

Aplicaciones informáticas de software libre en archivos: una visión integradora

Silvana Temesio

Montevideo, Uruguay

Documentación digital y aplicación de las TIC

silvana.temesio@fic.edu.uy

Abalos 3148, Montevideo, Uruguay

Uruguay. Universidad de la República. Facultad de Información y Comunicación. Instituto de Información.

Eje tematico: 3. Procesos, tecnologías e innovaciones

Resumen

El mundo de los archivos cambió rápidamente la tipología de usuarios a los que asiste y la manera como éstos se enfrentan la información, con el advenimiento de la mentalidad participativa de la web 2.0, el desarrollo de las TIC, el gobierno electrónico y las leyes de acceso a la información pública. Los archivos tienen que dar solución a disponer su información en la web, o a la interna de las instituciones y habilitar capacidades de búsqueda de información e interacción crecientes.

Se plantea una solución informática para archivos, entendida como una aplicación o un conjunto de aplicaciones que actúan integradas y que resuelven:

- Registros de ISAD(G), ISAAR (CPF), ISDIAH, ISDF
- Puntos de acceso normalizados e integrados a otras facetas de la información
- Documentos digitales y su preservación digital
- Ingesta desde otras aplicaciones como expediente electrónico o gestores documentales
- Ingesta desde sistemas legados como winisis o planillas electrónicas

Lo que se detalla es la solución práctica a estos requerimientos a través de una serie de aplicaciones que se proponen para trabajar articuladamente:

- ICAATOM, TEMATRES, ARCHIVEMATICA, ALFRESCO

Se analiza la aplicación de estándares informáticos y archivísticos y los aspectos de interoperabilidad de la información que resultan sustantivos en una sociedad de información en red.

Palabras clave: software libre, archivos, estándares, ICAAtOM, Tematres, Archivematica, Alfresco

Abstract

World of archives has quickly changed the type of users who attends and also how the information they are faced. That occur with the advent of participatory mentality of Web 2.0, the development of ICT, e-government and access laws to public information. Archives must solve how to get its records available on the web, or in the private institutions and how to enable search capabilities over information and also increasing interaction from users. A software solution for archives defined as an application or set of applications and with integrated operation arises:

- Records of ISAD (G), ISAAR (CPF), ISDIAH, ISDF
- Standardized and integrated Access Points into to other facets of information
- Digital documents and digital preservation
- Ingest from other applications such as electronic file or document management
- Ingest from legacy systems as winisis or spreadsheets

What is discussed is how to address these requirements through a range of software applications and how they can work integrated:

- ICAATOM, tematres, ARCHIVEMATICA, ALFRESCO

The application of computer and archival standards, and interoperability aspects of information, items that are substantive in a networked information society, is analyzed.

Keyword: free software, archives, standards, ICAAtOM, Tematres, Archivematica, Alfresco

Introducción

Los cambios tecnológicos como la rápida expansión de las redes de telefonía, la velocidad de conexión de Internet, la proliferación de dispositivos y su integración a los servicios y a los sistemas de información establecen una nueva forma de relacionarse con la información. La información es accesible desde múltiples interfaces y dispositivos. Estos cambios han sacudido a los archivos, a los usuarios, y a la manera como éstos interaccionan con la información archivística. El archivo custodia, pero también exhibe la información que custodia en la vitrina de la web, y luego aún la tendrá que disponer en múltiples dispositivos y deberá adaptarse a los requerimientos de los niveles de servicio e interacción que los usuarios exijan. Este es un camino que está en su inicio y que da señales de un largo y vertiginoso trayecto.

Esta situación plantea múltiples miradas, una de ellas es que información archivística disponer en la web. La elección de los fondos a exhibir conlleva una postura ideológica ya que ninguna visión es ajena a una determinación subjetiva. La selección será seguramente la opción practicable y si bien se puede tener como criterio realizar una descripción exhaustiva de los recursos archivísticos ésta será un objetivo con hitos a ir completando. No solo deberá decidirse que mostrar, sino con que profundidad. Los recursos digitales pueden mostrarse todos, pero habrá que decidir que resulta ilustrativo o conveniente mostrar, eligiendo con algún criterio una porción representativa o especial de la documentación que se dispone. También se deberán sopesar la implantación de servicios que puedan ofrecer el envío o disponibilización de aquellos recursos que no se pongan en exhibición.

Otros aspectos a considerar del punto de vista tecnológico son los procesos que garanticen la calidad e integración de la información archivística con otras piezas de información no archivísticas con las que opere la Institución.

En la trastienda de este escaparate archivístico en la web o a nivel de la intranet institucional es necesario garantizar un acoplamiento semántico

armónico y de calidad de la información. Los canales informativos, los flujos de datos, la integración de servicios producen una interacción que es transparente para el usuario pero que debe tejerse desde la planificación del servicio de exhibición de los archivos, para lograr una integración que sea consistente.

Finalmente como último punto de esta trastienda, es necesario elegir las aplicaciones y los estándares archivísticos, informáticos y de comunicación e integración para sustentar este andamiaje.

Software libre

La elección de aplicaciones tiene un punto de evaluación desde las prestaciones y la adecuación a las tareas que desarrolla el software, pero tiene también un posicionamiento ideológico. La tecnología no es neutral y los presupuestos sobre los que se basa son significativos a una visión ética.

La visión del software libre se alinea con libertad del conocimiento, con el acceso abierto, y con la visión colaborativa aplicada al software.

El software libre brinda la libertad de conocer lo que realiza la aplicación informática exponiendo su código fuente. El software es un conjunto de sentencias de un lenguaje de programación (código fuente) que especifican las acciones que realiza la aplicación. Estas sentencias están escritas en un idioma extranjero –un lenguaje de programación-, no todos conocemos la lengua, pero existen diccionarios –los manuales de referencia de los lenguajes- y es posible leer este libro tan especial si el libro se publica. En el caso del software libre es requisito la publicación del código fuente de la aplicación. El software libre puede ser copiado y compartido, se puede estudiar y cambiar ajustándolo a nuestros propósitos.

Una aplicación es software libre si los usuarios tienen las cuatro libertades esenciales:

- La libertad de ejecutar el programa como se desea, con cualquier propósito (libertad 0).

- La libertad de estudiar cómo funciona el programa, y cambiarlo para que haga lo que usted quiera (libertad 1). El acceso al código fuente es una condición necesaria para ello.
- La libertad de redistribuir copias para ayudar a su comunidad (libertad 2).
- La libertad de distribuir copias de sus versiones modificadas a terceros (libertad 3). Esto le permite ofrecer a toda la comunidad la oportunidad de beneficiarse de las modificaciones. El acceso al código fuente es una condición necesaria para ello.

Para saber si un software es libre o no debemos fijarnos en su licencia. La licencia de un software libre está claramente especificada junto con el software ya que es un punto importante para la comunidad establecer el licenciamiento. La licencia más conocida y emblemática es GNU-GPL (General Public License), pero existen otras licencias y puede verificarse en la página de la Free Software Foundation (FSF) ¹.

Otro aspecto a considerar son los formatos. Los formatos son codificaciones de datos legibles por aplicaciones, por ejemplo un archivo .docx es interpretable por una aplicación de procesador de texto, en este caso microsoft word. Para poder visualizar un formato .docx es necesaria la aplicación microsoft word. La aplicación en este caso no es software libre, se trata de software propietario. Los formatos que emanan de aplicaciones de software libre están documentados, se conocen sus especificaciones y aún en la situación de no contar con el software que nos permite gestionarlos, como el formato está documentado, se puede desarrollar una herramienta para poder visualizar el contenido de un documento digital.

Una directiva de preservación digital es justamente por estas razones expuestas, el uso de formatos abiertos.

En lo que sigue se comentarán distintas aplicaciones de software libre y la manera como pueden articularse para trabajar integradamente.

¹ <http://www.gnu.org/licenses/license-list.html#Introduction>

ICAATOM

ICA-AtoM es acrónimo de "International Council on Archives - Access to memory" (Consejo Internacional de Archivos - Acceso a la Memoria). Es una aplicación de software libre con arquitectura web para descripción archivística normalizada, en un entorno multilingüe y multi-repositorio disponible en: <https://www.ica-atom.org/>

Esta aplicación nace de un proyecto de la International Council of Archives (ICA) "Open Source Archival Information System " (OSARIS) y luego la UNESCO se une apoyando un proyecto sobre derechos humanos para crear una guía online sobre la violación de los derechos humanos.

El propósito del proyecto fue proporcionar una aplicación de software libre²:

- que permitiera a las instituciones disponer sus fondos archivísticos online, especialmente aquellas instituciones que no tuvieran otra manera de disponerlas
- que las descripciones archivísticas estuvieran de acuerdo con los estándares de la International Council of Archives (ICA)
- que se proporcionara una interface multilingüe y opciones para realizar traducciones
- que tuviera una arquitectura web, tomara en cuenta los aspectos de usabilidad y mejores prácticas.
- que fuera flexible y personalizable
- que fuera útil para pequeñas y grandes instituciones
- que proporcionara implementación tanto para un único repositorio o múltiples repositorios

² https://www.ica-atom.org/download/ICA-AtoM_JBushey.pdf

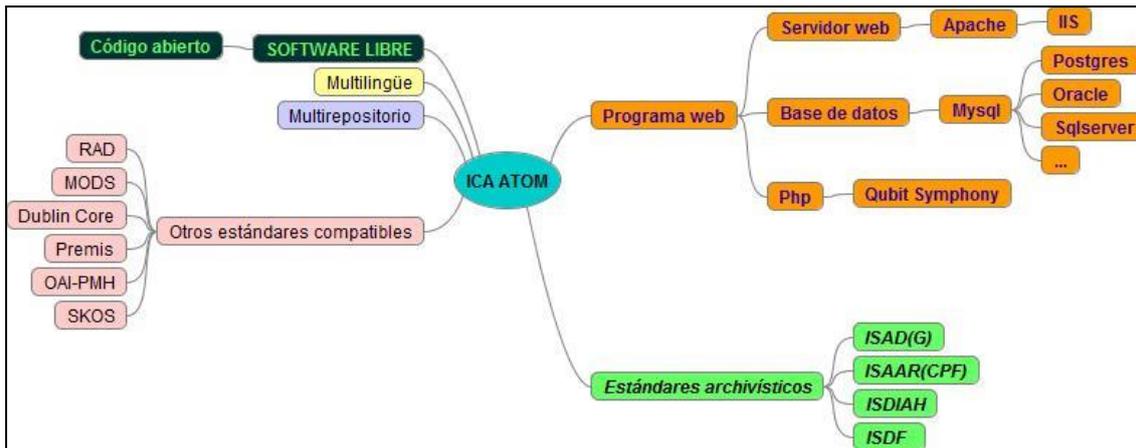


Figura 1. Mapa mental características técnicas ICAATOM

ICAATOM es una aplicación web –una aplicación que corre en un navegador web-.

Cabe aclarar que la aplicación se puede instalar

- en forma local en un equipo y usarse solo en ese equipo
- en un equipo y habilitarse el acceso en una red interna, con lo cual se accederá desde cualquier equipo en la red interna
- en un equipo publicado en Internet y accederse por la web. El equipo puede estar en la Institución o puede instalarse en un datacenter en la nube

En cualquiera de estos casos la aplicación es web, corre en un navegador web, que puede ser mozilla, internet explorer, chrome, etc.

La aplicación guarda sus datos en una base de datos, por defecto se instala con mysql, pero admite otras bases de datos libres como postgres o propietarias como oracle.

Los archivos digitales que se adosen a los datos se guardan en el sistema de archivos.

Para que una aplicación pueda ser ejecutada en un navegador es necesario un programa que sea servidor web donde será instalada la aplicación. El servidor web por defecto es Apache, aunque admite otros.

Icaatom está programado con php, por lo que debe habilitarse este lenguaje en el servidor web y los entornos usados son Qubit y symfony.

Todas las interacciones del usuario con el sistema (crear, ver, buscar, actualizar y borrar) tienen lugar a través del navegador web del usuario. Los usuarios acceden a las páginas HTML en el servidor web; al clicar botones se ejecuta código php que envía comandos a la base de datos y retorna la salida en respuesta al navegador web.

ICA-AtoM está construido con herramientas de software libre (Apache, MySQL, PHP, symfony) no con software propietario. El conjunto de herramientas Qubit Toolkit fue desarrollado por el proyecto ICA-AtoM y es de código abierto, es decir que el código está disponible para usar o modificar. (Licencia GPL versión 2).

No tiene costo alguno descargar cualquier software requerido para ejecutar la aplicación ICA-AtoM.

ICA-AtoM está construido en torno a los estándares descriptivos del Consejo Internacional de Archivos (CIA)

- Norma Internacional General De Descripción Archivística (ISAD(G) - segunda edición, 1999
- Norma Internacional de Registro de Autoridad (Organizaciones, Personas, Familias) (ISAAR(CPF)) - 2nd edition, 2003
- Norma Internacional para la descripción de instituciones con material de archivo (ISDIAH) - 1st edition, March 2008
- Norma Internacional para describir funciones (ISDF) - 1st edition, May 2007

ICA-AtoM está diseñado para adaptarse a otros estándares como Dublin Core, un estándar de metadatos de propósito general, y otras normas archivísticas como RAD o MODS. También es compatible con otros estándares como SKOS (estándar para tesauros), OAI-PMH (protocolo para compartir información) y Premis (metadatos de preservación).

Todos los elementos de la interfaz de usuario y el contenido de la base de datos puede ser traducido a múltiples idiomas. La versión actual de ICA-AtoM soporta varios idiomas: Árabe, Holandés, Inglés, Farsi, Francés, Alemán, Griego, Islandés, Italiano, Japonés, Coreano, Portugués, Esloveno y Español.

ICA-AtoM puede ser utilizado por una institución concreta para sus propias descripciones o puede ser instalado como un multi-repositorio, una especie de "catálogo colectivo" que acepte descripciones de instituciones participantes.

En el sitio se puede acceder a abundante documentación sobre instalación, actualización, manual de usuario, manual de administración y preguntas frecuentes. También se pueden revisar casos de uso de instituciones archivísticas que lo han adoptado.

El foro de soportes es una herramienta de uso común en el software libre y constituye una ayuda para todo tipo de problemas, inquietudes, comentarios que puedan realizarse sobre la aplicación. Este foro es muy activo y contiene mucho material valioso de situaciones planteadas.

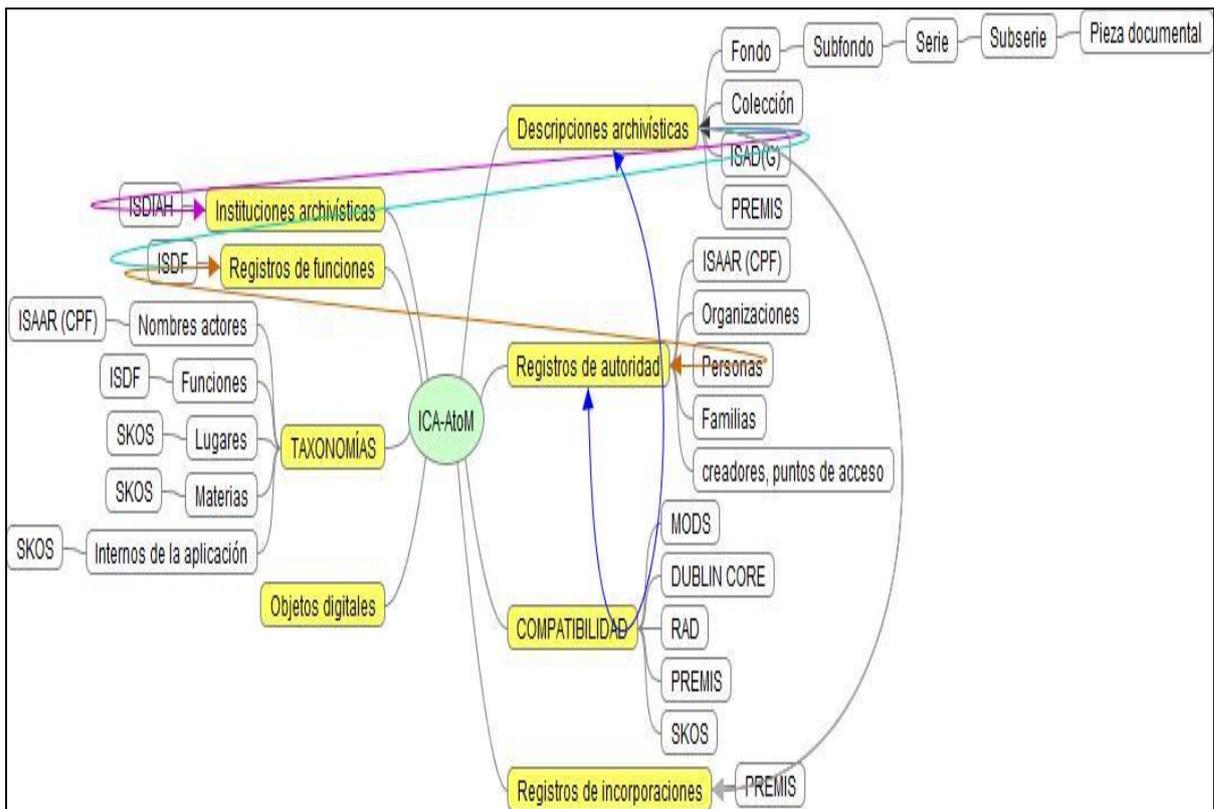


Figura 2. Mapa mental características generales ICAATOM

La aplicación permite el ingreso de registros de descripción archivística, explicitando la estructura jerárquica: fondo, subfondo, serie, subserie, pieza documental, etc., de acuerdo a las normas de descripción archivística ISAD(g).

Los registros de autoridad que pueden corresponder a organizaciones, personas, familias o actores que tienen relación con la descripción archivística y que son susceptibles de ser accedidos como puntos de acceso se describen de acuerdo a la norma ISAAR(CPF)

Los registros correspondientes a la entidad custodiadora, la institución archivística, se describe de acuerdo a la norma ISDIAH.

Cada descripción archivística tiene una relación con uno o varios registros de autoridad y con una institución archivística y es posible navegar desde un elemento a otro.

Las funciones desarrolladas en una institución archivística se describen con un registro ISDF y es posible navegar hacia los otros elementos relacionados como la descripción archivística y el registro de autoridad.

En las últimas versiones se ha incorporado un registro de incorporaciones que son aquellos fondos recibidos por alguna vía como la donación entre otras y que luego recibirán una descripción archivística y que ya pueden consignar los metadatos de preservación PREMIS.

No abundaremos aquí en el modelo de datos y las relaciones que corresponden al dominio archivístico y que ICAATOM implementa y que son objeto de análisis en un artículo que puede consultarse.³

Algunos aspectos de la aplicación que son de importancia para la interoperabilidad y la recuperación de la información son: los puntos de acceso y la compatibilidad con Dublin Core respectivamente.

En la descripción archivística se incorporan los puntos de acceso por lugares, nombres y materias. En ICAATOM al ingresar puntos de acceso por nombre se genera en forma automática un registro de autoridad, implementando de esa forma una normalización de los puntos de acceso de nombres. En el caso de los puntos de acceso por lugares o materias se agregan en una taxonomía.

Una taxonomía es un sistema de clasificación sobre un conjunto de ítems que permite su ordenación. Se estructura en un árbol jerárquico. Se parte de un concepto global de cierta complejidad y se subdivide en conceptos progresivamente más específicos.

Este concepto de taxonomía sistematiza los términos usados, estableciendo el término adoptado y consecuentemente remitiendo a éste desde términos similares. También se modelan las relaciones de granularidad (términos específicos) y las relaciones generales (términos relacionados).

ICAATOM modela también taxonomías de uso interno que pueden personalizarse como es el caso de los niveles de descripción (fondo, subfondo,

³ <http://informatio.eubca.edu.uy/ojs/index.php/Infor/article/view/125>

etc.) sistematizando de esta manera los elementos flexibles para que se usen en forma estandarizada.

Las taxonomías están regidas por el estándar Skos-Core (Simple Knowledge Organizations System)⁴ al cual la aplicación adhiere.

Las descripciones archivísticas pueden ingresarse con una plantilla de ISAD(G) o con una plantilla de Dublin Core y sucede lo mismo con la exportación que puede realizarse en ambos formatos. Eso es posible porque existe a nivel del modelo conceptual de ICAATOM la implementación de un mapa cruzado de las concordancias entre ambos sistemas de metadatos. La correlación no es absoluta porque Dublin core tiene menos expresividad que ISAD(G), pero para aquellos elementos en que hay concordancia ésta se establece. La importancia de tener una forma de ingreso o de salida con Dublin Core tiene que ver con la posibilidad de relacionarse con otras aplicaciones o fuentes de datos que manejan Dublin Core ya que es uno de los estándares de metadatos para objetos digitales de mayor uso.

A modo de muestra de un mapa cruzado se expone el área de contenido y estructura, con las equivalencias entre isad(g) y ead con los campos equivalentes de ambos en ICAATOM:

⁴ <http://www.w3.org/2004/02/skos>

3.3 CONTENT AND STRUCTURE AREA				
ISAD(G)	ICA-AtoM Field(s)	EAD (Tag Library Crosswalk)	AtoM Export EAD (1.3)	Comments
3.3.1 Scope and content	Scope and content	<scopecontent >	<scopecontent encodinganalog="3.3.1"> < p >	Comments
3.3.2 Appraisal, destruction and scheduling information	Appraisal, destruction and scheduling	<appraisal>	<appraisal encodinganalog="3.3.2"> < p >	
3.3.3 Accruals	Accruals	<accruals>	<accruals encodinganalog="3.3.3"> < p >	
3.3.4 System of arrangement	System of arrangement	<arrangement>	<arrangement encodinganalog="3.3.4"> < p >	

Figura 3. Parcial de mapa cruzado

Al tener implementadas las concordancias entre distintos estándares es posible incorporar piezas documentales provenientes de gestores documentales o aplicaciones de expediente electrónico, que tengan a su vez una salida en Dublin Core, lo cual es una buena práctica.

También usando Dublin Core es posible compartir metadatos remitiendo a los objetos digitales a través del protocolo OAI-PMH, para así poder formar portales generales de archivos que puedan ofrecer sus metadatos para la búsqueda en un único sitio pero que cuando tienen que mostrar el recurso remiten al sitio de origen.

ICAATOM implementa los metadatos PREMIS para la preservación digital que se ingresan en las descripciones archivísticas o pueden heredarse desde el registro de incorporaciones (accession records).

Cuando no se contaba con aplicaciones archivísticas adecuadas los registros se llevaban en aplicaciones legadas como winisis o planillas electrónicas, de resultas de lo cual muchos registros se encuentran actualmente de esta manera y será necesario migrarlos.

ICAATOM ofrece dos opciones una importación en XML o en .csv. Tanto desde winisis como de planillas electrónicas no se podrán obtener registros xml adecuados en forma sencilla, por lo cual la manera más práctica de realizar la ingesta de estos datos será a través del formato .csv. El formato .csv es un archivo de texto plano separado por comas y puede visualizarse y transformarse fácilmente al formato de una planilla electrónica, de modo que a los efectos prácticos se trabajará en una planilla electrónica para luego que los datos estén prontos para ser migrados transformarse nuevamente en .csv. Desde winisis habrá que realizar una exportación a planilla electrónica de todos los datos y luego vendrá la tarea de ver que columna corresponde a cada uno de los datos del registro ISAD(G). ICAATOM documenta el nombre de cada uno de los campos del registro ISAD(G) que tomará para la importación de los datos de modo que es necesario encabezar las columnas de esa planilla electrónica con los nombres de los campos y poner los datos adecuados en cada columna provenientes de la exportación de winisis. Se puede realizar la migración del registro EAD con todos los niveles jerárquicos anidados, solo es necesario establecer el vínculo a través del campo específico que detalla la anidación jerárquica

TEMATRES

TemaTres es una aplicación web de software libre con interface multilingüe (italiano, alemán, español, inglés, francés y portugués). Su utilidad es la gestión de lenguajes documentales, vocabularios controlados, tesauros, taxonomías y otros modelos de representación formal del conocimiento. Implementa el desarrollo de tesauros jerárquicos con las relaciones jerárquicas (término específico/ término general), relaciones asociativas (término

relacionado) y relaciones de equivalencia (usado por). Permite generar notas de alcance de distinto tipo. Adhiere a los estándares de competencia como Skos-Core (Simple Knowledge Organizations System) Zthes, TopicMap, Dublin Core, MADS, BS8723-5 (Structured Vocabularies for Information Retrieval), RSS, SiteMap, TopicMaps (ISO/IEC 13250:2003), IMS VDEX Scheme (Vocabulary Definition and Exchange), JSON y JSON-LD.

Es un programa web que funciona con un servidor web (Apache, IIS), una base de datos (mysql, postgres, Oracle, etc.) y está desarrollado en php.

Presenta amplia documentación y un sitio de demo para realizar pruebas⁵.

TemaTres ofrece una interfaz de servicios web (web services) que permite y facilita su integración o articulación con otras plataformas de gestión o el desarrollo de servicios derivados basados en la explotación de vocabularios controlados.

Existe una serie de herramientas asociadas entre las que se destaca, visualvocabulary que realiza representaciones visuales de los términos del tesaurus y plugins para usar en wordpress (blogs) y en open journal systems (revistas electrónicas).

Como tematres e ICAATOM adhieren a SKOS es posible trasvasar las taxonomías de ambas aplicaciones entre sí. La razón por la que se plantea usar una aplicación específica para el vocabulario controlado de una organización y no hacerlo directamente en ICAATOM o no hacerlo solamente en ICAATOM, es que los archivos no son los únicos repositorios de información que precisan ser indizados en una organización. Dentro de la organización existirán gestores documentales, bibliotecas digitales a predominio textual o a predominio gráfico, gestores de contenidos, expediente electrónico, datos geográficos, aplicaciones informáticas, colecciones de fotografías, videos, etc. La Institución a la que pertenece el archivo tiene necesidades diversas de recuperación de información, pero tiene que establecer un lenguaje consensuado para todos sus objetos, que sea único. La información precisa

⁵ <http://www.r020.com.ar/tematres/wiki/doku.php?id=:en:Inicio>

constituye un insumo para la toma de decisiones, para planificar, proyectar y establecer políticas y acciones empresariales. Para extraer conocimiento de la información contenida en la documentación generada por el organismo es necesario realizar una recuperación adecuada. Los puntos de acceso son herramientas importantes para la recuperación de la información, por lo cual deben estar previstos y ser homogéneos en su ámbito de uso. Cada institución tiene su terminología específica en la cual se asientan las definiciones de trabajo implícitas en su utilización, éstas luego pueden establecerse como un nexo cuando se relacionen con otras instituciones generando enlaces interinstitucionales con interoperabilidad semántica. La organización, primero hacia la interna debe tener un lenguaje claramente establecido y documentado. Este lenguaje será usado para la recuperación de objetos diversos: documentos, gráficos o multimedia, aplicaciones etc. Este lenguaje de la organización debe elaborarse en forma participativa con el aporte de los usuarios técnicos que conocen la jerga de uso y su alcance y los profesionales de información que manejan el conocimiento de la arquitectura de la información. Este lenguaje necesita además estar claramente documentado y ser ampliamente difundido, conocido y manejado en la organización para que resulte de utilidad. En este caso lo que se estandariza a nivel de la organización es el valor del punto de acceso, aportando un nombre único y un significado claro y establecido que elimine ambigüedades y contribuya a recuperar la información adecuada que se requiera.

La fundamentación y análisis de esta propuesta puede consultarse⁶ y también su implementación en un archivo gubernamental que la ha adoptado⁷.

Preservación digital

Los objetos digitales están sujetos a problemas de obsolescencia por cambios tecnológicos o problemas con los medios.

⁶ <http://revistaacervo.an.gov.br/seer/index.php/info/article/view/615>

⁷ http://archivo.mtop.gub.uy/index.php/?sf_culture=es
<http://tesauro.mtop.gub.uy>
<http://tesaurovisual.mtop.gub.uy>

La preservación digital consiste en la ejecución de una serie de actividades que aseguren el acceso a los materiales digitales tanto como sea necesario en el tiempo. Se debe entonces, garantizar la autenticidad, exactitud y funcionalidad del objeto digital.

La preservación digital es un aspecto poco visible para los cuadros jerárquicos en una organización. Cuando un objeto digital sea requerido y no pueda brindarse por resultar el medio inaccesible, recién entonces se comprenderá cabalmente la necesidad de establecer una política de preservación digital.

Es necesario entonces realizar una tarea de difusión del tema y generar proyectos de preservación digital que puedan irse transitando antes de que se pierdan los activos documentales de la organización.

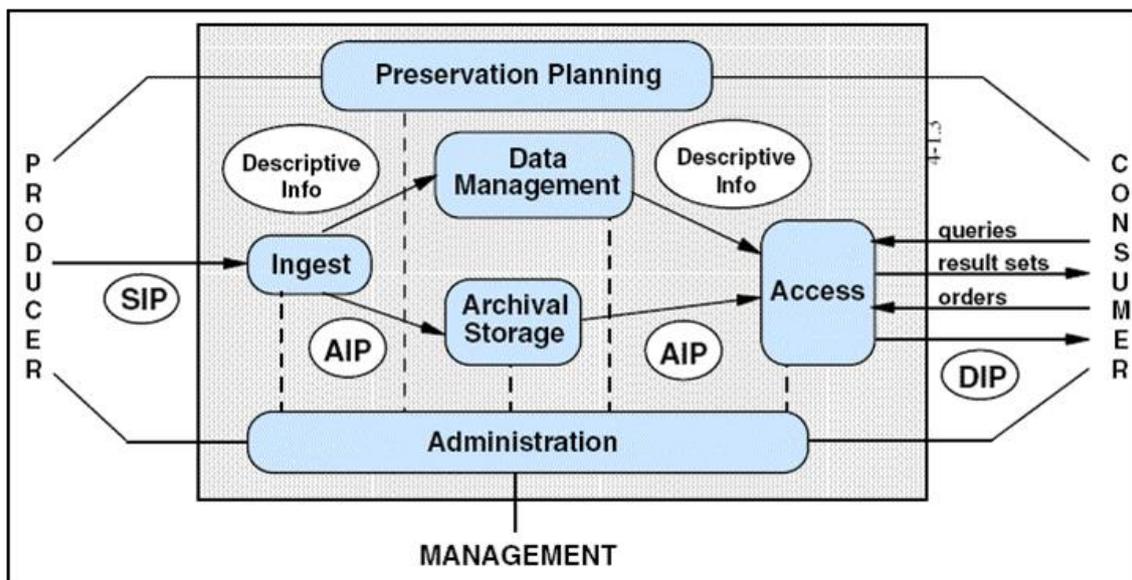


Figura 4. Esquema OAIS

El OAIS (Open archival information system -ISO 14721) es un modelo conceptual de referencia para preservación de largo plazo. No especifica detalles funcionales o técnicos, sino que constituye una guía de los procesos, los actores y sus roles, las responsabilidades y los métodos que involucran la preservación a largo plazo.

Tenemos un recurso digital que tiene un formato y está guardado en un medio y es accedido por un programa que lo disemina. Para fijar ideas, digamos que

tenemos un archivo .doc adosado a la descripción de una pieza documental simple en ICAATOM y es guardado en el disco de la red. El modelo OAIS nos proporciona un marco conceptual para poder preservar este archivo a largo plazo. Para ello se vale de la idea de paquete que es una unidad de información autodescriptiva que encapsula la interpretación de la información contenida para lidiar con los cambios en la tecnología y la representación. El paquete es el conjunto del documento y sus metadatos (descriptivos, contextuales, de preservación) y existen tres tipos de paquetes en el modelo, el paquete que es ingestado al sistema de preservación (SIP), proveniente del productor; el paquete para almacenar (AIP) como copia de preservación y el paquete para diseminar y presentar ante las búsquedas del consumidor (DIP). Cada paquete tiene una función, un proceso y es manejado por distintos actores con distintos roles.

Este paquete de ingesta que contiene el documento .doc y sus metadatos ISAD(G) o dublin core y premis, se migra a un formato de preservación por ej. open office y se encapsula en un paquete de almacenamiento AIP y también se genera un paquete de diseminación DIP en un formato PDF con algunos metadatos que se consignan como relevante para ser desplegados al usuario final.

Para realizar estos procesos que plantea el modelo OAIS podemos usar Archivematica. Archivematica es una aplicación de software libre diseñada para brindar acceso a largo plazo a colecciones de objetos digitales a través del uso de estándares.

Cuenta con la colaboración de la UNESCO a través del proyecto Memoria del mundo y de otras instituciones como el archivo municipal de la ciudad de Vancouver y la biblioteca de British Columbia en Canadá, entre otras.

Utiliza micro servicios que proporcionan una serie integrada de herramientas que permiten a los usuarios procesar objetos digitales desde la ingesta hasta el acceso de acuerdo al modelo funcional ISO-OAIS.

Los procesos se controlan a través de un tablero web que indican la situación de los paquetes y detallan los procesos a ser cumplidos. Utiliza METS, PREMIS, Dublin Core y otros estándares y buenas prácticas. Implementa planes de preservación de medios basados en características significativas de los formatos de archivo a través de un plan de preservación ⁸.

El código está disponible para ser descargado en <http://www.archivematica.org>. Archivematica actúa en forma integrada con ICAAtOM e interactúa con otras aplicaciones como DSpace e Islandora, entre otras.

Expediente electrónico o gestor documental

Existen dos situaciones diferentes en cuanto a procesos documentales: el expediente electrónico y el gestor documental.

El advenimiento del gobierno electrónico como proceso global a nivel de los gobiernos y el avance tecnológico respecto a la firma digital han popularizado aplicaciones de expediente electrónico. Estas aplicaciones, con todas sus posibles variantes, básicamente siguen un flujo de procesos, con usuarios y roles especificados y van generando documentos que se van adhiriendo al camino que sigue el proceso. Cuando el expediente se archiva y ya no fluye más a través de la organización, el documento adosado o generado es susceptible de integrarse a una aplicación de descripción archivística como ICAAtOM y desde allí ser consultado y preservado.

Esto se puede hacer de muchas maneras. Si es posible extraer metadatos de la aplicación de expediente electrónico y estos metadatos son estándares como por ejemplo Dublin Core, entonces se puede ingestar el documento digital y sus metadatos. Si esto no es posible o si se decide simplemente guardar la serie, o el nivel que fuera, lo que se puede hacer es ubicar los documentos digitales en el sistema de archivos y realizar la descripción de ese nivel en ICAAtOM y referirla al sitio del sistema de archivos, para que de esta manera sea alcanzado por la aplicación de preservación digital.

La gestión documental tiene que ver con la práctica de trabajo con los documentos a nivel de la organización en otro marco. En este caso hay son grupos de trabajo sobre documentos, con diferentes roles, eventualmente flujos

⁸ https://www.archivematica.org/wiki/Media_type_preservation_plans

(v.g. revisión, aprobación, versionado) y finalmente el documento se publica o se hace algún proceso con él, pero está provisionalmente terminado.

La aplicación de gestión documental en forma muy simplificada se ocupa de centralizar el documento en un único sitio accesible a todos los roles que van a actuar sobre éste. Respecto a los roles están los usuarios que podrán modificar y crear los documentos y los usuarios que podrán solamente leerlos. El documento puede ser público o confidencial y su visualización será controlada de acuerdo a los roles establecidos.

El documento va recorriendo un proceso de creación colectiva o individual y se generan distintas versiones. Eventualmente el documento se integra a un flujo, donde por ejemplo un grupo elabora un documento que es luego revisado o aprobado por otro de acuerdo a lo que se defina. Así el documento va siguiendo un proceso y en cierto sentido tiene analogías con el expediente electrónico, pero en este caso, el objeto central es el documento, no el proceso, aunque el documento pueda moverse en un proceso. En el caso del gestor documental es importante consignar aspectos como versionado y la participación en una escritura grupal, que no se toman en cuenta en el expediente electrónico y aquí son medulares. Un caso emblemático de gestión documental, son los procesos de calidad, que conllevan una serie de documentos que deben ser elaborados por distintas personas y luego aprobados en un ciclo definido, quedando disponibles para una auditoría.

Si bien se pueden forzar las utilidades de una u otra aplicación, los enfoques son diferentes.

Alfresco es un sistema de gestión de contenidos de código abierto que permite gestionar los contenidos críticos empresariales, para cualquier tipo de procesos. Lleva versionado de documentos, tiene un motor de flujo de trabajo configurable para adecuar a los procesos específicos, desde un proceso de revisión y aprobación hasta un proceso más complejo de varias fases. Se integra con las aplicaciones de procesamiento de texto, locales o en la nube. Proporciona una serie de metadatos estándares (Dublin Core) y permite realizar una extensión de esos metadatos.

Los aspectos de búsqueda y recuperación de documentos están bien diseñados y permiten rápidamente ubicar los documentos de acuerdo a sus

metadatos. Se puede usar en forma local o como servicio en la nube y se adapta a todo tipo de dispositivos móviles como tablets y smartphones.

Alfresco tiene una versión libre de la comunidad distribuido con la licencia LGPLv3, que puede descargarse de forma gratuita⁹.

Los documentos generados con Alfresco luego de cumplido su flujo de proceso pueden ingestarse en el ICAAtóm con sus metadatos dublin core.

Alfresco corre sobre un servidor de aplicaciones que puede ser Tomcat o JBoss, está desarrollado en java y utiliza una base de datos que puede ser postgresSQL. Los documentos que gestiona pueden estar en diversos formatos y trabaja con una serie de aplicaciones para los tipos de documentos como OpenOffice, ImageMagick, aunque soporta otras.

Existen otros gestores de contenido de software libre de prestaciones similares entre los que se puede mencionar a Nuxeo¹⁰.

Será necesario definir si estos documentos que fluyen por la organización van a conservarse y donde. Puede ser en el archivo con una descripción archivística, pero puede considerarse agruparlos además en una colección digital o biblioteca digital donde puedan ser fácilmente consultados. El archivo convive con diferentes necesidades, situaciones, formatos, y habrá que analizar en función de las necesidades actuales o proyectadas de los usuarios, los procesos de ingesta y los formatos digitales, un tratamiento doble para algunos recursos digitales especiales.

Si se trata de una colección de fotos, o documentos gráficos existen aplicaciones como Omeka¹¹ o Resource Space¹². Estas aplicaciones pueden categorizarse como de gestión de recursos digitales (DAM-Digital Asset Management). Permiten agrupar los recursos digitales por colecciones y disponerles metadatos Dublin Core y extenderlos. Ambas son aplicaciones web LAMP (Linux, apache, mysql, php). Tienen la posibilidad de definir usuarios que suban sus propios recursos al estilo de un repositorio. Ambas se especializan en colecciones gráficas aunque admiten otro tipo de recursos como texto,

⁹ <http://www.alfresco.com/es/descargue-e-instale-alfresco-community>

¹⁰ <http://www.nuxeo.com/>

¹¹ <http://omeka.org/>

¹² <http://www.resourcespace.org/>

sonido y video. Cuando se sube el recurso automáticamente se genera la resolución en miniatura para mostrar en la colección, una resolución media para mostrar cuando es elegida para desplegarse y el archivo original en su resolución original para ser descargado. La búsqueda por metadatos es muy eficiente y es posible hojear las colecciones de miniaturas con mucha sencillez.

Si la colección digital fuera a predominio textual, digamos resoluciones administrativas que son distribuidas y consultadas en la organización, o reglamentos o leyes, entonces existe un gestor de colecciones digitales que tiene gran versatilidad con la recuperación textual: Greenstone¹³. Greenstone también admite otros formatos y es una aplicación más versátil y flexible, pero tiene una gran eficiencia para la recuperación de texto. Permite ingestar desde sitios web, hasta archivos digitales, pasando por recolección a través del protocolo OAI-PMH. Tiene un editor de metadatos que ofrece algunos metadatos de uso generalizado y la posibilidad de extenderlos o generar el propio. Greenstone también realiza una extracción de metadatos en forma automática. Fue producido por la Universidad de Waikato en Nueva Zelanda, se desarrolló y distribuyó en colaboración con la Unesco bajo la licencia libre GPL. En América Latina existe un foro de usuarios y técnico muy activo¹⁴ y gran cantidad de colecciones digitales disponibles en la web.

Conclusiones

Los archivos tienen un papel muy relevante en una organización de cualquier índole, constituyen la memoria, la traza de los acontecimientos, los procesos, las actividades, dan testimonio, son una herramienta para la auditoría, el control, la evaluación y la planificación estratégica. Pero los recursos son siempre escasos y los archivos encuentran dificultades para asegurar sus requerimientos. El software libre constituye una oportunidad de realizar un trabajo profesional que pueda ser visible y permita mostrar las potencialidades de la recuperación de la información de una manera eficiente a través de las

¹³ http://www.greenstone.org/index_es

¹⁴ http://www.freelists.org/list/greenstone_es

TIC. Debe resultar claro que el software libre y en el caso de los mencionados aquí, gratuitos, tienen costos ocultos que debemos evaluar.

Todas las aplicaciones que se mencionaron son estables y cumplen criterios de calidad de aplicaciones, pero unas más, otras menos requieren capacitación y conocimiento. Existe un costo de aprendizaje, en tiempo, en esfuerzo, y económico también. Existe un costo de recursos informáticos capacitados para la instalación, plataformas adecuadas en donde realizarlas y soporte para los procesos de preservación, respaldo y contingencia. Estos costos no son altos, porque toda la documentación y el conocimiento está libre, abierto, disponible, pero son recursos requeridos si se desea llevar a buen puerto un proyecto de estas características.

Los archivos tienen la responsabilidad de coordinar con otras fuentes de información de la organización que no pertenecen al archivo y establecer vínculos, proyectos limítrofes y una terminología estandarizada y consensuada. Una responsabilidad hacia adentro de la organización de un lenguaje común y el aseguramiento de que la información de diversas fuentes pueda ser navegable y buscable sin tropiezos.

Pero muchos archivos también se abrirán a Internet y encontrarán Instituciones hermanas con las que compartir información. Existe una manera federada de compartir información a través del protocolo OAI-PMH. Este protocolo permite que un sitio web agrupe toda la información respecto a un tema, el sitio no es productor de información sino que recolecta de los productores los metadatos que sirven para la búsqueda y cuando el internauta requiere un recurso el protocolo enlaza con el sitio de producción y realiza la entrega. Podría ser el caso de un Archivo Nacional, que disponga fondos propios, pero también ofrezca la información de otros archivos bajo esta modalidad.

Cómo última reflexión, lo que nos facilita la integración de la información y procesos es la adherencia a estándares: estándares archivísticos, terminológicos, de comunicación, informáticos. Los estándares son la lengua común por la cual tenemos que transitar para poder acoplarnos y entendernos.

El trabajo pretende mostrar que existe una solución de software libre para cada nicho de actividad y que esas soluciones pueden y deben integrarse a través de una reflexión prospectiva y la adherencia a las buenas prácticas y los estándares.

Bibliografía consultada

Gonzales Aguilar, Audilio, Ramírez Posada, Diego Ferreyra, María. Tematres: software para gestionar tesauros (2012). *El profesional de la información*, 21(3), 319-325. Recuperado de http://www.academia.edu/3768287/Tematres_Software_para_Gestionar_Tesauros

Temesio, Silvana (2013). Interoperabilidad de la información en el gobierno electrónico. *Acervo revista do arquivo nacional*. 26 (2), 189-206. Recuperado de <http://revistaacervo.an.gov.br/seer/index.php/info/article/view/615>

Temesio, Silvana (2012). Modelo conceptual de archivos: caso de estudio ICAAtoM. *Informatio*, 17, 9-31. Recuperado de <http://informatio.eubca.edu.uy/ojs/index.php/Infor/article/view/125/173>

Temesio, Silvana . (2010). El software libre, el profesional de la información y la EUBCA. En Segundo Congreso Uruguayo de Bibliotecología e Información/Congreso Regional de Bibliotecología e Información (CUBI-II/CORBI). Asociación de bibliotecólogos de Uruguay, Montevideo, Uruguay. <http://www.abu.net.uy/sites/default/files/documentos/cubi2/TemesioS.pdf>