

Libre acceso a la información científica de la Universidad Federal de Rio Grande do Sul - Brasil

CATERINA GROPOSO PAVÃO, BEATRIZ HELENA PIRES DE SOUZA CESTARI,
CARLA METZLER SAATKAMP, DENISE RAMIRES MACHADO,
JANISE SILVA BORGES DA COSTA, MANUELA KLANOVICZ FERREIRA,
ZAIDA HOROWITZ, ZITA PRATES DE OLIVEIRA
Universidad Federal de Rio Grande do Sul - UFRGS

Resumen: Este trabajo describe el diseño, planificación, aplicación y gestión del repositorio institucional de acceso abierto de la Universidad Federal de Rio Grande do Sul. Aborda los recursos de software y hardware utilizado y los estándares empleados para la representación de la información, con vistas a su difusión, uso, preservación de contenidos e interoperabilidad con otros repositorios. Más ampliamente se constituyen como un verdadero sistema de gestión de contenido, porque además de los propios documentos, Lume - Repositorio Digital de la Universidad Federal de Rio Grande do Sul, ofrece a la comunidad académica un conjunto de servicios para la gestión de su producción científica.

Palabras clave: Repositorios institucionales; Repositorios digitales; Iniciativa de archivos abiertos; metadatos; DSpace.

Digital repository: free access to information at the
federal university of rio grande do sul - brazil

Abstract: This paper describes the design, organization and management of Lume the open access institutional repository of Federal University of Rio Grande do Sul (UFRGS). It shows the resources of soft-

ware and hardware and the patterns of representation concerning the information used in its implementation, aiming at the preservation and dissemination of its content and interoperability with other digital repositories. As a system of contents management Lume, beyond digital documents, offers a variety of services to the management of the scientific production of UFRGS community.

Keywords: Digital repositories; Institutional repositories; Open Archives Initiative; Metadata; *DSpace*.

1 Introducción

Desde 1989, cuando se implantó la automatización en su sistema de bibliotecas, la Universidad Federal de Rio Grande do Sul (UFRGS) ha trabajado en la reunión y registro de la producción intelectual (PI) de sus docente e personal técnico-administrativo para garantizar el control bibliográfico institucional y difundir los documentos producidos en su ámbito. En 2001, implantó su Biblioteca Digital de Tesis y Disertaciones y, desde entonces, ha desarrollado estudios hacia los repositorios institucionales, considerando los avances en las áreas de tecnología de la información y comunicación.

La tecnología de la Información tiene un papel estratégico en la UFRGS, no sólo a respecto de las posibilidades de expansión de los servicios bibliotecarios y como un instrumento fundamental para la formación presencial y a distancia, sino también como una fuente de indicadores gerenciales y de integración con otros repositorios de formación e investigación en el país. La gran cantidad de producción de conocimientos en las universidades hace necesario, además de su difusión y uso, su preservación.

Los repositorios iniciaron y se están extendiendo apoyados por tres proposiciones básicas: auto-archivo de la producción científica, la interoperabilidad¹ entre los diferentes sistemas de publicación y el libre acceso a los documentos y/o archivos de la producción científica.

1 La interoperabilidad es la capacidad de un sistema de hardware o de software para comunicarse y trabajar eficazmente en el intercambio de datos con otro sistema, generalmente de un tipo diferente, diseñado y producido por un proveedor distinto (ONLINE, 2010).

Al utilizar metodologías de archivos abiertos (*Open Archives Initiative*), los repositorios digitales promueven el libre acceso a la información científica y académica y, sobre todo, permiten reunir, preservar y divulgar la producción intelectual, maximizar el impacto y la visibilidad de las investigaciones desarrolladas en la institución. (CAMARGO; VIDOTTI, 2006).

La gestión del conocimiento está directamente relacionada con los proyectos de desarrollo de repositorios institucionales en las universidades. Estos cooperar para mejorar la gestión y maximizar el acceso al conocimiento, estimulando así, la generación de nuevos conocimientos.

Sobre este enfoque fue desarrollado Lume, designación del repositorio digital de la Universidad Federal de Rio Grande do Sul - que significa la manifestación de conocimientos, luz, brillo.

2 LUME - Repositório digital de la ufrgs

2.1 Antecedentes

Con la automatización de rutinas y servicios básicos de las bibliotecas en una etapa avanzada, la Universidad busco ampliar el uso de la información especializada disponible en las colecciones del Sistema de Bibliotecas de la UFRGS (SBU) y también el producido por sus docente y personal técnico-administrativo.

La Biblioteca Digital fue el vehículo elegido para divulgar y preservar una colección en continuo crecimiento, por su capacidad para ampliar los límites de acceso y uso de información ultrapasando las barreras de hora, ubicación geográfica y espacio físico de las bibliotecas tradicionales. Diferente a un catálogo en línea, que presenta sólo una descripción de los documentos bibliográficos, la biblioteca digital agrega el contenido del documento a la descripción bibliográfica y temática, demandando atención diferenciada para garantizar su integridad, preservación y, también, el cumplimiento de las leyes de derecho de autor.

La Biblioteca Digital de Tesis y Disertaciones de la UFRGS (BDTD) fue creada en 2001. Desde su implantación, debido a su intensa utili-

zación, la Universidad identificó la necesidad de reunir y dar acceso a otras colecciones digitales existentes en el SBU y otras tantas dispersas por la Universidad (p. ex. artículos de periódicos, comunicaciones en eventos, archivos audio y video, fotos, partituras) optimizando su organización, gestión, mantenimiento y compartiendo recursos. A los usuarios finales, la reunión de los archivos digitales permite búsquedas en un único portal de información.

En este contexto, en enero de 2008 Lume fue puesto a disposición, como una iniciativa conjunta del Centro de Procesamiento de Datos (CPD) y la Biblioteca Central (BC), a partir de la migración de la Biblioteca Digital de Tesis y Disertaciones para el software DSpace,² que permite almacenar distintos tipos de documentos en formato digital, resultado de las actividades de formación e investigación de profesores, investigadores y personal técnico-administrativo de la Universidad.

2.2 Objetivos

LUME pretende dar una mayor visibilidad a la producción científica de la UFRGS; preservar la memoria institucional; reunir colecciones digitales producidas y dispersas en los distintos departamentos académicos y centros de investigación de la Universidad y garantizar el acceso fiable y permanente a los objetos digitales, maximizando el uso de estos recursos.

La preservación de documentos digitales puede ser identificada como una de las principales tareas del repositorio, pero también proporcionar a la institución, transparencia y accesibilidad, por la disponibilidad que ofrece a la comunidad en general de su producción científica, permite maximizar el impacto de la investigación, aumentando la posibilidad de recaudación de fondos, obtención de premios y prestigio compartida por investigadores y por la institución (CAFÉ, 2003). Por medio de los repositorios institucionales, un investigador tiene acceso a la información que se podrá combinar y reagrupar de manera a añadir valor y nuevos conjuntos de informaciones, generando un nuevo conocimiento explícito (LEITE, 2006).

2 <<http://www.dspace.org/>>.

2.3 Arquitectura y tecnología

Lume fue concebido usando el *DSpace Institutional Digital Repository System* (DSpace, 2008), software libre, proyectado y elaborado en colaboración por el MIT (Massachusetts Institute of Technology, EUA) y HP (Hewlett-Packard Company, EUA), orientado a la creación de repositorios institucionales y la conservación digital de su producción intelectual.

En la UFRGS el DSpace está instalado en un servidor Linux. El software, escrito en Java, está apoyado por un conjunto de herramientas de código abierto:

- Bases de datos relacionales PostgreSQL, donde están almacenados los metadatos (datos sobre datos);
- Servidor de aplicaciones TOMCAT Servlet Engine;
- Motor de búsqueda LUCÈNE, para la indización y recuperación de los metadatos y documentos textuales;
- Apache;
- Bibliotecas Java (JavaBeans Activation Framework, Java Servlet e JavaMail API) y
- Compilador Ant.

DSpace permite organizar la información en comunidades, subcomunidades y colecciones. La estructura puede ser distinta para cada comunidad. Bajo las colecciones están los contenidos digitales, que son recuperados por los usuarios. Puede haber uno o más objetos digitales - texto, imagen, audio y video - asociados con un documento digital.

Para garantizar acceso permanente a los documentos digitales se utiliza *Handle System*, servicio ofrecido por CNRI - *Corporation for National Research Initiatives* (CORPORATION FOR..., 2008), que consiste en asignar un identificador persistente a cada documento digital garantizando que, incluso si la dirección del servidor del repositorio digital cambia, los recursos puede seguir siendo citados unívocamente. Usa el software Samba para compartir con seguridad archivos que están en servidores Windows en red.

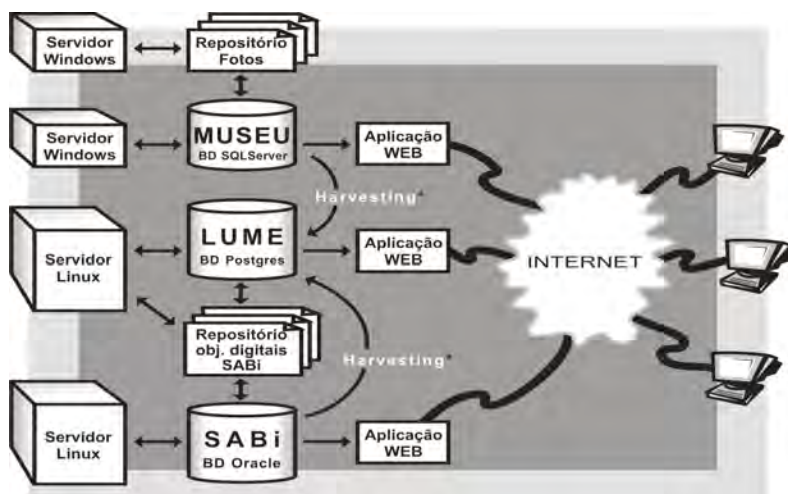
El sistema y el banco de datos están en un equipo HP Proliant ML370, con sistema operativo Linux Red Hat Enterprise.

Actualmente, Lume contiene metadatos y documentos digitales provenientes de la recolección automática de otros sistemas (*harvesting*), permitiendo la integración de los distintos procesos de depósito necesarios en un sistema multidisciplinario:

- El sistema MUSEO contiene la colección de fotografías del Museo de la UFRGS y la colección histórica de la Escuela de Educación Física. Los datos se encuentran en un BD SQL Server - servidor Windows, y los documentos digitales en otro servidor Windows;
- El sistema SABi, Sistema de Automatización de Bibliotecas de la UFRGS, usa el software Aleph500, hace la gerencia de las actividades y servicios ofrecidos por las bibliotecas para la comunidad usuaria. Las informaciones del sistema son almacenadas en un BD Oracle, en una máquina Linux, donde, también se encuentran los documentos digitales.

La *Figura 1* ilustra la estructura del sistema.

Figura 1 - Arquitectura de Lume.



*harvesting incremental: recolección automática de metadatos a cada 24 horas.

2.4 Organización e interfaz gráfica

El DSpace está organizado en comunidades, subcomunidades y colecciones. Hasta el momento, están disponibles las siguientes comunidades:

- Acervo Fotográfico;
- Artículos de periódico;
- Libros y capítulos de libros;
- Tesis y disertaciones;
- Trabajo de curso de especialización;
- Trabajo de conclusión de grado, y
- Comunicaciones en eventos.

Cada comunidad determina su propia política de información, proceso de depósito (*workflow*) y niveles de acceso a los documentos, y puede ser dividida en subcomunidades y colecciones. Cada colección contiene metadatos, como título, autor, fecha, palabra clave, resumen, tamaño en *bytes*, entre otros, y uno o más archivos en formatos variados (pdf, jpg, avi, MP3, html, etc.) llamado *bitstreams*, con contenidos digitales. Estos datos pueden ser visualizados en formato breve o completo. En Lume, los metadatos se definen de manera diferente, según las necesidades y circunstancias particulares de los documentos de cada comunidad.

Desde la implantación del proyecto piloto varios cambios se hicieron en la interfaz de DSpace original, con el uso de la herramienta Manakin, donde fue posible programar cambios estructurales directamente desde el código fuente de la aplicación, personalizando totalmente el sistema. Además de las alteraciones visuales, se llevó a cabo la traducción, del inglés al portugués, de diversas etiquetas de los campos del menú, de textos y otras ayudas. Con la migración a la versión 1.5.2 de DSpace la personalización de la interfaz fue optimizada gracias a la incorporación de Manakin en el propio DSpace.

Es posible recuperar información por filtros genéricos predefinido, en forma de listas, en los metadatos de título, autor, palabra clave o año, en los filtros específicos de una comunidad, sujeto a las particularidades de los objetos digitales pero, mediante un único por-

tal de información. Las búsquedas pueden ser perfeccionadas con el uso de los operadores booleanos (*AND*, *OR* o *NOT*). Pueden hacerse búsquedas genéricas, que abarcan el conjunto del repositorio. Además de los metadatos es posible hacer búsquedas en el contenido de los documentos digitales almacenados, gracias a los mecanismos de indexación adoptados.

Los documentos de Lume están, en su mayoría, disponibles en texto completo, salvo algunas excepciones que, a petición del autor, son revelados parcialmente porque se tratan de artículo de periódicos sometidos a publicación, porque contienen datos sigilosos de empresas o pacientes que participan en la investigación, entre otros motivos.

Hay, además, los documentos digitales de acceso restringido, debido a los derechos de autor, para los cuales sólo los usuarios de la UFRGS tienen acceso. La autenticación de la identificación y contraseña de estos usuarios se hace de manera segura mediante criptografía SSL.

Tanto los usuarios internos como los usuarios externos tienen la posibilidad de crear un perfil donde, mediante la asignación de colecciones, identifican sus necesidades de información. De esta forma, cuando se incluyen nuevos documentos en las colecciones asignadas, el usuario recibirá el informe por correo electrónico.

El *link* Estadísticas disponible en todos los niveles del repositorio, desde las colecciones hasta el documento en sí, muestra el número de accesos y de *downloads*. También es posible obtener informaciones generales a respecto de los documentos más consultados en cada comunidad, así como el número de *downloads* general o por comunidad y el número de usuarios internos y externos registrados.

2.5 Metadatos e interoperabilidad

Lume es compuesto de metadatos e información digital. Metadatos son información estructurada con semántica estándar y se utilizan para representar la información digital de forma bibliográfica. Siguen la norma Dublin Core - *Dublin Core Metadata Initiative* (DUBLIN CORE..., 2008), con 15 elementos básicos utilizados para identificar y describir documentos digitales y pueden ser asignados calificadores, de conformidad con las características y necesidades de la institución.

La organización de la información en el contexto de Internet, con vistas a la recuperación calificada de la información, ha sido constante desafío para superar, dada su complejidad, representada por el volumen de información que reúne y el constante y indiscriminado crecimiento. Hay varios servicios disponibles libremente a la comunidad, permitiendo a los usuarios la creación de sus textos y hacerlos accesibles. También son innumerables los intentos de sistematización del contenido en el ambiente en red, mediante el establecimiento de estándares para la descripción de los recursos, sin embargo, estos también están proliferando rápidamente.

Transponiendo la reflexión para el universo de los repositorios digitales, también, se puede notar el uso de diversos estándares. En la UFRGS, para el desarrollo de Lume, se adoptaron estándares ya difundido nacional e internacionalmente, con el fin de facilitar la consulta a los usuarios, ya que tienen una estructura común de organización de la información.

La utilización de Dublin Core permite detallar de forma más refinada el contenido de los documentos digitales, con características tan peculiares, normalmente no incluidas en un catálogo en línea, lo que beneficia de los usuarios. El Apéndice A relaciona los metadatos utilizados en cada comunidad.

El uso de estándares nacionales e internacionales para el tratamiento de recursos de información es una práctica hace mucho tiempo adoptada en la Universidad centrándose, sobre todo, en la integración y compartir la información con otros sistemas.

Para asegurar la interoperabilidad con otros repositorios digitales se usa el protocolo de recolección de metadatos de la iniciativa de los archivos abiertos OAI-PMH - *Open Archives Initiative Protocol for Metadata Harvesting (THE OPEN...*, 2004), que permite la transferencia de datos entre repositorios digitales.

Este protocolo está construido sobre el protocolo *http* y se basa en aplicaciones del tipo *request-response* definido utilizando el lenguaje de descripción XML, que proporciona compatibilidad prácticamente universal.

2.6 Poblar el repositorio

Poblar el repositorio es uno de los principales desafíos para lograr el éxito esperado. Las dificultades para gestionar aspectos relacionados con la propiedad intelectual y los derechos de autor (*copyright*), tradiciones y valores diversos en diferentes comunidades científicas, la inercia o sobrecarga de trabajo de los autores y por último las dificultades y barreras tecnológicas pueden ser factores relevantes para impedir el crecimiento del repositorio.

La definición de estrategias de comunicación y promoción interna y externa del repositorio y la definición de una política institucional pueden minimizar las dificultades. Esta política debe proporcionar una visión clara de los principales actores involucrados en el contexto del repositorio.

En la UFRGS el diferencial que se puede apuntar como responsable por el éxito de la población de Lume nos remite a la política de control bibliográfico de la IP de la Universidad que tuvo inicio en los años de 1970, cuando se llevaron a cabo los primeros esfuerzos de automatización de los servicios bibliotecarios tras las normas internacionales para el registro de la información.

La UFRGS, mediante su Sistema de Bibliotecas, ha priorizado, a lo largo de los años, la recolección, registro y difusión de la producción intelectual institucional en el SABi, Sistema de Automatización de Bibliotecas, base de datos que compila las colecciones bibliográficas de las 33 unidades que integran el SBU.

La automatización del proceso de descripción bibliográfica en el SBU comenzó en 1975 y la recolección de la PI se restringía a tesis y disertaciones. Desde 1989, con la implantación del SABi, la recogida y registro de PI empezó a realizarse de forma continuada y sistemática para todos los tipos de documento. El proceso fue facilitado con la actitud tomada por la Administración Superior que determinó la obligatoriedad de depósito, en las bibliotecas, de un ejemplar de todo documento producido en la Universidad.

Con el objetivo de uniformizar y normalizar la entrada de datos de la PI y utilizando la estructura de campos, subcampo y los indicadores propuestos por MARC, se creó el campo 909 para el control de la pro-

ducción científica, técnica, artísticas y administrativas de la UFRGS. También se prepararon documentos auxiliares (glosarios y tablas de códigos) que se utilizan para llenar este y otros campos, destinado a hacer más completa la descripción bibliográfica y el acceso al texto completo de los documento de la producción intelectual de la UFRGS. (OLIVEIRA et al., 2004). El uso de este campo es obligatorio para los documentos producidos por los docentes, el personal técnico-administrativo y estudiantes, siempre que guiado por profesores de la Universidad. El campo 909 incluye Unidad/Departamento/organismos, tipo de producción, Programa/Curso de postgrado, curso de especialización, curso de grado, curso de formación profesional y organismos de promoción.

La recolección de la PI por las bibliotecas y su registro en SABi son prioritarias en el SBU. La labor sistemática de los bibliotecarios en este sentido ha fortalecido y consolidado el SABi como instrumento de control de la PI y está permitiendo la poblado rápida y consistente del repositorio. Debido a que los metadatos de las comunidades de periódico artículos, libros y capítulos de libros, tesis y disertaciones, trabajo de cursos de especialización, trabajos de grado y comunicaciones en eventos son extraídos automáticamente del SABi.

El depósito de los objetos digitales en Lume se celebra en la medida que son solicitadas y firmada por los autores y/o editores las autorizaciones para tanto, ya que en ningún momento son dejadas de lado las cuestiones de derecho de autor y tampoco, la atribución de licencias para el uso de los documentos.

3 Conclusión

Las instituciones de educación e investigación están utilizando, cada vez más, la tecnología digital para proporcionar información sobre sus logros y para almacenar, preservar y dar acceso a los contenidos informativos en medio digital, utilizando como instrumento para este propósito los repositorios.

Lume se consolida como un proyecto permanente de la UFRGS y, en este sentido, en 2010 se constituyó el Comité Gestor y se definió la política institucional de información para el repositorio.

Considerando la ampliación de su cobertura de aplicación para la totalidad de las colecciones digitales de la Universidad hay planes para la compra de equipo dedicado exclusivamente al sistema. Se están desarrollando mecanismos estadísticos mejorados con el objetivo de proporcionar informaciones más detalladas sobre el acceso a los objetos digitales que permitan gestionar de manera más eficiente las investigaciones realizadas en la Universidad.

Hasta ahora la experiencia adquirida nos permite afirmar que Lume es una herramienta poderosa de organización, difusión, preservación y visibilidad de las informaciones de la Universidad y, sobre todo, de planificación y gestión del conocimiento producido en la Universidad.

Referencias

CAFÉ, Lígia et al. **Repositórios institucionais**: nova estratégia para publicação científica na Rede. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO (INTERCOM), 26., 2003, Belo Horizonte. **Anais eletrônicos...** Disponível em: <http://dspace.ibict.br/dmdocuments/EN-DOCOM_CAFE.pdf>. Acesso em: 12 jan. 2011.

CAMARGO, Liriane Soares de Araújo; VIDOTTI, Silvana Ap. Borseti Gregorio. Elementos de personalização em repositórios institucionais. In: CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA, 1., 2006, Brasília. **Anais eletrônicos...** Disponível em: <http://dici.ibict.br/archive/00001077/01/cipecc_liriane.pdf>. Acesso em: 14 jan. 2011.

CORPORATION FOR NATIONAL RESEARCH INITIATIVES. **The Handle System**. 2008. Disponível em: <<http://www.handle.net/>>. Acesso em: 21 jan. 2011.

DSPACE. **DSpace Institutional Digital Repository System**. Disponível em: <<http://www.dspace.org/>>. Acesso em: 18 jan. 2011.

DSPACE System Documentation: configuration and customization. In: THE DSPACE FOUNDATION. DSPACE 1.5.2 Manual. 2009. cap.5. Disponível em:

< http://www.dspace.org/1_5_2Documentation/ch05.html >. Acesso em: 18 jan. 2011.

DUBLIN CORE METADATA INITIATIVE. **Using Dublin Core**. Disponível em: <<http://www.dublincore.org/documents/usageguide/>>. Acesso em: 18 jan. 2011.

LEITE, Fernando César de Lima; COSTA, Sely. Repositórios institucionais como ferramenta de gestão do conhecimento científico no ambiente acadêmico. **Perspectivas em ciência da informação**, v.11, n.2, p.206-219, maio/ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a05.pdf>>.

Acesso em: 20 jan. 2011.

OLIVEIRA, Zita Prates de, PAVÃO, Caterina Groposo, COSTA, Janise Silva Borges da et al. O uso do campo MARC 9xx para o controle bibliográfico institucional. **Ciência da informação**, v.33, n.2, p.179-186, maio/ago. 2004. Disponível em:

<<http://www.scielo.br/pdf/ci/v33n2/a19v33n2.pdf>>. Acesso em: 12 jan. 2011.

ONLINE Dictionary for Library and Information Science - (ODLIS). 2004. Disponível em: <<http://lu.com/odlis/>>. Acesso em: 12 jan. 2011.

THE OPEN Archives Initiative Protocol for Metadata Harvesting. 2008. Disponível em:

<<http://www.openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: 12 jan. 2011.

Comunidades				Contenido
Acervo Fotográfico1	Artículos de Periódicos	Eventos2	Libros e Capítulos de libros	TD3, TCCE4, TCC5
Metadatos				
		contributor:advisor	contributor:advisor	contributor:advisor
				contributor:advisor-co
contributor.author	contributor.author	contributor.author	contributor.author	contributor.author
	contributor.other	contributor.other	contributor.other	
		contributor.event		
coverage.spatial				
coverage.temporal				
date.accessioned	date.accessioned	date.accessioned	date.accessioned	date.accessioned
date.issued	date.issued	date.issued	date.issued	date.issued
				Fecha de publicación
				Fecha de defensa



Comunidades				Contenido
Acervo Fotográfico1	Artículos de Periódicos	Eventos2	Libros e Capítulos de libros	TD3, TCCE4, TCC5
Metadatos				
				degree.department
				degree-grantor
				degree.level
				degree.local
				degree.program
				degree.specialization
description				description
	description.abstract		description.abstract	Resumen/Abstract
			description.edition	Edición
description.historic				Histórico
format.extent	format.extent	format.extent	format.extent	Tamaño del archivo
format.medium				Características del original
format.mimetype	format.mimetype	format.mimetype	format.mimetype	Tipo de archivo



Comunidades				Contenido
Acervo Fotográfico01	Artículos de Periódicos	Eventos2	Libros e Capítulos de libros	TD3, TCCE4, TCC5
Metadatos				
identifier.codmuseu				Número de identificación
	identifier.isbn	identifier.isbn	identifier.isbn	ISBN
	identifier.nrb	identifier.nrb	identifier.nrb	ISSN
identifier.uri	identifier.uri	identifier.uri	identifier.uri	Número de sistema
	language.iso		language.iso	URI
			publisher	Idioma
	relation.ispartof	relation.ispartof	relation.ispartof	Editor
subject	subject	subject	subject	Cita
title	title	title	title	Palabra clave
	title.alternative	title.alternative	title.alternative	Título
type	type	type	type	Título en otro idioma
				Tipo de documento