

Estándares y metadatos en repositorios: Buenas prácticas

Víctor Leonardo Guerrero Sarmiento
leonardo.guerrero@infotegra.com

Los repositorios son una herramienta para la comunicación académica y pueden formar parte de un sistema mayor que puede ser del orden nacional, regional o global; es por ello que la interoperatividad es un factor a tener en cuenta, y es mediante la adopción de estándares que puede ser lograda.

Los repositorios son una herramienta para la comunicación académica, empleados por universidades, organismos y centros de investigación. Generalmente gestionados por las bibliotecas. Dado que pueden formar parte de un sistema mayor de orden nacional, regional o global la interoperatividad y la unificación de criterios de calidad resultan fundamentales para el aprovechamiento de sus contenidos.

Aspecto - Usuarios

- Capacitar a los usuarios que van a alimentar el repositorio y/o van a realizar labores de revisión.
- Gestión adecuada de permisos y grupos de usuarios. Auditorias a permisos y accesos.
- Establecer procedimientos y manuales para el manejo de flujos de trabajo (Sobre todo si se realiza Autoarchivo).

Aspecto - Metadatos

- Definir un perfil de metadatos a utilizar.
- Crear campos nuevos solo en caso de ser necesario y no antes de haber realizado una correcta revisión bibliográfica.
- Observar que el uso de los campos sea correcto y conforme a sus definiciones.

Aspecto - Metadatos [2]

- Idioma: dc.language
 - Según DCMI usar RFC 4646
 - Según OpenAIRE usar ISO 639-x, uso recomendado ISO 639-3
 - Según RedCol usar ISO 639-3 y cualificador dc.language.iso
- Cobertura geográfica: dc.coverage.spatial o datacite.geoLocation
 - Uso de vocabularios controlados de nombres geográficos.
 - Si se reportan coordenadas solicitar que estén en WGS84 y expresadas en grados decimales.
- Fechas: dc.date
 - Usar formato ISO 8601 <https://www.w3.org/TR/NOTE-datetime>

Aspecto - Metadatos [3]

- Tipología documental: dc.type o oaire:resourceType:
 - Uso de vocabularios controlados:
 - i. DCMI TYPE <https://dublincore.org/specifications/dublin-core/dcmi-terms/>
 - ii. Driver <https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-Publicationtypes>
 - iii. COAR https://vocabularies.coar-repositories.org/documentation/resource_types/

- Estado de publicación: dc.type.version o oaire:version
 - Uso de vocabularios controlados, tales como:
 - i. DRIVER https://guidelines.openaire.eu/en/latest/literature/field_publicationversion.html
 - ii. COAR https://vocabularies.coar-repositories.org/version_types/

Aspecto - Metadatos [4]

- Formato: dc.format.
 - Uso de vocabulario de tipos de medios.
 - Para el caso de dc.format.mimetype:
IANA <https://www.iana.org/assignments/media-types/media-types.xhtml>
- Descriptores. dc.subject
 - Incluir descriptores normalizados (local, lemb, dewey, etc...)

Aspecto - Metadatos [5]

- Derechos de acceso: dc.rights o datacite:rights
 - Uso de vocabularios controlados:
 - i. OpenAIRE:
<https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights>
 - ii. COAR: https://vocabularies.coar-repositories.org/documentation/access_rights/
- Condición de licencia: dc.rights o oaire:licenseCondition
 - Recomendación: Inclusión de licencias internacionales.

Aspecto - Metadatos [5]

- Extensión y tamaño del recurso: dc.format
 - Ser explícitos en las unidades empleadas.
 - Para extensión (dc.format.extent): Cantidad de páginas, duración en tiempo.
 - Para tamaño (dc.format.size): Dimensiones físicas del objeto (Unidades de distancia para objetos físicos y en bytes si se trata de un archivo digital).

Aspecto - Metadatos [6]

- El perfil de metadatos se debe revisar y actualizar, de acuerdo los cambios observados en las directrices nacionales, regionales o internacionales que la institución se haya trazado cumplir.
- Asegurarse del correcto uso de los distintos campos de metadatos y la calidad de los mismos.
 - Completitud
 - Precisión
 - Conformidad con expectativas
 - Consistencia y coherencia lógica
 - Oportunidad
 - Accesibilidad
 - Contexto adecuado
 - Contenido coherente
 - Uso de vocabularios estandarizados
 - Consistencia

Calidad de los metadatos [1]

- **Compleitud:** Se debe seleccionar un conjunto de elementos que permitan que los recursos sean descritos tan completamente como sea posible con el mínimo de elementos y ese conjunto seleccionado debe ser aplicado tan completamente como sea posible.
- **Precisión:** Los metadatos deben ser correctos y conformes con la sintaxis de los conjuntos de elementos en uso.
- **Conformidad con expectativas:** Elementos de metadatos, uso de vocabularios controlados y robustez deben llenar las expectativas de una comunidad en particular.

Calidad de los metadatos [2]

- **Consistencia y coherencia lógica:** Uso de elementos que coincide con las definiciones y la aplicación coherente de estos elementos.
- **Oportunidad:** Mantener los metadatos al día con los cambios de los recursos que describen.
- **Accesibilidad:** La asociación apropiada entre los metadatos con los recursos que describen y la legibilidad para los usuarios a los que están destinados.

Calidad de los metadatos [3]

Conjuntos de metadatos que se pueden compartir

- **Contexto adecuado:** Cada registro debe contener el contexto necesario para entender de que se trata el recurso que describe, sin necesidad de información externa.
- **Contenido coherente:** Los registros deben tener la información suficiente para que los registros tengan sentido por sí mismos, excluyendo información que solo tiene sentido en un entorno local.

Calidad de los metadatos [4]

Conjuntos de metadatos que se pueden compartir

- **Uso de vocabularios estandarizados:** Su uso mejora la integración de los registros de metadatos de una fuente con los registros provenientes de otras fuentes.
- **Consistencia:** Todas las decisiones sobre la aplicación de elementos, sintaxis y uso de vocabularios controlados deben ser consistentes con el conjunto de metadatos definido, de forma tal que se puedan aplicar reglas de transformación sin tener inconsistencias en el proceso.

Aspecto - Organización estructural del repositorio

- La organización del repositorio debe orientarse a facilitar la búsqueda. Se pueden usar diferentes estrategias de agrupación de contenidos según sea la orientación de la institución.
 - Por ejemplo:
 - Orientados a tipologías documentales
 - Orientados a niveles académicos
 - Orientados al organigrama institucional.
- Evitar colecciones con muy pocos ítems.

Aspecto - Preservación de la información

- Emplear URL persistentes para cada registro (Handle).
- Definir los formatos digitales que permitan la correcta preservación de los objetos digitales.
- Ejecución periódica de los procesos que garantizan el correcto funcionamiento de índices de búsqueda, sitemaps, interfaz OAI, entre otros.
- Definir políticas de copias de seguridad.

Aspecto - Visibilidad

- W3C – Garantizar la visibilidad para el acceso abierto e inclusivo.
- Realizar registro en directorios de repositorios de documentos de acceso abierto (ROAR - OpenDOAR).
- Registrar el repositorio con proveedores de servicio OAI-PMH (recolectores).

Aspecto - Políticas

- Definición de una política de contenidos, donde se indique alcance y forma de los objetos a ser incluidos en el repositorio.
- Definir política para depósito.
- Definir política para la preservación.
- Definir política de uso de metadatos.
- Definir política de uso de datos.

OAI-PMH

Open Archive Initiative

Es un proyecto para promover el desarrollo y uso de estándares para la interoperabilidad en internet.

OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting)

OAI PMH ha sido ampliamente adoptado desde su introducción en 2001, ha demostrado ser una herramienta útil para mover y agregar metadatos desde diversas instituciones.

OAI-PMH - Conceptos básicos [1]

Recurso (Resource): Es un objeto al rededor del cual los metadatos son creados.

Registro (Record): Es un conjunto de declaraciones en un esquema de metadatos, acerca de un recurso.

Identificador único (Unique Identifier): Es un identificador para un recurso o registro. Debe ser un URL.

Repositorio (Repository): Es una colección de registros de metadatos.

OAI-PMH - Conceptos básicos [2]

Recolector/Cosechador (Harvester): Es un algoritmo que hace peticiones a un repositorio y recolecta los registros de metadatos que son retornados por el repositorio.

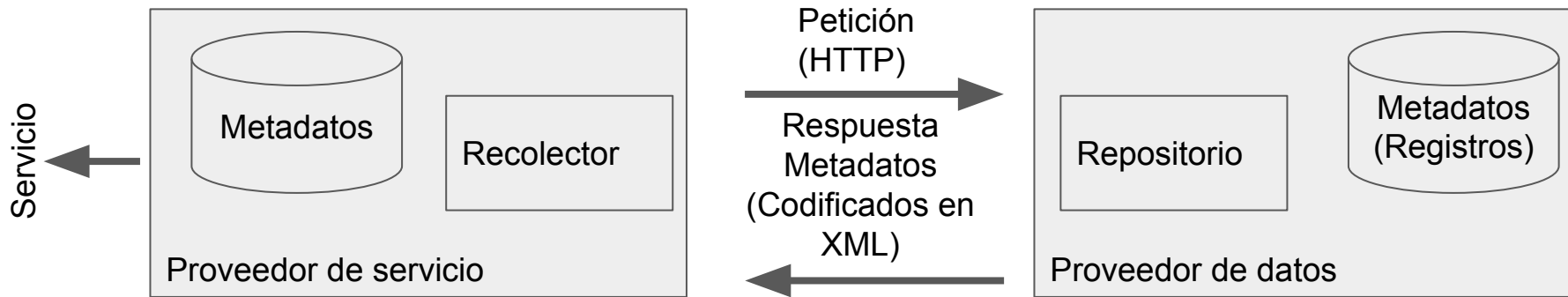
Proveedor de datos (Data provider): Es una institución, organización o individuo que expone metadatos, de uno o más recursos, por medio de OAI-PMH.

OAI-PMH - Conceptos básicos [3]

Proveedor de servicios (Service provider): Utiliza OAI-PMH para recolectar los metadatos de los proveedores de datos. El proveedor de servicios generalmente recolecta metadatos de diferentes proveedores de datos y crea una base de datos que agrega todos los metadatos.

La intención de recopilar los metadatos es la de proveer a los usuarios de un punto de búsqueda que le permita encontrar la información de múltiples fuentes.

Diagrama del funcionamiento de OAI-PMH



Buenas prácticas OAI [1] - Ciclo de vida del repositorio

- **Registro como proveedor de datos:** Se debe prestar atención a cambios en el repositorio que potencialmente puedan afectar a los proveedores de servicio.

<https://www.openarchives.org/Register/ValidateSite>

- **Finalización del ciclo de vida de repositorio:**
 - Todos los recursos descritos, por los metadatos, dejarán de estar disponibles o de existir. (Reportar y marcar registros como borrados)
 - El proveedor de datos dejará de mantener el repositorio, porque no quiere o no puede seguir manteniéndolo. (Reportar y buscar alternativas)

Buenas prácticas [2] - Identificación del repositorio [1]

Elementos obligatorios de la respuesta “Identify”

- **<repositoryName>**: Nombre del repositorio. Usado por proveedores de servicio para identificar potenciales repositorios a recolectar.

Ej.

American Journal of Engineering Education (AJEE)

Archivo Español de Arqueología

Buenas prácticas [3] - Identificación del repositorio [2]

<baseURL>: Es la URL del repositorio OAI. Esta URL debe estar actualizada.

<protocolVersion>: Es la versión del protocolo OAI-PMH que es soportada por el repositorio. La versión actual es 2.0, la cual fue liberada a mediados del 2002.

(Las versiones 1.0 y 1.1 son experimentales, la versión 2.0 es una versión estable)

Buenas prácticas [4] - Mantenimiento del repositorio

- **<earliestDatestamp>**: Es el límite inferior de todas las fechas que registren cambios, modificaciones o eliminaciones en el repositorio.
- **<deletedRecord>**: Los valores posibles son: no, transient (transitorio), persistent (persistente). El valor recomendado es “persistent”.
- **<granularity>**: Granularidad de la marca temporal (nivel de detalle).
- **<adminEmail>**: Dirección de correo electrónico de un administrador del repositorio.

Buenas prácticas [5] - Identificadores OAI

- **<oai-identifier> (sampleIdentifier):** El identificador debe ser un nombre de dominio registrado a la organización que controla el repositorio OAI.

La buena práctica es usar el nombre de dominio donde el servidor OAI se encuentra alojado.

(Se desaconseja el uso de direcciones IP).

Recursos

- Marco de referencia de buenas prácticas en repositorios:
<https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/>
- Guía para la evaluación de repositorios institucionales de investigación:
https://www.recolecta.fecyt.es/sites/default/files/contenido/documentos/2021GuiaEvaluacionRecolecta_vf.pdf
- Lineamientos OpenAIRE:
<https://guidelines.openaire.eu/en/latest/literature/index.html>
- Perfil de aplicación de metadatos RedCol:
<https://redcol.readthedocs.io/es/latest/descGenPer.html>

