

Bozkurt, A., Jung, I., Xiao, J., Vladimirschi, V., Schuwer, R., Egorov, G., Lambert, S., Al-Freih, M., Pete, J., Olcott, Jr., D., Rodes, V., Aranciaga, I., Bali, M., Alvarez, A. J., Roberts, J., Pazurek, A., Raffaghelli, J. E., Panagiotou, N., de Coëtlogon, P., Shahadu, S., Brown, M., Asino, T. I., Tumwesige, J., Ramírez Reyes, T., Barrios Ipenza, E., Ossiannilsson, E., Bond, M., Belhamel, K., Irvine, V., Sharma, R. C., Adam, T., Janssen, B., Sklyarova, T., Olcott, N., Ambrosino, A., Lazou, C., Mocquet, B., Mano, M., & Paskevicius, M. (2020). A global outlook to the interruption of education due to COVID-19 pandemic: Navigating in a time of uncertainty and crisis. *Asian Journal of Distance Education*, 15(1), pp. 1-126. http://asianjde.org/ojs/index.php/AsianJDE/article/view/462.

Briand, L., Morasca, S. y Basili, V. (1996). Property-based software engineering measurement. *Software Engineering, IEEE Transactions on*, 22, pp. 68-86. https://ieeexplore.ieee.org/document/481535.

Bustos, V., Broisin, J., Munoz-Arteaga, J., y Guzman, J. (2016). Extension of IEEE LOM Standard for Describing Educational Interactive Application: An Accessibility approach. *IEEE Latin America Transactions*, 14 (8). pp. 3847-3855. ISSN 1548-0992. https://hal.archives-ouvertes.fr/hal-01445133/document.

Butcher, N., Asha, K. y Stamenka, U. (2015). *Guía Básica de Recursos Educativos Abiertos*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura – Unesco. https://unesdoc.unesco.org/ark:/48223/pf0000232986.

Cabrera, M. (2016). *Conceptualización y generación de rdf de los recursos bibliográficos de la biblioteca "Benjamín Carrión" de UTPL*. Tesis de Maestría. Máster universitario en inteligencia artificial, Escuela técnica superior de ingenieros informáticos, Universidad Politécnica de Madrid. http://oa.upm.es/40329/.

Calderón, M. (2006). *Calificadores del estándar de metadatos Dublin Core (Dublin Core Qualifiers)*. Centro de investigación de la Web. Universidad de Chile. http://www.ciw.cl/proyectos/calificadoresDC.html.

California State University Center for Distributed Learning - CSU-CDL. (2020). *Merlot*. https://www.merlot.org/merlot/.

Canabal, M., Sarasa, A. y Sacristán, J. (2008). *LOM-ES: Un perfil de aplicación de LOM.* Universidad Pontificia de Salamanca, España. http://www.web.upsa.es/spdece08/contribuciones/110_Spedece2008-lom-es.pdf.

Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. American Library Association, Chicago. ISBN: 978-0-8389-0847-1.

Cappiello, C., Di Noia, T., Marcu, B. y Matera, M. (2016). A Quality Model for Linked Data Exploration. *International Conference on Web Engineering ICWE 2016*. https://link.springer.com/chapter/10.1007/978-3-319-38791-8_25.

Carcaño, F. *¿Qué es OSINT y qué son fuentes abiertas?.* FCD Intelligence. 2018. https://www.fcd-intelligence.com/2018/09/que-es-osint-y-que-son-fuentes-abiertas/.

Castro, I. (2008). *Muestreo Aleatorio Simple*. http://matematicas.unex.es/~inmatorres/teaching/muestreo/assets/Cap_3.pdf.

Castro, M., Castillo, E., Lillo, P., Quiroz, A. y Riquelme, P. (2009). *Guía para la creación de metadatos usando Dublin Core*. Comité de Metadatos de la Biblioteca Nacional de Chile. https://rdu-demo.unc.edu.ar/handle/123456789/800.

CECC - Coordinación Educativa y Cultural Centroamericana. (2020). *Coordinación Educativa y Cultural Centroamericana*. http://ceccsica.info/.
Cepal - Comisión Económica para América Latina y el Caribe. (2019). *Qué son los Metadatos?*. Gestión de Datos de Investigación. https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/metadatos.
Cepal - Comisión Económica para América Latina y el Caribe. (2020). *Que son los Datos Abiertos?*. https://biblioguias.cepal.org/EstadoAbierto/datospublicos.
Ceriotto, P. y Testa, P. (2011). *LUCIS - Directrices y recomendaciones para la aplicación de metadatos en repositorios institucionales*. Sistema integrado de documentación. Universidad Nacional de Cuyo. https://bdigital.uncu.edu.ar/objetos_digitales/4123/directrices-lucis-bd-2.pdf.
Cetis. (2020). *Metadata for Learning Opportunities (MLO)*. http://www.cetis.org.uk/inloc/MLO.
Ckan. (2020). *CKAN, the world's leading Open Source data portal platform*. https://ckan.org/.
CLARA - Cooperación Latino Americana de Redes Avanzadas. (2020). *Red CLARA*. https://www.redclara.net/index.php/es/.
Clarke, P. (2014). *A framework for the analysis and user-driven evaluation of trust on the Semantic Web*. Master Dissertation, Dublin Institute of Technology. https://arrow.dit.ie/scschcomdis/53/.
CO. (2020). *Postdata - Más Allá de los Datos*. Gobierno de Colombia. https://www.postdata.gov.co/.
Codina, L. y Pedraza, R. (2017). *Taxonomías y ontologías: qué son y cómo se aplican en medios de comunicación*. https://www.lluiscodina.com/taxonomias-ontologias/.
COE - Council of Europe. (2019). *Global Education Guidelines*. North-South Centre of the Council of Europe in consultation with the Global Education Network. https://rm.coe.int/prems-089719-global-education-guide-a4/1680973101.
COL - Commonwealth of Learning. (2015). *Understanding Open Educational Resources*. https://www.colvee.org/course/technology-enabled-learning/understanding-open-educational-resources.
Colombia Aprende. (2020). *Colombia Aprende – La red de conocimiento*. http://aprende.colombiaaprende.edu.co/es/contenidoslo.
Cornell - Biblioteca de la Universidad de Cornell. (2003). *Metadatos. Departamento de Preservación y Conservación*. http://preservationtutorial.library.cornell.edu/tutorial-spanish/metadata/metadata-01.html.
Cortés, P. (2019). *El uso de metadatos en webs institucionales*. Máster en Sistemas de información Digital. Universidad de Salamanca. https://gredos.usal.es/bitstream/handle/10366/140477/TFM_SistemasInfoDigital_CortesGarcia_Pedro_SI_80_2018-2019.pdf?sequence=1&isAllowed=y.
Coyle, K. (2009). *Guidelines for Dublin Core™ Application Profiles*. Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/profile-guidelines/.
Crunchbase. (2020). *Semantic Crunch Base*. https://www.crunchbase.com/organization/semantic.
Curriki. (2020). *Curriki*. https://www.curriki.org/.
Chan, L. y Zeng, M. (2006). Metadata Interoperability and Standardization – A Study of Methodology. Achieving Interoperability at the Schema Level. *D-Lib Magazine*, 12(6), ISSN 1082-9873. http://www.dlib.org/dlib/june06/zeng/06zeng.html.
Da Silva, W. (2019). *Softwares para repositorios digitales: Evaluación para la gestión y preservación de los materiales bibliográficos y de los documentos archivísticos de la Biblioteca Nacional de Brasil*. Trabajo de Fin de Master, Universidad de Salamanca. https://gredos.usal.es/bitstream/handle/10366/140478/TFM_SistemasInfoDigital_DaSilva_Wellington_SI_85_2018-2019.pdf?sequence=1&isAllowed=y.
Datasketch. (2020). *Datasketch*. https://www.datasketch.co/.

DB-Engines. (2020). *System Properties Comparison Apache Jena - TDB vs. RDF4J vs. Virtuoso*. Knowledge Base of Relational and NoSQL Database Management Systems. https://db-engines.com/en/system/Apache+Jena+-+TDB%3BRDF4J%3BVirtuoso.

DBpedia. (2020). *DBpedia*. https://wiki.dbpedia.org/.

DCMI. (2020). *DCMI Metadata Terms*. Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Qing Yu, H., Giordano, D., Marenzi, I. y Pereira Nunes, B. (2009). *Interlinking educational Resources and the Web of Data – a Survey of Challenges and Approaches*. https://stefandietze.files.wordpress.com/2009/01/dietze-et-al-linkededucation-survey-final.pdf.

Dietze, S., Yu, H., Giordano, D., Kaldoudi, E., Dovrolis, N., y Taibi, D. (2012). Linked education: interlinking educational resources and the web of data. *The 27th ACM Symposium on Applied Computing (SAC-2012)*, Special Track on Semantic Web and Applications. http://oro.open.ac.uk/31077/1/dietze-2012_sac_linked-education-cam-ready.pdf.

DNB. (2020). *La Biblioteca Nacional de Alemania*. https://www.dnb.de/EN/Home/home_node.html.

Donohue, T. (2018). *DSpace 6.x Documentation*. Lyrasis. https://wiki.lyrasis.org/display/DSDOC6x/DSpace+6.x+Documentation.

Driver – Digital Repository Infraestructure Vision for European Research. (2020). OpenAIRE - Open Access Infrastructure for Research in Europe. http://www.expania.es/driver-digital-repository-infrastructure-vision-for-european-research/.

DrugBank (2018). *DrugBank Database*. http://www.drugbank.ca/.

Drupal. (2020). *Drupal.* https://www.drupal.org/.

EBI - The European Bioinformatics Institute. (2019). *Chemical Entities of Biological Interest (ChEBI)*. http://www.ebi.ac.uk/chebi/.

EdMatrix. (2020). *Table of Standards Organizations*. Directory of Learning Standards. https://www.edmatrix.org/matrix.html.

EDP – European Data Portal. (2019). Open Data Maturity. https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf.

Esri (2019). *Datos Abiertos*. Esri Colombia. https://esri.co/datos-abiertos/.

Europeana Foundation. (2020).*Europeana*. https://www.europeana.eu/es/.

Evans, A., Sammut, P. y Willans, J. (2003). Metamodelling for MDA. *First International Workshop York*. Departamento de Ciencias de la Computación. Universidad de York. https://www.cs.york.ac.uk/metamodel4mda/onlineProceedingsFinal.pdf.

Evangelia, M., Taibi, D., Giordano, D., Dietze, S., Yu, H., Bamidis, P., Bratsas, C. y Woodham, L. (2011). Connecting medical educational resources to the Linked Data cloud: the mEducator RDF Schema, store and API. *Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, 8th Extended Semantic Web Conference (ESWC2011).* http://oro.open.ac.uk/29124/.

Farias-Lóscio, B., Burle, C. y Calegari, N. (2019). *Data on the Web Best Practices. W3C Editor's Draft 2019.* https://w3c.github.io/dwbp/bp.html.

Fayyaz, N., Ullah, I. y Khusro, S. (2018). On the Current State of Linked Open Data: Issues, Challenges, and Future Directions. *International Journal on Semantic Web and Information Systems (IJSWIS),* 14(4), pp. 110-128. https://www.igi-global.com/article/on-the-current-state-of-linked-open-data/210655.

Fermoso García, A., Sánchez Alonso, S. y Sicilia, M. (2008). *Una ontología en OWL para la representación semántica de objetos de aprendizaje*. V Simposio Pluridisciplinar sobre Diseño y Evaluación de Contenidos Educativos Reutilizables. http://www.web.upsa.es/spdece08/contribuciones/176_Fermoso_Sanchez_Sicilia_LOMOWL.pdf.

Fernández, C. (2014). El soporte de ORCID en Dspace 5 (y superiores). Hablando de Dspace. https://www.arvo.es/dspace/tag/xmlui/.

Fernández-Pampillón, A. (2017). The Role of E-Vocabularies in the Description and Retrieval of Digital Educational Resources. *Education Sciences*, 7(33). https://doi.org/10.3390/educsci7010033.

Friesen, N. (2004). *Technical Evaluation Report - 40. The International Learning Object Metadata Survey*. CanCore Initiative. http://www.irrodl.org/index.php/irrodl/article/view/195/277.

Galin, D. (2004). *Software Quality Assurance - From theory to implementation*. Pearson Education ISBN 0-201-70945-7.

Gao, Q. y Houben, G. (2010). A framework for trust establishment and assessment on the web of data. *19th international conference on World Wide Web (WWW '10)*. ACM. http://dx.doi.org/10.1145/1772690.1772822.

García Díaz, V. (2012). *Ingeniería Dirigida por Modelos*. Universidad de Oviedo. https://es.slideshare.net/vicegd/introduccin-a-la-ingeniera-dirigida-por-modelos.

GeoNames. (2020). *GeoNames*. https://www.geonames.org/.

Giménez-García, J., Thakkar, H. y Zimmermann, A. (2016). Assessing trust with PageRank in the Web of Data. *13th International Conference on Semantic Web, ESWC 2016*. Pp. 293-307 Lecture Notes in Computer Science http://dx.doi.org/10.1007/978-3-319-47602-5_45.

Giordano, D., Kavasidis, I., Spampinato, C. y Bamidis, P. (2011). Developing controlled vocabularies for educational resources sharing: A case study. *CEUR workshop proceedings*. https://pdfs.semanticscholar.org/2af7/3195cd549c9f272e7a27c0676aa311b0cbb4.pdf.

Giri, K. (2011). Role of Ontology in Semantic Web. *DESIDOC Journal of Library & Information Technology*, 31(2), pp. 116-120. https://www.researchgate.net/publication/285450608_Role_of_Ontology_in_Semantic_Web.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. http://dx.doi.org/10.1006/knac.1993.1008.

Gutiérrez, M. (2019). El manejo y uso de los metadatos para archivos: análisis comparativo de los principales modelos y estándares. *NSU - no solo usabilidad: revista sobre personas, diseño y tecnología*, Numero 18. ISSN 1886-8592. http://www.nosolousabilidad.com/articulos/manejo_y_uso_de_los_metadatos_para_archivos.htm.

GODI - Global Open Data Index (2019). *Tracking the State of Open Government Data*. Open Knowledge Foundation. https://index.okfn.org/.

Graves, A. (2012). *Cómo crear Aplicaciones Basadas en Linked Data?*. https://es.slideshare.net/alangrafu/como-crear-aplicaciones-basadas-en-linked-data-usando-lods-pea-kr.

Guy, M., d'Aquin, M., Dietze, S., Drachsler, H., Herder, E. y Parodi, E. (2014). *LinkedUp: Linking Open Data for Education*. Ariadne. http://www.ariadne.ac.uk/issue/72/guy-et-al/.

Guzmán, J., Motz, R. y Rodríguez da Silva, A. (2011). Hacia la Publicación Abierta de Objetos de Aprendizaje. *Cuadernos de Informática*, 6(1). En VI Congreso Ibero-americano de Telemática (CITA 2011). https://seer.ufrgs.br/cadernosdeinformatica/article/view/v6n1p301-304/11822.

Guzmán, J., López, M. y Torres, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica*, 2(50). Universidad Tecnológica de Pereira. ISSN 0122-1701. https://doi.org/10.22517/23447214.6693.

Hallo, M., Martínez, M. y De la Fuente, P. (2012). Las tecnologías de Linked Data y sus aplicaciones en el gobierno electrónico. *Scire: representación y organización del conocimiento*. 18(1), 49-61. https://www.ibersid.eu/ojs/index.php/scire/article/view/3923.

Hasan, R., Endris, K. y Gandon, F. (2014). SPARQL query result explanation for linked data. *Third International Conference on Semantic Web Collaborative Spaces* (SWCS'14). Vol. 1275, pp 38-52. https://dl.acm.org/citation.cfm?id=2878805.

Hausenblas, M. (2011). *Linked Data life cycles*. Linked Data Research Center. https://es.slideshare.net/mediasemanticweb/linked-data-life-cycles.

Heath, T. y Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology, pp. 1-136. Morgan & Claypool. http://linkeddatabook.com/editions/1.0/.

Hausenblas, M. (2009). *Linked Data Applications*. DERI Digital Enterprise Research institute. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.303.219&rep=rep1&type=pdf.

Hasan, R. y Gandon, F. (2012). Linking justifications in the collaborative semantic web applications*. 21st International Conference on World Wide Web (WWW '12 Companion).* ACM, 1083-1090. https://doi.org/10.1145/2187980.2188245.

Haslhofer, B. (2008). *A Web-based Mapping Technique for Establishing Metadata Interoperability*. Tesis Doctoral. Doctorado en Ciencias Técnicas. Universidad de Wien. https://eprints.cs.univie.ac.at/307/.

Haslhofer, B. y Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys*. 42(2). https://doi.org/10.1145/1667062.1667064.

HCLS - Semantic Web Health Care and Life Sciences Community Group. (2019). *Semantic Web in Health Care and Life Sciences Community Group*. W3C Community and Business Groups. https://www.w3.org/community/hclscg/.

Herrera-Cubides, J., Romero, X. y Chaparro, C. (2014). Uso de las TIC en la construcción de OVA, como estrategia para inclusión de personas con discapacidad visual. Actas de las *Jornadas Virtuales de Colaboración y Formación Virtual USATIC 2014, Ubicuo y Social: Aprendizaje con TIC*. ISBN: 978-84-686-5393-8. España. http://www.bubok.es/libros/234490/Actas-de-las-Jornadas-Virtuales-de-Colaboracion-y-Formacion-Virtual-USATIC-2014-Ubicuo-y-Social-Aprendizaje-con-TIC.

Herrera-Cubides, J. F., Gaona-García, P. A. y Sánchez-Alonso, S. (2016). The web of data: Past, present and ¿future?. *2016 XI Latin American Conference on Learning Objects and Technology (LACLO)*. https://ieeexplore.ieee.org/document/7751802.

Herrera-Cubides, J., Gaona-García, P. y Gordillo-Orjuela, K. (2017). A View of the Web of Data. Case Study: Use of Services CKAN. *Ingeniería*. Universidad Distrital Francisco José de Caldas*,* 22(1), pp. 111-124. https://doi.org/10.14483/udistrital.jour.reving.2017.1.a07.

Herrera-Cubides, J., Alonso-Echeverri, J., Riaño-Vargas, K., Gómez-Acosta, A. y Gaona-García, P. (2018a). A fuzzy logic system to evaluate levels of trust on linked open data resources. *Revista Facultad de Ingeniería Universidad de Antioquia*, No. 86. https://revistas.udea.edu.co/index.php/ingenieria/article/view/328937.

Herrera-Cubides, J., Gaona-García, P. y Sánchez-Alonso, S. (2018b). Linked Data: what happens to heterogeneity and interoperability. *Scientia et. Technica*, 23(2), pp. 230-240. https://doi.org/10.22517/23447214.16791.

Herrera-Cubides, J., Gaona-García, P. y Sánchez-Alonso, S. (2020). Open-Source Intelligence Educational Resources: A visual perspective analysis. *The Applied Sciences Journal*, 10 (21), 7617. https://doi.org/10.3390/app10217617.

Hilera, J. y Hoya, R. (2010). *Estándares de e-learning: guía de consulta*. Universidad de Alcalá, España. ISBN: 978-84-693-0263-7. http://www.cc.uah.es/hilera/GuiaEstandares.pdf.

Hogan, A., Harth, A., Passant, A., Decker, S., and Polleres, A. (2010). Weaving the pedantic web. In *LDOW-2010 Linked Data on the Web 2010.* http://ceur-ws.org/Vol-628/ldow2010_paper04.pdf.

Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A. y Decker, S. (2012). An empirical survey of Linked Data conformance. *Journal of Web Semantics*, Volume 14, Pages 14-44, ISSN 1570-8268. https://doi.org/10.1016/j.websem.2012.02.001.

Hogan, A. (2013). *Linked Data and the Semantic Web Standards*. Capítulo de Libro. Linked Data Management. ISBN 9781466582408. http://aran.library.nuigalway.ie/xmlui/handle/10379/4386.

Holst, T. y Höfig, E. (2013). Investigating the Relevance of Linked Open Data Sets with SPARQL Queries*. IEEE 37th Annual Computer Software and Applications Conference Workshops*. https://ieeexplore.ieee.org/document/6605794.

HP Labs. (2020). *Hewlett Packard Labs*. https://www.labs.hpe.com/.

Hummel, H., Manderveld, J., Tattersall, C. y Koper, R. (2004). Educational modelling language and learning design: new opportunities for instructional reusability and personalised learning.

International Journal of Learning Technology, 1(1), p.p. 111–126. https://pdfs.semanticscholar.org/7647/88365d381aaf1553ba59e7b2129667abfa67.pdf.

Hyland, B., Atemezing, G. y Villazón-Terrazas, B. (2014). *Best Practices for Publishing Linked Data*. W3C Working Group Note. https://www.w3.org/TR/ld-bp/.

IASA. (2011). IASA. *Directrices para la producción y preservación de objetos digitales de audio*. Ed. by Kevin Bradley. Second edition 2009. Standards, Recommended Practices and Strategies, IASA-TC 04. International Association of Sound and Audiovisual Archives. https://www.iasa-web.org/tc04-es/la-produccion-y-preservacion-de-objetos-digitales-de-audio.

IDECA - Infraestructura de Datos Espaciales para el Distrito Capital. (2015). *Web Semántica, Ontologías y Linked Data: nuevas tendencias para la gestión de la información geográfica*. https://www.ideca.gov.co/sites/default/files/presentaciones/idecaseminariotallerwebsemanticaontologiasylinkeddata.pdf.

IDLab. (2020). *IDLab Turtle Validator*. http://ttl.summerofcode.be/.

IDEE – Infraestructura de datos espaciales española. (2005). *Núcleo español de metadatos (NEM v1.0)*. Subgrupo de Trabajo del Núcleo Español de Metadatos. https://www.idee.es/resources/documentos/NEMv1.2.pdf.

Idehen, K. (2018). *What is a SPARQL Endpoint, and why is it important?*. Medium. https://medium.com/virtuoso-blog/what-is-a- SPARQL-endpoint-and-why-is-it-important-b3c9e6a20a8b.

IEEE - Institute of Electrical and Electronics Engineers. (2002). *Estándar para Metadatos de Objetos Educativos*. Comité de Estandarización de Tecnologías Educativas del IEEE. https://www-gist.det.uvigo.es/~lanido/LOMes/LOMv1_0_Spanish.pdf.

IGI (2018). *Information retrieval and management: concepts, methodologies, tools, and applications.* Hershey: IGI Global, Information Resources Management Association (USA). ISBN10: 1522551913. https://doi.org/10.4018/978-1-5225-5191-1.

Isaac, A. y Pastor, J. (2018). *SKOS: overview and modeling of controlled vocabularies*. Dublin Core Metadata Initiative. https://www.dublincore.org/webinars/2018/skos_overview_and_modeling_of_controlled_vocabularies/.

Jano. (2020). *Sistema de gestión para concurso de méritos*. Oficina Asesora de Sistemas. Universidad Distrital Francisco José de Caldas. https://funcionarios.portaloas.udistrital.edu.co/jano/index.php?data=Ikj0z-v94jvkHPz2bqjHNVI7pHIG00-CRHd_i80bJ5fAamAdKghN9VXrBb1uelvKTUZNtxh6DBlQ_SlVCall_w.

Jis. (2020). *OpenDOAR Statistics*. https://v2.sherpa.ac.uk/view/repository_visualisations/1.html.

Junar. (2020). *Junar Data Platform*. https://www.junar.com/.

Knap, T. (2013). *Towards Trustworthy Linked Data Integration and Consumption*. Disertación de Tesis Doctoral. Faculty of Mathematics and Physics, Charles University. Prague. https://dspace.cuni.cz/handle/20.500.11956/53271.

Kyron. (2017). *Sistema de gestión de información docente – kyron (módulo de hojas de vida)*. Oficina Asesora de Sistemas. Universidad Distrital Francisco José de Caldas. http://repository.udistrital.edu.co/bitstream/11349/5765/3/AcevedoArceJorgeArmando2017Anexo%203.pdf.

Király, P. (2015). A Metadata Quality Assurance Framework. *8th International Conference on Qualitative and Quantitative Methods in Libraries*. https://pkiraly.github.io/metadata-quality-project-plan.pdf.

Kleppe, A., Warmer, J. y Bast, W. (2003). MDA explained - the Model Driven Architecture: practice and promise. Addison-Wesley Professional. ISBN: 9780321194428.

Klímek, J., Skoda, P. y Necaský, M. (2018). Survey of Tools for Linked Data Consumption. *IOS Press Content Library*, pp. 1-57. https://content.iospress.com/articles/semantic-web/sw316.

Koutsomitropoulos, D., Alexopoulos, A., Solomou, G. y Papatheodorou, T. (2010). The Use of Metadata for Educational Resources in Digital Repositories: Practices and Perspectives. *D-Lib Magazine*, 16(1). http://www.dlib.org/dlib/january10/kout/01kout.html.

Kuhn, T., Chichester, C., Krauthammer, M., Queralt-Rosinach, N., Verborgh, R., Giannakopoulos, G., Ngonga Ngomo, A., Viglianti, R. y Dumontier, M. (2016). Decentralized provenance-aware publishing with nanopublications. *PeerJ PrePrints*. https://doi.org/10.7287/peerj.preprints.1760v1.

Kujane (2020). Minciencias. Nodo Colombia – Proyecto LA Referencia. [Online] Disponible en: http://kujane.minciencias.gov.co/vufind/.

LAR. (2019). *LA Referencia*. Red de Repositorios de Acceso Abierto a la Ciencia. http://www.lareferencia.info/es/.

Laufer, C. (2015). *Semantic Web Guideline*. Gobierno del Estado de Sao Paulo. Brasil. http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/05/Book-Semantic-Web-Guideline.pdf.

Le Ngoc, L., Tireau, A., Venkatesan, A., Neveu, P. y Larmande, P. (2016). Development of a knowledge system for Big Data: Case study to plant phenotyping data. WIMS: Web Intelligence, Mining and Semantics, *WIMS '16 Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, pp.1-9. https://doi.org/10.1145/2912845.2912869.

Li, Z., Wu, S., Zhou, H., Zou, S. y Dong, T. (2019). Analytic Model and Assessment Framework for Data Quality Evaluation in State Grid. *Journal of Physics: Conference Series*, 1302(2), IOP Publishing Ltd. https://iopscience.iop.org/article/10.1088/1742-6596/1302/2/022083.

LinkedGeoData. (2018). *LinkedGeoData - Adding a spatial dimension to the Web of Data*. http://linkedgeodata.org/About.

LOC. (2020). *Biblioteca del Congreso de los Estados Unidos*. https://www.loc.gov/.

LOD2 (2019). *LOD2: Creating Knowledge out of Interlinked Data*. Semantic Web Company. https://semantic-web.com/project/lod2-the-linked-data-technology-stack-for-enterprises/.

LOV. (2020). *Linked Open Vocabularies*. Ontology Engineering Group. https://lov.linkeddata.es/dataset/lov.

LRES. (2020). *Learning Resource Exchange for Schools*. http://lreforschools.eun.org/web/guest.

LRMI. (2013). *Learning Resource Metadata Initiative (LRMI)*. Dublin Core Metadata initiative. https://www.dublincore.org/specifications/lrmi/lrmi_1/.

Lyrasis. (2020a). *Dspace*. https://duraspace.org/dspace/.

Lyrasis. (2020b). *Dspace Technical Specifications*. https://duraspace.org/dspace/resources/technical-specifications/.

Martínez, F., Perea Reyes, N., Ortiz Ancona, D. y Palma Peña, J. (2019). Harmonization and interoperability of metadata schemes used at the National Autonomous University of Mexico (UNAM) repositories: the IIBI-UNAM Repository. *IFL WLIC 2019: The International Federation of Library Associations and Institutions – World Library and Information Congress*. http://library.ifla.org/2466/1/208-martinez-en.pdf.

Martínez, F., Pastor-Sánchez, J. y López, R. (2020). Linked open data en bibliotecas: estado del arte. *IR Information Research*, 25(2). http://informationr.net/ir/25-2/paper862.html.

Marzal, M., Calzada, J. y Cuevas, A. (2006). Desarrollo de un esquema de metadatos para la descripción de recursos educativos: el perfil de aplicación Mimeta. *Revista Española de Documentación Científica*, 29(4). ISSN 0210-0614. http://redc.revistas.csic.es/index.php/redc/article/view/308.

Massachusetts Institute of Technology - MIT. (2020). *Massachusetts Institute of Technology Libraries*. https://libraries.mit.edu/.

Massart, D. y Shulman, E. (2011). *Metadata and Controlled Vocabularies. Learning Resource Exchange Metadata Application Profile - Version 4.7*. http://lreforschools.eun.org/web/guest/metadata.

McCrae, J. (2020). *The Linked Open Data Cloud*. https://lod-cloud.net/.

Mehta, Y. y Buch, S. (2016). Semantic proximity with linked open data: A concept for social media analytics. *International Conference on Computing, Communication and Automation (ICCCA)*, pp. 337-341. https://doi.org/10.1109/CCAA.2016.7813741.

MEN – Ministerio de Educación Nacional. (2007). *Banco Nacional de Objetos de Aprendizaje e Informativos*. Colombia Aprende http://portalapp.mineducacion.gov.co/drupalM/.

MEN – Ministerio de Educación Nacional. (2009). *Qué es un Objeto de Aprendizaje*. Colombia Aprende. https://aprendeenlinea.udea.edu.co/lms/men/oac1.html.

MEN - Ministerio de Educación Nacional. (2012). *Recursos Educativos Digitales Abiertos*. Graficando Servicios Integrados. https://aprende.colombiaaprende.edu.co/sites/default/files/naspublic/libroreda_0.pdf.

MEN - Ministerio de Educación Nacional. (2013). *Especificación perfil de metadatos CEM 2.0*. http://aprende.colombiaaprende.edu.co/sites/default/files/naspublic/especificacion_cem.pdf.

MEN - Ministerio de Educación Nacional. (2015). *Manual de metadatos y directrices de interoperabilidad*. SNAAC - Sistema Nacional de Acceso Abierto al Conocimiento. http://snaac.renata.edu.co:8080/documents/21409/22448/Manual+de+Metadatos+e+Interoperabilidad+enero+29+de+2016.pdf/46595bf1-d97d-43c2-8669-9bbab45712d0.

MEN - Ministerio de Educación Nacional. (2020). Plataforma Aprender Digital Contenidos para Todos. https://contenidos.colombiaaprende.edu.co/.

Mendes, P. (2012). *Conceptual model and best practices for high-quality metadata* publishing. PalnetData. https://docplayer.net/40848695-D2-1-conceptual-model-and-best-practices-for-high-quality-metadata-publishing.html.

Miao, F., Mishra, S. y McGreal, R. (2016). *Open Educational Resources: Policy, Costs and Transformation*. United Nations Educational, Scientific and Cultural Organization (UNESCO) and Commonwealth of Learning. ISBN 978-1-894975-75-9. http://oasis.col.org/handle/11599/2306.

Miao, F., Mishra, S., Orr, D. y Janssen, B. (2019). *Directrices para la elaboración de políticas de recursos educativos abiertos*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura – UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000373558.

Miller, K. (2016). Data's Identity Crisis: The Struggle to Name It, Describe It, Find It, and Publish It. *Biomedical Computation Review*. http://www.bcr.org/content/data%E2%80%99s-identity-crisis-struggle-name-it-describe-it-find-it-and-publish-it.

Mintic - Ministerio de Tecnologías de la Información y las Comunicaciones. (2015). *Estudio del grado de preparación para la apertura de datos del sector educativo en Colombia*. https://estrategia.gobiernoenlinea.gov.co/623/articles-9407_datos_sector_educativo.pdf.

Mintic - Ministerio de Tecnologías de la Información y las Comunicaciones. (2016). *Proyecto de Datos Abiertos*. Estrategia de Gobierno en Línea. República de Colombia. https://estrategia.gobiernoenlinea.gov.co/623/w3-article-9407.html.

Montenegro, C., Gaona García, P., Cueva Lovelle, J. y Sanjuán Martínez, O. (2011). Aplicación de ingeniería dirigida por modelos (MDA), para la construcción de una herramienta de modelado de dominio especifico (DSM) y la creación de módulos en sistemas de gestión de aprendizaje (LMS) independientes de la plataforma. *DYNA*, 78(169), pp. 43-52. ISSN 2346-2183. https://revistas.unal.edu.co/index.php/dyna/article/view/20390.

Montenegro, C. (2017). *Ingeniería dirigida por modelos (MDA) y casos prácticos* (pp. 36-38). Doctorado en Ingeniería. Universidad Distrital Francisco José de Caldas. ISBN: 978-958-8972-89-3.

Mora, B., Ruiz, F., García, F. y Piattini, M. (2006). *Definición de lenguajes de modelos MDA vs DSL*. Departamento de Tecnologías y sistemas de Información. Universidad de Castilla – La Mancha. https://studylib.es/doc/1669785/definici%C3%B3n-de-lenguajes-de-modelos-mda-vs-dsl.

Mortera, F. (2010). Implementación de Recursos Educativos Abiertos (REA) a través del portal TEMOA (Knowledge Hub) del Tecnológico de Monterrey, México. *Formación Universitaria*, 3(5), 9-20. https://dx.doi.org/10.4067/S0718-50062010000500003.

Mouromtsev, D. y d'Aquin, M. (2016). Open Data for Education - Linked, Shared, and Reusable Data for Teaching and Learning. *Lecture Notes ComputerState-of-the-Art Surveys*. https://www.springer.com/gp/book/9783319304922.
MusicBrainz. (2020). *MusicBrainz*. https://musicbrainz.org/.
MVNRepository. (2014). Apache Jena Fuseki A SPARQL 1.1 Server. https://mvnrepository.com/artifact/org.apache.jena/jena-fuseki/1.0.1.
Nahhas, S., Bamasag, O., Khemakhem, M. y Bajnaid, N. (2018). Added Values of Linked Data in Education: A Survey and Roadmap. *Computers*, 7(45). https://doi.org/10.3390/computers7030045.
National Center for Education Statistics – NCES. (2019). *Distance Learning*. https://nces.ed.gov/fastfacts/display.asp?id=80.
Narváez, E. y Piedra, N. (2018). Un enfoque de Linked Data para garantizar la interoperabilidad semántica e integridad de datos académicos universitarios. *IWSW 2018 - International Workshop on Semantic Web*. http://ceur-ws.org/Vol-2096/.
Navarrete, R. y Luján, S. (2015). Use of Linked Data to Enhance Open Educational Resources. *14th International Conference on Information Technology Based Higher Education and Training (ITHET 2015)*, p. 1-6. ISBN: 978-1-4799-1756-3. http://desarrolloweb.dlsi.ua.es/oer/use-linked-data-enhance-open-educational-resources.
Newfield, C. (2020). Going online due to COVID-19 this fall could hurt colleges' future. *The Conversation*. https://theconversation.com/going-online-due-to-covid-19-this-fall-could-hurt-colleges-future-138926.
Nilsson, M., Powell, A. y Johnston, P. (2008). *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dc-rdf/.
NLM - US National Library of Medicine. (2019). *DailyMed*. http://dailymed.nlm.nih.gov/dailymed/index.cfm.
Note, M. (2011). *Managing Image Collections*. ISBN 978-1-84334-599-2. https://www.sciencedirect.com/book/9781843345992/managing-image-collections.
Noy, N. y McGuinness, D. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory. Stanford University. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf.
OAS. (2020). *Oficina Asesora de Sistemas*. Universidad Distrital Francisco José de Caldas. https://portalws.udistrital.edu.co/oas/#.
OBS - Business School. (2019). *El mercado del e-Learning crecerá cada año un 7.6% hasta 2020.* https://www.obs-edu.com/int/blog-investigacion/e-learning/el-mercado-del-e-learning-crecera-cada-ano-un-76-hasta-2020.
Ocampo, G., Sartor, N. y Favieri, A. (2015). Declaración de París sobre Recursos Educativos Abiertos. *Cuarto Congreso Virtual Mundial de E-Learning*. https://doi.org/10.13140/RG.2.1.3959.2720.
ODC – Open Data Charter. (2020). *Carta Internacional de Datos Abiertos*. https://opendatacharter.net/principles-es/.
OECD - Organization for Economic Co-operation and Development. (2007). *Giving Knowledge for Free. The Emergence of Open Educational Resources.* Centre for Educational Research and Innovation. http://www.oecd.org/education/ceri/38654317.pdf.
OER-C (2014). *OER Commons Learning Object Metadata Element Set, Based on IEEE Learning Object Metadata Information Model*. ISKME - The Institute for the Study of Knowledge Management. https://www.micollegesonline.net/uploads/ckeditor/files/OERCv4-MetadataApplicationProfile.pdf.
OER-C. (2020). *OER Commons*. https://www.oercommons.org/.
OERWM. (2020). *The OER World Map*. https://oerworldmap.org/.
Oer schema. (2020). *Open Educational Resource Schema*. http://oerschema.org/.
OGoov. (2020). *OGoov: solución para una Plataforma de Gobierno Abierto*. https://www.ogoov.com/es/.

O'Hara, K., Alani, H., Kalfoglou, Y. y Shadbolt, N. (2004). Trust Strategies for the Semantic Web. *Workshop on Trust, Security, and Reputation on the Semantic Web*. https://pdfs.semanticscholar.org/c771/5a87e1dc757abbcd62f4afc7c70c88b8d9e2.pdf.

OKF – Open Knowledge Foundation. (2019). *Open Data Handbook*. http://opendatahandbook.org/.

OMG - Object Management Group. (2014a). *MDA Guide rev. 2.0*. https://www.omg.org/cgi-bin/doc?ormsc/14-06-01.

OMG - Object Management Group. (2014b). Meta Object Facility. https://www.omg.org/spec/MOF/2.4.2/About-MOF/.

OMG - Object Management Group. (2019). *MDA® - The architecture of choice for a changing world*. https://www.omg.org/mda/.

OpenCourseWare Consortium. (2020). *OpenCourseWare.* Open Education Consortium – The Global Network for Open Education. https://www.oeconsortium.org/.

OpenData. (2015). *5 Stars Open Data*. http://5stardata.info/en/.

Paquette, G. y Miara, A. (2014). Managing Open Educational Resources on the Web of Data. Presenting COMETE, an RDF-based Resource Manager. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 5(8). https://r-libre.teluq.ca/490/1/Managing%20OER%20with%20RDF.pdf.

Pattanaphanchai, J. (2011). DC Proposal: Evaluating Trustworthiness of Web Content Using Semantic Web Technologies. *The Semantic Web – ISWC 2011*. Lecture Notes in Computer Science, Volume 7032. https://doi.org/10.1007/978-3-642-25093-4_25.

Pereira, C., Matsui, S., Nunes, B. y Dietze, S. (2018). Linked Data in Education: A Survey and a Synthesis of Actual Research and Future Challenges. *IEEE Transactions on Learning Technologies*, 11(3), pp. 400-412. https://doi.org/10.1109/TLT.2017.2787659.

Pérez, C. (2002). Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del español (ELiEs)*, Volumen 18 (2002), ISSN: 1139-8736. http://elies.rediris.es/elies18/531.html.

Pérez, G. (2008). *Un aporte gráfico a las herramientas para transformación entre modelos*. Tesis de Grado. Licenciatura en Informática, Universidad Nacional de La Plata. http://sedici.unlp.edu.ar/handle/10915/3979.

Pérez, N., Díaz, M. y Rincón, N. (2015). Web semántica y su aporte a la estrategia de datos abiertos del estado colombiano. *Revista Científica*, 23(3), pp. 124 – 132. https://doi.org/10.14483/23448350.9796.

Piedra, N., Chicaiza, J. y López, J. (2010). Design study of OER-CC ontology: A semantic web approach to describe Open Educational Resources. *10th IEEE International Conference on Advanced Learning Technologies*. https://core.ac.uk/download/pdf/80739232.pdf.

Piedra, N., Chicaiza, J., López, J. y Tovar, E. (2013). Using linked open data to improve the search of open educational resources for engineering students. *IEEE Frontiers in Education Conference (FIE)*, pp. 558-560, DOI: 10.1109/FIE.2013.6684885.

Pinkel, C. (2013). *Building Linked Data Applications*. EUCLID – EdUcational Curriculum for the usage of Linked Data. http://euclid-project.eu/modules/chapter5.html.

Planestic. (2020). *Planestic-UD*. Universidad Distrital Francisco José de Caldas. https://planestic.aulasvirtuales.udistrital.edu.co/.

Planestic. (2019). *Revisión modelo de datos para producción académica*. Universidad Distrital Francisco José de Caldas. https://github.com/udistrital/campus_cliente/issues/170.

Pons, D., Hilera, J. y Pagés, C. (2011). ISO/IEC 19788 MLR: Un Nuevo Estándar de Metadatos para Recursos Educativos*. IEEE-RITA,* 6(3). http://rita.det.uvigo.es/201108/uploads/IEEE-RITA.2011.V6.N3.A8.pdf.

PORT - Postgraduate Online Research Training. (2019). *Linked Data Handbook*. School Advanced Study, University of London. https://port.sas.ac.uk/mod/book/view.php?id=2210&chapterid=2156.

PostgreSQL. (2020). *PostgreSQL: The World's Most Advanced Open Source Relational Database*. https://www.postgresql.org/.

Poyato, M. (2014). *Extensión Semántica de OML (Semantic-OML)*. Trabajo de Fin de Master. Universidad Autonoma de Madrid. https://repositorio.uam.es/bitstream/handle/10486/662572/Poyato_Pino_Mario_tfm.pdf?sequence=1.

Pressman, R. (2010). *Ingeniería de Software, Un enfoque práctico*.McGraw-Hill, Septima Edición. ISBN: 978-0-07-337597-7.

Procomún. (2020). *Red de recursos educativos en abierto*. http://procomun.educalab.es/es.

Prud'hommeaux, E. y Seaborne, A. (2008). *SPARQL Query Language for RDF*. W3C - World Wide Web Consortium – Recommendation. https://www.w3.org/TR/rdf- SPARQL-query/.

Qing Yu, H., Pedrinaci, C., Dietze, S. y Domingue, J. (2012). Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning. In *IEEE Transactions on Learning Technologies*, 5(2), pp 130-142. https://doi.org/10.1109/TLT.2012.1.

Quintero, J. y Anaya, R. (2007). MDA y el papel de los modelos en el proceso de desarrollo de software. *Revista EIA*, (8), pp. 131-146. Escuela de Ingeniería de Antioquia. http://www.scielo.org.co/pdf/eia/n8/n8a11.pdf.

Rajabi, E., Rodríguez, C., Sánchez, S. y Sicilia, M. (2014). *Exposición de metadatos de objetos de aprendizaje como datos enlazados: el caso del proyecto Open Discovery Space*. En Latin-American Community of Learning Objetcs - LACLO 2014.

Ramesh, P., Vivekavardhan, J. y Bharathi, K. (2015). Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. *DESIDOC Journal of Library & Information Technology*, 35(3), pp. 193-199. https://doi.org/10.14429/djlit.35.3.8074.

Rayón, A. (2016). *Del Open Data al Linked Open Data: sacando valor de los datos enlazados*. Deusto Data. https://blogs.deusto.es/bigdata/del-open-data-al-linked-open-data-sacando-valor-de-los-datos-enlazados/.

Rector, A. y Noy, N. (2006). *Defining N-ary Relations on the Semantic Web*. The World Wide Web Consortium (W3C). https://www.w3.org/TR/swbp-n-aryRelations/.

RedCol - Red Colombiana de Información Científica. (2020). *Directrices para repositorios institucionales de investigación de la RedCol*. https://redcol.readthedocs.io/es/latest/.

Red Nacional Académica de Tecnología Avanzada – RENATA. (2020). *RENATA*. https://www.renata.edu.co/.

Riley, J. (2017). *Understanding metadata: what is metadata, and what is it for?: A Primer*. University of Horth Texas. https://digital.library.unt.edu/ark:/67531/metadc990983/.

Ríos-Hilario, A., Martín-Campo, D. y Ferreras-Fernández, T. (2012). Linked data y Linked Open Data: su implantación en una biblioteca digital. El caso de Europeana. *Profesional de la Información*, 21(3). http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/10.html.

Ríos-Hilario, A. (2014). Linked Open Media Data: La tecnología de datos vinculados al servicio de los medios de comunicación. *Cuadernos de Documentación Multimedia*, Vol 25, pp. 118-140, ISNN: 1575-9733. https://doi.org/10.5209/rev_CDMU.2014.v25.47470.

RIUD - Repositorio Institucional Universidad Distrital. (2016). *Instructivo Catalogador Metadatos y Publicaciones de Documentos*. Universidad Distrital Francisco José de Caldas, Bogotá. http://sistemadebibliotecas.udistrital.edu.co:8000/documentos/riud/CatalogadorV2-2016.pdf.

RIUD. (2020). *Repositorio Institucional Universidad Distrital*. Universidad Distrital Francisco José de Caldas, Bogotá. http://repository.udistrital.edu.co/.

Rocha Amorim, R., Rabelo, T. y Gomes, D. (2012). *Open Educational Resources Ontology*. Universidad del Estado de Bahia, Brasil. https://www.br-ie.org/pub/index.php/wcbie/article/viewFile/1887/1651.

Rodríguez, I. (2009). *El Estándar de Metadatos RDF*. Escuela Técnica Superior de Ingenieros, Universidad de Sevilla. http://bibing.us.es/proyectos/abreproy/11731/fichero/PFC_Final.pdf.

Rodríguez, L. (2017). *Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios*. Tesis Doctoral. Departamento de Informática, Universidad de Oviedo. https://dialnet.unirioja.es/servlet/tesis?codigo=179170.

Romero, B. (2018). *Una propuesta metodológica para garantizar la usabilidad y la accesibilidad en el ECRM centrada en persona con discapacidades y adultos mayores.* Tesis Doctoral, Doctorado en Tecnologías de la Información y la Comunicación, Universidad de Granada. https://digibug.ugr.es/handle/10481/55787.

Ruckhaus, E., Vidal, M., Castillo, S., Burguillos, O. y Baldizan, O. (2014). Analyzing Linked Data Quality with LiQuate. *The Semantic Web: ESWC 2014 Satellite Events*. https://doi.org/10.1007/978-3-319-11955-7_72.

Rühle, S., Baker, T. y Johnston, P. (2017). *Publishing Metadata*. Dublin Core Metadata Initiative. https://www.dublincore.org/resources/userguide/publishing_metadata/.

Ruiz, I. (2012). *Ingeniería Dirigida por Modelos y Calidad de Software*. Trabajo de Investigación. Escuela Superior de Ingeniería. Universidad de Cádiz. España. http://rodin.uca.es/xmlui/bitstream/handle/10498/15744/2012-Ingenier%C3%ADa%20dirigida%20por%20modelos%20y%20calidad%20de%20software.pdf.

Sánchez, J. (2014). *Linked Data para la Generación de Conocimiento Financiero a partir de la Extracción de Información Semiestructurada*. Tesis Doctoral. Departamento de Informática, Universidad Carlos III de Madrid. https://e-archivo.uc3m.es/bitstream/handle/10016/21861/tesis_jose-luis_sanchez_cervantes_2015.pdf?sequence=1&isAllowed=y.

Sánchez, N. (2016). *Metamodelo para la definición e implantación de los procesos de planificación y control de proyectos en pequeños equipos de desarrollo*. Tesis de Maestría. Universidad Tecnológica de la Mixteca. http://jupiter.utm.mx/~tesis_dig/12998.pdf.

Sánchez-Alonso, S., Sicilia, M. y Pareja, M. (2007). Mapping IEEE LOM to WSML: an ontology of learning objects. In Proceedings of the *ITA'07 - 2nd International on Internet Technologies and Applications*, pp. 92 - 101. http://www.cc.uah.es/ssalonso/papers/ITA2007.pdf.

Sánchez Yopazá, W. (2017). *Guía de metadatos - Guía para la formulación de un esquema de metadatos para la gestión de documentos*. Archivo General de la Nación. https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/GuiaDeMetadatos.pdf.

Sarcia, S. (2010). Is GQM+ Strategies really applicable as is to non-software development domains?. *ACM-IEEE International Symposium on Empirical Software Engineering and Measurement.* https://doi.org/10.1145/1852786.1852844.

Schmachtenberg, M., Bizer, C. y Paulheim H. (2014). Adoption of the Linked Data Best Practices in Different Topical Domains. Mika P. et al. (eds) *The Semantic Web – ISWC 2014*. Lecture Notes in Computer Science, Volume 8796. Springer, Cham. https://doi.org/10.1007/978-3-319-11964-9_16.

Sena - Servicio Nacional de Aprendizaje. (2017). *Orientaciones para realizar el inventario de recursos educativos*. Proceso gestión de formación profesional integral. http://compromiso.sena.edu.co/mapa/descarga.php?id=1316.

Senso, J. y Arroyo, W. (2018). La publicación en Linked Data de registros bibliográficos: modelo e implementación. *Revista Española de Documentación Científica*, 41(4). https://doi.org/10.3989/redc.2018.4.1535.

Serrano, E. (2013). *Metamodeo para la mejora del proceso de gestión de requisitos*. Tesis doctoral. Universidad Politecnica de Madrid. http://oa.upm.es/21988/1/ARIEL_EDGAR_SERRANO_RICO_C.pdf.

Singal, H. y Kohli, S. (2016). Trust Necessitated through Metrics: Estimating the Trustworthiness of Websites. *Procedia Computer Science*, Volume 85, pp 133-140. https://doi.org/10.1016/j.procs.2016.05.199.

Siegel, J. (2014). *MDA Guide revision 2.0.* Object Management Group – OGM. https://www.omg.org/cgi-bin/doc?ormsc/14-06-01.

Sixto, A., Aleixandre, R., Vidal, A., Lucas, R. y Castelló, L. (2019). *DATA SHARING: qué son y cómo se pueden compartir los datos de investigación. Manual de recomendación para*

gestores de la información. SEDIC. ISBN 978-84-09-17458-4. https://biblioguias.unex.es/ld.php?content_id=32767777.
Smith-Yoshimura, K. (2018). Linked data implementations - who, what, why?. *Semantic Web in Libraries (SWIB18).* https://swib.org/swib18/slides/2_smith-yoshimura_linked-data-implementations.pdf.
Smothers, V. y Giordano, D. (2015). *Healthcare Learning Object Metadata - Specifications and Description Document*. MedBiquitous Consortium. https://www.medbiq.org/sites/default/files/files/HealthcareLOMSpecifications_pointrelease.pdf.
Sommerville, I. (2011). *Software Engineering*. Ninth Edition. Addison-Wesley. ISBN-13: 978-0-13-703515-1, pp. 41.
Song, D., Luo, Y. y Heflin, J. (2017). Linking Heterogeneous Data in the Semantic Web Using Scalable and Domain-Independent Candidate Selection. *IEEE Transactions on Knowledge and Data Engineering*, 29(1). https://dl.acm.org/citation.cfm?id=3024196.
Styles, R. y Shabir, N. (2008). Academic Institution Internal Structure Ontology (AIISO). https://vocab.org/aiiso/.
Sulé, A., Centelles, M., Franganillo, J. y Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. *Revista española de Documentación Científica*, 39(1), ISSN 1988-4621. http://dx.doi.org/10.3989/redc.2016.1.1268.
Sunlight Foundation. (2017). Ten principles for opening up government information. https://sunlightfoundation.com/policy/documents/ten-open-data-principles/.
Tauberer, J. (2014). *Open Government Data: The Book*. Second Edition. https://opengovdata.io/.
Telefónica. (2020). *Miríadax*. https://miriadax.net/home.
Temesio, S. (2015). Metadatos para recursos educativos. *Memoria Académica*, 5(1). Facultad de Humanidades y Ciencias de la Educación (FaHCE), Universidad Nacional de La Plata. http://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.6885/pr.6885.pdf.
The Open University. (2020). *OpenLearn*. https://www.open.edu/openlearn/.
Tian, J. (2005). *Software Quality Engineering*. IEEE Computer Society. Published by John Wiley & Sons, Inc.
Torres, A. (2008). *Método de pruebas de sistema basado en modelos navegacionales en un contexto MDWE*. Memoria de investigación presentada al Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla. http://www.lsi.us.es/docs/doctorado/memorias/MemoInvestigArturoHTorres.pdf.
Truyen, F. (2006). *The Fast Guide to Model Driven Architecture. The Basics of Model Driven Architecture (MDA)*. Cephas Consulting Corp. https://www.omg.org/mda/mda_files/Cephas_MDA_Fast_Guide.pdf.
Unesco - Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2015). *Directrices para los recursos educativos abiertos (REA) en la educación superior*. https://unesdoc.unesco.org/ark:/48223/pf0000232855.
UNESCO - United Nations Educational, Scientific and Cultural Organization. (2019). *Recursos Educativos Abiertos*. https://es.unesco.org/themes/tic-educacion/rea.
Unión Europea. (2020). *Portal Europeo de Datos*. https://www.europeandataportal.eu/es.
Universidad de Campiñas. (2020). M3 – *Recursos Educativos Multimedia para Matemáticas de Enseñanza Media.* https://m3.ime.unicamp.br/.
Universidad Técnica Particular de Loja. (2020). *Serendipity – Linked OpenCourseWare Data Faceted Search*. http://serendipity.utpl.edu.ec/.
Van Solingen, R. y Berghout, E. (1998). *Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill Education. ISBN-10: 0077095537.
Vila-Viñas, D., Araya, D. y Bouchard, P. (2015). *Educación - Recursos educativos abiertos*. Buen Conocer - FLOK Society. https://book.floksociety.org/wp-content/uploads/2015/05/1_1_-_Educacion.pdf.

Villalobos, J. (2020). *Modelos y Metamodelos*. Lecturas para Arquitectos de Negocio, No. 4 – Versión 1.0. Universidad de los Andes. https://www.studocu.com/co/document/universidad-de-los-andes-colombia/arquitectura-empresarial/resumenes/modelos-y-metamodelos-1/8837064/view.

W3C - The World Wide Web Consortium. (2006). *Validation service*. https://www.w3.org/RDF/Validator/.

W3C – The World Wide Web Consortium. (2009). *LODD Data DataSet Evaluation*. http://www.w3.org/wiki/HCLSIG/LODD/Data/DataSetEvaluation.

W3C – The World Wide Web Consortium. (2012). *Linking Open Drug Data (LODD)*. http://www.w3.org/wiki/HCLSIG/LODD.

W3C – The World Wide Web Consortium. (2013). *SPARQL 1.1 Query Language W3C Recommendation*. https://www.w3.org/TR/ SPARQL11-query/.

W3C – The World Wide Web Consortium. (2015). *Linked Data*. Semantic Web. https://www.w3.org/standards/semanticweb/data.

W3C – The World Wide Web Consortium. (2017). *Linking Open Data*. W3C SWEO Community Project. https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

W3C - The World Wide Web Consortium. (2020). *Data Catalog Vocabulary (DCAT) - Version 2*. The World Wide Web Consortium. https://www.w3.org/TR/vocab-dcat-2/.

Wason, T. (2000). Guide to Using Classification: A Dr. Tom Guide. http://www.tomwason.com/drtomclassification.html.

Wenige, L. y Ruhland, J. (2015). Scalable Property Aggregation for Linked Data Recommender Systems. *3rd International Conference on Future Internet of Things and Cloud (FICLOUD'15)*. IEEE Computer Society, pp. 451-456. https://doi.org/10.1109/FiCloud.2015.30.

Wikipedia. (2011). Learning object metadata. https://en.wikipedia.org/wiki/Learning_object_metadata.

Willer, M. y Dunsire, G. (2013). *Bibliographic Information Organization in the Semantic Web*. Chandos Publishing. ISBN: 978-1-84334-731-6. https://www.elsevier.com/books/bibliographic-information-organization-in-the-semantic-web/willer/978-1-84334-731-6.

Wood, D. (2011). *Linking Government Data*. Springer. ISBN 978-1-4614-1766-8. http://dx.doi.org/10.1007/978-1-4614-1767-5.

Wood, D., Zaidman, M., Ruth, L. y Hausenblas, M. (2014). Linked Data: Structured data on the Web. Manning. ISBN 978-1-6172-9039-8. https://www.manning.com/books/linked-data.

WWWF – World Wide Web Foundation. (2019). *The Open Data Barometer.* https://opendatabarometer.org/?_year=2017&indicator=ODB.

Xie, I. y Matusiak, K. (2016). *Metadata y Controlled Vocabulary*. Discover Digital Libraries, Science Direct. https://www.sciencedirect.com/topics/computer-science/controlled-vocabulary.

Yaghouti, N., Kahani, M. y Behkamal, B. (2015). A metric-driven approach for interlinking assessment of RDF graphs. *International Symposium on Computer Science and Software Engineering (CSSE)*, pp. 1-8. https://ieeexplore.ieee.org/document/7369244.

Yamada, T. y Morimoto, Y. (2010). Utilization of metadata for quality assurance in the international sharing and reuse of open educational resources. *Asian Association of Open Universities Journal*, 5(1). https://doi.org/10.1108/AAOUJ-05-01-2010-B004.

Yu, L. (2014). *A Developer's Guide to the Semantic Web*. Springer Berlin Heidelberg. ISBN: 978-3-642-15970-1. https://www.springer.com/gp/book/9783662437957.

Zahedi, M. y Fathian, A. (2016). Proposed metrics for data accessibility in the context of linked open data. *Program: electronic library and information systems*, 50(2), pp. 184-194. https://doi.org/10.1108/PROG-01-2015-0007.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. y Auer, S. (2012). Quality Assessment Methodologies for Linked Open Data. A Systematic Literature Review and Conceptual Framework. *Semantic Web Journal,* Volume 1, pp. 1-33. http://www.semantic-web-journal.net/system/files/swj414.pdf.

Zaveri, A. (2015). *Linked Data Quality Assessment and its Application to Societal Progress Measurement*. Disertación Doctoral. Leipzig. https://cris.maastrichtuniversity.nl/en/publications/linked-data-quality-assessment-and-its-application-to-societal-pr.

Zeinstra, M. y Keller, P. (2016). Open Linked Data and Europeana. Europeana Foundation. https://mzeinstra.nl/wp-content/uploads/2016/06/Open-Linked-Data.pdf.

Zemanta. (2019). *Zemanta - The Art & Science of Content Promotion*. https://www.zemanta.com/.

Zhao, J., Miles, A., Klyne, G. y Shotton, D. (2009). Linked data and provenance in biological data webs. Briefings in *Bioinformatics*, 10(2), pp. 139–152. https://doi.org/10.1093/bib/bbn044.

Zuiderwijk, A., Jeffery, K. y Janssen, M. (2012). The potential of metadata for linked open data and its value for users and publishers. *JeDeM - eJournal of eDemocracy and Open Government*, 4(2), ISSN 2075-9517. https://pdfs.semanticscholar.org/9487/48173aabd89fa01a7c72c36d82fd0d2b816c.pdf.

Anexo 1 Investigaciones en Datos Enlazados.

Iniciativa	Autor(es)	Descripción
Linking Justifications in the Collaborative Semantic Web Applications	Hasan y Gandon (2012)	Presenta el concepto de justificaciones enlazadas en el espacio semántico colaborativo. Proporciona esquemas para codificar y enlazar justificaciones centradas en las plataformas colaborativas, usando Ratio4TA vocabulary.
SPARQL Query Result Explanation for Linked Data	Hasan *et al.* (2014)	Presenta un enfoque para explicar los resultados de la consulta SPARQL para datos enlazados mediante el uso de la razón de procedencia. Presenta un algoritmo basado en no-anotaciones para generar por qué-procedencia y mostrar su factibilidad para Linked Data.
Provenance and Linked Data in Biological Data Webs	Zhao *et al.* (2009)	Destaca la importancia de mantener la información de procedencia sobre los vínculos entre los elementos de datos de diferentes fuentes, y propone el uso de named graphs para hacer una declaración de procedencia sobre cada par de elementos de datos enlazados y cada liberación de una web de datos
Decentralized provenance-aware publishing with nanopublications	Kuhn et al. (2016)	El enfoque de la nano publicación se reduce a las ideas de subdividir los resultados científicos en aserciones atómicas, representadas en RDF, adjuntando información de procedencias en el nivel de afirmaciones individuales y tratando cada una de estas diminutas entidades como una publicación individual.
A Framework for Linking Educational Medical Objects: Connecting Web2.0 and Traditional Education	Al Fayez y Joy (2014)	El modelo de datos propuesto se centra en la posibilidad de descubrimiento, dado que su objetivo principal es exponer y enlazar los objetos educativos médicos recolectados (harvested) de diversas fuentes que tienen diferentes estándares o esquemas de metadatos.
Analyzing Linked Data Quality with LiQuate	Ruckhaus *et al.* (2014)	Cómo las consultas a una Red Bayesiana que modela datos RDF y dependencias entre propiedades, pueden usarse para estudiar problemas de calidad relacionados con la incompletitud de los enlaces y las ambigüedades entre etiquetas y enlaces. Para realizar esta verificación, se propone The LiQuate System Architecture
A Framework for Trust Establishment and Assessment on the Web of Data	Gao y Houben (2010)	Plantea un framework que hace uso de 3 diferentes estrategias para la generación de confianza: a) vocabulary-based, b) triple-based y c) interlinking-based.
Assessing Trust with PageRank in the Web of Data	Giménez-García *et al.* (2016)	Plantea la evaluación de la confianza en los datasets mediante la medición de su popularidad basada en la reutilización de los recursos de un dataset en otro dataset.
DC Proposal: Evaluating Trustworthiness of Web Content using Semantic Web Technologies	Pattanaphanchai (2011)	Plantea evaluar la confiabilidad de la información basada en la autoridad, la actualidad, la precisión y la relevancia.
Using linked open data to improve the search of	Piedra *et al.* (2013)	Presenta un motor de búsqueda facetada (faceted search engine) que consume datos

open educational resources for engineering students		desde LOCWD: Serendipity, versión basada en http://flamenco.berkeley.edu
Semantic Proximity with Linked Open Data: A Concept for Social Media Analytics	Mehta y Buch (2016)	Propone un framework que maneja un proceso de minería de datos basado en la proximidad semántica, para el análisis de las tendencias de empleo a través de LinkedIn
Scalable Property Aggregation for Linked Data Recommender Systems	Wenige y Ruhland (2015)	Propone un sistema de recomendación de datos enlazados escalable que calcula las preferencias en varias dimensiones de propiedad.
Exposición de metadatos de objetos de aprendizaje como datos enlazados: el caso del proyecto Open Discovery Space	Rajabi *et al.* (2014)	Plantea un método mediante el cual se exponen los metadatos de los objetos de aprendizaje del proyecto ODS, como datos enlazados
Trust Necessitated through Metrics: Estimating the Trustworthiness of Websites	Singal y Kohli (2016)	Describe una herramienta basada en la web que estima la confianza (de sitios web) impuesta por las métricas web recopiladas a través de similarweb.com.
A Framework for the Analysis and User-Driven Evaluation of Trust on the Semantic Web	Clarke (2014)	Evalúa las dimensiones de Procedencia y Licenciamiento, Reputación y Credibilidad, Verificabilidad.
Towards Trustworthy Linked Data Integration and Consumption	Knap (2013)	Presenta una herramienta de consulta y gestión de datos enlazados (ODCleanStore), capaz de proporcionar datos enlazados a los consumidores de datos, los cuales son limpiados, enlazados correctamente, integrados y de confianza, de acuerdo con las necesidades subjetivas del consumidor

Anexo 2 Otros Estándares y Esquemas de Metadatos.

Estándar de Metadatos	Descripción
Describing Archives: A Content Standard (DACS)	DACS facilita la descripción de materiales de archivo y creadores de materiales de archivo.
EXIF (Exchangeable Image File Format)	EXIF7 especifica los formatos que se pueden escribir y leer desde un archivo digital de imágenes fijas, imágenes, sonido y etiquetas auxiliares utilizadas por cámaras digitales, escáneres y otros sistemas que pueden manejar archivos de imagen y sonido.
e-GMS (e-Government Metadata Standard)	El e-GMS, como parte del Marco de Interoperabilidad del Gobierno Electrónico (e-GIF), define cómo el contenido de los organismos del sector público del Reino Unido (como páginas web y documentos) se administra, encuentra y comparte.
ISO 19115: 2003(E) — Geographic Information: Metadata	Describe información geográfica y servicios que proporcionan información sobre la identificación, el alcance, la calidad, el esquema espacial y temporal, la referencia espacial y la distribución de datos geográficos digitales.
PREMIS	Proporciona un diccionario de elementos básicos de metadatos utilizados para apoyar la preservación de objetos digitales y garantizar su usabilidad a largo plazo.
Resource Description and Access (RDA)	Estándar de catalogación y contenido de datos que se utiliza para crear y cambiar datos bibliográficos y proporciona orientación sobre cómo ingresar en esquemas de metadatos basados en bibliotecas.
Machine Readable Catalogue (MARC)	Estándar de metadatos muy extenso y formalizado para describir material bibliográfico y material no relacionado con libros o colecciones de archivos de la biblioteca.
Dublin Core Metadata Initiative (DCMI)	Los términos en vocabularios DCMI se usan en combinación con otros vocabularios compatibles en el contexto de los perfiles de aplicación y en base al Modelo abstracto DCMI [DCAM]. Dublin Core Metadata Initiative. Uno de los esquemas más generales existentes y también el más utilizado. http://dublincore.org/
Text Encoding Initiative (TEI)	Estándar para describir y codificar textos literarios, utilizado por bibliotecas, museos, editoriales y académicos individuales para presentar textos para investigación, enseñanza y preservación en línea.
Metadata Encoding and Transmission Standard (METS)	Desarrollado por la Biblioteca del Congreso puede estructurar o "empaquetar" otros metadatos o datos para su intercambio o entrega mediante la incorporación o enlace a otros metadatos basados en XML
Metadata Object Description Schema (MODS)	Conjunto de elementos bibliográficos utilizado para una variedad de propósitos, particularmente en bibliotecas. Consiste en un subconjunto de campos MARC y utiliza etiquetas basadas en el lenguaje, reagrupando elementos del formato bibliográfico MARC 21
MIDAS Heritage: the UK Historic Environment Data Standard	Estándar de datos marco para registrar todo tipo de activos digitales de patrimonio histórico y cultural (por ejemplo, edificios, sitios arqueológicos, naufragios, áreas de interés, artefactos y ecofactos), sus actividades relacionadas con la gestión, representación de mapas (SIG) y fuentes de información.
Encoded Archival Description (EAD)	Se utiliza para codificar ayudas para la búsqueda, como inventarios, índices o guías creados por repositorios de archivos y manuscritos para proporcionar información sobre colecciones específicas o materiales de archivo, para su uso en un entorno en red. Disponible en: http://www.loc.gov/ead/
IEEE Learning Object Metadata (LOM)	Modelo de datos abiertos desarrollado por el Instituto de Ingenieros Eléctricos y Electrónicos, codificado en XML, que se utiliza para describir un objeto de aprendizaje digital y recursos. Admite la

Estándar de Metadatos	Descripción
	reutilización de los objetos de aprendizaje, para ayudar a la detección y para facilitar su interoperabilidad, en los sistemas de gestión de aprendizaje en línea (LMS). Disponible en: http://ltsc.ieee.org/wg12/
Functional Requirements for Bibliographic Records (FRBR)	Modelo conceptual internacional publicado por IFLA, para describir los recursos de información como entidades y sus relaciones dentro del contexto de una biblioteca. FRBR no es un esquema de metadatos, sino que se utiliza para analizar esquemas existentes o influir en nuevos esquemas o estándares de contenido.
NISO Technical Metadata for Digital Still Images	Es un "diccionario de datos" de categorías de datos utilizados para describir los aspectos técnicos de las imágenes digitales basadas en ráster (por ejemplo, TIFF, JPEG, GIF) y permitir a los usuarios desarrollar, intercambiar e interpretar archivos de imágenes digitales.
PBCore (Public Broadcasting Metadata Dictionary)	Se utiliza para describir medios digitales y analógicos, para su uso por televisión, radio y emisoras web para recuperar y compartir fácilmente entre colegas, sistemas de software, instituciones, socios comunitarios y de producción, particulares y educadores.
SMPTE (Society of Motion Picture Technical Experts) Data Dictionary	Usado para describir datos de video, audio y multimedia en sus diversas formas.
SPECTRUM	Utilizado para documentar los procedimientos y objetos de un museo
TV-Anytime	Especificación de metadatos utilizada en las Guías electrónicas de programas (EPG) o en páginas web para describir contenido multimedia.
XMP (Extensible Metadata Platform)	Estándar abierto de Adobe basado en XML, utilizado dentro del software de imágenes de Adobe y otras herramientas.
Cataloguing Cultural Objects (CCO)	Estándar de contenido de datos, guías para ingresar datos en esquemas relacionados con objetos culturales, a saber, VRA Core y CDWA Lite.
P/Meta	Estándar de metadatos flexible que se utiliza para intercambiar medios entre organizaciones profesionales de transmisión de medios.
MODS: Metadata Object Description Schema.	Conjunto de elementos bibliográficos utilizado para una variedad de propósitos, particularmente en Bibliotecas. Consiste en un subconjunto de campos MARC y usa etiquetas basadas en el lenguaje, reagrupando elementos del formato bibliográfico MARC 21. Disponible en: http://www.loc.gov/standards/mods/
IFLA: Metadata Resources for Digital Libraries.	Disponible en: http://www.ifla.org/II/metadata.htm
CIMI: Computer Interchange of Museum Information.	Disponible en: http://www.cimi.org/ (El Consorcio cerró sus operaciones en 2003).
MCF: Meta Content Framework.	Disponible en: http://www.textuality.com/mcf/NOTE-MCF-XML.html y http://www.w3.org/TR/NOTE-MCF-XML/ (para colecciones de información en red usando XML)
UDEF: Universal Data Element Framework.	Disponible en: http://www.udef.org/
MPEG-7: Multimedia Content Description Interface.	Disponible en: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm (Metadatos para contenido audiovisual, esto es, para describir contenido multimedia)
GILS: US Governmment Information Locator Service.	Disponible en: http://www.usgs.gov/gils/
CSDGM: Content Standard for Digital Geospatial Metadata.	Disponible en: http://www.fgdc.gov/metadata/contstan.html
W3C Metadata activity.	Disponible en: http://www.w3.org/Metadata/Activity.html

Estándar de Metadatos	Descripción
W3C Semantic Web activity.	Disponible en: http://www.w3.org/2001/sw/

Tomado de:

a. https://aprendizbibliotecologa.wordpress.com
b. Metadata diversity - An overview on Metadata Standards and Schemas. P. Ramesh, V. Lalitha, J. Vivekavardhan.
c. Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. P. Ramesh, J. Vivekavardhan and K. Bharathi. DESIDOC Journal of Library & Information Technology, Vol. 35, No. 3, May 2015, pp. 193-199 DOI: 10.14429/djlit.35.3.8074. 2015.

Anexo 3. Entrevistas y Diálogos realizados.

a. ***William Toloza Castillo***: Productor de contenidos digitales en proyectos E-learning con experiencia de 15 años en UNAB Virtual:

   - Facultad de Comunicación Social y Artes audiovisuales,
   - Docente virtual de la Maestría en Tecnología educativa (UNAB-Tec de Monterrey),
   - Docente presencial en la Maestría en Educación de la UNAB).

   Emprendedor TIC y Mentor Experto del Ministerio de las TIC en el programa Apps.co. Asesor de 10 Universidades en educación virtual con el Ministerio de Educación Nacional: Universidad Politécnico Grancolombiano, Universidad Manuela Beltrán, Universidad Externado de Colombia, Universidad Corporativa de Colombia, Universidad de la Costa, Unisangil, Universidad de Ibagué, Universidad de Nariño. Consultor en Producción e-Learning: Telefónica Educación Digital y Universidad Externado de Colombia. Comunicador social y Maestro en Tecnología Educativa. Socio y Director Creativo en Arte y Animación S.A.S. Cofundador de Edumagicos.co. Certificado en Design thinking y capacitador corporativo en Innovación, Emprendimiento, Lean LaunchPad, Business Model Canvas, Creatividad, Ponente Internacional en Educación y Medios Digitales. Cofundador y compositor y guitarrista con la banda de Rock Zendas.

b. ***Tatiana Valencia Molina***: Directora CRAI - Centro de Recursos para el Aprendizaje y la Investigación, Vicerrectoría Académica. Pontificia Universidad Javeriana – Cali.
c. ***Ana Lucila Herrera Silva***: Jefe Sección de Análisis de Información, Biblioteca General Alfonso Borrero Cabal, S.J. Pontificia Universidad Javeriana.
d. ***Viky Julieta Arias***: Ingeniera de Proyectos, Biblioteca General Alfonso Borrero Cabal, S.J. Encargado Proyectos Repositorios y Plataformas. Pontificia Universidad Javeriana.
e. ***Alejandro Montes Briceño***: Coordinador de Ambientes Educativos Virtuales, Centro Ático. Pontificia Universidad Javeriana.
f. ***Pablo Jaramillo***: Profesional en Ingesta y Archivo, Coordinación de Producción, Centro Ático. Pontificia Universidad Javeriana.
g. ***David Felipe Farias Cárdenas***: Profesional en Análisis de Información, Biblioteca General Alfonso Borrero Cabal, S.J. Encargado Proyecto Metadatos. Pontificia Universidad Javeriana.
h. ***Wilson Leandro Pardo***: Profesional de Vicerrectoría Académica, Pontificia Universidad Javeriana – 2019. Vicerrector académico, Corporación Universitaria de Asturias – 2016. Decano de la Escuela de Ingenierías, CUN Corporación Unificada Nacional de Educación Superior – 2015. Coordinador general de proyectos, Universidad de los Andes – 2014. Coordinador de Producción de Contenidos, Universidad del Rosario – 2011.
i. ***Luis Ali Ortiz Martínez***: Repositorio Institucional-RIUD, Universidad Distrital Francisco José de Caldas.
j. ***Enith Mireya Zarate Peña***: Coordinadora Biblioteca-UDFJC, Universidad Distrital Francisco José de Caldas.
k. ***Liliana Cadena Montenegro***: Diseñadora Pedagógica en PlanEsTic- UD, Universidad Distrital Francisco José de Caldas. Magister en Educación en Tecnología, UDFJC. Experiencia en docencia universitaria, formación de profesores, diseñadora instruccional de ambientes virtuales de aprendizaje, tutora virtual. Experiencia en el acompañamiento a la creación de nuevos programas académicos y en procesos de acreditación.

Anexo 4 Cuadro comparativo de estandartes de metadatos.

Grupo	Tag	Sena	LOM-Co	LOM-ES	LOM-UK	CEM	LRMI	MLR
General	Identificador	O		O	O	O		O
	Título del Recurso	O	O	O	O	O	O	O
	Idioma	O	O	O	O	O	O	O
	Descripción (Abstract)	O	O	O	O	O	O	O
	Palabras Claves	O	O	O		O		O
	Cobertura	R						O
	Estructura General	O				O		
	Nivel de Agregación	R		O		O		
	Subtitulo	R						
	Catálogo General	O						
	Entrada General	O						
Ciclo de vida	Versión	O	O			O		
	Estado	O				O		
	Rol de Ciclo de Vida	O	O	O	O	O	O	O
	Entidad de Ciclo de Vida	O	O					O
	Fecha de Ciclo de Vida	O	O				O	O
Metametadata	Identificador	R			O	O		
	Catálogo de Metadatos	R						
	Entrada de Metadatos	O						
	Rol de Metadatos	R		O	O			
	Entidad de Metadatos	R		O				
	Fecha de Metadatos	R		O				
	Esquema de Metadatos	R		O	O	O		
	Lenguaje	R		O	O	O		
Técnica	Formato	O	O			O	O	O
	Tamaño	O	O			O		
	Ubicación (uri)	O	O	O	O	O	O	O
	Tipo (Requisitos)	R	O	O		O		
	Nombre Técnico (software necesario)	R	R			R		
	Versión Mínima	O		O			O	
	Versión Máxima	P						
	Otros requerimientos de plataforma	P	O	O				
	Duración	R						
	Uso	R						
Educativo	Tipo de Interactividad	O	O			O	O	

	Tipo de Recursos de Aprendizaje	O	O	O		O		O
	Nivel de Interactividad	O	O			O		
	Población Objetivo (Rol de usuario final previsto)	R	O	O		O	O	O
	Contexto Educativo	R	O	O		O	O	O
	Densidad Semántica	R						
	Rango Típico de Edad	R		O			O	O
	Dificultad	R						
	Tiempo de Aprendizaje	R					O	
	Descripción	O		O		O		O
	Idioma	O		O			O	O
	Contexto	R					O	

Derechos	Contribuyentes	O						
	Costo	R	O			O		
	Licenciamiento y otras restricciones	O	O	O	O	O		
	Descripción	O					O	
	Derecho del recurso	O	O		O		O	O
	Disponibilidad	O						

Relación	Tipo	P						O
	Identificador	P						
	Descripción	P						
	Catálogo de relaciones	P						
	Entrada de relaciones	P						

Anotación	Persona de Anotación	P		O		O		O
	Fecha	P		O		O		
	Descripción	P	O	O		O		O

Clasificación	Propósito	P						O
	Origen	P	O					O
	Taxón	P	O	O		O		O
	Identificador	P						
	Entrada de clasificación	P		O				
	Descripción de la clasificación	P		O				
	Palabras clave	P		O		O		
	Nombre de Programa	O						
	Código de Programa	O						
	Facultad	O						

Anexo 5. Perfil de Aplicación PA-REA.

Perfil de Aplicación para Recursos Educativos Abiertos (PA-REA)

Para garantizar que la generación de modelos para el enlace de recursos educativos abiertos, una de las actividades principales fue diseñar un Perfil de Aplicación que se utilizaría para describir los recursos educativos.

OBJETIVO
Establecer lineamientos para almacenar, preservar, reutilizar, enlazar y facilitar el acceso y uso de los recursos educativos abiertos; garantizando de esta manera la integridad, usabilidad, accesibilidad e interoperabilidad de la información.

ALCANCE
El alcance de este documento es documentar el perfil de aplicación de metadatos para los recursos educativos abiertos. Un perfil de aplicación se define como un esquema que consta de elementos de datos extraídos de uno o más espacios de nombres, combinados y optimizados para una aplicación local en particular. El esquema resultante combinan los esquemas existentes para ajustarse mejor a las necesidades de describir y compartir recursos de aprendizaje.

ELEMENTOS DEL PERFIL DE APLICACIÓN: DESCRIPCIÓN Y PAUTAS DE CATALOGACIÓN

La siguiente tabla proporciona una descripción general de los elementos incluidos en el Perfil de la Aplicación. También incluye información breve sobre los vocabularios controlados utilizados y si es obligatorio o no.

Metadatos:

Categoría	Metadatos
General	Identificador (ISSN, ISBN, Cita, URI, handle), Titulo, Subtitulo, Palabras Claves, Descripción, Idioma
Ciclo de vida	Versión, Contribuyentes (Rol, Entidad), Fecha de versión inicial, fecha de publicación
Técnica	Formato, Tamaño, Ubicación (uri), Tipo, Versión mínima
Educativo	Tipo de Interactividad, Nivel de Interactividad, Tipo de recurso, Población o usuario final, Contexto Educativo, Nivel Educativo, Rango de edad, Descripción Educativa (Competencias), Idioma
Derechos	Licenciamiento, Descripción (derechos del recurso), Costo, fecha del Licenciamiento
Anotación	Persona, Fecha, Descripción.
Clasificación	Institución, Organización, Grupo, División, Taxón, Curso

Diccionario de Datos de PA-REA.

Elemento	Descripción	Criterio
Identificador	**GENERAL.IDENTIFIER - LOM 1.1** Definición: Característica irrepetible que identifica de forma exclusiva el recurso en sí mismo y no una versión, formato o copia particular de este recurso. Pautas: • Los identificadores (números estándar) son números tomados del ISSN, ISBN, URL, números de trabajo y otros números no asignados por la agencia de catalogación. Puede estructurarse de la siguiente manera: - identificador de catálogo (es decir, el nombre del esquema de identificación o catalogación del identificador), y - una entrada (es decir, el valor del identificador dentro del esquema de identificación que identifica el recurso).	Ob
Titulo	**GENERAL.TITLE - LOM 1.2** Definición: Nombre del recurso. Pautas: • Puede contener múltiples valores, pero cada uno debe estar en un idioma diferente y semánticamente equivalente. • Cuando los títulos no tienen sentido, son incompletos, ambiguos, engañosos o demasiado generales, se recomienda el enriquecimiento de títulos. Deben agregarse palabras (enriquecimiento) que corregirán las deficiencias y reflejarán el contenido del documento.	Ob
Subtitulo	**GENERAL.DESCRIPTION.SUBTITLE - LOM 1.4** Definición: Información secundaria relacionada con el título. Pautas: • Debe usarse con moderación y solo junto con el elemento Título. • Debe usarse para ingresar: elementos de título que no pertenecen adecuadamente al título y subtítulo pero que se anexan a él. • El subtítulo amplía el área del conocimiento del recurso. • Se diligenciará en minúsculas a excepción de los nombres propios	Op
Palabras Claves	**GENERAL.KEYWORD - LOM 1.5** Definición: Una lista clasificada de términos o palabras clave, en un campo particular, para su uso en indexación y recuperación de información. Pautas: • Usar "LOM 9.2 Taxon Path" en lugar de General.Keyword cuando los valores se derivan de un Tesauro. • Permite diez diferentes palabras claves, en minúsculas, y se admiten palabras compuestas.	Ob

Elemento	Descripción	Criterio
	• La interfaz de usuario de etiquetado debe diseñarse de modo que brinde la posibilidad de ingresar palabras clave individuales en diferentes idiomas en un elemento. • Deben usarse los términos más específicos para describir el contenido del recurso • Cada término o frase debe usar un elemento de palabra clave separado y se deben evitar las frases largas. • Es recomendable utilizar tesauros como UNESCO para su definición	
Cita	**GENERAL.DESCRIPTION.CITE - LOM 1.4** Definición: cita del recurso publicado. Pautas La cita del recurso, para aquellos recursos que apliquen, debe contener como mínimo: • Autores • Título del recurso • Título del contenedor donde está publicado • ISSN/ISBN, • Volumen, • Número o Serie, • Página inicial - final	Op
Descripción (Abstract)	**GENERAL.DESCRIPTION - LOM 1.4** Descripción (Abstract) Definición: Resumen o breve descripción del contenido del recurso. Pautas • Cada descripción puede tener hasta 2.000 caracteres de longitud. • No debe haber más de una descripción por idioma. • La información específica de una versión, formato o copia particular de este objeto no debe estar presente en esta descripción. • Un identificador de versión: identificador único de una versión del recurso. Cada versión debe tener un identificador, el cual debe ser diferente al identificador del recurso. • Una descripción de versión: describe los elementos o mejoras claves de la versión • No es necesario que la descripción este en el lenguaje y los términos apropiados para los usuarios del recurso que se describe. La descripción debe estar en lenguaje y términos apropiados para aquellas personas que deciden si el recurso es o no utilizable. • Se recomienda construirla incluyendo los siguientes elementos: - Tipo de recurso según la tipología de Recursos Educativos, la cual contiene categorías no excluyentes, condición que debe reflejarse. - Presentación de la temática. - Áreas de conocimiento que abarca el recurso; cuando se considere pertinente, se debe nombrar la asignatura.	Ob
Idioma General	**GENERAL.LANGUAGE - LOM 1.3** Definición: Idioma de la descripción del recurso.	Ob

Elemento	Descripción	Criterio
	Pautas • Las versiones deben tener “Language” como atributo de descripción • Para el contenido de esta característica se puede usar un vocabulario controlado como ISO 639-1 (2 letras para el idioma) o ISO 639-2 (3 letras para el idioma). • Si el recurso no tiene contenido lingual, el valor apropiado para este elemento es "ninguno".	
Versión	**LIFE CYCLE.VERSION – LOM 2.1** Definición: Nombre o número de la última versión del recurso. Pautas: • Debe indicarse la fecha a la que corresponde la versión	Ob
Fecha de Version Inicial	**LIFE CYCLE.CONTRIBUTE.DATE – LOM 2.3.3** Definición: Fecha de la versión inicial. Pautas: • Se debe escribir como año, mes, día • Año de 4 dígitos, mes y día de dos dígitos	Ob
Contribución	**LIFE CYCLE.CONTRIBUTE – LOM 2.3** Definición: Describe quién ha contribuido al recurso. Pautas • El "LOM 2.3 Life Cycle.Contribute" es un elemento contenedor. • Si se usa este elemento, en los elementos ‘LOM 2.3.1 Life Cycle.Contribute.Role’ y ‘LOM 2.3.2 Life Cycle.Contribute.Entity’ deben existir valores, es decir, se convierten en elementos obligatorios. Se recomienda el uso de "LOM 2.3.3 Life Cycle.Contribute.Date", pero no es obligatorio.	Re
Rol	**LIFE CYCLE.CONTRIBUTE.ROLE – LOM 2.3.1** Definición: Describir el rol del contribuyente. Pautas Para su diligenciamiento se puede usar el siguiente vocabulario: • Autor: Una entidad principalmente responsable de hacer el contenido del recurso. Un autor puede ser una persona, institución, grupo u otra entidad. Si un equipo ha creado el recurso, será necesario identificar roles distintos de "Autor" (por ejemplo, "Diseñador gráfico", "Implementador técnico", etc.) • Proveedor de contenido: La persona o entidad que proporciona contenido para el recurso. • Validador educativo: La persona o entidad responsable de confirmar la integridad educativa del recurso. • Diseñador gráfico: El especialista o entidad responsable de la construcción de los elementos visuales de un recurso. • Financiador: La persona, institución o agencia de financiación originalmente responsable de iniciar el proceso de desarrollo. • Diseñador instruccional: El especialista o entidad responsable de aplicar los principios basados en la investigación al diseño del recurso.	Ob

Elemento	Descripción	Criterio
	• Licenciante: La persona o entidad que posee los derechos vinculados al recurso. • Editor: El individuo u organización responsable de hacer que el recurso esté disponible en su forma actual, como una editorial, un departamento universitario o una entidad corporativa. • Guionista: La persona o entidad responsable de la creación de un texto leído o realizado en un recurso tipo audio, video y / o interactivo. • Experto en la materia: La persona o entidad que es experta en el dominio perteneciente al recurso • Implementador técnico: El especialista o entidad responsable de la construcción de los elementos técnicos de un recurso (generalmente un programador de software). • Validador técnico: La persona o entidad responsable de confirmar la integridad técnica del recurso. • Validador: La persona o entidad responsable de confirmar la integridad general del recurso. • Desconocido En general: • El nombre de las personas debe ingresarse en la siguiente secuencia: apellido, nombre (s) inicial (es). • Este campo también puede incluir la Organización o un grupo de personas identificadas por un nombre responsable de crear o hacer contribuciones al contenido intelectual del recurso.	
Entidad	**LIFE CYCLE.CONTRIBUTE.ENTITY – LOM 2.3.2** Definición: Identificación e información sobre entidades (es decir, personas, organizaciones) que contribuyen al recurso Pautas • Información mínima sobre una persona: nombre (nombre y apellido) y afiliación. Información mínima sobre la organización: nombre.	Re
Fecha del Ciclo de Vida	**LIFE CYCLE.CONTRIBUTE.DATE – LOM 2.3.3** Definición: Fecha de la contribución, o Fecha en que el recurso se puso a disposición del público. Pautas: • Ingrese el año de publicación como un número de cuatro dígitos.	Ob
Formato Técnico	**TECHNICAL.FORMAT – LOM 4.1** Definición: El formato de los materiales o el material físico en el que se pueden almacenar o registrar los datos. Pautas: Para su diligenciamiento se puede usar el siguiente vocabulario: • CD-Rom / DVD: en forma física. • Documento en papel: tinta en papel, no electrónica. • Documento electrónico: se utiliza para cualquier tipo de texto electrónico u hoja de cálculo, p. Microsoft Word, Adobe Acrobat PDF o Microsoft Excel.	Ob

Elemento	Descripción	Criterio
	• Compresión de archivos. Los formatos de archivo son: zip, tar, rar. • Diapositivas: se usa para la presentación de diapositivas, p. Ej. en formato Microsoft PowerPoint. • Imágenes animadas: Los formatos de archivo de ejemplo son: gif • Imágenes: Los formatos de archivo de ejemplo son: gif, jpg, png, etc. • Sitio web: se utiliza para una colección completa de páginas web y otra información, independientemente de las tecnologías utilizadas para desarrollarlo. • Audio: se utiliza para cualquier tipo de archivos de audio. Los formatos de archivo de ejemplo son: ".mp3", ".wav", ".mp4", ".midi", ".mpga", etc. • Video: se usa para cualquier tipo de archivos de video. Los formatos de archivo de ejemplo son: ".avi", ".wmv", ".mpeg", ".mov", etc. • Animación/multimedia. Los formatos de archivo de ejemplo son: Flash, Shockware, AutorWare. Dreamweaver • Otro	
Tamaño del Formato	**TECHNICAL.SIZE – LOM 4.2** Definición: Tamaño del recurso digital en bytes para el caso de recursos textuales, aplicaciones o herramientas. Para el caso de recursos tipos sonoros, visuales o audiovisuales, esta propiedad se expresa en tiempo de duración. Pautas: • Peso del archivo en la nomenclatura internacional, Byte, GB,MB	Re
Ubicación Técnica	**TECHNICAL.LOCATION – LOM 4.3** Definición: Ubicación del recurso. Pautas: Indica la ubicación del recurso dentro de las siguientes opciones: • Ubicación: Descripción textual acerca de cómo localizar el recurso. • Web. Corresponde a la ubicación en la web del recurso, puede ser una URI o una URL. • Fuente. Corresponde a la ubicación en la web de los archivos fuentes del recurso, puede ser una URI o una URL.	Ob
Tipo Técnico	**TECHNICAL.REQUIREMENT.ORCOMPOSITE.TYPE – LOM 4.4.1.1.** Definición: Describir los requisitos técnicos obligatorios para el uso del recurso. Pautas: • Para cumplir cada requisito se pueden describir varias opciones. • En caso de necesitar un software específico. • Tecnología requerida para utilizar el recurso como sistema operativo o navegador.	Re

Elemento	Descripción	Criterio
Versión Mínima	**TECHNICAL.REQUIREMENT.ORCOMPOSITE.MINIMUNVERSION – LOM 4.4.1.3.** Definición: Describir en caso de requerir un software específico. Pautas: • Indicar nombre tecnología para utilizar recurso.	Re
Tipo de interactividad	**EDUCATIONAL.INTERACTIVITYTYPE – LOM 5.1** Definición: Tipo de interactividad que ofrece el recurso Pautas Seleccionar el tipo de interactividad. Para su diligenciamiento se puede usar el siguiente vocabulario: • Alto. recursos que en su contenido inducen a la participación directa por parte de los usuarios finales. (interactuar, introducir información semánticamente significativa, tomar decisiones o realizar algún tipo de actividad productiva) • Descriptivo: recursos que en su contenido inducen al usuario final a asimilar conceptos que le son expuestos. No solicita ningún tipo de acción interactiva. • Mixto. Combina las características de los tipos alto y descriptivo de los recursos. • Sin definir	Re
Nivel de interactividad	**EDUCATIONAL.INTERACTIVITYLEVEL – LOM 5.3** Definición: El nivel de comunicación entre el recurso y el usuario, determina la influencia que el usuario puede tener sobre el comportamiento del recurso. Pautas: Indique el grado en que el alumno puede influir en el aspecto o comportamiento del componente. Para su diligenciamiento se puede usar el siguiente vocabulario: - Muy bajo: recurso expositivo - Bajo: recurso expositivo con poca interactividad - Medio: interactividad activa - Alto: interactividad activa con funcionalidades - Muy alto: interactividad activa con múltiples actividades	Op
Tipo de Recurso	**EDUCATIONAL.LEARNING RESOURCE TYPE – LOM 5.2** Definición: Tipo identifica la naturaleza o género del contenido del recurso. Pautas: Se debe seleccionar el tipo de la lista analizando el contenido del recurso. A menudo, el tipo es evidente en el título del recurso. Para su diligenciamiento se puede usar el siguiente vocabulario: • Audio: sonido grabado. • Datos: información visual, fáctica o numérica, sin procesar u organizada en listas, tablas, etc. • Imagen: cualquier imagen fija o representación gráfica de objetos (reales o no).	Ob

Elemento	Descripción	Criterio
	• Texto: una unidad de escritura. • Video: cualquier imagen en movimiento, incluida la animación. • Aplicación: Software informático diseñado para permitir a los usuarios finales realizar una tarea específica o un grupo de tareas. • Evaluación: su propósito principal es la evaluación de la capacidad, comprensión, habilidades, desempeño o progreso del alumno. • Broadcast: grabación de cualquier programa de radio o televisión transmitido por las ondas o sus equivalentes digitales (podcasts, webcasts, screencasts, etc.) • Estudio de Caso: una descripción detallada y análisis de un solo proyecto en una ubicación o región específica. Puede contener lecciones aprendidas que probablemente ayuden a modificar y mejorar futuras actividades. • Curso: Una colección estructurada de materiales de enseñanza y aprendizaje (incluidos tutoriales) destinados a lograr una amplia gama de objetivos durante un período prolongado de tiempo. • Lección: diseñadas para enseñar un tema específico y pueden incluir instrucciones de enseñanza (metas educativas, objetivos de aprendizaje, procedimientos, etc.) o recursos para el uso directo de los estudiantes, desde texto imprimible hasta actividades interactivas en línea. Incluye cualquier grabación de audio o video que pueda incluir una demostración, presentación o una conferencia. • Ejercicios y prácticas: el ejercicio es una tarea que se realiza como parte del plan de aprendizaje u otra unidad más grande de instrucciones destinadas a desarrollar o evaluar habilidades y conocimientos particulares. Por ejemplo, prueba, cuestionario, examen, experimento, actividad. • Juego educativo: diseñado para enseñar a los alumnos sobre un tema determinado o para ayudarlos a aprender una habilidad mientras juegan o se entretienen. • Actividad orientada a la indagación: Material para una actividad en la que parte o la totalidad de la información con la que interactúan los estudiantes proviene de recursos en Internet (webquest, caza del Tesoro ...) • Experimento: se realizan acciones u operaciones (como un procedimiento científico) para descubrir algo desconocido, probar una hipótesis, establecer o ilustrar alguna verdad conocida. • Exploración: anima a los alumnos a explorar y realizar sus propias investigaciones para comprender un fenómeno, proceso o actividad o para resolver un problema específico. Algunos de estos recursos pueden tener un mínimo de instrucciones sobre cómo proceder y el énfasis está mucho más en el "aprendizaje por descubrimiento" en lugar de la capacitación didáctica. • Glosario: colección de términos especializados y sus significados generalmente organizados en un orden específico. • Directrices: proporcionan instrucciones y consejos para realizar una tarea y sugieren posibles enfoques. Ejemplos de pautas incluyen guías de instalación, guías de planificación, plan de estudios, plan de estudios, manuales, etc.	

Elemento	Descripción	Criterio
	- Guía: manual o tutorial que proporciona orientación sobre un tema en particular y que, por lo general, también se debe tener a mano como referencia. - Plan de estudios: proporciona al profesor una descripción detallada de todos los elementos necesarios para impartir con éxito una lección específica (objetivos de la lección, tiempo requerido, recursos, etc.). • Actividad abierta: Material para proyectos artísticos y ejercicios creativos. Estos son proyectos y ejercicios que no son muy limitados o limitados. Los juegos complicados que requieren más que lógica simple pertenecen a esta categoría. • Otro: Un tipo de recurso que no está en la lista. • Proyecto: Cualquier recurso que describe ideas y procedimientos para el trabajo del proyecto. • Material promocional: cualquier material o comunicación creada para la sensibilización del público con fines educativos. Esto normalmente incluye la publicación en cualquier periódico, revista o medio similar, o material creado para su transmisión por televisión, radio u otro medio electrónico. Algunos ejemplos comunes son volantes, carteles, hojas informativas y boletines. • Referencia: Base de datos, diccionario, enciclopedia, glosario, etc. que es una fuente general de información o que proporciona información específica sobre un tema o actividad. • Módulo: secuencia de materiales de enseñanza y aprendizaje destinados a lograr una amplia gama de objetivos, a menudo durante un período prolongado de tiempo. • Manual de capacitación: recurso diseñado para enseñar al lector cómo hacer algo, como usar un software, cultivar una planta, administrar una actividad, etc. • Técnicas de monitoreo y evaluación: proporcionan un proceso formal para evaluar el desempeño y el impacto utilizando indicadores que ayudan a medir el progreso hacia las metas definidas. • Mejores prácticas: proceso, técnica o metodología que se ha adoptado con éxito en la entrega de un resultado particular y se reconoce que mejora el rendimiento y la eficiencia en áreas específicas. • Juego de rol: Material para una actividad que implica la participación activa de un alumno en una situación concreta. El alumno debe desempeñar un papel a través del cual debe comprender mejor el contenido o el tema tratado en el recurso. • Simulación: explica un procedimiento, proceso o principio, no mostrando objetos o acciones reales, sino representándolos, por ejemplo, mediante una animación. • Libro de texto: Proporciona materiales completos para temas específicos (uso para: capítulos u otros componentes típicos del libro). • Herramienta: Editores y otros tipos de programas para producir algo. Los editores pueden procesar, p. texto o imágenes y se pueden usar para crear y editar otros LO. Las herramientas también pueden realizar cálculos o conversiones. • Sitio web: Una colección de páginas web (que contiene una colección de documentos y archivos con hipervínculos) típicamente identificados por una dirección IP común en la World Wide Web en Internet.	

Elemento	Descripción	Criterio
	• Redes sociales: plataforma para compartir marcadores. Una colección de marcadores alojados en una plataforma que facilita el intercambio y el etiquetado de marcadores en línea. • Plataforma para compartir imágenes: Una imagen o colección de imágenes alojadas en una plataforma que facilita el intercambio y el etiquetado de imágenes generadas por el usuario. • Plataforma de intercambio de referencias: Una colección de referencias alojada en una plataforma que facilita el intercambio y el etiquetado de citas bibliográficas. • Plataforma para compartir sonido: Una grabación de sonido o música o una colección de grabaciones alojadas en una plataforma que facilita compartir y etiquetar sonidos o música generados por el usuario. • Plataforma de intercambio de videos: Un video o colección de videos alojados en una plataforma que facilita las transmisiones y el etiquetado del contenido de video generado por el usuario. • Weblog: Una publicación en línea (a veces llamada simplemente "blog") que consiste principalmente en una serie de artículos, escritos regularmente por un individuo o un grupo. • Wiki: Un tipo de sitio web que permite a los usuarios agregar y editar contenido fácilmente y es especialmente adecuado para la escritura colaborativa.	
Rol Usuario Final	**EDUCATIONAL.INTENDED END USER ROLE - LOM 5.5** Definición: Usuario (s) principal para el cual se diseñó este recurso. Pautas: Los elementos se ordenan de manera que el primer valor sea el más utilizado. Para su diligenciamiento se puede usar el siguiente vocabulario: • Consejero asesor como parte del 'servicio de orientación • Aprendiz: Alumno que trabaja con un recurso. • Gerente: gestiona la entrega de un recurso • Acudiente: Persona que ejerce las funciones padre o tutor. • Profesor: Persona que ejerce las funciones de maestro o instructor • Otro: ninguno de los anteriores Autor, Estudiante, Gerente y Maestro pertenecen al vocabulario de LOM y los otros tres (Consejero, Acudiente y Otro) son extensiones.	Ob
Nivel Educativo	**EDUCATIONAL.CONTEXT - LOM 5.6** Definición: El entorno principal o nivel educativo dentro del cual se pretende que tenga lugar el aprendizaje y el uso del recurso. Pautas • Se debe usarse junto con el elemento "LOM 5.7 Rango de edad típico" para expresar el contexto completo. • Al seleccionar el nivel de audiencia apropiado para el material, los usuarios que realicen búsquedas podrán reducir los materiales apropiados para aquellos a quienes están enseñando. • Para su diligenciamiento se puede usar el siguiente vocabulario: - Preescolar	Re

Elemento	Descripción	Criterio
	- Básica Primaria - Media - Media profesional - Superior - Técnico - Superior - Tecnólogo - Superior - Profesional - Superior - Posgrado - Otro: nivel educativo que no es uno de los anteriores.	
Contexto Educativo	**CLASSIFICATION.DESCRIPTION – LOM 9.2** Definición: Hace referencia al contexto en cual se van a utilizar los recursos. Pautas: Para su diligenciamiento se puede usar el siguiente vocabulario: • Educación formal: Educación regular y otra educación después del jardín de infantes y antes de la educación superior. • Educación continua: La educación de adultos / continua no está relacionada con la capacitación laboral en este contexto. • Educación a distancia: Entrega de instrucción que no restringe al estudiante a estar físicamente presente en la misma ubicación que el instructor. • Administración educacional: Gestión y administración de instituciones educativas y de formación. • Desarrollo profesional: Formación o educación relacionada con la mejora de las habilidades profesionales. • Educación especial: Diseñado o provisto para personas que tienen necesidades educativas especiales que les impiden recibir (en su totalidad) educación general. • Educación vocacional (para el trabajo): Formación o educación perteneciente o relacionada con una vocación u ocupación.	Re
Rango de Edad	**EDUCATIONAL.TYPICAL AGE RANGE - LOM 5.7** Definición: edad típica del usuario del recurso. Pautas • Rango de edad mínimo a máximo expresado en años integrales y separados por un guion (Mínimo-Máximo). • El valor mínimo o máximo se puede establecer en U (indefinido), lo que significa que el rango se extiende de esa manera. p.ej. "10 -U" significa desde el año 10 en adelante; "U-12" significa 12 años o menos. • Si es necesario, se puede proporcionar información equivalente en lenguaje natural	Re
Descripción Educativa	**EDUCATIONAL.DESCRIPTION – LOM 5.10** Definición: Proporcionar una descripción textual de los usos educativos del recurso, en cuanto a la intencionalidad educativa o meta a alcanzar a través del contenido del recurso. Pautas En su redacción se debe tener en cuenta a quién va dirigido, qué se espera que haga y qué contenido (temáticas) se espera dominar al finalizar (Competencias)	Ob

Elemento	Descripción	Criterio
Licenciamiento y Otras Restricciones	**RIGHTS.COPYRIGHT AND OTHER RESTRICTIONS - LOM 6.2** Definición: Indica si se aplican derechos de autor u otras restricciones al recurso. (Tipo de licencia: publica o restringida) Pautas • Vocabulario (Si o No) • Valores: SI - Los derechos de autor y / u otras restricciones se aplican al recurso. NO - No se aplican derechos de autor u otras restricciones al recurso (El recurso pertenece al dominio público) • Si el valor de este elemento es "sí", entonces el elemento " 'LOM 6.3 Rights.Description'" se convierte en un elemento obligatorio.	Ob
Descripción de los Derechos	**RIGHTS.DESCRIPTION - LOM 6.3** Definición: Descripción textual de los derechos de autor u otras restricciones que se aplican al recurso. Pautas: • Es obligatorio incluir una URL de licencia (Creative Commons u otra) en este elemento. • Cuando un recurso se distribuye bajo una licencia de derechos de autor (copyright), es obligatorio poner la URL en la declaración de licencia o términos de uso en una cadena de descripción • Por lo general, los Derechos contendrán una declaración de gestión de derechos para el recurso o harán referencia a un servicio que brinde dicha información. • La información sobre derechos a menudo abarca los derechos de propiedad intelectual, derechos de autor y varios derechos de propiedad. Esto proporciona información sobre los derechos de autor y los derechos de acceso de los usuarios al recurso.	Ob
Costo de los Derechos	**RIGHTS.COST - LOM 6.1** Definición: Indica si el uso del recurso requiere algún pago. Pautas • Vocabulario (Si o No) • Valores: SI - el uso del recurso requiere de pago. NO - El uso del recurso no requiere pago • Si el valor de este elemento es "sí", el elemento "Derechos LOM 6.3. Descripción" debe contener más información sobre el costo, por lo que el elemento "Derechos LOM 6.3. Descripción" se vuelve obligatorio.	Re
Fecha del Licenciamiento	**RIGHTS.DATE - LOM 6** Definición: Fecha en que se registró el Licenciamiento Pauta • Se debe escribir como año, mes, día • Año de 4 dígitos, mes y día de dos dígitos	Ob
Persona	**ANNOTATION.ENTITY – LOM 8.1** Definición: Persona que realizó la anotación	Op

Elemento	Descripción	Criterio
Fecha	**ANNOTATION.DATE – LOM 8.2** Definición: Fecha en que se realizó la Anotación Pauta • Se debe escribir como año, mes día • Año de 4 dígitos, mes y día de dos dígitos	Op
Descripción	**ANNOTATION.DESCRIPTION – LOM 8.3** Definición: Breve declaración, anotación, comentario y / o aclaración sobre cualquier aspecto del recurso.	Op
Institución	**LIFECYCLE.CONTRIBUTOR.INSTITUTION – LOM 8.3** Definición: Tipo de organización principal que provee el recurso educativo. Pautas: • Para su diligenciamiento se puede usar el siguiente vocabulario: - Universidad - Institución Universitaria - Institución tecnológica - Institución técnica - Institución de educación para el trabajo - Fundación - Corporación - Colegio - Escuela - Instituto	Ob
Organización	**LIFECYCLE.CONTRIBUTOR.ORGANIZATION – LOM 8.3** Definición: Nombre de la institución principal que provee el recurso	Ob
Grupo	**LIFECYCLE.CONTRIBUTOR.GROUP – LOM 8.3** Definición: Grupo de personas reconocidas por una organización, que conforman un grupo cohesivo. Pautas: • Para su diligenciamiento se puede usar el siguiente vocabulario: - Facultad - Programa - Departamento - Grupo de investigación.	Op
División	**LIFECYCLE.CONTRIBUTOR.DIVISION – LOM 8.3** Definición: Nombre del grupo	Op
Taxón	**CLASSIFICATION.TAXON PATH.TAXON – LOM 9.2.2** Definición: Equivale a "Área".	Ob
Curso	**CLASSIFICATION.PURPOSE – LOM 9.1** Definición: Equivale a "Curso" o materia para la cual se desarrolló el recurso.	Op

Anexo 6. Crosswalk PA-REA.

Crosswalk PA-REA

Definición:
Documento que detalla el mapea de las relaciones entre campos de diferentes esquemas de metadatos.

Planteamiento:
Para llevar a cabo el proceso de identificación de vocabularios a utilizar para el mapeo de los metadatos del perfil de aplicación PA-REA, se tuvieron en cuenta los siguientes considerandos:

- El perfil de aplicación se basa en el estándar de metadatos LOM.
- El estándar LOM tiene elementos y calificadores definidos para describir recursos, pero no posee un namespace ni vocabulario definido para mapear estos metadatos (Dietze et al., 2009). Razón por la cual, su mapeo debe usar vocabularios definidos.
- Se han intentado trabajar iniciativas para definir vocabularios LOM, tales como:
    - Recommendation on exposing IEEE LOM as Linked Data 1.0 (second version) ODS Recommendation Draft 20 May 2014 (http://data.opendiscoveryspace.eu/ODS_LOM2LD/ODS_SecondDraft.html)
    - IEEE Learning Object Metadata RDF binding (http://kmr.nada.kth.se/static/ims/md-lomrdf.html)

  Iniciativas que usaban namespaces externos como:

    - rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#.
    - rdfs: http://www.w3.org/2000/01/rdf-schema#.
    - dc: http://purl.org/dc/elements/1.1/.
    - dcterms: http://purl.org/dc/terms/.
    - dctype: http://purl.org/dc/dcmitype/.
    - vCard: http://www.w3.org/2001/vcard-rdf/3.0#.
    - lom: http://ltsc.ieee.org/rdf/lomv1p0/lom#
    - lomvoc: http://ltsc.ieee.org/rdf/lomv1p0/vocabulary#
    - lomterms: http://ltsc.ieee.org/rdf/lomv1p0/terms#
    - foaf: http://xmlns.com/foaf/0.1/

  Adicionalmente, cada categoría de LOM tenía su propio namespace:

    - lom: http://ltsc.ieee.org/2002/09/lom-base#.
    - lom-gen: LOM general: http://ltsc.ieee.org/2002/09/lom-general#.
    - lom-life: LOM life cycle: http://ltsc.ieee.org/2002/09/lom-lifecycle#.
    - lom-meta: LOM meta-meta-data: http://ltsc.ieee.org/2002/09/lom-metametadata#.
    - lom-tech: LOM technical: http://ltsc.ieee.org/2002/09/lom-technical#.
    - lom-edu: LOM educational: http://ltsc.ieee.org/2002/09/lom-educational#
    - lom-rights: LOM rights: http://ltsc.ieee.org/2002/09/lom-rights#.
    - lom-rel: LOM relation: http://ltsc.ieee.org/2002/09/lom-relation#.
    - lom-ann: LOM annotation: http://ltsc.ieee.org/2002/09/lom-annotation#.
    - lom-cls: LOM classification: http://ltsc.ieee.org/2002/09/lom-classification#

  Pero también se identifica que estas iniciativas no han tenido actualizaciones, sus namespaces no están activos y sus enlaces ya aparecen inactivos.

- Por otro lado, Dublin Core Metadata Initiative tiene una de las mayores aplicaciones en vocabularios para describir elementos del esquema de metadatos. Además es uno de los vocabularios más usados en plataformas para la gestión de recursos y en cosechas de datos.
- Bajo Dublin Core Metadata Initiative maneja los namespace /terms/ y /elements/1.1/. Adicionalmente gestiona The Learning Resource Metadata Initiative (LRMI).
- Adicionalmente, Dublin Core tiene un proceso definido para su mapeo correspondiente a RDF.

Por las razones descritas anteriormente, y considerando que Metadata Crosswalks (mapeos de los elementos, semántica y sintaxis de un esquema de metadatos a otro) facilitan aún más el intercambio de metadatos, PA-REA busca identificar alternativas de mapeo en diferentes esquemas, para los atributos de su perfil de aplicación. El grado en que el crosswalk es exitoso a nivel de elemento depende de la similitud de los esquemas, la granularidad de los elementos en el esquema de destino en comparación con el de la fuente, y la compatibilidad de las reglas de contenido utilizadas para completar los elementos de cada uno (Note, 2011).

Mapeo Crosswalk PA-REA.

A continuación se describe las alternativas de mapeo identificadas, considerando Dublin Core como vocabulario principal para ejecutar esta tarea.

Elemento PA-REA	dc	dcterms	lrmi	AIISO	FOAF	Observaciones
Identificador **GENERAL.IDENTIFIER – LOM 1.1**	dc:identifier dc:identifier:uri dc:identifier:issn dc:identifier:isbn	dcterms:identifier				
Titulo **GENERAL.TITLE - LOM 1.2**	dc.title	dcterms:title				
Subtitulo **GENERAL.DESCRIPTION.SUBTITLE - LOM 1.4**	dc.title.alternative	dcterms:alternative				dcterms:alternative: Nombre alternativo para el recurso
Palabras Claves **GENERAL.KEYWORD - LOM 1.5**	dc.subject	dcterms:subject				DC.Subject: Los tópicos del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso.
Descripción (Abstract) **GENERAL.DESCRIPTION - LOM 1.4**	dc:description dc.description.abstract	dcterms:abstract				
Cita **GENERAL.DESCRIPTION.CITE - LOM 1.4**	dc.identifier.citation	dcterms:bibliographicCitation				
Idioma General **GENERAL.LANGUAGE - LOM 1.3**	dc.language dc.language.iso	dcterms:language				
Fecha de Version Inicial **LIFE CYCLE.CONTRIBU**	dc.date.created	dcterms:created				

Elemento PA-REA	dc	dcterms	lrmi	AIISO	FOAF	Observaciones
TE.DATE – LOM 2.3.3						
Versión **LIFE CYCLE.VERSION**	dc:description:version	dcterms:hasversion				dcterms:hasversion: un recurso relacionado que es una versión, edición o adaptación de un recurso descrito
Contribución: **LIFE CYCLE.CONTRIBUTE – LOM 2.3** Esta subdividida en: Rol: **LIFE CYCLE.CONTRIBUTE.ROLE – LOM 2.3.1** Entidad: **LIFE CYCLE.CONTRIBUTE.ENTITY – LOM 2.3.2** (que corresponde al contenido almacenado en ROL)	dc.creator dc.contributor dc.publisher dc.other dc:contributor:Author dc:contributor:Editor dc:contributor:Illustrator dc:contributor:Other dc:contributor:Advisor dc:description:provenance dc: sponsorship	dcterms:contributor dcterms:creator dcterms:publisher dcterms:provenance				DC.Creator: La persona u organización responsable de la creación u autoría del contenido intelectual del recurso. dcterms:provenance: Una declaración de cualquier cambio en la propiedad y custodia del recurso desde su creación que sea significativo para su autenticidad, integridad e interpretación.
Fecha del Ciclo de Vida **LIFE CYCLE.CONTRIBUTE.DATE – LOM 2.3.3**	dc:date:available dc:date:copyright dc:date:created dc:date:issued dc:date:submitted dc:date:updated dc:date:accessioned	dcterms:created dcterms:date dcterms:dateAccepted dcterms:dateCopyrighted dcterms:dateSubmitted dcterms:available				dcterms:availabl: fecha en que el recurso está disponible dcterms:created: fecha de creación del recurso dcterms:dateCopyrighted: fecha de licenciamiento dcterms:dateSubmitted: fecha de envío del recurso
Formato Técnico **TECHNICAL.FORMAT – LOM 4.1**	dc.format	dcterms:format dcterms:isformatof				dcterms:isformatof: un recurso relacionado que es substancialmente el mismo recurso descrito, pero en otro formato
Tamaño del Formato **TECHNICAL.SIZE – LOM 4.2**	dc:format:extent	dcterms:extent				dcterms:extent: tamaño o duración del recurso
Ubicación Técnica **TECHNICAL. LOCATION – LOM 4.3**	dc:source dc:source:uri dc:description:uri					bibo:uri
Tipo Técnico **TECHNICAL.REQUIREMENT.ORCOMPOSITE.TYPE – LOM 4.4.1.1**		dcterms:requires				dcterms:requires: Un recurso relacionado que es requerido por el recurso descrito para soportar su función, entrega o coherencia.
Versión Mínima **TECHNICAL.REQUIREMENT.ORCOMPOSITE.MINIMUNVERSION – LOM 4.4.1.3**						

Elemento PA-REA	dc	dcterms	lrmi	AIISO	FOAF	Observaciones
Tipo de interactividad **EDUCATIONAL.INTERACTIVITYTYPE – LOM 5.1**		dcterms:Instructional Method	lrmi:interactivityType			
Nivel de interactividad **EDUCATIONAL.INTERACTIVITYLEVEL – LOM 5.3**						
Tipo de Recurso **EDUCATIONAL.LEARNING RESOURCE TYPE – LOM 5.2**	dc.type	dcterms:type	lrmi. learningResourceType			
Rol Usuario Final **EDUCATIONAL.INTENDEDENDUSERROLE - LOM 5.5**		dcterms:audience				
Nivel educativo **EDUCATIONAL.CONTEXT - LOM 5.6**		dcterms:educationLevel	lrmi:EducationalLevel			
Contexto Educativo **CLASSIFICATION. DESCRIPTION – LOM 9.1**		dcterms:mediator		Aiiso: KnowledgeGrouping		dcterms:mediator: Una entidad que media el acceso al recurso y para quien el recurso está destinado o es útil. KnowledgeGrouping: representa una colección de recursos, objetivos de aprendizaje, horarios y otros materiales.
Rango de Edad **EDUCATIONAL.TYPICAL AGE RANGE - LOM 5.7**			lrmi:typicalAgeRange			
Descripción Educativa **EDUCATIONAL.DESCRIPTION – LOM 5.10**		dcterms:description				
Licenciamiento y Otras Restricciones **RIGHTS.COPYRIGHT AND OTHER RESTRICTIONS - LOM 6.2**	dc.rights dc:rights:license	dcterms:Access Rights				dcterms:AccessRights: información sobre quién puede acceder al recurso o una indicación de su estado de seguridad.
Descripción de los Derechos **RIGHTS.DESCRIPTION - LOM 6.3**	dc:rights:uri	dcterms:license dcterms:rights dcterms:RightsHolder	Lrmi:useRightsUrl			dcterms:license: documento legal que da permiso oficiales para hacer algo con el recurso dcterms:rights: Information about rights held in and over the resource dcterms:RightsHolder: persona u

Elemento PA-REA	dc	dcterms	lrmi	AIISO	FOAF	Observaciones
						organización propietaria o gestora de los derechos del recurso
Costo de los Derechos **RIGHTS.COST - LOM 6.1**						
Fecha del Licenciamiento **RIGHTS.DATE - LOM 6**	dc.date.copyright	dcterms:dateCopyrighted				
Persona **ANNOTATION.ENTITY – LOM 8.1**						
Fecha **ANNOTATION.DATE – LOM 8.2**						
Descripción **ANNOTATION.DESCRIPTION – LOM 8.3**						
Institución **LIFECYCLE.CONTRIBUTOR.INSTITUTION – LOM 8.3**				Aiiso:Institution		
Organización **LIFECYCLE.CONTRIBUTOR.ORGANIZATION – LOM 8.3**					Foaf:organization	
Grupo **LIFECYCLE.CONTRIBUTOR.GROUP – LOM 8.3**					Foaf:group	
División **LIFECYCLE.CONTRIBUTOR.DIVISION – LOM 8.3**				Aiiso:division		
Taxón **CLASSIFICATION.TAXONPATH.TAXON – LOM 9.2.2**			lrmi.EducationalSubject			Lrmi. Educational Subject: define el contexto de la asignatura en el cual el recurso es descrito
Curso **CLASSIFICATION.PURPOSE – LOM 9.2**				Aiiso:course		

Fuente:

- DCMI Metadata Terms. (https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-3)
- Lrmi Metadata Terms: https://www.dublincore.org/specifications/lrmi/lrmi_terms/, https://www.dublincore.org/specifications/lrmi/concept_schemes/
- Aiiso: http://vocab.org/aiiso/schema#

Después de identificar las diferentes alternativas de Metadata Crosswalk, y considerando los siguientes aspectos:

- RDF como modelo de datos, en su estructura básica acepta tripletas binarias.
- El mapeo realizado a RDF en los procesos de serialización mapea elementos de primer nivel en el vocabulario. Los calificadores de segundo y tercer nivel pertenecientes al elemento, son mapeados bajo la denominación del elemento. Por esta razón, para la selección de los elementos del esquema que describirá los recursos educativos, los elementos del crosswalk tendrán solamente un primer nivel.
- Organizaciones que hacen uso Dublin Core, adaptan el modelo adicionando elementos o calificadores faltantes al esquema de datos, según sea sus necesidades.
- Por último, después de revisar las alternativas de vocabularios, no se identifican elementos de mapeo para 8 elementos, razón por la cual se hace necesario plantear un namespace local que defina el mapeo de estos elementos.

Los prefijos propuestos para el mapeo son:

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix parea: <http://parea.com/parea#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix lrmi: <http://purl.org/dcx/lrmi-terms/> .
@prefix aiiso: <http://vocab.org/aiiso/schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ja: <http://jena.hpl.hp.com/2005/11/Assembler#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

Se plantea la siguiente tabla de mapeo de Metadata Crosswalk PA-REA:

Elemento	Elemento DC	Mapeo PA-REA
Identificador	dc.identifier.doi dc.identifier.uri dc.identifier.isbn dc.identifier.issn	Bibo.doi Bibo.handle Bibo.issn
Título	dc:title	dcterms:title
Subtítulo	dc.title.alternative	dcterms:alternative
Palabras Claves	dc:subject	dc:subject
Descripción (Abstract)	dc.description.abstract	dcterms:abstract
Cita	dc.identifier.citation	dcterms:bibliographicCitation
Idioma General	dc.language.iso	dc:language
Fecha de Version Inicial	dc.date.created	dcterms:created
Versión	dc.description.version	dcterms:hasversion
Contribución Esta subdividida en: Rol, Entidad	dc.contributor.author dc.creator dc.contributor dc.contributor.advisor dc.description. sponsorship dc.publisher	dcterms:creator dc:creator (uri del autor) dc: advisor dc: sponsorship dc:publisher
Fecha del Ciclo de Vida	dc.date.available dc.date.updated dc.date.accessioned	dcterms:available dcterms:modified dcterms:issued
Formato Técnico	dc:format	dc:format
Tamaño del Formato	dc:format:extent	dcterms:extent
Ubicación Técnica	dc.description.uri	Bibo:uri
Tipo Técnico	dc.relation.requires	dcterms:requires
Versión Mínima	dc.relation.minimunversion	parea:minimunversion
Tipo de interactividad	dc.description. interactivityType	lrmi:interactivityType

Elemento	Elemento DC	Mapeo PA-REA
Nivel de interactividad	dc.description.interactivitylevel	parea:interactivitylevel
Tipo de Recurso	dc:type	dc:type
Rol Usuario Final	dc.description.audience	dcterms:audience
Nivel Educativo	dc.description.educationlevel	dcterms:educationLevel
Contexto Educativo	dc.description. KnowledgeGrouping	Aiiso:KnowledgeGrouping
Rango de Edad	dc.description. typicalAgeRange	lrmi:typicalAgeRange
Descripción Educativa	dc:description	dcterms:description
Licenciamiento y Otras Restricciones	dc:rights	dc:rights
Descripción de los Derechos	dc.rights.uri	dcterms:rights
Costo de los Derechos	dc. rights.cost	parea: rightscost
Fecha del Licenciamiento	dc.date.copyright	dcterms:dateCopyrighted
Persona (anotación)	dc. relation. annotationentity	parea:annotationentity
Fecha (anotación)	dc. relation. annotationdate	dcterms:date
Descripción (anotación)	dc.relation. annotationdesc	parea:annotationdesc
Institución	dc.relation.institution	Aiiso:Institution
Organización	dc.relation. organization	Foaf:organization
Grupo	dc.relation. group	Foaf:group
División	dc.relation.division	Aiiso:division
Taxón	dc.relation. EducationalSubject	lrmi.EducationalSubject
Curso	dc.relation.course	Aiiso:course

Anexo 7. Descripción del Vocabulario PAREA.

PA-REA: Perfil de Aplicación de Recursos Educativos Abiertos

El Perfil de Aplicación de Recursos Educativos Abiertos (PAREA) proporciona clases y propiedades para complementar la descripción de un recurso educativo abierto. PEREA está diseñado para trabajar en asociación con Dublin Core Metadata Initiative (http://purl.org/dc/elements/1.1/ y http://purl.org/dc/terms/) para describir propiedades que complementan la descripción de un recursos educativo abierto.

Namespace
La URI de este vocabulario es http://parea.com/parea#

Al abreviar términos, el prefijo sugerido es “parea”.
Cada clase o propiedad en el vocabulario tiene una URI construido agregando un nombre de término a la URI del vocabulario. Por ejemplo:

http://parea.com/parea#minimunversion

Resumen de términos

Término	URI	Descripción
minimunversion	http://parea.com/parea#minimunversion	Versión Mínima de la aplicación de software, necesaria para que el recurso educativo abierto pueda funcionar adecuadamente.
interactivitylevel	http://parea.com/parea#Interactivitylevel	Nivel de relación reciproca que puede tener un usuario con el recurso educativo abierto.
rightscost	http://parea.com/parea#Rightscost	Costo de reproducción, si este se encuentra definido.
annotationentity	http://parea.com/parea#Annotationentity	Entidad que realiza las anotaciones y comentarios sobre el material educativo
annotationdesc	http://parea.com/parea#Annotationdesc	Anotaciones y comentarios sobre el material educativo