

HACIA UN MODELO DE CALIDAD DE DATOS Y METADATOS PARA REPOSITORIOS INSTITUCIONALES ABIERTOS

Carlos Salgado¹, Mario Peralta¹, Alberto Sanchez¹, Javier Saldarini², Claudio Carrizo², Julio Trasmontana², Silvana Armando²

¹Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales Universidad Nacional de San Luis, Ejército de los Andes 950. San Luis, C.P. 5700.
e-mail: {csalgado, mperalta, alfanego}@unsl.edu.ar

²Universidad Tecnológica Nacional, San Francisco, Córdoba, Argentina.
e-mail: {saldarinijavier, cjarrizo77, julio.trasmontana, silvana.armando}@gmail.com

INTRODUCCIÓN

Las universidades nacionales generan diversos tipos de documentos (revistas, artículos de investigación, tesis, etc.), por lo que siempre ha sido un desafío poner toda esta producción intelectual a disposición de la comunidad científica de manera abierta (Testa P., Degiorgi H., 2013). A partir de la Declaración Internacional de Acceso Abierto (Budapest Open Access Initiative, 2002), comenzaron a surgir los primeros repositorios institucionales abiertos. Lynch. C. (2003) se refiere específicamente a repositorios institucionales como el ámbito en el cual las universidades otorgan visibilidad a su producción intelectual y menciona a actores, tareas, condiciones, organización e inclusive tecnología y preservación, como partes constituyentes de los mismos.

El Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina, realizó en el año 2008 un relevamiento de los repositorios existentes, en donde convocó a los responsables de estos para que aporten su experiencia y forma un Comité de Expertos en Repositorios Digitales. En el año 2010 se crea el Sistema Nacional de Repositorios Digitales (SNRD, 2012) cuyo propósito es “conformar una red interoperable de repositorios digitales en ciencia y tecnología, a partir del establecimiento de políticas, estándares y protocolos comunes a todos los integrantes del Sistema”. Se consideran repositorios digitales a aquellas colecciones digitales de la producción científico-tecnológica de una institución, en las que se permite la búsqueda y la recuperación de información para su posterior uso nacional e internacional. Un repositorio digital contiene mecanismos para importar, identificar, almacenar, preservar, recuperar y exportar un conjunto de objetos digitales, normalmente desde un portal web. Esos objetos son descritos mediante metadatos que facilitan su recuperación, los cuales están descritos por el SNRD en las directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales del Ministerio de Ciencia, Tecnología e Innovación Productiva (Directrices, 2015), cuyo objetivo es “constituirse en una norma que, con su cumplimiento, garantice la interoperabilidad nacional e internacional de nuestros repositorios digitales”. Las directrices SNRD son un perfil de aplicación de metadatos Dublin Core (Dublin

Core Metadata Initiative, 2007), que deberán ser implementados, para poder participar del Sistema Nacional de Repositorios.

En Argentina existen diversos repositorios institucionales abiertos y a fin de que los objetos digitales provistos por los mismos sean confiables y aptos para el uso, es necesario no sólo una gestión integral de datos y metadatos, sino también asegurar la alta calidad de los mismos. De acuerdo a un relevamiento bibliográfico realizado, no existen trabajos que aborden la evaluación de calidad de datos y metadatos para repositorios institucionales abiertos, desde la perspectiva de una organización o institución en el rol de productora y proveedora de datos.

El objetivo de este trabajo consiste en elaborar un modelo de calidad para datos y metadatos asociados desde la perspectiva de la organización como productora y proveedora de datos, tomando como punto de partida la norma ISO/IEC 25012:2008 para la elaboración del modelo compuesto por características inherentes de calidad de datos y la norma ISO 25024:2015 para la elaboración de métricas aplicadas a cada una de las características del modelo.

MÉTODOS

Para cumplir con el objetivo propuesto se establecen los pasos que se describen a continuación:

Paso 1: Análisis e Identificación de las directrices para metadatos propuestas por el SNRD para la construcción de repositorios institucionales abiertos.

Paso 2: Estudio de las características de calidad de datos de norma ISO/IEC 25012.

Paso 3: Estudio de las métricas de calidad de datos ofrecidas por la norma ISO/IEC 25024.

Paso 4: Análisis e identificación de características de calidad de datos ofrecidas por la norma ISO/IEC 25012, desde la perspectiva de la organización como productora y proveedora de datos.

Paso 5: Análisis e identificación de métricas e indicadores asociados a las características seleccionadas.

Paso: 6: Elaborar un modelo compuesto por características de calidad de datos y metadatos.

Paso 7: Validar el modelo de calidad elaborado en el repositorio institucional abierto de la UTN.

RESULTADOS

En una primera etapa se realizó el relevamiento, análisis e identificación de directrices para proveedores de contenido que provee el Sistema Nacional de Repositorios Digitales, en la Tabla 1 se puede observar los siguientes campos definidos por el SNRD para repositorios institucionales abiertos:

Tabla 1. Campos asociados a directrices del SNRD

Campo SNRD	Uso
Título	Obligatorio
Autor	Obligatorio
Materia	Obligatorio si es aplicable
Descripción	Obligatorio si es aplicable
Filiación del Autor	Obligatorio
Editor	Obligatorio si es aplicable
Colaborador	Opcional excepto para las Tesis
Fecha de Publicación	Obligatorio
Fecha Finalización Embargo	Obligatorio
Tipo de Resultado Científico	Obligatorio
Versión de la publicación	Obligatorio
Formato	Obligatorio
Identificador del recurso	Obligatorio
Fuente	Recomendado
Idioma	Obligatorio
Relación	Obligatorio si es aplicable
Publicación referenciada	Recomendado
Identificador alternativo	Recomendado
Conjunto datos relacionados	Obligatorio si es aplicable
Identificador del Proyecto	Obligatorio si es aplicable
Cobertura	Recomendado
Nivel de accesibilidad	Obligatorio
Condiciones de uso	Obligatorio

También se realizó el estudio de las características inherentes de calidad de datos de la norma ISO 25012:2008, en donde hasta el momento se identificaron algunas características tomando como referencia los campos descriptos en las directrices del SNRD. También se realizó un estudio de las métricas que ofrece la norma ISO/IEC 25024:2015, en donde se identificaron algunas métricas asociadas a las características de calidad de datos identificadas. En la tabla 2 se pueden observar las características del modelo de calidad de datos junto con sus métricas asociadas:

Tabla 2: Modelo de Calidad de Datos y Metadatos con asociación de métricas

Características	Métricas
Exactitud	Exactitud sintáctica y semántica de los datos Exactitud de la metadata
Complejidad	Complejidad de valores de datos Complejidad de la metadata
Consistencia	Consistencia Semántica Consistencia del formato de datos
Accesibilidad	Accesibilidad del usuario Accesibilidad del formato de datos

CONCLUSIONES

Hasta el momento se logró la elaboración de un modelo de calidad de datos con características y métricas asociadas, que están alineadas con las directrices descriptas por el SNRD.

Actualmente se están analizando otras características y métricas que podrían ser incorporadas al modelo de calidad propuesto, algunas de ellas puede que no estén incluidas en las normas ISO/IEC 25012 e ISO 25024, por lo que se obtendría también una especie de extensión de dichas normas o bien definir un modelo mixto.

En trabajos futuros, se pretende obtener el modelo de calidad de datos y metadatos final para ser validados o contrastados en al menos 2 repositorios institucionales abiertos de diferentes universidades.

REFERENCIAS

- Budapest Open Access Initiative. (2002). Disponible en: <http://www.soros.org/openaccess/read.shtml>
- Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales (2015). Disponible en: http://repositorios.mincyt.gov.ar/vufind/themes/snrd/images/files/Directrices_SNRD_2015.pdf
- Dublin Core Metadata Initiative. (2007). Disponible en: <http://dublincore.org/>.
- International Organization for Standardization. (2008). ISO/IEC 25012:2008 - Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model
- International Organization for Standardization. (2015). ISO/IEC 25024:2015-Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of data quality
- Lynch, C. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. Disponible en: http://scholarship.utm.edu/21/1/Lynch,_IRs.pdf.
- Sistema Nacional de Repositorios Digitales (2012). Disponible en: <http://repositoriosdigitales.mincyt.gov.ar>
- Testa P., Degiorgi H. "Esquemas de metadatos para los repositorios institucionales de las universidades nacionales argentinas". Universidad Nacional de Cuyo. Facultad de Educación Elemental y Especial (2013). (1983).