
ITINERARIOS METODOLÓGICOS COMUNES Y DIFERENCIADOS ENTRE REPOSITORIOS INSTITUCIONALES Y REPOSITORIOS DE PRESERVACIÓN: LA EXPERIENCIA EN CAICYT-CONICET

Tatiana M. CARSEN
tcarsen@caicyt.gov.ar

Mariano COLLADO
mcollado@caicyt.gov.ar

Mirna Silvina PRIETO
mprieto@caicyt.gov.ar

CAICYT-CONICET: Saavedra 15, 1er. Piso (1083) – Ciudad Autónoma de Buenos Aires

RESUMEN: CAICYT como Laboratorio de Información realiza experiencias orientadas a formalizar metodologías, buenas prácticas y la aplicación de lineamientos internacionales con el objeto de capitalizar estas habilidades para transferirlas a la comunidad de especialistas en Ciencias de la información.

En este sentido se trabaja en dos líneas: el Repositorio Institucional y el de Preservación.

Se presentarán:

- a. Proyectos en curso: aspectos conceptuales y bases de implementación
- b. Repositorio institucional: metodologías de modelización de los procesos de población e ingesta
- c. Repositorio de preservación: metodologías de abstracción de los procesos de gestión, entidades y objetos intervinientes y esquemas de metadatos pertinentes
- d. Aspectos comunes y diferenciados de los casos antes mencionados

PALABRAS CLAVES: Buenas prácticas en conservación y preservación, Metadatos de preservación, Modelo conceptual OO, Modelo E-R, Población de repositorios, Políticas de selección documental, Procesos de preservación.

ABSTRACT: As an Information Laboratory, CAICYT (Argentina) works on methodologies and best practices, following international guidelines. Its aim is to capitalize this knowledge and transfer it to the community of specialists in Information Science.

CAICYT maintains two repositories: the Institutional and the Preservation Repository. In this paper, it will be presented:

- a. Current projects: conceptual and implementation bases
- b. Institutional repository: modeling of feeding and loading methodologies
- c. Preservation repository: abstraction of management processes methodologies, involved entities and objects, and appropriate metadata schemes.
- d. Common and distinct aspects of the these cases

KEYWORDS: Best practices in conservation and preservation; preservation metadata; OO conceptual data modeling; E-R data modeling; repositories feeding; document selection policies; preservation processes.

LABORATORIO DE INFORMACIÓN

Los cambios y las innovaciones en el campo tecnológico y normativo documental son constantes y obligan a una dinámica de ritmo elevado y a una interacción constante entre operadores, infraestructuras de información, documentos, usuarios, con diversos grados de complejidad.

Un laboratorio de información supone la experimentación y pruebas “bajo condiciones controladas para normalizar las prácticas” (BOSCH, 2015), la discusión de los resultados que puede derivar en la consolidación de metodologías e instrumentos de trabajo, cuya resultante puede ser: buenas prácticas, protocolos, estándares, programas e infraestructuras mejoradas y posibilitar una ulterior transferencia de los conocimientos así acumulados.

Esto implica un cambio en el tipo de investigación en bibliotecología y documentación, menos descriptiva o fenomenológica y más dirigida a la investigación y desarrollo.

Experiencia en CAICYT-CONICET

En este contexto se promueve, en los procesos de trabajo, el análisis de problemas a resolver en documentación científica y técnica o en la implementación de infraestructuras de información, utilizando herramientas de trabajo colaborativo presencial y virtual entre sus profesionales o junto a profesionales de otros centros o especialistas.

Por supuesto, que esto está en consonancia con los objetivos estratégicos que CAICYT se propone. Dos de los cuales se refieren a:

1. Apoyar a la edición científica
2. Constituirse en un centro de investigación además de su tradicional función como centro de servicios

En el primer caso, el apoyo a la edición científica se realiza a través de brindar una infraestructura al editor científico que facilite, por un lado, su registro en sistemas internacionales y le proporcione asistencia en la aplicación de normas de calidad editorial aceptadas internacionalmente.

En torno a estas actividades de apoyo a la edición se generan una serie de prácticas administrativas, las cuales producen documentación de respaldo para el proceso de

evaluación de las publicaciones científicas argentinas. Esta documentación debe ser resguardada y conservada debidamente, siguiendo criterios propios de la archivística y con herramientas informáticas ya probadas en diversos archivos y museos. CAICYT ha contado con la posibilidad de constituir un equipo dedicado a las tareas de conservación y preservación digital, el cual interviene en los procesos mencionados. Lo hace probando metodologías en las que se incorporen buenas prácticas de gestión documental. Esto conduce a la necesidad de contar con un repositorio de preservación, cuyo itinerario experiencial se explica más adelante.

El segundo aspecto que ha encarado CAICYT es el de evolucionar desde un centro de servicios hacia un centro de investigación. Y un primer atributo de un centro de investigación es contar con producción intelectual propia.

Aquí entra en escena la consolidación de una metodología de diseño y población de un repositorio institucional de pequeño tamaño, desde la selección del software a usar, la construcción de un corpus documental que poblará el repositorio, hasta el conjunto de metadatos seleccionados.

Resulta interesante observar el itinerario metodológico seguido en cada uno de estos casos, ya que varía en cada repositorio debido a sus objetivos particulares.

REPOSITORIO INSTITUCIONAL

Itinerario *Bottom-Up*: respondiendo a las situaciones que se presentan

La metodología seguida (“*bottom-up*”) ofrece la ventaja de una gran flexibilidad al comienzo y a lo largo del proceso, involucrándose con el caso real. Esto conlleva una construcción progresiva del repositorio COMCIENT; la particularidad del método es la de volver sobre pasos ya transitados.

Al plantearse la necesidad de exponer al público la producción intelectual del CAICYT, y dar cumplimiento, de este modo, a lo dispuesto por la Ley Nacional de Repositorios Digitales, hubo que definir:

1. Objetivos del repositorio institucional
2. Qué plataforma de exposición de datos usar
3. A quienes estaría destinada
4. Criterios de población del repositorios y metadatos iniciales

Ante las características del material seleccionado para poblar COMCIENT surgió la necesidad de adecuar el conjunto de metadatos a aquellas y considerar qué servicios deberá ofrecer el repositorio (en proceso de definición).

1. Objetivos del repositorio institucional

Dar acceso al público a la producción intelectual de investigadores, profesionales y técnicos del CAICYT, tanto comunicaciones institucionales, como producción original o resultados de investigación. Dichos documentos podrán también ser compartidos libremente en soporte digital, con el único requisito de respetar la integridad de la obra. También se hará una digitalización progresiva de los trabajos creados en formato analógico.

2. Qué plataforma de exposición de datos usar

El repositorio se llamará ComCient (Comunicación Científica) debido a que el campo de trabajo de CAICYT-CONICET es la documentación y comunicación científica y tecnológica. Se decidió utilizar OMEKA¹ como plataforma de exposición de datos, ya que ofrece las siguientes ventajas:

- Es apto para colecciones de dimensiones diversas, grandes y pequeñas
- Permite mostrar las colecciones cual si fuera una exposición personalizada

¹ **Omeka** es una plataforma de publicación web libre, flexible y de código abierto, pensada para mostrar colecciones digitales y exposiciones virtuales de bibliotecas, archivos y museos. Se trata de un proyecto del [Roy Rosenzweig Center for History and New Media](#). Se encuentra liberado bajo una Licencia de software libre (GPLv3), con lo cual su distribución, uso y modificación son libres.

- Puede ser cosechado por OAI-PMH ya que, por defecto, tiene set de metadatos Dublin Core, que pueden ser editados, ampliados o ajustados a requerimientos propios
- De rápido aprendizaje y adaptación a las necesidades del repositorio
- Para el editor su uso es amigable e intuitivo. Para el público, muestra una interfaz atractiva
- Es integrable a Wordpress y articulable a plataformas como TemaTres y otras
- Finalmente, se seleccionó OMEKA por su amplia y activa comunidad de usuarios, quienes no necesariamente tienen formación informática

3. Destinatarios del repositorio

- Comunidad bibliotecológica
- Comunidad científica en general
- Sistema nacional de repositorios y otros
- Comunidad CONICET

4. Criterios de población del repositorio

Para poblar el repositorio se determinaron una serie de parámetros, a saber:

- Producción de investigadores, profesionales y/o técnicos de CAICYT-CONICET a lo largo de su trayectoria y con una temática propia de este centro
- Documentos en los que CAICYT es el editor, tales como actas de conferencias, etc.
- También se incluyen los documentos de gestión administrativa generados por CONICET que afectan explícitamente a CAICYT, ya sea por su misión estratégica o sus actividades
- Pueden incluirse diferentes tipos documentales: artículos, monografías, comunicaciones a congresos, materiales de cursos, seminarios, jornadas, resoluciones, etc.
- Se optó por realizar una ingesta retrospectiva de registros, desde el presente hasta los inicios de la institución

5. Qué tipo de metadatos se requieren

La entrada de autor y la identificación de títulos se hará siguiendo la *ISO 690*, los metadatos Dublin Core, y las recomendaciones y directrices del SNRD.

Se trabajó en la adaptación de algunos metadatos a la metodología SNRD versión 2012, la cual, a la fecha ha sido reformulada para algunos metadatos. Esto obliga a volver sobre los propios pasos para ajustar nuevamente la estructura de metadatos de COMCIENT.

Experiencia COMCIENT con OMEKA, algunos obstáculos epistemológicos:

La dificultad inicial fue la de cambiar el ámbito habitual de trabajo, de bases de datos del tipo jerárquico a una plataforma en línea orientada a la visualización final.

Una vez superada la resistencia de pasar de una interfaz de carga otra, con maneras de etiquetar diferentes, se realizó una importación masiva de datos, que están siendo curados registro por registro.

Decisiones al poblar COMCIENT:

1. Modalidad de ingesta: archivo mediado por el administrador del repositorio.
2. Metodología para la curación de los datos: carga masiva y nombramiento de los archivos, importación y posterior revisión interna de los datos.
3. Pedido de control y adecuaciones por parte de los autores en caso de considerarse necesarias.
4. Durante el proceso de se observó la necesidad de agregar metadatos siguiendo especificaciones de SNRD al conjunto base propuesto por Dublin Core. Esto queda escrito en un instructivo interno que sirva para documentar los criterios de normalización a emplear.
5. Los archivos se almacenan en formato PDF/A para conservación y preservación.

REPOSITORIO DE PRESERVACIÓN

Itinerario *Top-Down*: del concepto a la práctica

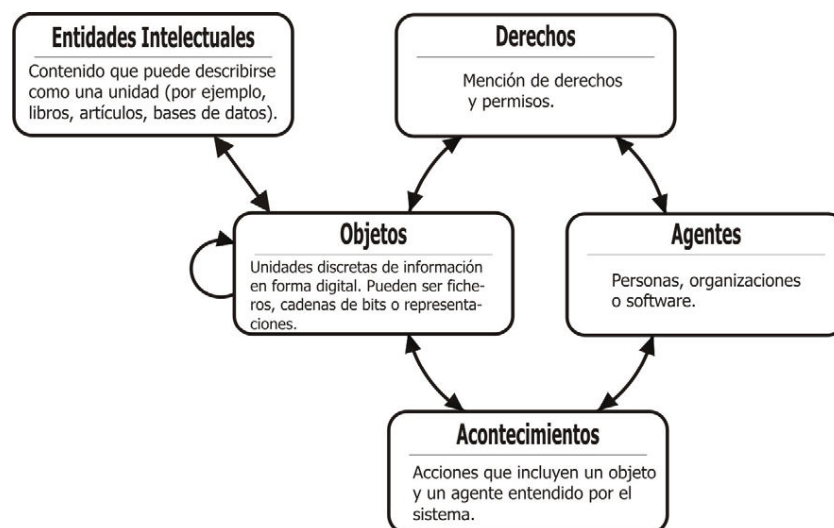
Teniendo presente que *guardar/archivar/poblar un repositorio* no significa automáticamente estar realizando un acto de *preservación documental a largo plazo*, CAICYT se abocó a la elaboración de una metodología que satisfaga los requerimientos necesarios para enfrentar los desafíos de la obsolescencia de software y hardware, que hace peligrar la disponibilidad de los documentos en un futuro no muy lejano.

En este sentido, el mundo anglosajón tiene ya un largo historial en el abordaje de esta problemática, por tanto la primera acción fue la de ir a beber de las fuentes existentes. Algunas de estas fuentes, como se verá, ya han pasado a formar parte del acervo internacional, tomando la forma de Estándares como, por ejemplo, las distintas Normas ISO.

De esta forma, se asimiló el marco teórico *Open Archival Information System OAIS* (ISO 14721:2003), que brinda un modelo de organización y gestión para la preservación de documentos digitales; como así también, el modelo *PREservation Metadata: Implementation Strategies PREMIS*, que propone la utilización de una serie de metadatos específicos para enfrentar los problemas de autenticidad, integridad, fiabilidad y disponibilidad de los documentos electrónicos.

GRÁFICO 1: El modelo de datos PREMIS

fuelle: Caplan, P (2009)



Asimismo, considerando las *Directrices del Sistema Nacional de Repositorios Digitales SNRD 2012*, adoptadas por parte de los Repositorios Institucionales del ámbito de ciencia y técnica argentinos, y que cumplen con el protocolo de interoperabilidad **OAI-PMH**, se ha decidido incorporarlas al proyecto de esquema de metadatos para Repositorios de Preservación.

Por último, se adoptaron como guía las viejas normas de descripción de documentos de archivo *General International Standard Archival Description ISAD(G)*, y la **ISO 15498** de gestión de documentos, ya que, después de todo, se está tratando con documentos de archivo.

Recorrido Experimentado

Teniendo en cuenta las necesidades y las propiedades de los documentos digitales que produce CAICYT, es decir, en su gran mayoría documentos textuales, se elaboró un perfil de aplicación acorde, tomando como referencia los campos o elementos propuestos por los esquemas ISAD(G), PREMIS y SNRD. Como primera valoración, se han incorporado al esquema de metadatos todos aquellos elementos calificados como obligatorios y, a posteriori, fueron agregados algunos de los no obligatorios, por ser considerados de utilidad. Este proceso se ha esquematizado en la siguiente tabla:

TABLA 1: Esquema de metadatos del perfil de aplicación de preservación

ELEMENTO	NORMA		
	ISAD(G)	PREMIS	SNRD
Identificador <i>nombre</i>	1	1.1.1	10
Identificador <i>valor</i>	1	1.1.2	10
Título	2	1.6	1
Autor	6		2
Editor			5
Fecha	3		7
Tipo			8
Materia			3
Idioma	16		12
Descripción			4
Nivel de descripción	4	1.2	
Relación		1.10	13
Tamaño	5	1.5.3	
Formato <i>nombre</i>	17	1.5.4.1.1	9
Formato <i>versión</i>	17	1.5.4.1.2	9
Aplicación creadora <i>nombre</i>	17	1.5.5.1	
Aplicación creadora <i>versión</i>	17	1.5.5.2	
Mensaje cifrado <i>tipo</i>		1.5.2.1	
Mensaje cifrado <i>valor</i>		1.5.2.2	
Evento <i>tipo</i>		2.2	
Evento <i>fecha</i>		2.3	
Evento <i>detalle</i>		2.4	
Derechos de acceso	15	1.13	15
Almacenamiento		1.7	

Nota: los números corresponden al ítem con que cada norma designa al elemento en cuestión

Referencias:

Metadatos administrativos

Metadatos descriptivos

Metadatos técnicos

ISAD (G): *General International Standard Archival Description*

PREMIS: *Preservation Metadata: Implementation Strategies*

SNRD: *Sistema Nacional de Repositorios Digitales*

Como puede apreciarse en la TABLA 1, los elementos de tipo técnico están destinados a favorecer la preservación a largo plazo de los documentos, haciendo mención al entorno tecnológico, y tomando en cuenta los cambios y actualizaciones que van experimentando a través del tiempo, como es el caso de las migraciones de formato.

El paso siguiente a la constitución de la TABLA 1, de los elementos que conforman los metadatos, es la elaboración de un Diccionario de Datos con la definición y descripción de cada uno de los ítems de la lista. Este paso es fundamental en vistas a la implementación de un Repositorio de Preservación, ya que transcribe el modelo descriptivo/conceptual en un modelo lógico aplicable a un software de gestión documental.

Como complemento necesario a la utilización de estos metadatos, es clave el cumplimiento de determinadas buenas prácticas que aporten eficacia al sistema. Una de estas prácticas es la unificación y normalización de los formatos de los artefactos digitales a preservar. En CAICYT se decide unificar los formatos procedentes de diferentes aplicaciones convirtiéndolos en archivos de tipo PDF/A, como paso previo al ingreso en el sistema automatizado. Este tipo de formato está considerado internacionalmente como un formato de preservación (ISO 19005-1:2005), ya que es de carácter autocontenido y capaz de ejecutarse de forma satisfactoria en diversas herramientas de visualización.

Este proceso de investigación en torno a la preservación digital, que se lleva a cabo en CAICYT, se encuentra aún en su primera etapa, en su estadio teórico/conceptual. El paso siguiente es su conversión a un lenguaje lógico, capaz de ser interpretado por un sistema informático. Y por último, la implementación y puesta en práctica del modelo en condiciones de laboratorio controladas.

CONCLUSIONES

Es importante el desarrollo de un marco conceptual para cada caso:

En lo referente a la metodología *Botton-Up*, dicho marco se hizo explícito al momento de la implementación de la herramienta y se fue documentando para su continuidad metodológica.

Para el itinerario *Top-Down* se ha partido de un modelo conceptual fuertemente orientado a los requerimientos visionados para el producto final, la meta a alcanzar.

En ambos itinerarios es una práctica imprescindible la documentación de los pasos seguidos, para poder consolidar el conocimiento tácito adquirido en el proceso y convertirlo en un conocimiento explícito que pueda ser transferido a nuevos integrantes de los respectivos equipos, o bien aplicables por grupos o instituciones que encaren una tarea semejante dentro de la comunicad científica.

Estas formas de documentación pueden tomar diferentes modalidades:

- Documentos de Trabajo
- Plataformas de gestión de conocimiento (wiki institucional)
- Otras formas de comunicación que sirvan de soporte para la transferencia y articulación de conocimiento intra e inter institucional

En tal sentido ambos perfiles de aplicación forman parte de un paquete metodológico propio que CAICYT ofrece a su comunidades de usuarios con los modos de transferencia previstos por la institución: Consultorías, Formación, Tutorías...

BIBLIOGRAFÍA

Alonso, J.A.; García-Alsina, M.; Lloveras, R. *La norma ISO 15489 : un marco sistemático de buenas prácticas de gestión documental en las organizaciones*. Ítem : revista de biblioteconomía i documentació, 2007, n. 47, pp. 41-70. Disponible en:

<http://eprints.rclis.org/12263/>

Argentina. Ministerio de Ciencia, Tecnología e Innovación Productiva. Directrices SNRD: Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales. Buenos Aires: 2015. Consultado el 11 de agosto de 2015:

http://repositorios.mincyt.gob.ar/pdfs/Directrices_SNRD_2015.pdf

Caplan, P. (2009). *Entender PREMIS*. Consultado el 11 de agosto de 2015:

http://www.loc.gov/standards/premis/UnderstandingPREMIS_espanol.pdf

Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS)*. Junio de 2012.

<http://public.ccsds.org/publications/archive/650x0m2.pdf>

PREMIS (2015). *PREMIS Data dictionary for preservation metadata, version 3.0*. Consultado el 11 de agosto de 2015:

<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

UNE-ISO 23081-1:2008 *Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1: Principios*.

UNE-ISO 19005-1:2008 *Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1)*.

Unesco (2003). *Directrices para la preservación del Patrimonio digital*. Preparado por la Biblioteca Nacional de Australia. Marzo de 2003. Consultado el 11 de agosto de 2015:

<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>