

Proyecto de Grado

Bibliotecas Digitales para la Biblioteca Nacional y Facultad de Información y Comunicación

Autores

Florencia Borghini
Reyna Der Boghosian
Leonardo Casep

Tutora

Regina Motz

Carrera de Ingeniería en Computación
Facultad de Ingeniería
Universidad de la República
Montevideo – Uruguay
Diciembre de 2015

RESUMEN

Este proyecto se desarrolló bajo el marco del convenio realizado entre la Universidad de la República (UdelaR), Facultad de Información y Comunicación (FIC) y la Biblioteca Nacional del Uruguay (BNU).

Este convenio propone generar una Biblioteca Digital con el software de Repositorios DSpace para la BNU y otra para la FIC con objetos digitales documentales, manuscritos, imágenes, video, audio y otros, donde cada una tiene estructuras jerárquicas de información, interfaz gráfica de presentación y flujo de trabajo diferentes, según necesidades expresadas por cada institución, pero a su vez compartirán contenidos acordados entre las partes de las Publicaciones Periódicas a través de la sincronización automática de la Biblioteca Digital de la FIC hacia la Biblioteca Digital de la BNU.

El proyecto tiene como objetivo determinar las necesidades y requerimientos de ambas instituciones, realizando una propuesta de metadatos para cada objeto digital cumpliendo con los estándares Dublin Core, migrar hacia la Biblioteca Nacional la información existente de Publicaciones Periódicas del Uruguay perteneciente a la FIC así como también mantener sincronizados automáticamente estos contenidos con la Biblioteca Digital de la BNU.

Palabras clave:

Biblioteca Digital, Repositorio, Sincronización, Metadatos, Publicaciones Periódicas, DSpace

TABLA DE CONTENIDO

Resumen	2
Índice de Figuras	6
Índice de Tablas	8
Capítulo 1	9
1. Introducción.....	9
1.1. Objetivos.....	10
1.2. Alcance del Proyecto	10
1.3. Cronograma	10
1.4. Organización del documento	11
Capítulo 2	12
2. Estado del Arte	12
2.1. Concepto General.....	12
2.2. Relevamiento del trabajo de la BNU	13
2.3. Relevamiento del trabajo de digitalización	13
2.4. Estándares y Tecnologías	15
2.4.1. Metadatos	15
2.4.2. Repositorios Digitales y Bibliotecas Digitales	16
2.4.3. Software de Repositorios	18
2.5. Sincronización.....	24
Capítulo 3	27
3. DSpace.....	27
3.1. Arquitectura	27
3.2. Estructura de Base de datos.....	28
3.3. Estructura de almacenamiento (Bitstream Store)	28
3.4. Administración de DSpace	30
3.5. Flujo de Trabajo (Workflow)	32
3.6. Registro de formatos.....	33
3.7. Estructura de importación de ítems	34
3.8. Control de integridad y calidad.....	35
3.9. Interfaces de integración	36
3.9.1. XMLUI	36
3.9.2. JSPUI	37
3.9.3. RDF	37
3.9.4. SOLR	38
3.9.5. OAI.....	38

3.9.6.	LNI.....	38
3.9.7.	REST.....	38
3.9.8.	SWORD.....	39
3.10.	Seguridad.....	39
3.11.	Búsqueda avanzada, básica y filtros.....	40
3.11.1.	Búsquedas por Facetas (Sidebar Facet).....	41
3.11.2.	Búsquedas simples y avanzadas.....	41
3.11.3.	Búsqueda por texto completo.....	42
3.12.	Indexación.....	42
Capítulo 4	45
4.	Propuesta de Solución.....	45
4.1.	Propuesta de Metadatos.....	45
4.2.	Calidad Metadatos.....	56
4.3.	Estructura jerárquica BNU.....	57
4.4.	Estructura jerárquica FIC.....	59
4.5.	Propuesta de Flujo de Trabajo.....	60
4.5.1.	Propuesta de flujo Alternativa 1.....	61
4.5.2.	Propuesta de flujo Alternativa 2.....	62
4.6.	Complementos (Plugins).....	63
4.7.	Interfaz de usuario.....	64
4.8.	Sincronización.....	65
4.9.	Propuesta Autores del Uruguay.....	66
4.10.	Propuesta frontend prototipo inicial.....	69
4.11.	Propuesta frontend BNU.....	75
4.12.	Propuesta Frontend FIC.....	77
Capítulo 5	79
5.	Implementación.....	79
5.1.	Metadatos.....	79
5.1.1.	Metadatos en Ítems de Colecciones.....	79
5.1.2.	Metadatos en Colecciones.....	81
5.2.	Sincronización.....	84
5.3.	Complementos.....	86
5.3.1.	Complementos para imágenes:.....	86
5.3.2.	Complemento para PDF.....	88
5.3.3.	Complemento para Video.....	89
5.3.4.	Complemento para Audio.....	91
5.3.5.	Complemento Carrusel.....	92

5.3.6.	Personalización de la interfaz	93
5.3.7.	Personalización de la Interfaz BNU	93
5.3.8.	Personalización de la Interfaz FIC.....	97
5.3.9.	Proceso de conversión de información	104
Capítulo 6	106
6.	Pruebas Realizadas	106
6.1.	Carga masiva	106
6.2.	Sincronización.....	106
6.3.	Carga archivos grandes	106
Capítulo 7	108
7.	Conclusiones y Trabajo a Futuro	108
7.1.	Conclusiones.....	108
7.2.	Dificultades.....	108
7.3.	Aportes	110
7.3.1.	Bibliotecas Digitales	110
7.3.2.	Jornada Biblioteca Digital Universitaria	110
7.3.3.	Uruguay crece contigo	111
7.4.	Trabajo a futuro.....	111
7.4.1.	Agregar complementos para otros formatos de archivos	111
7.4.2.	Calidad.....	111
7.4.3.	Mejorar ingreso de metadatos	112
7.4.4.	Extracción automática de metadatos.....	112
Glosario	114
Referencias	116
Anexo Manual de Instalación DSpace 5.x en Ubuntu 14.02	120
Anexo Configuraciones Realizadas	124
Anexo – Reuniones	131

ÍNDICE DE FIGURAS

Figura 1 - Trabajo actual de digitalizadores de la FIC.....	14
Figura 2 – Proceso de digitalización de la BNU	14
Figura 3- Usabilidad a nivel mundial de repositorios digitales de acceso libre [39]	23
Figura 4 – Funcionamiento del Protocolo OAI-PMH.....	25
Figura 5 - Interacción entre componentes [41].....	26
Figura 6 - Arquitectura de DSpace [43]	28
Figura 7-Estructura de almacenamiento	29
Figura 8 - Nombre de archivo	29
Figura 9 - Menú de administración	30
Figura 10- Flujo de trabajo estándar de DSpace	32
Figura 11- Estructura de importación de archivo	34
Figura 12 – Interacción entre capas [47]	37
Figura 13- Ejemplo de búsqueda facetada.....	41
Figura 14 - Esquema de Estructura de la BNU.....	59
Figura 15 – Esquema de la estructura de la FIC	60
Figura 16- Flujo propuesto 1.....	61
Figura 17 - Flujo propuesto 2.....	62
Figura 18 - Últimos ítems agregados	63
Figura 19- Uso de interfaces [52].....	64
Figura 20- Esquema de sincronización OAI	66
Figura 21-Propuesta de solución del árbol de Autores del Uruguay.....	67
Figura 22- Flujo de trabajo propuesta.....	68
Figura 23 - Propuesta de FIC para Autores del Uruguay.....	68
Figura 24 - Localización del menú y texto introductorio en la versión por defecto de DSpace	69
Figura 25- Localización del menú y texto introductorio en la versión del prototipo.....	70
Figura 26- Comunidades en versión por defecto	70
Figura 27 - Comunidades y área de noticias en el prototipo	71
Figura 28- Pie de página versión por defecto	71
Figura 29 - Pie de página prototipo	71
Figura 30- Página de una subcomunidad - colección en versión por defecto	72
Figura 31 - Página de una subcomunidad en prototipo.....	72
Figura 32 - Página de despliegue de ítems de una colección en versión por defecto.....	73
Figura 33 - Despliegue de un ítem en versión por defecto.....	73
Figura 34 - Despliegue de ítems de una colección en prototipo	74
Figura 35 - Paleta de colores FIC.....	77
Figura 36 - Índice actual de Periódicas	78
Figura 37 - Pagina de OAI desde DSpace	84
Figura 38 - Ingreso de colección a cosechar	86
Figura 39 - Complemento Imagen y Zoom	87
Figura 40 - Complemento PDF	89
Figura 41- Complemento Video.....	90
Figura 42 - Complemento de audio	91
Figura 43 - Carousel para últimos ítems agregados.....	92
Figura 44 - Barra de navegación	94
Figura 45 - Portada de la biblioteca digital de la BNU	95
Figura 46 – Formulario de edición de metadatos para BNU	97
Figura 47 - Portada de Anáforas.....	99
Figura 48 - Subcomunidades y colecciones	100

Figura 49 - Ejemplo de un ítem con complemento PDF -	101
Figura 50 - Búsqueda en Anáforas.....	102
Figura 51 - Búsqueda por autor	103
Figura 52 - Ejemplo de ingreso de link para Autores del Uruguay	103
Figura 53 - Colecciones de Autores del Uruguay	104
Figura 54 - Registro de bitstream en versión por defecto	109
Figura 55 - Registro de bitstream corregido	109
Figura 56 - Arquitectura propuesta	112

ÍNDICE DE TABLAS

Tabla 1- Comparación de software de repositorios [38]	22
Tabla 2- Metadatos para Colecciones de Imágenes	47
Tabla 3 - Metadatos para Colecciones de Libros	50
Tabla 4 - Metadatos para colecciones de Mapas	52
Tabla 5- Metadatos a nivel ítem de Publicaciones Periódicas	54
Tabla 6- Metadatos a nivel de colección de Publicaciones Periódicas	56
Tabla 7- Soporte de video HTML5 de los navegadores [61]	89
Tabla 8 - Formatos adicionales de video	90
Tabla 9 - Soporte de audio HTML5 de los navegadores [61]	91
Tabla 10 - Formatos adicionales de audio	92

CAPÍTULO 1

1. INTRODUCCIÓN

Este proyecto se desarrolló bajo el marco de un convenio [1] entre la Universidad de la República y la Comisión Sectorial de enseñanza y Biblioteca Nacional que nace ante la necesidad de permitir el acceso libre, universal y gratuito al patrimonio documental nacional contribuyendo a la identidad cultural del país, la Biblioteca Nacional del Uruguay (BNU) dando cumplimiento a uno de sus cometidos específicos que es la difusión de sus colecciones digitalizadas en los últimos diez años, y la Facultad de Información y Comunicaciones de la UdelaR (FIC) con fines académicos y curriculares de docencia e investigación.

El objetivo general de este convenio es incluir en dos bibliotecas digitales los contenidos que se acuerdan entre las partes de las publicaciones periódicas digitalizadas por la FIC, que aparecen en el sitio “Periódicas” [2] y llegar a la consolidación de La Biblioteca Digital de Uruguay, contribuyendo asimismo a la mejora de la gestión de la documentación con valor patrimonial en posesión de la BNU.

Para cumplir con parte del convenio mencionado, becarios de la FIC se encuentran digitalizando dentro de la BNU procurando el acrecentamiento de la documentación disponible y mejorando la calidad de las imágenes. Esta información era volcada en el repositorio de Publicaciones Periódicas, en el software de publicaciones digitales Omeka [3]. Debido a que el software permite ingresar una mínima cantidad de metadatos básicos y trabajan con una versión del software no actualizada, lo cual dificulta el agregado de información útil y necesaria para lograr una buena calidad de información, decidieron utilizar otro software que se ajuste mejor a estas necesidades.

Al comienzo, el proyecto se nos presentó como la extensión de los metadatos de Dublin Core [4], esquema de metadatos utilizado por la BNU en su repositorio, implementado en DSpace [5], para reflejar la realidad de interés, especialmente en investigación, junto con un el grupo de la FIC y el desarrollo de un repositorio para la FIC con el contenido de Publicaciones Periódicas.

Relevando los requerimientos se vislumbró que no contaban aún con dicho repositorio, debido a falta de tiempo y problemas de desarrollo por parte de la BNU. Por lo que se extendió el proyecto a la realización de una biblioteca digital para la BNU, independiente de la biblioteca digital para la FIC.

En resumen, el proyecto consistió principalmente en la investigación de softwares de repositorios utilizados a nivel mundial y en particular del software seleccionado a través de común acuerdo por ambas instituciones, DSpace, y en el desarrollo de dos bibliotecas digitales para ambas instituciones. Ambas bibliotecas comparten los contenidos acordados entre las partes de las Publicaciones Periódicas, a través de un mecanismo de sincronización automática de la Biblioteca Digital de la FIC hacia la

Biblioteca Digital de la BNU. Además se debió definir en conjunto los metadatos para cada objeto digital.

Adicionalmente por parte de la FIC se pidió la posibilidad de incluir la información existente en el sitio Autores del Uruguay [6], alojada en una wiki. Debido a que la migración completa del sitio no era viable debido a la dificultad que la misma presenta, se proporcionó una propuesta alternativa de solución siendo esta aceptada la cual se detallara más adelante.

1.1. OBJETIVOS

En el marco de este proyecto los objetivos son:

- Realizar el estudio del estado del arte de los software de repositorios existentes y definir si el elegido por ambas instituciones cumple con las necesidades.
- Definir en conjunto con las partes los metadatos y la calidad para cada objeto digital.
- Customizar e implementar mejoras de funcionalidad al software seleccionado mediante complementos.
- Migrar en conjunto con las partes involucradas la información actualmente existente dentro del repositorio de Publicaciones Periódicas.
- Mantener sincronizados ambos repositorios automáticamente.
- Presentar una propuesta para las colecciones de Autores del Uruguay sin la migración de los datos existentes.

1.2. ALCANCE DEL PROYECTO

Como se menciona anteriormente el principal objetivo de este proyecto es la implementación de dos bibliotecas digitales, trabajando de forma independientes pero donde una de ellas debe consumir los datos de la otra, por este motivo el trabajo presentado tiene una fuerte componente de investigación acerca de software de repositorio utilizados a nivel mundial, específicamente el repositorio DSpace.

1.3. CRONOGRAMA

En el siguiente diagrama se muestra la distribución a lo largo del proyecto de algunas de las principales actividades

Mes	Abril				Mayo				Junio				Julio				Agosto				Setiembre				Octubre				Noviembre				Diciembre			
Semana	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Inicio del proyecto																																				
Relevamiento de requerimientos																																				
Relevamiento del estado del arte																																				
Instalación de ambiente y DSpace																																				
Pruebas de la plataforma																																				
Preparacion prototipo																																				
Presentacion prototipo																																				
Periodo de prueba por las instituciones																																				
Agregar complementos																																				
Customizacion FIC																																				
Customizacion BNU																																				
Pruebas de sincornizacion																																				
Preparacion y presentacion JBUD																																				
Documentación																																				

1.4. ORGANIZACIÓN DEL DOCUMENTO

El informe se encuentra estructurado en 8 capítulos. En los mismos, se pretende mostrar al lector una visión global de los aspectos más relevantes del proyecto. A continuación se realiza un resumen del contenido de cada uno de los capítulos:

Capítulo 2 - Desarrolla los temas relacionados al análisis del problema, mostrando una visión general del estado del arte.

Capítulo 3 - Investigación del repositorio DSpace utilizado para este proyecto.

Capítulo 4 - Propuesta de solución y decisiones tomadas.

Capítulo 5 - Implementaciones realizadas y mejoras de la interfaz del software.

Capítulo 6 - Pruebas realizadas.

Capítulo 7 - Finalmente, se describen los aportes, dificultades y conclusiones obtenidas y el trabajo futuro.

Al final de este documento se encuentran las **Referencias** citadas a lo largo del mismo.

CAPÍTULO 2

2. ESTADO DEL ARTE

La etapa de análisis en la fase inicial del proyecto, tiene como objetivo principal la visión general de los requerimientos del proyecto la definición del entorno de desarrollo del mismo.

En un principio, el estudio del estado del arte, se enfocó en la búsqueda de información referente a la detección de los cambios y optimización en los procesos que hasta el momento se estaban aplicando. Como resultado de esta actividad se encontró que por parte de la BNU no había propuestas concretas ni procesos que se estuviesen aplicando dentro de un repositorio digital. Los documentos digitalizados eran guardados en un servidor, sin difusión debido a no poseer un repositorio digital. Por otro lado, realizan la catalogación de documentos como ser libros, mapas, revistas en formato MARC 21 [7].

Sin embargo por parte de la FIC, se encuentran realizando el proceso de ingreso de los documentos digitalizados y un conjunto básico de metadatos, como ser título y fecha, en un software de repositorios llamado Omeka, para la descripción del objeto. Más adelante se analiza más en detalle este software utilizado.

Se identificaron que se necesitaban herramientas tecnológicas más actuales y que se ajusten mejor a los requisitos proporcionados para la realización del repositorio digital solicitado por ambas instituciones.

A continuación presentaremos los aspectos más relevantes del proyecto.

2.1. CONCEPTO GENERAL

Los siguientes conceptos formalizan lo mencionado en apartados posteriores.

Metadata - Definidos como “datos de los datos” que describen el contenido, calidad, condición y otras características de los datos. Son relevantes a la hora de ayudar a los usuarios a encontrar los datos deseados. Es el término usado en la era de Internet para la información que los bibliotecarios tradicionalmente habían puesto en los catálogos, y más comúnmente se refiere a información descriptiva sobre recursos de la Web.

Comunidad - DSpace está dividido en comunidades, las cuales pueden ser divididas en sub-comunidades, reflejando una estructura que encapsula la información. Dentro de cada comunidad puede existir un número ilimitado de sub-comunidades y un número ilimitado de colecciones.

Colección - Las Comunidades contienen Colecciones, los cuales son grupos de contenidos relacionados, ítems.

Ítem - Objeto digital, cada colección se compone de ítems, los cuales son los elementos básicos de archivo. Cada ítem es propiedad de una colección. Adicionalmente, un ítem puede aparecer en colecciones adicionales; sin embargo cada ítem tiene una y solo una colección a la cual pertenece.

2.2. RELEVAMIENTO DEL TRABAJO DE LA BNU

La BNU inició su proceso de automatización hacia 2005, adquiriendo por convenio con la UDELAR el Software de Gestión de Bibliotecas (SGB) Aleph [8] en 2006. Aleph es un sistema de gestión completo e integrado que maneja todos los aspectos de los servicios bibliotecarios, tanto para el personal como para los usuarios.

Los SGB son programas que permiten la gestión automatizada de los procesos bibliotecarios, principalmente catalogación y tematización, préstamo y adquisiciones. La cara visible al usuario final de estas aplicaciones es el llamado OPAC (acrónimo de Open Public Access Catalogue), que es básicamente una versión electrónica del antiguo catálogo en fichas de la biblioteca. Desde que las bibliotecas han agregado a sus colecciones objetos digitales éstos son catalogados y vinculados a los registros que los describen, generalmente mediante un enlace externo hacia la URL donde esté alojado el objeto.

Sería deseable por parte de los bibliotecólogos que a futuro los OPAC y los manejadores de Bibliotecas Digital (BD) y Repositorios digitales se fusionen en una única plataforma de software, como es el caso del nuevo OPAC de la Biblioteca del Congreso de los Estados Unidos, lanzada en 2014. De todas maneras la práctica estándar sigue siendo el de mantener la biblioteca virtual separada del OPAC, muchas veces con un nombre e imagen institucional distinto del resto de la biblioteca. Tal es el caso a nivel mundial de la Biblioteca Nacional de Francia (Gallica) [9], Biblioteca Nacional de España [10], y en la región las bibliotecas nacionales de Argentina (Trapalanda) [11], Brasil (BNDigital) [12] y Chile (Memoria chilena) [13].

Los motivos de esta separación son que las BDs están enfocadas únicamente a la presentación de contenidos digitales, muchas veces creados en forma colaborativa entre muchas instituciones distintas, en tanto que los OPAC presentan registros bibliográficos de la biblioteca y están orientados a la búsqueda.

La adopción de DSpace como plataforma para la publicación de las colecciones digitales de la BNU no va a desplazar el software SGB, ni el SGB podrá sustituir a una plataforma de software dedicada a mantener la futura BD.

2.3. RELEVAMIENTO DEL TRABAJO DE DIGITALIZACIÓN

El análisis efectuado determinó como es el proceso de digitalización de los textos por parte de los becarios de la FIC el cual describimos a continuación:

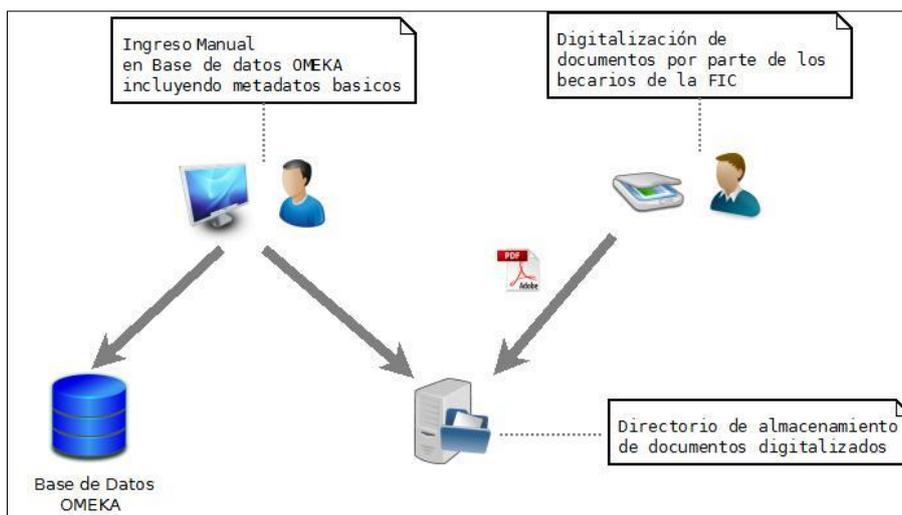


Figura 1 - Trabajo actual de digitalizadores de la FIC

Por parte de la FIC se observó que para el ingreso de objetos digitales cuentan con una base de datos basada en OMEKA, una plataforma de software libre cuya finalidad es la de crear exposiciones virtuales a partir de colecciones de documentos digitales. Algunas desventajas de este software es que brinda un conjunto básico de metadatos para descripción de objetos y no tiene soporte técnico.

Por otro lado se efectuó el análisis del proceso que realiza la BNU para la digitalización de objetos, la cual se representa en el siguiente diagrama:

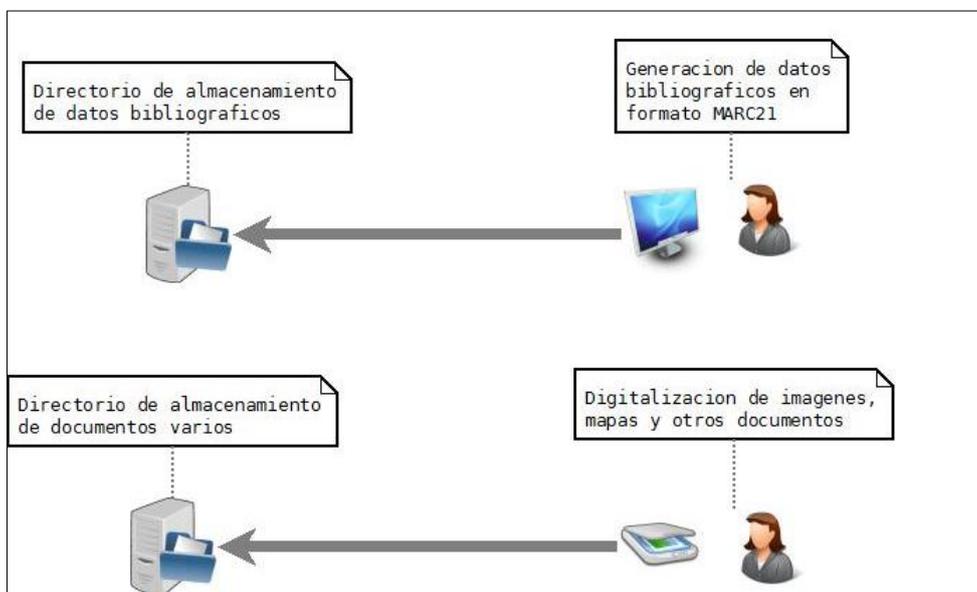


Figura 2 – Proceso de digitalización de la BNU

Los expertos de la BNU realizan la digitalización de imágenes, mapas y documentos que luego son almacenados en un directorio, pues aún no existe ningún tipo de repositorio digital como ya hemos mencionado.

A su vez en paralelo con dicho proceso se encuentran los bibliotecólogos quienes se encargan de la conversión manual de los metadatos en formato MARC 21 a Dublin Core para dichos objetos. Luego estos son almacenados en un directorio.

En el relevamiento realizado no se encontró ningún proceso que siga alguna norma de trabajo para procesar esta información.

2.4. ESTÁNDARES Y TECNOLOGÍAS

2.4.1. METADATOS

Con el crecimiento de la publicación electrónica y las bibliotecas digitales, la concurrente "sobrecarga de información" que resulta de los grandes volúmenes de datos digitales disponibles en línea, uno de los grandes objetivos del proyecto es mejorar la calidad y enriquecer la información a través de los metadatos, es por eso que una de las razones que brindó relevancia a la hora de analizar la plataforma DSpace fue como la misma maneja la calidad y los metadatos.

A partir del análisis que se efectuó en el proceso, se observaron una serie de aspectos que se destacan debido a la importancia a la hora de definir metadatos que sean más exhaustivos y completos a los que ya se tenían al momento para detallar al objeto digital.

Gran parte de los documentos digitalizados requieren una descripción más detallada y enriquecedora del mismo. Aportar al usuario la máxima cantidad de información posible es vital para ayudarles en la búsqueda siendo esta eficiente y exitosa. Por tal motivo el uso y detalle de ingreso de metadatos conforma una parte esencial de este proyecto.

MARC 21

MAchine-Readable Cataloging (MARC) [14] es una especificación de datos y un conjunto de estándares asociados, utilizado por las bibliotecas para codificar y compartir información sobre libros y otros documentos. Se creó en los años sesenta y sigue la especificación ISO 2709 [15] para definir la estructura de cada registro.

MARC 21 [7] es la norma técnica para codificar la información bibliográfica. Este formato fue creado por un equipo de bibliotecarios de la Biblioteca del Congreso (Estados Unidos) en combinación con Canadá. Permite estructurar e identificar los datos de tal forma que puedan ser reconocidos y manipulados por computadora.

Con el surgimiento de los metadatos, la Biblioteca del Congreso de Estados Unidos, encargada de mantener del formato MARC 21, desarrolló tres esquemas de metadatos basados en MARC: son el MARC XML [16], el MODS [17] y el MADS [18].

El MARC XML es una transposición literal del formato MARC 21 al lenguaje XML, en tanto MODS y MADS son versiones simplificadas del MARC 21, orientadas a una

semántica más sencilla: por ejemplo sustituye las etiquetas numéricas para la descripción de campos propias de MARC (100, 245) difíciles de entender para un no profesional por etiquetas textuales (name, titleInfo, etc). En tanto MARC XML y MODS están orientados a la gestión de información bibliográfica, MADS lo hace a la información de “autoridades”, es decir a la normalización de elementos como nombres de autor y temas.

El formato MARC21 es el utilizado por la BNU para catalogar sus documentos, en particular utilizan la herramienta Aleph mencionada en 2.2 Relevamiento del trabajo de la BNU.

Dublin Core

El estándar Dublin Core [19] fue creado para describir recursos de Internet y responder a la necesidad de crear una nomenclatura común de metadatos. Surgió en un encuentro realizado en Dublin (Ohio), en 1995.

Constituye un mecanismo básico de descripción que puede usarse en todos los dominios, para todo tipo de recursos, sencillo pero potente, que puede extenderse fácilmente y puede trabajar conjuntamente con otras soluciones específicas.

Está compuesto de quince elementos de descripción de documentos electrónicos (contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type) que facilita la información básica sobre los mismos [20].

El Dublin Core cualificado es una extensión del Dublin Core donde algunos de sus elementos son acompañados de un cualificador que los hace más restrictivos (por ejemplo Date.Created, Date.Modified, Date.Issued, Relation.IsPartOf) [21].

2.4.2. REPOSITORIOS DIGITALES Y BIBLIOTECAS DIGITALES

Repositorio Digital

Los Repositorios Digitales (RD) o también conocidos como Repositorios Institucionales (RI) son un medio de almacenamiento centralizado, el cual permite gestionar, almacenar, preservar, difundir y facilitar el acceso a los objetos digitales que alberga. Estos pueden ser de acceso público, o pueden estar protegidos y necesitar de una autenticación previa. Los repositorios digitales más conocidos son los de carácter académico e institucional y tienen por objetivo organizar, archivar, preservar y difundir la producción intelectual resultante de la actividad investigadora de la entidad.

A continuación veremos algunas definiciones de autores sobre los repositorios digitales:

Crow en el año 2002 [22] es uno de lo que introduce el concepto de RD definiéndolos como colecciones digitales que capturan y preservan la producción intelectual de las comunidades universitarias.

Van de Sompel junto con otros autores en el año 2004 [23] lo define como “un sistema de comunicación académica futura que incorpore de forma nativa la capacidad de grabar y exponer esa dinámica, las relaciones y las interacciones en la propia infraestructura de comunicación académica. El registro de este cuerpo de información es sinónimo de registrar la evolución de la academia en una granularidad fina. Esto permitirá rastrear el origen de las ideas específicas hasta sus raíces, analizando las tendencias en un momento específico en el tiempo y proyectando las futuras líneas de investigación”.

Bustos y Gonzalez en el año 2008 [24] definen los RD como "un archivo electrónico de la producción científica de una institución, almacenada en un formato digital, asimismo contiene mecanismos para importar, identificar, almacenar, preservar, recuperar y exportar un conjunto de objetos digitales, normalmente desde un portal web".

Biblioteca Digital

Luego de leer varias definiciones que datan desde 1997 hasta el presente, observamos que el concepto de Bibliotecas Digitales (BD) ha ido evolucionando en todos esos años y sobre todo con el crecimiento de la tecnología, representada por Internet, el software y el hardware. Algunas de estas definiciones son las siguientes:

El autor M. Lesk en el año 1997 [25] expresa: “Las bibliotecas digitales son colecciones organizadas de información digital. Combinan la estructura y concurrencia de la información, que siempre han tenido las bibliotecas y los archivos, con la representación digital que han hecho posible las computadoras”.

Borgman en el año 1999 [26] establece que las bibliotecas digitales son un conjunto de recursos electrónicos y de capacidades técnicas asociadas para crear, buscar y utilizar la información.

Candela junto a otros autores en el año 2007 [27] establecen un marco conceptual con tres niveles:

- Biblioteca digital es la organización que recopila, gestiona, preserva y ofrece contenidos digitales
- Sistema de biblioteca digital se refiere al software que proporciona la funcionalidad requerida por una biblioteca digital particular

- Sistema de gestión de biblioteca digital que se refiere a la plataforma: sistema operativo, bases de datos, interfaz de usuario.

Witten junto a otros autores en el año 2009 [28] afirman que las BD son colecciones enfocadas de objetos digitales (texto, vídeo y audio), así como los métodos de acceso y recuperación de datos, y para la selección, organización y mantenimiento de la colección. Tradicionalmente estos se dividen en búsqueda y consulta, aunque en realidad la distinción no es clara.

Por lo tanto no hay una única definición para lo que es una biblioteca digital, se puede decir que una biblioteca digital es un sistema de información que permite el acceso y transferencia de información digital, estructurada alrededor de colecciones de documentos digitales sobre los cuales se ofrecen servicios a los usuarios.

Diferencias entre un Repositorio Digital y una Biblioteca Digital

Estas son algunas de las diferencias más relevantes entre los RD y las BD:

- Los RD principalmente son utilizados para recoger, preservar y poner a disposición la producción académica de una institución, en cambio las BD pueden estar organizadas en: temas, disciplinas, o incluso tipos de documentos en particular.
- Los RD están organizados en relación a una comunidad institucional en particular, en cambio las BD están construidas en torno diferentes: tópicos, temas, disciplinas, etc. O sea tienen una estructura jerárquica más compleja.
- Los materiales de los RD suelen ser de origen digital, en cambio las BD por lo general tienen que digitalizar los viejos materiales.

En algunos casos una BD puede tomarse como un RD y viceversa, debido a que ambos ofrecen cosas similares y el uso de cada término depende del ámbito donde se aplica.

Según lo definido, nuestro proyecto se basa en realizar dos bibliotecas digitales utilizando un software de repositorios como lo es DSpace. Cabe aclarar que en momentos hablaremos de repositorio debido que al software utilizado es de repositorios, pero el objetivo final se basa en la conversión de repositorios a bibliotecas de contenido digital.

2.4.3. SOFTWARE DE REPOSITORIOS

Uno de los objetivos principales de este proyecto es disponer de una Biblioteca Digital que mejor se adecue a las necesidades de las instituciones involucradas en el mismo, por tal motivo se propuso por parte de las instituciones involucradas la plataforma DSpace.

La justificación de las instituciones para la elección del software es la siguiente:

- Número de instalaciones a nivel mundial, que lo transforman en el software más utilizado en bibliotecas de tamaño mediano para el mantenimiento de colecciones digitales. A nivel regional este software es usado por la Biblioteca Nacional Argentina y la Biblioteca Nacional de Brasil. A nivel nacional fue adoptado en 2014 por la UDELAR para el mantenimiento de su repositorio de publicaciones académicas.
- Ser software libre de fuente abierta, con una comunidad activa que participa de su desarrollo.
- Hay investigaciones a nivel nacional [29] que recomiendan su uso.
- Un buen equilibrio entre lo robusto de su tecnología, capaz de alojar colecciones de tamaño mediano y grande, y la facilidad de instalación y mantenimiento.

Las alternativas de software descartadas por las instituciones fueron las siguientes:

- OMEKA es un programa libre para la gestión de colecciones digitales. Es fácil de instalar y configurar. Incluye gran cantidad de complementos y utiliza MySQL como motor de datos. Lo descartamos dado que el volumen de datos que debemos manejar estaría medido en varios Terabytes, esto debido a que las base de datos relacionales utilizan para realizar las búsquedas el operador JOIN para cruzar las tablas, puede demorar un tiempo considerable la respuesta de la consulta si el volumen de datos es grande.
- Fedora [30] es un software orientado al manejo de grandes volúmenes de datos, se descartó por su complejidad de instalación y configuración.
- Greenstone [31] es un software neozelandés con antecedentes de uso en Uruguay. Hasta mediados de la década de los 2000 era el estándar en software para repositorios en la región. Sus principales defectos a nuestros fines son una visualización de los ítems un poco rígida y el hecho de las colecciones no pueden ser anidadas, también el hecho que su presentación fuera la de un buscador, no estando muy adaptado a la navegación dentro de las colecciones.

Los integrantes de ambas instituciones consideraban que el principal esfuerzo del proyecto estaría en la incorporación de complementos para la correcta visualización de los objetos digitales, de características bastante diversas, ya que incluía material textual e imágenes de diversos tipos, además de la posibilidad de contar en un futuro con archivos de audio y video.

Por nuestra parte además realizamos un análisis acerca de la usabilidad del software propuesto y de otras plataformas a modo de ver cuál es la que podría cumplir mejor con los requisitos y por qué DSpace si lo hace.

A continuación daremos una breve reseña de los repositorios tanto de código abierto como comerciales más utilizados a nivel mundial:

DSpace

DSpace [5] ha sido desarrollado en el año 2002 por el MIT (Massachusetts Institute of Technology) es uno de los repositorios digitales de código abierto más usados en el mundo académico por sus amplias funcionalidades y relativa facilidad de configuración.

Este repositorio digital permite organizar la información en comunidades que a su vez, se segmentan en colecciones de documentos. Es multiplataforma y es posible realizar mejoras a la interfaz (customizar) y extenderlo. La instalación es un tanto complicada por una serie de prerrequisitos.

DSpace almacena casi cualquier tipo de formato y documento, así como la catalogación de los mismos utilizando el estándar Dublin Core para los metadatos que describen los objetos en el repositorio.

EPrints

EPrints [32] es uno de los software libre más populares para la creación y gestión integral de repositorios digitales de Acceso Abierto. Sus características, prestaciones y versatilidad lo convierten en una aplicación informática muy interesante para poner en marcha repositorios institucionales OAI-compliant. Desarrollada en el año 2000 por la universidad Southampton School of Electronics and Computer Science esto la transforma en el repositorio digital más antiguo.

Fue diseñada principalmente para archivos como tesis de grado, documentos de búsqueda, para que el usuario pueda utilizarla como un repositorio de búsqueda de información orientada a estos tipos de documentos.

Actualmente Eprints es la segunda plataforma más utilizada a nivel mundial.

Es una herramienta muy flexible y brinda gran libertad para ampliar su funcionalidad, la misma puede manipular una gran variedad de archivos digitales, a diferencia de DSpace, EPrints no se encuentra orientado a comunidades ni colecciones estructuradas. Sin embargo permite creaciones de colecciones digitales a partir de un metadatos como por ejemplo creador, año, tema etc.

Greenstone

Greenstone [31] es un conjunto de programas de software diseñado para crear y distribuir colecciones digitales, proporcionando así una nueva forma de organizar y publicar la información a través de Internet o en forma de CD-ROM.

El objetivo de este software es dar el potencial de construir sus propias bibliotecas digitales a los usuarios, especialmente en universidades, bibliotecas y otras instituciones de servicio público.

Es un software de código abierto en varios idiomas distribuido conforme a los términos de la Licencia Pública General GNU. En comparación con DSpace es un poco débil en la administración de usuarios.

Fedora

Fedora [30] es un sistema de repositorio robusto y modular, de código abierto para la gestión y difusión de contenidos digitales. Es especialmente adecuado para las bibliotecas y archivos digitales, tanto para el acceso y la preservación.

También se utiliza para proporcionar acceso especializado a grandes y complejas colecciones digitales de materiales históricos y culturales, así como datos científicos.

Su esquema de metadatos es altamente extensible y permite la creación de relaciones entre objetos. Requiere de altos conocimientos para su instalación y configuración y no tiene interfaz de usuario.

Archimede

Archimede [33] es un software de repositorios Canadiense, desarrollado por la Universidad de Laval, el cual está inspirado en el modelo DSpace usando comunidades y colecciones de contenido. Este software tiene una limitada difusión institucional.

Open Repository

Open Repository [34] proporciona un servicio de software comercial para crear Repositorios Institucionales. Está basado en el Software DSpace, pero el mismo no es de código abierto.

Bepress

Bepress [35] es un software comercial con licenciamiento, desarrollado por Berkley Electronic Press, el mismo es utilizado para crear y alojar los repositorios de sus clientes.

CONTENTdm

CONTENTdm [36] es una herramienta para la organización, dirección, publicación y búsqueda de colecciones digitales. Dispone de una versión gratuita de prueba, pero es un software comercial con licenciamiento.

Comparativa entre las diferentes opciones

Para la evaluación y comparativa de los diferentes opciones existentes actualmente en el mercado utilizamos como principal referencia los paquetes de recomendaciones ofrecidos por OpenAIRE (Open Access Infrastructure for Research in Europe) [37].

Éstas se refieren a los siguientes aspectos:

- **Implementación de Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH):** define características obligatorias y recomendadas para solucionar problemas en distintas implementaciones de repositorios de acceso abierto. Esto es importante para la sincronización entre repositorios.

- **Recursos textuales:** contempla los requerimientos que deben cumplir los recursos. Destacamos la obligación de utilizar “sets” (agrupaciones) que definen las colecciones accesibles a texto completo.
- **Metadatos:** marca y define los elementos Dublin Core obligatorios y recomendados.

En la siguiente tabla se presenta una comparativa que realizamos de las características de los softwares mencionados:

Repositorio	Descripción	Disponibilidad	Características
Archimede	Software de código abierto para RI	Software gratuito, de código abierto, proporcionado por la GNU, licencia de público general.	Utiliza un Dublin Core metadata set. Cumple con OAI Inspirado en el modelo DSpace usando comunidades y colecciones de contenido. El buscador está basado en el código abierto Lucene, usando LIUS. Tiene una perspectiva multilingüe (inglés, francés y español). La interfaz del software es independiente y no está dentro del código.
Bepress	Software Licenciado con alto costo que incluye, la implementación de clientes, infraestructura, prácticas, servidor, la conservación, apoyo a distancia, apoyo técnico y actualizaciones de software.	Software comercial, pago de licencia y honorarios de suscripción.	OAI compliant. Exportación de datos XML. Personalización mediante API, plantillas. Búsqueda de texto completo. Ofrece software cliente EdiKit para añadir contenido al repositorio.
CONTENT	Proporciona herramientas desde la organización y dirección a la publicación y búsqueda de colecciones digitales en internet. También ofrece herramientas para archivar colecciones de cualquier tamaño.	Disponible una prueba gratuita de 60 días. Software comercial. El precio depende del tamaño de la colección y tiene costo adicional por mantenimiento anual.	Todo tipo de contenidos aceptados Cumple con OAI. Dublin Core metadatos. XML data export. Búsqueda de colección múltiple
DSpace	Está diseñado para capturar, almacenar, ordenar, conservar y redistribuir la producción intelectual y de investigación de una Universidad en formato digital.	Software gratuito de código abierto Distribuido por BSD licencia de código abierto	Todo tipo de contenidos aceptados. Dublin Core metadatos standard. Interfaz web personalizable. Cumple con OAI. Proceso de flujo de trabajo para envío de contenido. Capacidades de importación / exportación. Proceso de envío descentralizado. Extensible para Java API. Búsqueda de texto completo utilizando Lucene o Google. Base de datos: PostgreSQL, o SQL que apoyan transacciones como Oracle, MySQL.
EPrints	Software gratuito diseñado para RI de edición electrónica	Distribuido por la GNU general public licence	Cualquier tipo de contenido aceptado. Puede usar cualquier esquema de metadatos. Interfaz basado en la web. Características del workflow: el contenido para por un proceso de moderación para su aprobación, rechazo, o a otro autor para su revisión. Base de datos MySQL. Extensible a través de API usando lenguaje de programación PERL. Búsqueda de texto completo.
Fedora	Sirve para crear bibliotecas digitales, RI y otros sistemas para que sean interoperables basadas en la web.	Gratuito de código abierto. Distribuido por la licencia de código abierto Mozilla	Cualquier tipo de contenido aceptado. Dublin Core metadatos. OAI compliant Aportación de datos y almacenaje XML Extensibilidad: APIs para el manejo, acceso, servicios web. Versiones de contenido. Utilidad de migración.
Greenstone	Es un software para crear y distribuir colecciones de bibliotecas digitales	Software gratuito de código abierto distribuido por GNU general public licence	Multilingüe: los cuatro idiomas centrales son inglés, francés, español y ruso Incluye una colección de pre-creación en versión demo Ofrece una característica "Export to CDROM"
Open Repository	Proporciona un servicio de software comercial para crear RI. Basado en el software de DSpace ofrece múltiples ediciones con distintos niveles de apoyo y servicio.	Se ofrecen 3 ediciones: Standard, Silver y Gold	Acepta amplia variedad de formatos de contenido. Utilidad de conversión para crear PDFs. Metadatos basados en OAI.

Tabla 1- Comparación de software de repositorios [38]

El siguiente gráfico ilustra a nivel mundial que repositorio digital es el más utilizado. Se puede observar que DSpace es el preferido con una amplia diferencia por sobre otros repositorios.

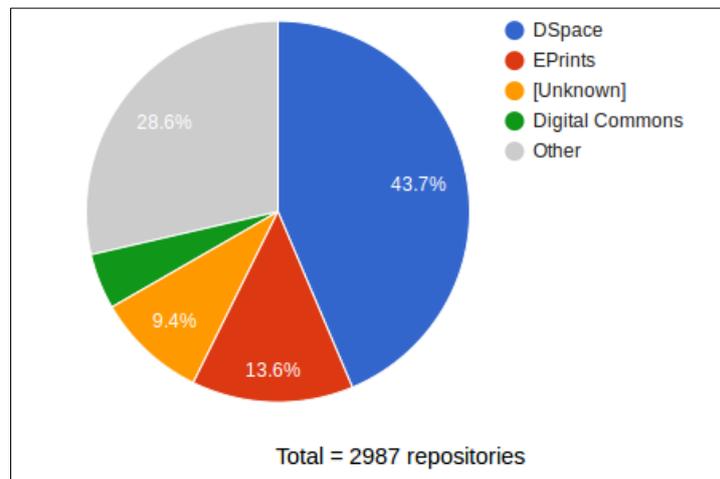


Figura 3- Usabilidad a nivel mundial de repositorios digitales de acceso libre [39]

Conclusiones

De acuerdo al estudio realizado, las herramientas disponibles tienen muchas características en común, pero el software de repositorios DSpace, se ajusta mejor a las necesidades de las instituciones pues la organización jerárquica que presenta la misma, como ser comunidades y colecciones permite una gran facilidad y organización a la hora de jerarquizar la información disponible.

Además destacamos una investigación realizada en un proyecto de grado realizado en el año 2012 [40], referente a la investigación de software de repositorios institucionales utilizados a nivel mundial. En la misma recomiendan el uso del software DSpace para la realización de repositorios institucionales.

El software Eprints por ejemplo no presenta jerarquización de información lo cual no sería una herramienta adecuada para los requerimientos de ambas instituciones.

Greenstone también se descarta debido a que la visualización de los ítems es poco flexible y las colecciones no pueden ser anidadas.

Fedora presenta dificultades para instalación y mantenimiento, además que no se encuentra dentro de los softwares más utilizados.

Open Repository, Bepress y CONTENTdm se descartan por no ser de código abierto.

En resumen las ventajas de utilizar DSpace:

- **Visibilidad:** Permite la difusión de material con gran facilidad.
- **Flexibilidad:** Permite el ingreso de múltiples formatos de archivo incluidos documentos de texto, programas, videos, etc.

- **Almacenamiento seguro:** Preservación y calidad de la documentación almacenada.
- **Control de acceso:** Manejo de control de acceso por parte de los administradores para los ítems digitales, con la posibilidad de variar esa condición en el transcurso de su publicación.

2.5. SINCRONIZACIÓN

Uno de los puntos más relevantes del proyecto se basa en la sincronización entre las bibliotecas digitales FIC – BNU. Para ello se realizó una ardua investigación acerca del mecanismo que se utiliza para la sincronización de bibliotecas digitales.

El estudio se basó en el protocolo OAI-PMH, cuyo objetivo es desarrollar y promover normas de interoperabilidad que tienen por objeto facilitar la difusión eficaz de los contenidos.

Provee de una interfaz muy simple a través de la cual se acceden a los metadatos de contenidos en formato XML provenientes de diferentes fuentes y plataformas. El principal objetivo de este protocolo es que logra estándares de interoperabilidades entre archivos de diferentes formatos permitiendo su facilidad de difusión, intercambio y accesibilidad entre los mismos. Este permite almacenar en un lugar los metadatos en los cuales se realizaran diferentes consultas, el protocolo se ocupa únicamente de la gestión de la información. A través de la interfaz que proporciona se exponen los servicios que este brinda:

- **Identify:** Se utiliza para recuperar información sobre el repositorio.
- **ListIdentifiers:** Se utiliza para recuperar los encabezados del registro desde el repositorio.
- **ListRecords:** Utilizados para capturar los registros completos desde el repositorio.
- **ListSets:** Se utiliza para recuperar la estructura de conjunto del repositorio.
- **ListMetadataFormats:** Son las listas de metadatos del repositorio, las cuales se puede difundir.
- **GetRecord:** Se utilizan para recuperar un registro individual del repositorio.

Los metadatos a transmitir vía OAI-PMH deberán codificarse en *Dublin Core* cuyo propósito consiste en minimizar los problemas derivados de las conversiones entre múltiples formatos.

El protocolo OAI se basa en un modelo cliente-servidor entre un proveedor de datos al cual se le denomina proveedor de datos OAI y un proveedor de servicios OAI server.

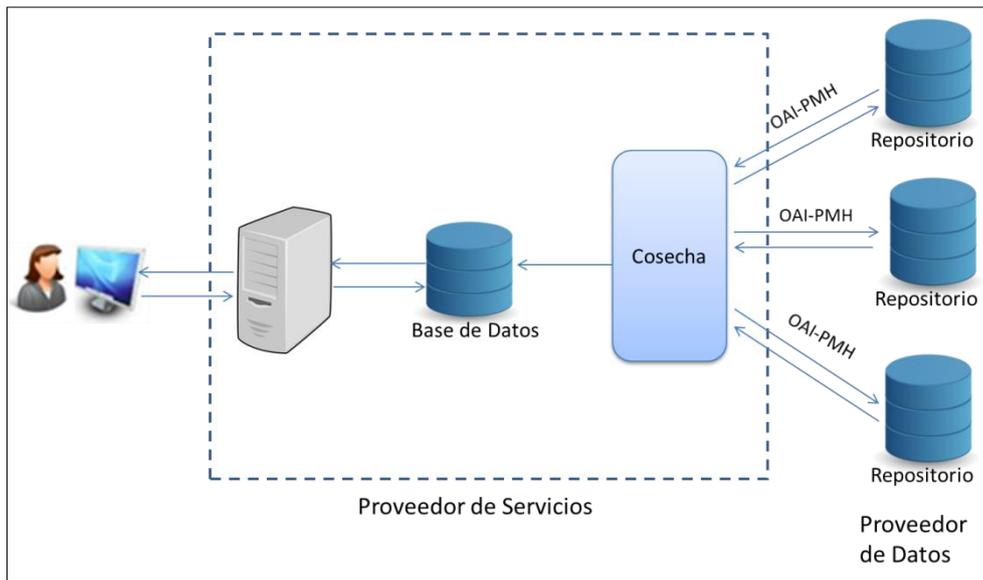


Figura 4 – Funcionamiento del Protocolo OAI-PMH

Los “Proveedores de datos” son los que proveen de información a los “cosechadores” así es el nombre que se le proporciona en la comunidad OAI a quienes copian los datos de diferentes repositorios tal como se muestra en la figura. Llevando este escenario a la situación actual de nuestro proyecto el cosechador sería la BNU y el proveedor de datos la FIC.

Arquitectura OAI

Presentaremos una descripción acerca de la arquitectura del protocolo OAI a modo de comprender mejor como es que se realiza el mecanismo de sincronización a través de este.

Las componentes de la arquitectura OAI son:

- **Analizador Sintáctico de Argumentos** - Valida las peticiones OAI.
- **Generador de Errores** - Genera respuestas XML con los mensajes de error codificados
- **Consulta a Base de Datos / Extracción de Metadatos Locales** – Su función es la recuperación de los metadatos del repositorio, de acuerdo al formato requerido.
- **Generador XML / Creación de la Respuesta** - Genera respuestas XML con la información de los metadatos codificada.
- **Control de Flujo** - Utiliza la señal de reanudación como mecanismo de control.

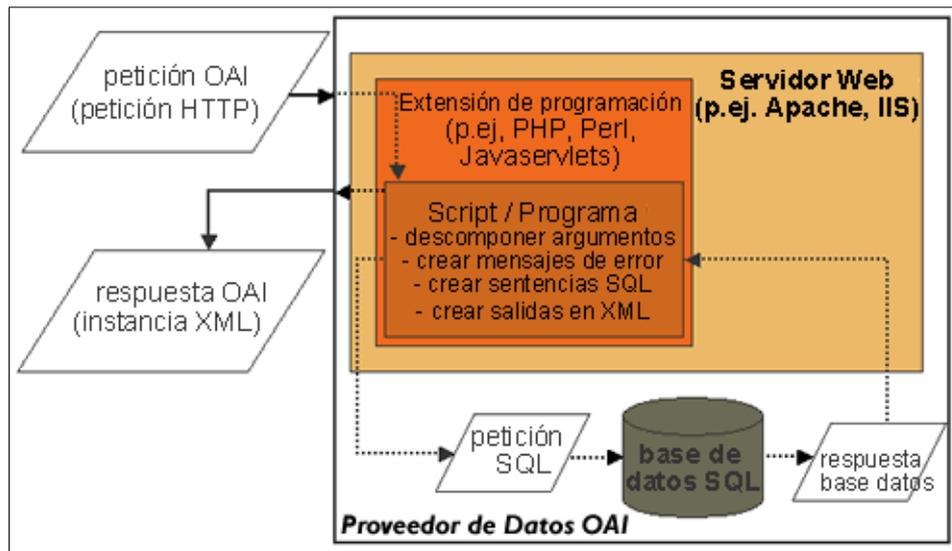


Figura 5 - Interacción entre componentes [41]

En el esquema de la Figura 5 se puede observar como es la interacción entre las componentes de la arquitectura, la respuesta del servidor OAI se expone en una instancia XML conteniendo la información que luego el cosechador a través del protocolo HTTP consume.

De esta forma BNU cosecharía los datos expuestos por la institución FIC de forma de mantener ambas bibliotecas digitales con la información en forma de espejo.

En Uruguay no hemos encontrado un ejemplo de sincronización entre repositorios que se haya realizado con el software DSpace anteriormente. Tampoco hemos encontrado a nivel regional repositorios donde realicen sincronización con otro repositorio de forma programada como lo haremos nosotros. Asimismo como también la escasa documentación existente al respecto contribuyo en la dificultad de investigación lo cual conllevo a la realización de múltiples pruebas “ensayo-error” hasta dar con la correcta configuración así como también con la comprensión del mecanismo.

En cuanto a la difusión de la información es deseable registrar el sitio que proporcionara la misma para que otras posibles cosechas en diferentes repositorios tengan la posibilidad de acceder a la información con el propósito de difundirla. Este registro se realiza en el sitio Open Archives [42].

CAPÍTULO 3

3. DSPACE

DSpace es un repositorio digital de código abierto que proporciona herramientas suficientes para el manejo y administración de colecciones, comunidades y objetos digitales.

Su diseño permite una distribución jerárquica de tareas entre un equipo de personas. En este capítulo se describen en detalle sus funcionalidades así como su arquitectura y diseño.

3.1. ARQUITECTURA

DSpace posee una arquitectura de 3 capas y cada una consta de varios componentes. La capa de almacenamiento es responsable del almacenamiento físico de los metadatos y el contenido. La capa de negocios gestiona el contenido de archivos, los usuarios (conocidos como e-personas en DSpace), los flujos de trabajo y las autorizaciones. Por último, en la capa de aplicación contiene componentes que ofrecen el acceso a los servicios de las demás capas.

Cada componente en las capas de almacenamiento y lógica de negocio tiene una API pública definida.

Aunque la capa lógica es la encargada de autorizar las acciones de la capa de negocio, el sistema se basa en aplicaciones individuales en la capa de aplicación para autenticar correctamente y de forma segura a los usuarios.

El código fuente está organizado para adherirse muy estrictamente a esta arquitectura de tres capas. El compilador de Java ayuda a asegurar que el código fuente se ajusta a la arquitectura.

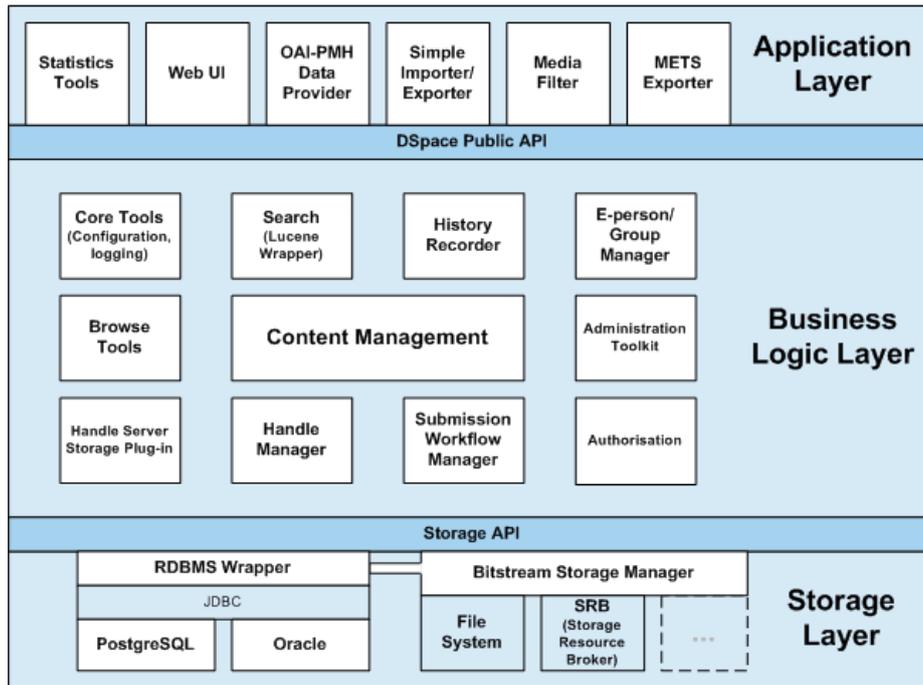


Figura 6 - Arquitectura de DSpace [43]

3.2. ESTRUCTURA DE BASE DE DATOS

DSpace utiliza una base de datos relacional para almacenar toda la información sobre la organización de contenidos, metadatos, flujos de trabajo, información sobre los usuarios y autorizaciones. También utiliza la base de datos relacional para mantener los índices para que los usuarios puedan realizar búsquedas.

La mayor parte de las funcionalidades que utiliza DSpace pueden ser ofrecidas por cualquier gestor de base de datos SQL estándar que soporta transacciones. Sin embargo en este momento, las APIS de DSpace utilizan algunas características específicas para PostgreSQL [44] y Oracle [45], por lo que para utilizar otra base de datos se tendrían que realizar algunas modificaciones al código.

3.3. ESTRUCTURA DE ALMACENAMIENTO (BITSTREAM STORE)

Mientras que el modelo de datos de DSpace está soportado por base de datos como PostgreSQL u Oracle, los contenidos de los ítems se almacenan bajo el directorio denominado assetstore.

Ofrece dos formas de almacenamiento de contenido, almacenamiento en el sistema de archivos y almacenamiento utilizando SRB (Storage Resource Broker). El SRB se puede utilizar en lugar de sistema de archivos del servidor o en conjunto al sistema de archivos.

El gestor de almacenamiento de bitstream (BitstreamStorageManager) proporciona acceso de bajo nivel a los bitstreams almacenados en el sistema. Además proporciona tres métodos: uno para almacenar, otro para recuperar y otro para borrar.

Los bitstreams son identificados por su ID, que es la columna bitstream_id, clave principal en la base de datos.

Un ejemplo de cómo queda estructurada la información se puede ver en la siguiente imagen.

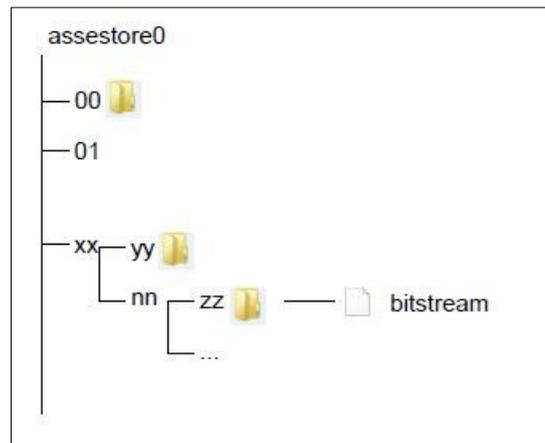


Figura 7-Estructura de almacenamiento

Siguiendo con este ejemplo, la referencia de un ítem a sus archivos se encuentra en la tabla bitstream, campo internal_id.

Un ejemplo de identificador es 110832826281924074367996140570931140204, este bitstream, se encuentra buscando los seis primeros dígitos del identificador, que indica en que subdirectorio de tercer nivel está el ítem (11 >> 08 >> 32) y el nombre del archivo será 826281924074367996140570931140204.

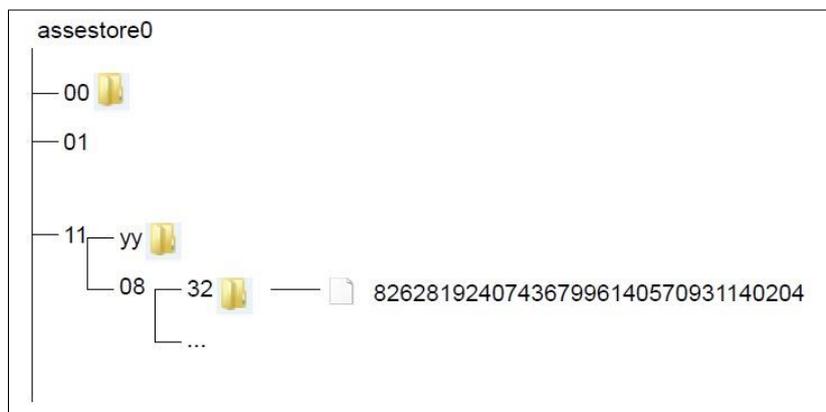


Figura 8 - Nombre de archivo

3.4. ADMINISTRACIÓN DE DSPACE

DSpace cuenta con una interfaz gráfica para el acceso a los usuarios, así como un control de administración para que aquellos usuarios que son administradores de la plataforma.

Estos pueden crear usuarios, roles, definir y adjudicar permisos, modificar la estructura jerárquica de la información, así como también definir los procesos de trabajo (workflow) para el proceso de ingreso de la información al repositorio.



Figura 9 - Menú de administración

El repositorio DSpace organiza la información en forma jerárquica con estructura de árbol

- **Comunidades** Dentro de estas pueden estar contenidas sub comunidades y/o colecciones.
- **Sub comunidades** Es una subcategoría dentro de una comunidad, puede contener a su vez otras sub comunidades y/o colecciones.
- **Colecciones** Las colecciones son categorías que contienen a los ítems, estas además no pueden contener comunidades ni sub comunidades.
- **Ítems** Representa a los documentos digitales que se ingresan en el repositorio. Estos son identificados por un número denominado Handle.

Usuarios

DSpace nombra a los usuarios como E-Personas (E-Persons) y maneja tres tipos de usuarios. Estos son:

- **Usuario anónimo:** Se trata de un usuario que no ha iniciado sesión en el repositorio tiene acceso solo para navegar y acceder a leer el contenido del repositorio.

- **Usuario conectado:** Usuario registrado en la plataforma, el administrador puede asignar permisos.
- **Usuario administrador:** Es el que administra el repositorio accediendo a todas las funciones de los contenidos.

Sólo el administrador puede conceder privilegios a los usuarios dentro de las colecciones, comunidades, etc.

Autorización (Permisos)

El sistema de autorización DSpace está basado en la asociación de acciones con los objetos y con las listas de personas (E-Personas) que pueden llevarlas a cabo. Las asociaciones se llaman Políticas de Recursos, y las listas de E-Personas se llaman grupos.

Hay dos grupos específicos:

- “Los administradores”
- “Anónimos”

Los administradores conforman un grupo de personas que tienen acceso a todas las funcionalidades sobre colecciones, ítems y comunidades. Mientras que los usuarios anónimos solo tienen un grupo reducido de acciones que pueden ejercer sobre las colecciones, etc. En principio este grupo solo tiene permiso de lectura sobre los ítems esto es el mínimo que pueden proporcionarle. Puede que si el administrador lo desee se le puede dar algún otro tipo de permiso como el de modificar u eliminar dentro de la estructura jerárquica.

Grupos

Los grupos en DSpace tienen como objetivo el fácil manejo de usuarios. Crear un grupo, puede ser útil si se tiene una comunidad, subcomunidad o colección cuyo propósito es brindar determinados permisos a un grupo de usuarios dentro de la misma.

Los Grupos que vienen creados por defecto en DSpace son:

- **Grupo Administrador** En este grupo se encuentra la persona administradora del repositorio, éste tiene todos los privilegios del repositorio. Adicionalmente se pueden añadir de forma ilimitada administradores al mismo. Existen dos tipos de administradores, los que lo manejan todos los permisos y funcionalidades del repositorio y los que sólo administran una comunidad, sub comunidad o colección.
- **Grupo Anónimo** Indica que cualquier usuario que ingrese al repositorio tiene privilegio únicamente de lectura sobre los ítems dentro del repositorio.
- **Grupo Envío** Personas que son miembros de un grupo que solo tiene autorización para realizar envíos a la colección y/o comunidad en cuestión

Además de los tres grupos principales expuestos, es posible crear una cantidad ilimitada de grupos brindándole a cada grupo un privilegio correspondiente.

3.5. FLUJO DE TRABAJO (WORKFLOW)

Cuando se crea una colección en DSpace, se tiene la opción de seleccionar el flujo de trabajo que tendrá la misma. El mismo ayuda a controlar el contenido subido al repositorio, pasando por una serie de revisiones previas a la publicación del ítem.

En el siguiente esquema se presenta como es el flujo de trabajo estándar propuesto por DSpace.

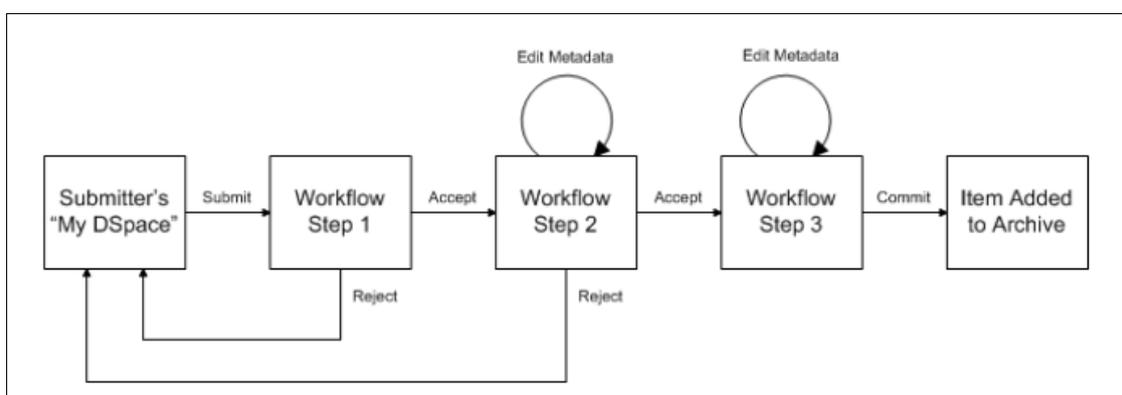


Figura 10- Flujo de trabajo estándar de DSpace

Descripción de los pasos:

- **Paso Inicial (colaborador)** - Se ingresa la información correspondiente a un ítem de una determinada colección, puede ser un ítem o una colección la idea es que se ingrese la información necesaria para una descripción básica del objeto.
- **Paso 1** - En este paso se generó una tarea pendiente del paso Inicial para la correspondiente aprobación, en esta etapa se evaluará el documento y se verificará la correctitud de la información en caso de que no sea aprobado por el experto en cuestión se puede devolver el documento al colaborador si el mismo no satisface con las condiciones mínimas.
- **Paso 2** - En esta etapa se puede agregar más metadatos en forma de detalle. Además el experto que se encuentre revisando este documento puede rechazarlo y al rechazarlo hace que al colaborador notifique para que el mismo pueda volver a corregir así lo pueden validar con el fin de publicarlo.
- **Paso 3** - En esta etapa se puede agregar más metadatos así como también controlar la calidad del mismo. Además el experto que se encuentre revisando este documento no puede rechazarlo ya en esta etapa lo cual hace que se

defina si descarta el documento o eventualmente lo envía para su publicación (No puede pedir re-enviar el documento digital por ejemplo)

- **Paso 4** - Confirmación y publicación del ítem en el repositorio.

3.6. REGISTRO DE FORMATOS

DSpace permite bitstreams de una gran variedad de formatos de archivos. Cuando se sube un fichero a DSpace, dependiendo del formato inferido de la extensión del mismo, se le asignará uno de los tres niveles de soporte siguientes:

- **Soportado:** El formato es soportado por el repositorio.
- **Conocido:** Indica que el formato está declarado en el registro como reconocible, pero no se garantiza soporte a efectos de preservación. Esto podría ser por ejemplo para formato de propietarios como Microsoft.
- **Desconocido:** El formato no está declarado en la lista de formatos de DSpace, estos archivos aparecen como “Unknown” o “application/octet-stream”

El administrador es el encargado de agregar formatos de archivos conocidos o soportados, a través de la interfaz web o modificando el archivo de configuración `bitstream-formats.xml` y un ejemplo de ingreso de registro es el siguiente:

```
<bitstream-type>
  <mimetype>image/png</mimetype>
  <short_description>image/png</short_description>
  <description>Portable Network Graphics</description>
  <support_level>1</support_level>
  <internal>>false</internal>
  <extension>png</extension>
</bitstream-type>
```

Dónde:

- `<mimetype>` - Identificador de tipo MIME (Multipurpose Internet Mail Extensions)
- `<short_description>` - Descripción mínima de este formato, es la que se le mostrará al usuario
- `<support_level>` - Nivel de soporte (0 Desconocido, 1 Conocido, 2 Soportado)
- `<description>` - Descripción del formato
- `<internal>` - Los formatos marcados como “internal”, se usan por el sistema y no se representan a los usuarios
- `<extension>` - Extensión habitual de filename

3.7. ESTRUCTURA DE IMPORTACIÓN DE ÍTEMS

DSpace permite realizar la importación masiva de ítems tanto desde la interfaz (usuario administrador) como por línea de comando. El método de importación es mediante el formato simple de archivo, o sea con una estructura de directorio, subdirectorios y archivos predeterminada.

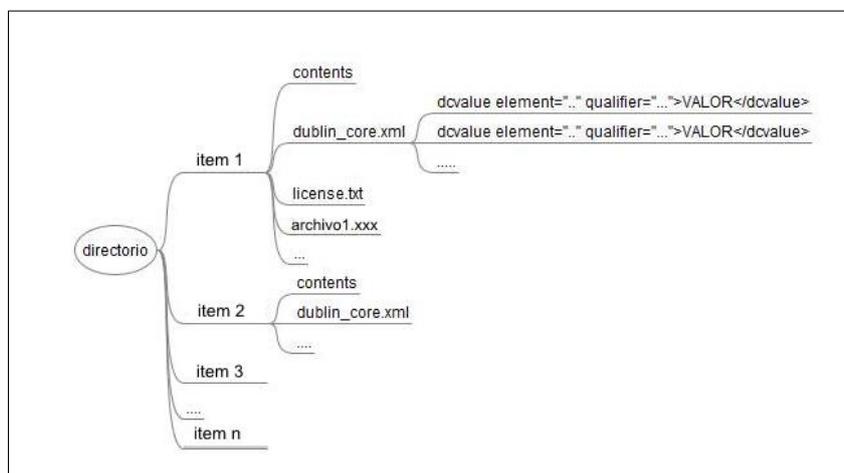


Figura 11- Estructura de importación de archivo

Cada directorio “ítem x” de la imagen corresponde a un ítem, los archivos que deberán ir obligatoriamente para la importación son dublin_core.xml y contents.

A continuación detallaremos el contenido de cada archivo:

- **dublin_core.xml:** Contiene los metadatos a asignar al ítem, tiene la siguiente estructura:

```
<dublin_core>
  <dcvalue element="contributor" qualifier="author">Nombre</dcvalue>
  <dcvalue element="language" qualifier="iso">es_ES</dcvalue>
  <dcvalue element="title" qualifier="none">Título</dcvalue>
  ...
</dublin_core>
```

- **contents:** el archivo contents enumera los archivos que están el subdirectorio, junto con indicaciones del bundle en el que deben ir. Los bundles son grupos de ficheros dentro del ítem que separan los diferentes tipos de archivos para que DSpace pueda tomarlos de forma diferente. Los posibles valores para un bundle son los siguientes: ORIGINAL, LICENSE, LICENSE_CC, THUMBNAIL entre otros.

Un ejemplo de contenido para un archivo contents es el siguiente:

```
license.txt bundle:LICENSE
archivo1.pdf bundle:ORIGINAL
archivo2.pdf bundle:ORIGINAL
```

- **archivo1.xxx:** Es el archivo en sí, se debe agregar una línea por cada archivo que se encuentre en el subdirectorio.

- **license.txt:** Archivo de licencia, es opcional, si se agrega debe ir especificado en el archivo contents.

3.8. CONTROL DE INTEGRIDAD Y CALIDAD

A medida que aumenta el volumen de datos en un repositorio, también aumenta la necesidad de tener mecanismos de control de metadatos y archivos de forma automática, a modo de simplificar el trabajo de control de calidad que deben realizar los usuarios administradores.

Para control de la calidad, DSpace ofrece los controles llamados “Curation Tasks” o en su traducción al español sanidad del sistema. Los Curation Tasks permiten evaluar y modificar características sobre los recursos del repositorio y es aplicable tanto a una colección específica como a una comunidad

DSpace posee un módulo llamado CurationSystem y tiene el objetivo de ofrecer un mecanismo automático para la ejecución de controles y validaciones de manera periódica sobre el contenido del repositorio. Esta funcionalidad puede ser extendida y configurada según las necesidades del repositorio. Es responsable de la selección, organización y ejecución de las curation tasks los cuales son módulos de menor tamaño que realizan tareas acotadas y concretas sobre un conjunto de objetos.

Las siguientes curation task están disponibles en la versión por defecto de DSpace:

- **Chequeo de links:** Sirve para chequear links que ya no se encuentran disponibles. Busca en los metadatos URLs y realiza conexiones a las mismas para determinar el código de respuesta. Este se divide en dos: Chequeo básico de links el cual selecciona sólo los metadatos cuyo cualificador es “uri” y el chequeo de valores de metadatos el cual selecciona todos los metadatos asociados a un ítem y en cada caso verifica si su contenido comienza con las cadenas “http://” o “https://”.
- **Búsqueda de virus (ClamScan):** ejecuta un análisis en busca de virus sobre los bitstreams de cada ítem. Utiliza el software antivirus ClamAV1
- **Metadata requerida:** Analiza los ítems para determinar si tiene todos los metadatos configurados como obligatorios, según esté definido en el formulario de envíos.
- **Formato perfilado (Format profiling):** se encarga de recorrer los ítems y elaborar una tabla que contiene todos los formatos de bitstreams encontrados junto a la frecuencia de ocurrencia.
- **Traductor Microsoft:** realiza traducciones automáticas de los metadatos utilizando la API Microsoft Translator.
- **Control de autoridad:** El Control de autoridades ayuda a mejora la calidad de contenidos y posibilita la interoperabilidad entre repositorios. Es un mecanismo

que por medio de directivas de configuración permite especificar que el contenido de ciertos metadatos deben ser obtenidos desde vocabularios controlados o desde servicios ajenos. Permite que los metadatos sean recuperados desde base de datos externas. De momento la única forma de conexión entre los servicios ajenos y el repositorio es por medio de un complemento. Por defecto DSpace trae el complemento Choice management el cual es capaz de leer y proporcionar términos desde un vocabulario controlado. Este módulo es suficiente cuando el vocabulario puede obtenerse en formato de archivo y si se trata de un vocabulario que tiene pocas actualizaciones. Para el caso por ejemplo de autores o instituciones este complemento no es suficiente.

3.9. INTERFACES DE INTEGRACIÓN

DSpace cuenta con varios módulos para integrar a la plataforma, se describen a continuación cada uno de ellos destacando sus funcionalidades y aspectos más relevantes.

3.9.1. XMLUI

Interfaz de usuario que proporciona DSpace, viene en el módulo XMLUI antes conocido como Manakin y para hacer uso de la misma se debe especificar en el servidor Tomcat que se utilizara esta. Se basa en Cocoon Framework [46], un modelo de capas divididas en tres fundamentales

- **Aspectos**
 - Son los generadores de contenido en XML
 - Implementan un conjunto de funcionalidades relacionadas
- **Esquema DRI (Digital Repository Interface)**
 - Representación abstracta en XML de la página de la plataforma
 - Es el esquema entrada/salida de los aspectos
- **Tema - definen como se visualiza el contenido**
 - Sitemap.xml: Archivo de configuración del tema
 - Hoja de estilo XSL: transforma el DRI Schema (XML) en HTML
 - Hoja de estilo CSS: determina la visualización final que verá el usuario

Los Aspectos pueden ser clases Java o archivos XSLT, son independientes entre sí se usan para manipular el DRI. Estos reciben un documento DRI como entrada y generan un documento DRI como salida tienen acceso a la API de DSpace. Es el mecanismo por el cual agrega funcionalidad a la interfaz de usuario, en la Figura 12 se muestra como las capas interactúan entre sí.

Los aspectos modifican el documento DRI y luego el tema transforma a este en el XHTML final. Así es como XMLUI genera el xhtml cara visible de la plataforma DSpace en caso de que se elija esta interfaz de usuario.

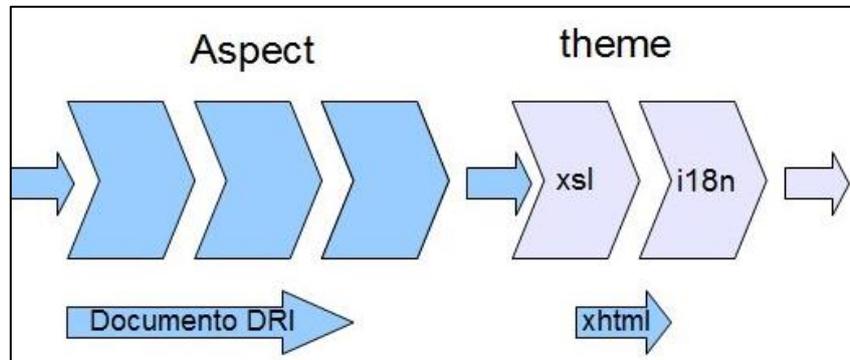


Figura 12 – Interacción entre capas [47]

3.9.2. JSPUI

Es una segunda opción de interfaz de usuario que brinda DSpace, viene en el módulo JSPUI para hacer uso de la misma se debe especificar en el servidor Tomcat [48] que se utilizara la misma.

Esta interfaz de usuario es la más antigua de las dos que DSpace ofrece.

La interfaz JSPUI se implementa basándose en Servlets JAVA los cuales implementan la lógica del negocio, y los Java Server Pages (JSP) producen las páginas HTML enviadas a un usuario final.

Esto hace posible adaptar de forma sencilla cualquier cambio en el estilo CSS u actualizar los archivos JSP.

3.9.3. RDF

DSpace cuenta con un módulo RDF cuyo objetivo se basa en agregar semántica a los objetos digitales dentro del repositorio, una de las formas para realizar dicho cometido es publicarlos como RDF triplas, estos objetos pueden ser Bitstreams, Bundle, Ítem, Colecciones, Comunidades etc.

Para tener una mejor visión de cómo es la sintaxis estas se componen de tres campos, el primero es el sujeto, luego el predicado y por último el objeto. Esta analogía en DSpace se comporta como el sujeto, el ítem por ejemplo /handle/1234.5/67890, el predicado es el metadato por ejemplo el campo dc.title y el objeto es el valor del campo. El objetivo del uso de RDF es la capacidad de darle al objeto una referencia a otras fuentes.

Como todos los demás módulos se incorpora la funcionalidad del mismo en el Apache Tomcat en el archivo server.xml.

3.9.4. SOLR

SOLR [49] es un proyecto de código abierto siendo este la evolución del Apache Lucene. Utilizado como motor de búsqueda, estadísticas, búsqueda facetada (permite representar una materia o tema combinando los conceptos que representan diferentes aspectos (facetas) de la misma), indexación, integración en bases de datos es uno de los más utilizados por los repositorios digitales.

En DSpace SOLR se utiliza para búsquedas y estadísticas, como en DSpace el manejo de grandes volúmenes de información es necesario se decidió incorporar este motor de búsqueda por su mecanismo de búsqueda rápido y eficiente, permitiendo búsquedas distribuidas y réplica de índices.

Al igual que los otros módulos para la utilización del mismo se incorpora desde el servidor web Apache Tomcat en el archivo server.xml.

3.9.5. OAI

El módulo OAI se encarga de la sincronización entre repositorios, este permite la recolección de los metadatos DSpace en otra plataforma, define estándares para intercambiar metadatos entre diferentes repositorios.

Al igual que los módulos anteriores para hacer uso del mismo desde nuestro servidor web en este caso tomcat7 debemos agregar la incorporación del mismo en el archivo server.xml del tomcat7.

Más adelante se detalla el uso y configuración de este protocolo para la realización de sincronización entre repositorios

3.9.6. LNI

Este módulo LNI Lightweight Network Interface se basa en el protocolo WebDAV diseñado para el acceso y modificación de las fuentes y metadatos en la web. Permite la integración de un sistema con DSpace.

Como todos los módulos anteriores para integrar el mismo desde el Tomcat debemos agregarlo al archivo server.xml.

3.9.7. REST

El propósito del desarrollo del mismo es proveer del servicio REST a la plataforma DSpace, habilitando a explotar las ventajas que el mismo provoca. Por ejemplo:

- Interacción entre DSpace y otros repositorios y/o bibliotecas digitales.
- Automatización de distintos procesos por ejemplo el envío de ítems.
- Integración de repositorios a nivel de flujos de trabajo entre otros sistemas.
- Interacción de diferentes sistemas o aplicaciones web tales como CMS, LMS, LCMS, VLS, AMS, etc.

3.9.8. SWORD

El módulo SWORD (Simple Web-service Offering Repository Deposit) que DSpace proporciona, tiene el propósito de alojar los ítems en distintos sistemas que utilicen este sin necesidad de que el usuario inicie sesión en la plataforma.

El protocolo SWORD se basa en atompub, Atom Publish Protocol, que define el depósito remoto de ítems en un repositorio por otras aplicaciones.

Por ejemplo si un autor desea enviar su artículo a un repositorio DSpace sin necesidad de iniciar sesión, lo realiza mediante una aplicación como por ejemplo Easy-deposit muy conocida por su fácil instalación y uso.

Utiliza SWORD como protocolo, envía el ítem al repositorio, se debe tener en cuenta que para esto DSpace debe tener activado el módulo SWORD y además autorizar que se pueda alojar cualquier ítem en este. Esto fomenta el incremento por parte de los usuarios a enriquecer de información dicho repositorio.

3.10. SEGURIDAD

A nivel de repositorios digitales algunas de las consideraciones a tomar con respecto a la seguridad son las siguientes:

- **Confidencialidad:** Esto significa mantener la información confidencial, es decir, que sólo es visto por las personas autorizadas. Por ejemplo, no permitir visualizar la contraseña al iniciar la sesión. La confidencialidad se implementa normalmente mediante cifrado.
- **Autenticación:** Es el proceso de probar la identidad de un usuario en el sistema, por ejemplo iniciar la sesión como E-Persona en DSpace. Es llevada a cabo por los módulos de Autenticación.
- **Autorización:** Esto significa garantizar los permisos a los recursos, o sea autorizar el acceso. Se basa en la autenticación para proporcionar la identidad del usuario, el sistema debe saber quién es el usuario para decidir que permitirle hacer.
- **Encriptación:** Un sistema de codificación que permite que dos partes (por ejemplo, el navegador web y el servidor DSpace) intercambiar datos sin que sea

legible a nadie más. En el contexto de las aplicaciones Web, el cifrado suele implicar el uso del protocolo HTTPS (HTTP sobre SSL).

Incluso si todos los materiales digitales almacenados en el repositorio DSpace están disponibles para acceso abierto, es necesario contar con seguridad. Es deseable que no todos los usuarios puedan añadir, modificar o eliminar contenidos, es más esta tarea debería ser realizada solo por unos pocos administradores.

Algunos archivos contienen materiales que sólo deben ser visibles para un público selecto. Para estos materiales se puede utilizar el mecanismo de DSpace de control de acceso para conceder privilegios de lectura a determinados usuarios y grupos. Incluso si el material es de acceso abierto, se puede considerar la seguridad del proceso de inicio de sesión, mediante E-Personas.

DSpace establece una cookie de inicio de sesión, que identifica a la E-Persona que inició la sesión. La cookie se transmite en las conexiones no cifradas, por lo que está sujeto a ser observada y "robada". Un atacante podría robar la cookie y conseguir acceso a DSpace, que es importante si se trata de un administrador. A partir de la versión 1.4 este riesgo se ha mitigado en gran medida. Si un cliente con una dirección IP utiliza la misma cookie de sesión, se rechaza respondiendo con un error.

DSpace sugiere dos mecanismos adicionales de seguridad, implementar seguridad a nivel de Servlets (ServletSecurity) y la utilización del conector Mod_jk para interactuar con Apache Tomcat. No entraremos en detalle de estos dos mecanismos.

3.11. BÚSQUEDA AVANZADA, BÁSICA Y FILTROS

Los tipos de búsqueda en DSpace pueden ser básicos o avanzados, o en las nuevas versiones se agregaron las búsquedas Discovery a través de SOLR (o como lo traduce DSpace al español Otras opciones relacionadas). A partir de la versión 4 de DSpace, las búsquedas utilizan por defecto el módulo Discovery.

Al principio DSpace realizaba las búsquedas basadas en el motor Jakarta Lucene, la misma es una API de código abierto para la recuperación de la información desarrollada por la empresa Doug Cutting. La lógica de este motor de búsqueda se centra en el concepto de Documento conteniendo campos de texto lo cual facilita la independencia del formato del archivo pudiendo ser indexados mientras se extrae información de los mismos. Esta opción sigue disponible en las nuevas versiones de DSpace pero debe ser activada.

El módulo Discovery habilita las búsquedas y las llamadas búsquedas facetadas o búsquedas guiadas (del inglés faceted searching). DSpace Discovery ofrece poderosas configuraciones para búsquedas que en versiones anteriores eran sólo es posible con la personalización de código.

3.11.1. BÚSQUEDAS POR FACETAS (SIDEBAR FACET)

El sistema de clasificación facetada permite representar una materia o tema combinando los conceptos que representan diferentes aspectos (facetas) de la misma. Este tipo de búsquedas en la versión traducida al español de DSpace aparece como "Otras opciones relacionadas", rompe con la tradicional búsqueda por múltiples categorías y le permiten al usuario restringir sus resultados de búsqueda basados en estas Facetas.

Otras opciones relacionadas		
Autor	Título	Fecha de lanzamiento
Flores Mora, Manuel (1923-1985) 1	Cine - historia 1	1970 - 1976 1
Garce, Arturo (1931) 1	Cuentos ingleses - S. XX 1	1960 - 1969 3
Greene, Graham (1904-1991) 1	Filosofía griega 1	1950 - 1959 2
Otero Mendoza, Jorge. 1	Genética 1	1930 - 1939 3
Payró, Julio E. (1899-1971) 1	Pintura 1	
Platschek, Hans (1923-2000) 1	Poesía uruguaya - S. XX 1	
Somers, Armonia (1914 - 1994) 1		

Figura 13- Ejemplo de búsqueda facetada

En el ejemplo hay tres facetas, una por autor, una por título y otra por fecha. Estas búsquedas pueden ser modificadas en el archivo de configuración de DSpace. Además la agrupación de búsqueda puede contener más de un metadato asociado, por ejemplo, para el autor incluye los valores de dc.contributor.author y el de dc.creador.

Otra propiedad importante de esta búsqueda es que sus contenidos se actualizan automáticamente en el contexto de la página.

3.11.2. BÚSQUEDAS SIMPLES Y AVANZADAS

Estas búsquedas permiten al usuario modificar la lista de resultados de búsqueda que aparecen especificando "filtros" adicionales que se aplicarán en la lista de resultados de búsqueda.

Para la búsqueda simple, se introduce cualquier palabra y el motor de búsqueda busca en los campos de los ítems de la plataforma. Por ejemplo autor, título, fecha, etc. devuelve como resultado cualquier coincidencia en los mismos con lo buscado.

Existe la posibilidad de realizar una búsqueda avanzada, esta consiste en introducir filtros en la misma para acotar el dominio de resultados. Este tipo de búsqueda puede configurarse en el archivo de configuración dspace.cfg, los campos donde buscar y combinarlos con el uso de los operadores lógicos AND, OR o NOT, depurando de esta manera los resultados de la búsqueda y adaptándolos a los intereses del usuario.

Para las búsquedas se utiliza por defecto el operador booleano OR, el cual requiere que al menos uno de dichos términos esté presente. Para modificar el operador lógico se debe indicar el mismo en la siguiente línea del archivo de configuración

```
search.operator = OR
```

Los filtros de la búsqueda se modifican en el archivo mencionado, aquí se puede observar los campos en donde se encuentran los metadatos asociados:

```
##### Campos indices de busqueda #####  
  
search.index.1 = author:dc.contributor.*  
search.index.2 = author:dc.creator.*  
search.index.3 = title:dc.title.*  
search.index.4 = keyword:dc.subject.*  
search.index.5 = abstract:dc.description.abstract
```

Cada índice tiene un metadato en cuestión, para la búsqueda por filtro se asocia estos metadatos y luego el motor realiza el en base a estos índices un macheo devolviendo los que cumplen con las condiciones indicadas.

En ambas tipos de búsquedas, DSpace ofrece la opción “Todo DSpace”, puede ser restringida a una Comunidad o Colección específica devolviendo los ítems que cumplan con las condiciones en toda la plataforma.

3.11.3. BÚSQUEDA POR TEXTO COMPLETO

DSpace permite la búsqueda por texto completo, o sea el texto ingresado en el campo de búsqueda también será buscado en el texto de los documentos archivados.

Las búsquedas por texto completo están permitidas para los siguientes formatos de archivo: Adobe PDF, Microsoft Word, Texto Plano, HTML.

Para activar las búsquedas por texto completo se deberá setear en el archivo de configuración de DSpace el atributo: `search.maxfieldlength`. Este campo especifica la cantidad máxima de palabras a indexar para cada documento, y por defecto se establece en las primeras 10.000 palabras. Para indexar un número ilimitado de palabras se debe ingresar el valor -1. Luego de realizar el cambio es necesario re-indexar el DSpace.

Para indexar el texto completo de los archivos añadidos es necesario ejecutar el comando para generar filtros de medio (media-filters). Por lo tanto si se ingresa un nuevo archivo, no se indexara por texto completo hasta que se ejecute el comando para generarlo. En el Anexo de Configuraciones realizadas se explica la generación de filtro de medios.

3.12. INDEXACIÓN

Como se mencionó en secciones anteriores SOLR es el motor de búsqueda que utiliza DSpace, siendo el mismo el encargado de la indexación de archivos para una búsqueda rápida y eficiente.

A través del archivo `discovery.xml`, se realizan las configuraciones para definir los índices de búsqueda que el motor utilizara.

Para cada búsqueda se definen los índices básicos, veremos a continuación que índices vienen definidos por defecto.

Búsqueda por facetas

- **searchFilterAuthor:** Grupo de campos de metadatos, vienen dados por dc.contributor.author y dc.creator. Ordenados por la cantidad de ocurrencias de cada uno
- **searchFilterSubject:** Agrupa todos los temas dados por el campo (dc.subject.*) ordenados por la cantidad de ocurrencias de cada uno
- **searchFilterIssued:** Contiene el metadato dc.date.issued que identifica el tipo de fecha y los ordena por el valor de la misma

Búsqueda por filtros

- **searchFilterTitle:** Contiene el metadato dc.title metadata
- **searchFilterAuthor:** Contiene los campo dc.contributor.author y dc.creator
- **searchFilterSubject:** Contiene el metadato dc.subject.*
- **searchFilterIssued:** Contiene el metadato dc.date.issued

En el archivo de configuración discovery.xml podemos ver como se declaran los índices de búsqueda para los filtros de la búsqueda avanzada los cuales se declaran en el archivo de la siguiente forma:

```
<property name="searchFilters">
  <list>
    <ref bean="searchFilterTitle" />
    <ref bean="searchFilterAuthor" />
    <ref bean="searchFilterSubject" />
    <ref bean="searchFilterIssued" />
    <ref bean="searchFilterVersion" />
  </list>
</property>
```

Luego por ejemplo para hacer referencia a uno de estos índices para ver que metadato en este caso se quiere asociar al índice de búsqueda en cuestión en el mismo archivo se declara:

```
<bean id="searchFilterIssued"
class="org.dspace.discovery.configuration.DiscoverySearchFilterFacet">
  <property name="indexFieldName" value="dateIssued"/>
  <property name="metadataFields">
    <list>
      <value>dc.date.issued</value>
    </list>
  </property>
  <property name="type" value="date"/>
  <property name="sortOrder" value="VALUE"/>
</bean>
```

Esto hace referencia al metadato que se le asocia al índice. Para el caso del índice searchFilterIssued se ve en la configuración anterior que el valor asociado es el metadato dc.date.issued, cuando el motor de búsqueda SOLR recibe una petición de búsqueda avanzada cuyo filtro es searchFilterIssued ya conoce que el índice de búsqueda será el metadato asociado al mismo, en este caso dc.date.issued.

Al definir este campo dentro del archivo, describiremos las propiedades que son importantes y obligatorias junto con las opcionales.

- **indexFieldName (Obligatorio):** El nombre de filtro de búsqueda es único, los metadatos se indexarán en SOLR bajo este nombre de campo
- **metadataFields (Obligatorio):** Lista de campos de metadatos necesarios para incluirlos en la búsqueda.
- **sortOrder (opcional):** El orden de los filtros puede tomar dos valores:
 - COUNT se ordenan según la cantidad de veces que aparece el ítem.
 - VALUE se ordenaran alfabéticamente
- **type(opcional) :** Puede ser de dos tipos:
 - text: tratados como tal
 - date: Sólo el año se almacenará en el índice SOLR. Estos se muestran automáticamente en rangos que se hacen más pequeños cuando se selecciona alguna.

SOLR se comunica a través de http enviando y recibiendo documentos XML, busca sobre sus índices y no sobre la base de datos. Esto es de suma importancia para realizar una búsqueda eficiente, la diferencia se encuentra en el modelo de datos.

Mientras que las bases de datos relacionales utilizan tablas para realizar búsqueda de datos entre varias tablas utilizando el operador JOIN para unir a las involucradas, lo cual hace que esta operación puede tardar bastante si el volumen de datos es muy grande. En SOLR se utiliza un modelo de datos orientado a documentos haciendo que sea una sola tabla, solamente un conjunto de campos.

CAPÍTULO 4

4. PROPUESTA DE SOLUCIÓN

4.1. PROPUESTA DE METADATOS

Para la descripción del material bibliográfico los bibliotecarios utilizan las llamadas “normas de catalogación”, estas normas prescriben reglas para la catalogación de libros; las normas actualmente utilizadas en las bibliotecas uruguayas son anteriores a la difusión de Internet (Reglas de Catalogación Anglo Americanas 2R - 1990), si bien ya estaban orientadas a la catalogación de libros utilizando computadoras. A estas reglas se superpone una segunda norma específica de las bases de datos bibliográficas, el llamado “Formato MARC”.

El formato MARC prescribe por un lado una estructura para los campos de un registro bibliográfico y establece un determinado formato de archivo para el intercambio de datos entre aplicaciones de bibliotecas, el llamado “MARC de comunicaciones” basado a su vez este último en la norma ISO 2709. En el caso de un registro bibliográfico la norma MARC se utiliza para etiquetar los campos, en tanto las normas de catalogación determinan las reglas de puntuación [50].

El estándar establecido en nivel mundial para las BD y repositorios es sin lugar a dudas Dublin Core cualificado.

Para la asignación de metadatos a los objetos de las Colecciones digitales de la BNU, se utilizó un criterio general para transcribir las descripciones bibliográficas de los equivalentes reales de los objetos digitales, es decir utilizar los mismos criterios que los bibliotecarios utilizan para describir documentos, dentro de la estructura de campos que ofrece Dublin Core (DC). Para realizarlo Biblioteca redactó en el 2015 una serie de normas y estándares para el uso de DC cualificado, divididas por tipo de material: se desarrolló una norma para Imágenes, otra para Mapas y otras para Libros y Publicaciones periódicas.

A continuación se detallan los metadatos a nivel de los ítems para cada colección, propuestos por BNU y la FIC. El esquema contiene la columna de metadatos en esquema MARC21 y su correspondencia en Dublin Core, así como también una breve descripción del significado de cada uno y si es repetible y/u obligatorio. Estas propuestas fueron validadas y ajustadas en conjunto.

Colección de Imágenes:

MARC 21	Metadata Dublin Core	Descripción
LDR/006 (Leader posición 06 Tipo - tipo de registro.)	Type - dc.type	Campo obligatorio Tipo de publicación - naturaleza del recurso, obligatorio para el registro de metadatos. Ejemplo <dc.type> imagen fija </dc.type> <dc.type> fotografía </dc.type> <dc.type> afiche </dc.type> <dc.type> grabado </dc.type>
N/A	Format - dc.format	Tipo MIME - formato de archivo Ejemplo: <dc.format>image/JPEG </dc.format>
008/35-35 - (008 Posiciones 35-37 Idioma)	Language dc.language	Ejemplo: <dc.language> es </dc.language>
856 (856 Subcampo u URL.)	dc.identifier.URI	Campo obligatorio identificación única del recurso <dc.identifier.URI> </dc.uri.identifier>
100, 700, 110, 710, 111, 711 (100 Autor principal - persona. 110 Autor principal - entidad. 111 Autor principal - congreso. 700 Autor secundario - persona. 710 Autor secundario - entidad. 711 Autor secundario - congreso.)	Creator - dc.creator	Campo obligatorio Autor personal / Autor institucional (R) - persona u organización responsable de la creación del contenido del recurso. Ejemplo: <dc.creator> Atget, Eugéne </dc.creator> <dc.creator> Montevideo Refescos S.A. </dc.creator>
245 (245 Título propio)	Title- dc.title	Campo obligatorio Título (NR) - En este campo se ingresa el título de la imagen. Ejemplo: <dc.title> Operarios en plena tarea </dc.title>
651, 650^z	Coverage - dc.coverage.spatial	Lugar (R) - lugar que representado en la imagen, en caso de existir Ejemplo: <dc.coverage.spatial> Buenos Aires </dc.coverage.spatial>
260\$b	Publisher - dc.publisher:	Campo repetible Editor - Editor responsable del recurso. Ejemplo: <dc.publisher>Bates</dc.publisher>
260\$c	Date - dc.date	Campo obligatorio Años de publicación - fechas de publicación de la imagen. Para el caso de fotografías suele ser el año en

		que la imagen fue tomada. Ejemplo: <dc:date> 1985 </dc:date>
300	Format - dc.format.extent	Descripción física - descripción física del documento, incluyendo dimensiones Ejemplo: <dc:format.extent> 1 foto : byn ; 18 x 12 cm. </dc:format.extent>
600, 610, 611, 630, 650, 651 (600 Materia - nombre persona. 610 Materia - nombre entidad. 611 Materia - nombre congreso. 630 Materia - título uniforme. 650 Materia. 653 Materia - Cronológica)	Tema - dc.subject	Campo repetible Temas (R) - descriptores temáticos distintos del nombre del mapa, que llevará una calificación específica. Se admiten dos tipos de valores posibles: se puede codificar una palabra clave o una tematización preestablecida. Se introducirá con mayúscula. Ejemplo: <dc:subject> TEMA </dc:subject>
500	Description - dc.description.note	Notas (R) - la totalidad de las notas correspondientes al campo 500 de MARC se consignan en esta etiqueta.
	BibliographicCitation - dcterms.bibliographicCitation	Descripción bibliográfica - cita bibliográfica de la publicación en formato ISBN.
490	Relation - dc.relation.isPartOf	Colección- título de la serie de la cual la imagen es parte, si hay número se consigna separado de título por comas Ejemplo <dc:relation.isPartOf> Primera colección Barrios Pintos </dc:relation.isPartOf/>
40	Source - dc.source	Fuente - colección de la BNU donde se encuentra el documento original Ejemplo <dc:source> Biblioteca Nacional del Uruguay, Materiales Especiales </dc:source>
N/A	License - dc.license	LICENCIA, DERECHOS - Licencia (Opcional) - licencia bajo la que se distribuye el objeto Ejemplo <dc:licence> Creative Commons 2.0 </dc:licence> (Optional)

Tabla 2- Metadatos para Colecciones de Imágenes

Colección Libros

MARC 21	Metadata Dublin Core	Descripción
LDR/006 (Leader posición 06 Tipo - tipo de registro.)	Type - dc.type	Campo obligatorio Tipo de publicación - naturaleza del recurso, obligatorio para el registro de metadatos. Ejemplo <dc:type> libro </dc:type>
N/A	Format - dc.format	Tipo MIME - formato de archivo Ejemplo: <dc:format>application/pdf </dc:format>
008/35-35 - (008 Posiciones 35-37 Idioma)	Language dc.language	Ejemplo: <dc:language> es </dc:language>
856 (856 Subcampo u URL.)	dc.identifier.URI	Campo obligatorio identificación única del recurso <dc.identifier.URI> </dc:uri.identifier>
20 (020 Subcampo a ISBN.)	Identifier- dc.identifier.isbn	ISBN - número de ISBN Ejemplo <dc.identifier.isbn> 9789974839601 </dc.identifier.isbn>
100, 700, 110, 710, 111, 711 (100 Autor principal - persona. 110 Autor principal - entidad. 111 Autor principal - congreso. 700 Autor secundario - persona. 710 Autor secundario - entidad. 711 Autor secundario - congreso.)	Creator - dc.creator	Campo obligatorio Autor personal / Autor institucional (R) - persona u organización responsable de la creación del contenido del recurso. Ejemplo: <dc.creator> Hernández, Felisberto (1902-1964) </dc.creator> <dc.creator> Universidad de la República (Uruguay). Facultad de Ingeniería. </dc.creator>

100, 700, 110, 710, 111, 711 (100 Autor principal - persona. 110 Autor principal - entidad. 111 Autor principal - congreso. 700 Autor secundario - persona. 710 Autor secundario - entidad. 711 Autor secundario	Contributor - dc.contributor	Otros roles de autoría - Para el caso de otros roles de autoría, como ilustradores, prologuistas, compiladores etc. se utilizará la etiqueta dc.contributor, seguida de un cualificador del tipo de contribución. Los cualificadores autorizados son: dc.contributor.illustrator dc.contributor.translator dc.contributor.editor dc.contributor.compiler dc.contributor.moderator En caso que el rol del autor no entre dentro de ninguna de estas categorías se utilizará la etiqueta sin calificación.
245 (245 Título propio)	Title- dc.title	Campo obligatorio Título (NR) - En este campo se ingresa el título de la imagen. Ejemplo: <dc.title> Operarios en plena tarea </dc.title>
260\$b	Publisher - dc.publisher:	Campo repetible Editor - Editor responsable del recurso. Ejemplo: <dc:publisher>Bates</dc:publisher>
260\$c	Date - dc.date	Campo obligatorio Años de publicación - fechas de publicación de la imagen. Para el caso de fotografías suele ser el año en que la imagen fue tomada. Ejemplo: <dc:date> 1985 </dc:date>
300	Format - dc.format.extent	Descripción física - descripción física del documento, incluyendo dimensiones Ejemplo: <dc:format.extent> 1 foto : byn ; 18 x 12 cm. </dc:format.extent>
600, 610, 611, 630, 650, 651 (600 Materia - nombre persona. 610 Materia - nombre entidad. 611 Materia - nombre congreso. 630 Materia - título uniforme. 650 Materia. 653 Materia - Cronológica)	Tema - dc.subject	Campo repetible Temas (R) - descriptores temáticos distintos del nombre del mapa, que llevará una calificación específica. Ejemplo: <dc:subject> TEMA </dc:subject>

654	Tema - dc.subject	Bajo esta etiqueta se pondrán también los encabezamientos temáticos de lugar Ejemplo: <dc:subject> Novela uruguaya S. XX.</dc:subject>
500	Description - dc.description.note	Notas (R) - la totalidad de las notas correspondientes al campo 500 de MARC
	BibliographicCitation - dcterms.bibliographicCitation	Descripción bibliográfica - cita bibliográfica de la publicación en formato ISBN.
490	Relation - dc.relation.isPartOf	Colección- título de la serie de la cual la imagen es parte, si hay número se consigna separado de título por comas Ejemplo <dc:relation.isPartOf> Primera colección Barrios Pintos </dc:relation.isPartOf/>
N/A	License - dc.license	LICENCIA, DERECHOS - Licencia (Opcional) - licencia bajo la que se distribuye el objeto Ejemplo <dc:licence> Creative Commons 2.0 </dc:licence> (Optional)

Tabla 3 - Metadatos para Colecciones de Libros

Colección Mapas

MARC 21	Metadata Dublin Core	Descripción
LDR/006 (Leader posición 06 Tipo - tipo de registro.)	Type - dc.type	Campo obligatorio Tipo de publicación - naturaleza del recurso, obligatorio para el registro de metadatos. Ejemplo <dc:type> mapa </dc:type>
N/A	Format - dc.format	Tipo MIME - formato de archivo Ejemplo: <dc:format>image/JPEG </dc:format>
008/35-35 - (008 Posiciones 35-37 Idioma)	Language dc.language	Ejemplo: <dc:language> es </dc:language>
856 (856 Subcampo u URL.)	dc.identifier.URI	Campo obligatorio identificación única del recurso <dc.identifier.URI> </dc:uri.identifier>
100, 700, 110, 710, 111, 711 (100 Autor principal - persona. 110 Autor principal - entidad. 111 Autor principal -	Creator - dc.creator	Campo obligatorio Autor personal / Autor institucional (R) - persona u organización responsable de la creación del contenido del recurso. Ejemplo: <dc.creator> Reyes, José María </dc.creator> <dc.creator> Uruguay. Comando General de la

congreso. 700 Autor secundario - persona. 710 Autor secundario - entidad. 711 Autor secundario - congreso.)		Armada. Servicio de Oceanografía, Hidrografía y Metereología. </dc.creator>
245 (245 Título propio)	Title- dc.title	Campo obligatorio En este campo se ingresa el título del mapa publicación periódica de acuerdo a su normalización Ejemplo: <dc.title>Carta geográfica de la República Oriental del Uruguay</dc.title>
260\$b	Publisher - dc.publisher:	Campo repetible Editor - Editor responsable del recurso. Ejemplo: <dc:publisher>SOHMA</dc:publisher>
260\$c	Date - dc.date	Campo obligatorio Años de publicación - fechas de publicación del mapa Ejemplo: <dc:date> 2006</dc:date>
300	Format - dc.format.extent	Descripción física - descripcion física del documento, incluyendo dimensiones Ejemplo: <dc:format.extent> 1 foto : byn ; 18 x 12 cm. </dc:format.extent>
600, 610, 611, 630, 650, 651 (600 Materia - nombre persona. 610 Materia - nombre entidad. 611 Materia - nombre congreso. 630 Materia - título uniforme. 650 Materia. 653 Materia - Cronológica)	Tema - dc.subject	Campo repetible Descriptores temáticos distintos del nombre del mapa, que llevará una calificación específica. Se admiten dos tipos de valores posibles: se puede codificar una palabra clave o una tematización preestablecida. Ejemplo: <dc:subject> Mapas </dc:subject>
654	Tema - dc.subject	Bajo esta etiqueta se pondrán también los encabezamientos temáticos de lugar Ejemplo: <dc:subject> Puerto de Montevideo – Mapas</dc:subject>
500	Description - dc.description.note	Ejemplo: <dc:descripton.note> Contiene referencias de datos de Marea. </dc:descripton.note>
	BibliographicCitation - dcterms.bibliographicCitation	Colección-título de la serie de la cual el mapa es parte

651, 650\$z	Coverage - dc.coverage.spatial	Lugar (R) - lugar representado por el mapa. También pueden figurar las coordenadas del mapa.
650\$y, 651	Coverage - dc.coverage.temporal	Fecha (R)- para el caso de mapas históricos su fecha, en caso de figurar en el título o en la descripción del mismo. Ejemplo <dc.coverage.temporal> 1880 </dc.coverage.temporal>
490	Relation - dc.relation.isPartOf	Colección- título de la serie de la cual la imagen es parte, si hay número se consigna separado de título por comas Ejemplo <dc.relation.isPartOf> Primera colección Barrios Pintos </dc.relation.isPartOf/>
N/A	License - dc.license	LICENCIA, DERECHOS - Licencia (Opcional) - licencia bajo la que se distribuye el objeto Ejemplo <dc:licence> Creative Commons 2.0 </dc:licence> (Optional)

Tabla 4 - Metadatos para colecciones de Mapas

Colección Publicaciones Periódicas

Para el caso de Publicaciones periódicas el uso de DSpace generaba ciertas dificultades, nacidas del hecho de que en DSpace la cardinalidad de relación entre metadatos y objetos es de 1 a 1. En el caso de las colecciones de publicaciones periódicas de FIC - BNU cada número de un diario o revista está escaneado en un PDF aparte; sin embargo los bibliotecarios, para describir este tipo de material, utilizan registros que describen o bien el título de la publicación como una totalidad (el conjunto de sus números) o bien unidades más pequeñas que el número, a saber los artículos que lo componen. Esto no se ajusta a la cardinalidad de la relación entre objetos y metadatos en DSpace.

Por otro lado algunos juegos de metadatos como MODS permiten hacer vínculos entre distintos registros, de manera que se podría llegar a tener un registro global para el número linkeado a los N registros correspondientes a cada artículo, pero DC no ofrece esta posibilidad.

Después de consultar otras BD en Internet la BNU optó por la siguiente solución:

- mantener un solo metadato para la totalidad del número
- agregar cada uno de los artículos en una nota de contenido, consignando autor del artículo, título del artículo y página inicial y final del mismo. Se trata de la forma en que se consigna el contenido de los libros y equivale más o menos a una transcripción del sumario de la publicación, cuando existe
- agregar todos los temas en el campo dc.description.subject

- agregar todos los autores principales en el campo dc.creator.author

Esta solución agregó trabajo a la conversión de registros bibliográficos en metadatos ya que era necesario unir todos los registros correspondientes a los artículos de un número determinado de un título en un solo registro.

La principal desventaja de la solución que adoptaron es no tener posibilidad de recuperar el objeto por el título de alguno de los artículos que contenga si la búsqueda se filtra por el campo dc.title.

Se detalla a continuación los metadatos propuestos, en los cuales se trabajó en conjunto con la BNU.

MARC 21	Metadata Dublin Core	Descripción
LDR/006 (Leader posición 06 Tipo - tipo de registro.)	Type - dc.type	Campo obligatorio Tipo de publicación - naturaleza del recurso, obligatorio para el registro de metadatos. Ejemplo <dc.type> revista </dc.type>
008/35-35 - (008 Posiciones 35-37 Idioma)	Language dc.language	Ejemplo: <dc:language> es </dc:language>
856 (856 Subcampo u URL.)	dc.identifier.URI	Campo obligatorio URI (NR) - URL de la colección en el repositorio de Periódicas del Uruguay. <dc.identifier.URI> </dc:uri.identifier>
20 (020 Subcampo a ISBN.)	Identifier- dc.identifier.isbn	ISBN - identificación única del recurso Ejemplo: <dc.identifier.isbn> ISSN:1361-3200 </dc.identifier.isbn>
110 Nombre Corporativo 710 Nombre Corporativo (Ent. Sec)	Creator - dc.creator	Editor institucional (R) - Es la entidad responsable de la creación del contenido del recurso, puede ser una persona (autor), organización o un servicio. Para el caso de periódicas se utilizar para registrar los editores institucionales.
245 (245 Título propio)	Title- dc.title	Campo obligatorio Título (NR) - En este campo se ingresa el título de la publicación periódica de acuerdo a su normalización. Ejemplo: <dc.title>La pluma</dc.title>
260\$b	Publisher - dc.publisher:	Campo repetible Editor - Editor responsable del recurso. Ejemplo: <dc:publisher>Universidad de la República (Montevideo) </dc:publisher>

260\$c	Publisher - dc:publisher	Lugar de edición - El lugar de edición deberá ser guardado después del nombre del editor, encerrado entre paréntesis. Ejemplo: <dc:publisher>Universidad de la República (Montevideo)</dc:publisher>
260\$c	Date - dc.date	Campo obligatorio Fechas extremas de la publicación. Campo obligatorio para las publicaciones periódicas. Ejemplo: <dc:date> 1947-1948</dc:date>
600, 610, 611, 630, 650, 651 (600 Materia - nombre persona. 610 Materia - nombre entidad. 611 Materia - nombre congreso. 630 Materia - título uniforme. 650 Materia. 653 Materia - Cronológica)	Tema - dc.subject	Campo repetible Descriptor temáticos Ejemplo: <dc.subject>Literatura uruguaya – Publicaciones periódicas</dc.subject> <dc.subject>Literatura -- Publicaciones periódicas</dc.subject>
654	Tema - dc.subject	Bajo esta etiqueta se pondrán también los encabezamientos temáticos de lugar Ejemplo: <dc.subject> Puerto de Montevideo – Mapas</dc.subject>
	BibliographicCitation - dcterms.bibliographicCitation	Descripción bibliográfica - cita bibliográfica de la publicación en formato ISBN. Ejemplo: <dc.bibliographicCitation> La Licorne. -- París : Éditions de la Licorne, 1947-1948. 2 v. ; 25 cm. -- Cahiers trimestriels de littérature publiés sous la direction de Susana Soca. - - Ed. bilingüe: francés y español. -- ISSN 1155-875X. -- Literatura uruguaya -- Publicaciones periódicas. -- Literatura -- Publicaciones periódicas. </dc.bibliographicCitation>

Tabla 5- Metadatos a nivel ítem de Publicaciones Periódicas

A continuación se presentan la correspondencia de metadatos a nivel de colección.

MARC 21	Metadata Dublin Core	Descripción
LDR/006 (Leader)	Type - dc.type	Campo obligatorio

posición 06 Tipo - tipo de registro.)		Tipo de publicación - naturaleza del recurso, obligatorio para el registro de metadatos. Ejemplo <dc:type> revista </dc:type>
008/35-35 - (008 Posiciones 35-37 Idioma)	Language dc.language	Ejemplo: <dc:language> es </dc:language>
856 (856 Subcampo u URL.)	dc.identifier.URI	Campo obligatorio URI (NR) - URL de la colección en el repositorio de Periódicas del Uruguay. <dc.identifier.URI> </dc:uri.identifier>
20 (020 Subcampo a ISBN.)	Identifier- dc.identifier.isbn	ISBN - identificación única del recurso Ejemplo: <dc.identifier.isbn> ISSN:1361-3200 </dc.identifier.isbn>
110 Nombre Corporativo 710 Nombre Corporativo (Ent. Sec)	Creator - dc.creator	Editor institucional (R) - Es la entidad responsable de la creación del contenido del recurso, puede ser una persona (autor), organización o un servicio. Para el caso de periódicas se utilizar para registrar los editores institucionales.
	BibliographicCitation - dcterms.bibliographicCitation	Descripción bibliográfica - cita bibliográfica de la publicación en formato ISBN. Ejemplo: <dc.bibliographicCitation> La Licorne. -- París : Éditions de la Licorne, 1947-1948. 2 v. ; 25 cm. -- Cahiers trimestriels de littérature publiés sous la direction de Susana Soca. - Ed. bilingüe: francés y español. -- ISSN 1155-875X. -- Literatura uruguaya -- Publicaciones periódicas. -- Literatura -- Publicaciones periódicas. </dc.bibliographicCitation>
260\$b	Publisher - dc.publisher:	Campo repetible Editor - Editor responsable del recurso. Ejemplo: <dc:publisher>Universidad de la República (Montevideo) </dc:publisher>

260\$c	Publisher - dc:publisher	Lugar de edición - El lugar de edición deberá ser guardado después del nombre del editor, encerrado entre paréntesis. Ejemplo: <dc:publisher>Universidad de la República (Montevideo)</dc:publisher>
260\$c	Date - dc.date	Campo obligatorio Fechas extremas de la publicación. Campo obligatorio para las publicaciones periódicas. Ejemplo: <dc:date> 1947-1948</dc:date>
600, 610, 611, 630, 650, 651 (600 Materia - nombre persona. 610 Materia - nombre entidad. 611 Materia - nombre congreso. 630 Materia - título uniforme. 650 Materia. 653 Materia - Cronológica)	Tema - dc.subject	Campo repetible Descriptor temáticos Ejemplo: <dc.subject>Literatura uruguaya – Publicaciones periódicas</dc.subject> <dc.subject>Literatura -- Publicaciones periódicas</dc.subject>
654	Tema - dc.subject	Bajo esta etiqueta se pondrán también los encabezamientos temáticos de lugar Ejemplo: <dc:subject> Puerto de Montevideo – Mapas</dc:subject>

Tabla 6- Metadatos a nivel de colección de Publicaciones Periódicas

4.2. CALIDAD METADATOS

DSpace provee un módulo denominado Curation System, proporciona como un mecanismo automático para la ejecución de controles y validaciones periódicas sobre el contenido. Se destaca su capacidad para ser extensible y configurado según las necesidades de cada institución. Para el control de calidad de los metadatos propusimos que se realicen las validaciones básicas puesto que DSpace es muy completo al respecto y puede ser una mejora a futuro. Por lo que nos pareció suficiente lo que se puede configurar para esta instancia.

Para el control de calidad de metadatos las curation tasks propuestos son las siguientes:

- Format profiling
- Required metadata
- Link checker

La funcionalidad de cada uno ya fue descrita en el capítulo referente a DSpace. Para una primera instancia estos tres controles de calidad nos parecieron suficientes dado que abarcan un análisis en profundidad de calidad sobre los ítems. Ambas instituciones estuvieron de acuerdo con lo propuesto.

4.3. ESTRUCTURA JERÁRQUICA BNU

Se detalla la propuesta de árbol de colecciones de la Biblioteca Digital para la BNU, la misma fue realizada junto con los colaboradores de la misma.

Las colecciones en un primer nivel quedan organizadas por tipo de material, salvo las publicaciones de la Biblioteca Nacional. La organización es la siguiente:

1. Publicaciones de la Biblioteca Nacional
2. Publicaciones Periódicas (sigue la estructura de Periódicas de la FIC)
3. Libros
4. Imágenes
5. Mapas
6. Archivo literario
7. Manuscritos históricos
8. Partituras

De estas comunidades, las correspondientes a Manuscritos Históricos, Partituras y Archivo Literario (quizás) no saldrían en producción en el lanzamiento original por lo que no será necesario definir la metadata correspondiente.

1. Publicaciones de la Biblioteca Nacional: Quedaría organizada por título, y dentro de las publicaciones periódicas, subdividida por fecha.
2. Publicaciones Periódicas: Se divide en las siguientes subcomunidades:
 1. Primeros impresos
 2. Diarios
 3. Semanarios
 4. Revistas

5. Boletines
 6. Otros impresos
3. Libros: Esta comunidad incluye libros y folletos.
 4. Imágenes: Tendría una primera subdivisión por tipo de material, la lista de categorías propuesta es la siguiente:
 1. Fotos
 2. Postales
 3. Monedas / medallas
 4. Afiches

Cada una de estas subcomunidades a su vez se dividirá por tópico (Aviación, Guerra del Paraguay, Montevideo), en su mayor parte por tópico geográfico, de acuerdo al perfil del material que se vaya subiendo.

5. Mapas: Quedarían divididos según criterio geográfico en las siguientes categorías:
 1. Uruguay
 2. Río de la Plata
 3. América
 4. Otras regiones
 5. Mapamundi

A su vez los mapas de Uruguay quedarían divididos en:

1. Mapas de Montevideo
2. Otros departamentos
3. Mapas del Uruguay

Para cada una de estas subdivisiones se hará una subdivisión cronológica por siglo (p.ej. Mapas del Río de la Plata del Siglo XVIII)

6. Archivo Literario: No será incluido actualmente. Quedaría dividido por fondos, y dentro de fondos por subfondos, según criterios archivológicos.
7. Manuscritos Históricos: No será incluido actualmente.
8. Partituras: No será incluido actualmente.

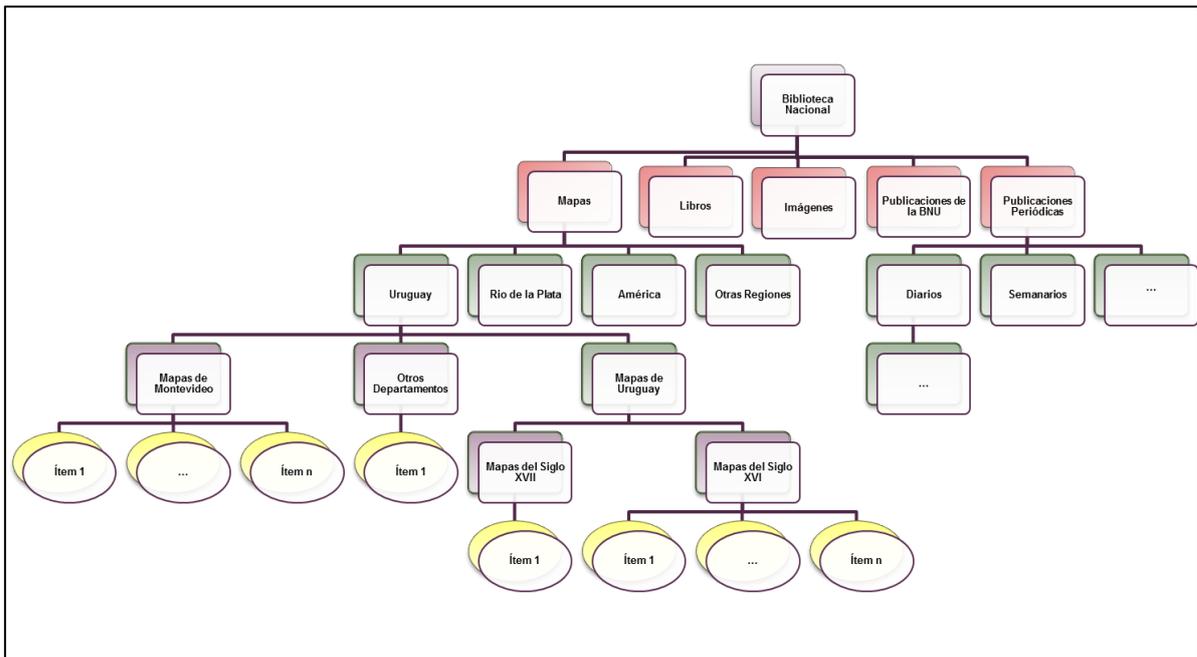


Figura 14 - Esquema de Estructura de la BNU

4.4. ESTRUCTURA JERÁRQUICA FIC

A continuación detallamos la estructura jerárquica para el repositorio de la FIC.

El primer nivel queda jerarquizado por las siguientes comunidades:

1. Autores del Uruguay
2. Publicaciones Periódicas
3. Figuras
4. Sobre la prensa

Estas se dividirán de la siguiente forma:

1. Autores del Uruguay: Contendrá los nombres de los autores como se especifica en la Propuesta de Autores del Uruguay más adelante
2. Publicaciones Periódicas:
 1. Diarios
 2. Semanarios
 3. Revistas
 4. Primeros impresos periódicos
 5. Boletines

6. Otros Impresos

3. Figuras: Actualmente solo tiene la información de Pedro Figari
4. Sobre la prensa

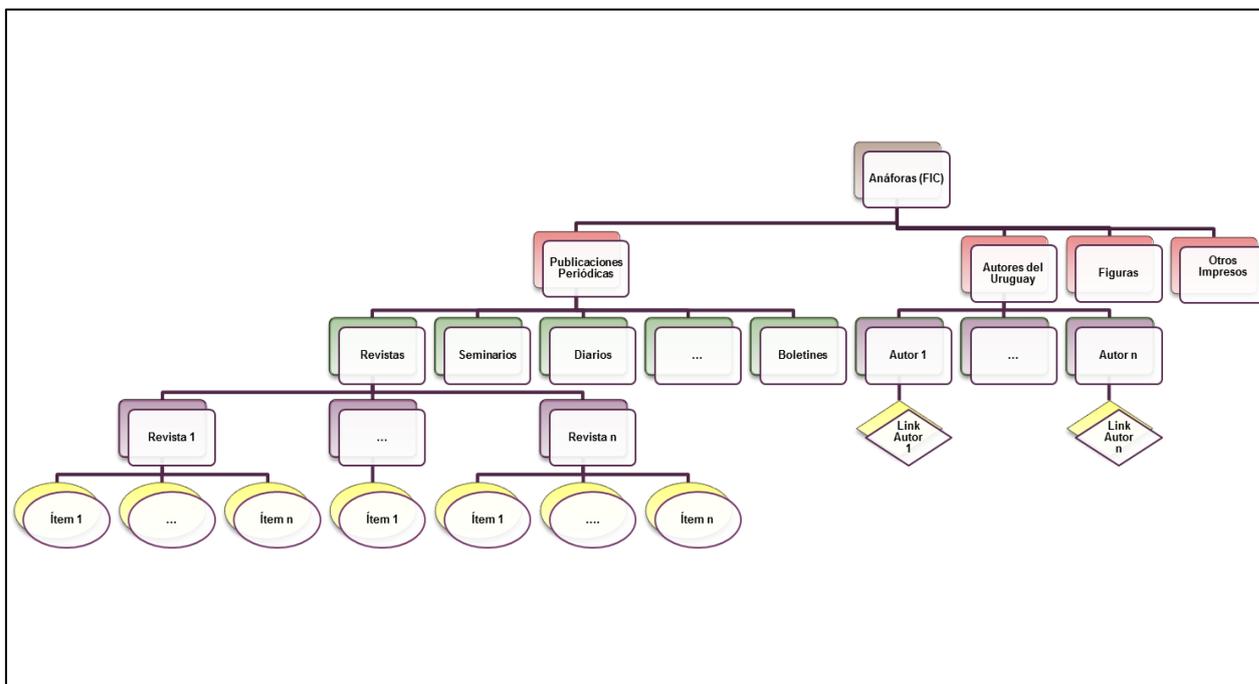


Figura 15 – Esquema de la estructura de la FIC

4.5. PROPUESTA DE FLUJO DE TRABAJO

A la hora de la publicación de un ítem DSpace proporciona alternativas de flujo de trabajo a seguir cuyo objetivo es que antes de la publicación del ítem el mismo sea revisado por los expertos para exponer la información de forma correcta y asegurar la calidad de la misma.

Para el diseño de la propuesta del flujo de trabajo se realizaron dos reuniones y conjunto a ambas instituciones surgieron dos propuestas de flujo según necesidades planteadas.

Se detalla el diagrama de cada uno junto con las alternativas brindadas, se puede observar los cambios realizados a partir del flujo que DSpace trae por defecto.

4.5.1. PROPUESTA DE FLUJO ALTERNATIVA 1

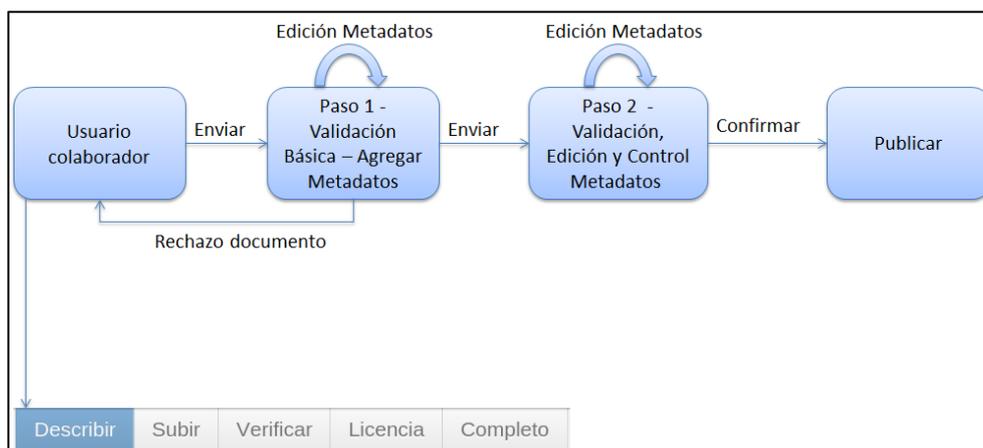


Figura 16- Flujo propuesto 1

Se realizó una posible alternativa que ajusta en un futuro a las necesidades de las instituciones. La misma consiste en la eliminación del primer paso del flujo de trabajo que propone DSpace originalmente, dejando solo el paso 2 y 3 de forma que resulte ágil el proceso de publicación.

- **Paso Inicial (colaborador)** - Se ingresa la información correspondiente a un ítem de una determinada colección, la idea es que se ingrese la información necesaria para una descripción básica del objeto.
- **Paso 1** - En este paso se generó una tarea pendiente del paso Inicial para la correspondiente aprobación, en esta etapa se evaluará el documento, se podrá agregar metadatos, verificarlos y se verificará la correctitud de la información en caso de que no sea aprobado por el experto en cuestión se puede devolver el documento al colaborador si el mismo no satisface con las condiciones mínimas.
- **Paso 2** - En esta etapa se puede agregar más metadatos en forma de detalle. El experto que se encuentre revisando este documento puede decidir descartar en caso de que no satisfaga con las condiciones necesarias para su correcta publicación.
- **Paso 3** - Confirmación y publicación del ítem.

4.5.2. PROPUESTA DE FLUJO ALTERNATIVA 2

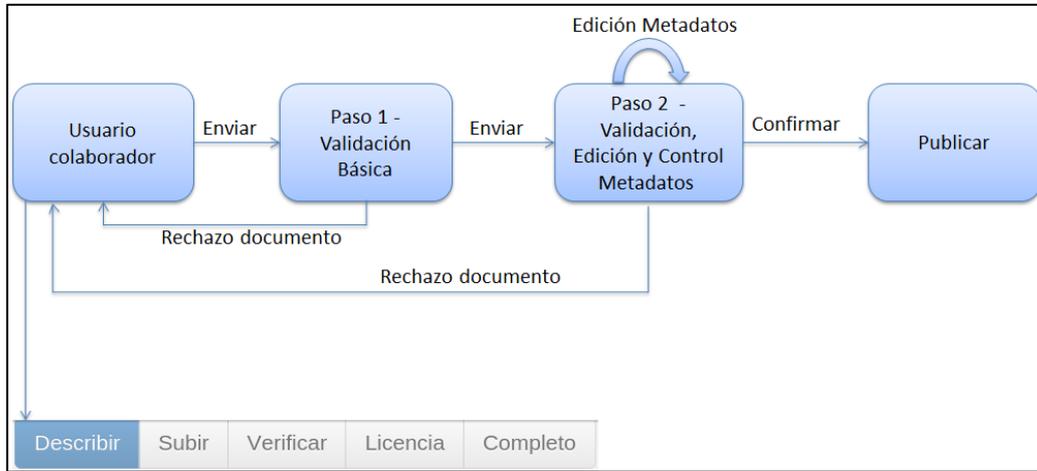


Figura 17 - Flujo propuesto 2

Esta alternativa propuso unificar los pasos 2 y 3 del flujo de trabajo original en uno solo, cuestión de que solo un experto realice la edición y control de metadatos ingresados en el objeto.

- **Paso Inicial (colaborador)** - Se ingresa la información correspondiente a un ítem de una determinada colección, la idea es que se ingrese la información necesaria para una descripción básica del objeto.
- **Paso 1** - En este paso se generó una tarea pendiente del paso Inicial para la correspondiente aprobación, en esta fase se evaluará el documento, se verificará la correctitud de la información en caso de que no sea aprobado por el experto en cuestión. Se puede devolver el documento al colaborador si el mismo no satisface con las condiciones mínimas.
- **Paso 2** - En esta etapa se puede agregar más metadatos en forma de detalle. El experto que se encuentre revisando este documento puede rechazarlo y al rechazarlo el proceso notifica al colaborador para que este pueda corregir las observaciones que le detallan.
- **Paso 3** - Confirmación y publicación del ítem.

Estas alternativas propuestas reflejan el flujo de trabajo “ideales” para la publicación de un ítem junto con los controles necesarios del mismo por parte del personal involucrado en cada fase. En la actualidad las instituciones involucradas no cuentan con la cantidad suficiente de recursos de personal como para llevarlos a cabo, por lo cual estas propuestas quedaron para un posible flujo de trabajo a futuro, por lo tanto no se tendrá un flujo de trabajo al menos al principio.

Adicionalmente la institución FIC propuso agregar un campo versión que indica la versión del documento, a modo que el usuario sepa qué versión del documento se encuentra consultando, si es versión final o en desarrollo.

4.6. COMPLEMENTOS (PLUGINS)

A la hora de desplegar un ítem, DSpace no ofrece ninguna vista previa de los archivos dentro de la página, sino que si se quiere visualizar el/los archivos correspondientes al ítem se debe descargar el archivo o visualizarlo en una pestaña aparte. Si trae una pre visualización para imágenes, activándola desde el archivo de configuración, pero con imágenes grandes como lo son los mapas no se llega a distinguir nada ya que crea una imagen más pequeña con menos resolución que la original.

Por parte de la FIC y BNU hicieron especial énfasis en que no quedara como algo plano como lo es un catálogo, donde el usuario tiene que acceder a un link para ver el material, sino mostrar al usuario el material de una forma más amigable y sin necesidad de descargárselo. Como se trabaja con imágenes, mapas y PDFs de gran tamaño, al tener que descargarlo o visualizarlo de la forma por defecto puede ser molesto tener que descargarse el archivo del ítem o esperar a que cargue completamente la imagen o el archivo para poder visualizarlo.

Se propuso agregar complementos de imágenes y PDF's para poder visualizar los archivos sin necesidad de descargarlos. Luego por parte de la FIC se pidió poder visualizar videos y escuchar archivos de audios.

Existen desarrollos por parte de una empresa llamada Atmire [51], la cual ofrece algunos complementos para imágenes, video y PDF para agregar a DSpace pero la mayoría no son de código abierto y son para la interfaz XMLUI. Cómo la incorporación de complementos de Atmire no era posible realizarla sin generar costos adicionales, no tomamos en cuenta la opción.

Por este motivo se agregaron en ambos repositorios complementos de terceros para una visualización más amigable al usuario de las imágenes, archivos PDFs y de los archivos de audio y video.

Otro pedido de las partes fue poder visualizar los últimos ítems agregados. DSpace trae por defecto esta configuración pero la visualización no era muy amigable, además solo se muestra el nombre del ítem, como se muestra en la Figura 18.



Figura 18 - Últimos ítems agregados

Nos pareció pertinente que sería más atractivo para el usuario agregar un carrousel de imágenes, como el que trae por defecto (Bootstrap carousel) solo permite utilizar una imagen por vez y las miniaturas son generadas con un tamaño pequeño quedarían perdidas en el mismo. Se optó por utilizar un complemento que permitiera mostrar varias imágenes al mismo tiempo.

En el próximo capítulo se detallan cuales fueron utilizados y las modificaciones realizadas para utilizar los mismos.

4.7. INTERFAZ DE USUARIO

DSpace trae dos posibles interfaces de usuario como ya se mencionó, JSPUI y XMLUI, si bien pueden convivir ambas, se tuvo que seleccionar una de ellas para customizar.

En las primeras versiones de DSpace solo existía la versión JSPUI y por tal motivo esta interfaz tiene amplia difusión.

XMLUI

XMLUI tiene una curva de aprendizaje mayor a JSPUI. Utiliza el framework Apache Cocoon, el cual es un Framework para la construcción de aplicaciones Web, con transformaciones XSL agregando complejidad a la comprensión de la misma. Permite a cada comunidad o colección tener su propia customización.

JSPUI

JSPUI está implementado utilizando Java Servlets que manejan la lógica de negocio y JavaServer Pages que produce los HTML. Los cambios son realizados sobrescribiendo los JSP originales. Para cambio de colores y tipografía utiliza hojas de estilo, las cuales pueden ser modificadas. Solo permite realizar cambios básicos a la interfaz, para realizar otro tipo de cambios se debe agregar código extra y/o modificar el existente.

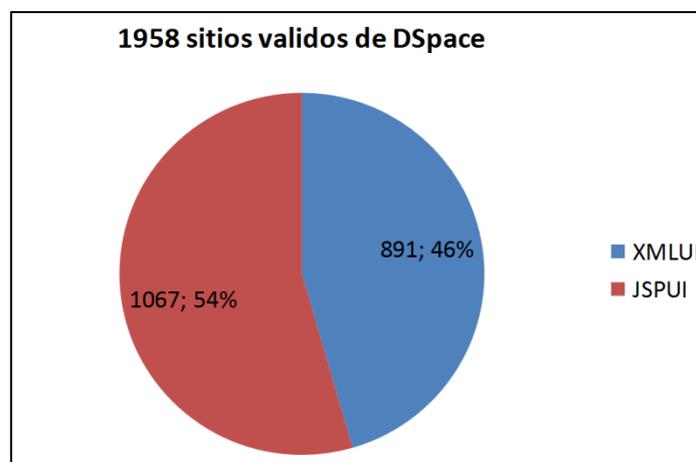


Figura 19- Uso de interfaces [52]

Se seleccionó la versión JSPUI para customizar porque consideramos que es la más flexible e interesante y la cual tiene menor curva de aprendizaje debido al tiempo y conocimiento con el que contábamos. Además JSPUI es la más usada.

4.8. SINCRONIZACIÓN

Uno de los objetivos más importantes del proyecto consiste en la realización de la sincronización entre la FIC y la BNU. Para ambas instituciones es de vital importancia mantener en forma de “espejo” la información en ambas bibliotecas. Para esta instancia analizamos tres posibles soluciones.

Solución 1

Esta fue la primera posible solución que se analizó, realizar una sincronización entre ambas bases de datos de las instituciones. La idea era correr un job a la noche en la cual la base de datos de la FIC se sincronizará con la de la BNU, y que copiara toda la información a esta de forma de mantener en ambas bases la información actualizada. Luego de investigar como DSpace almacenaba los datos vimos que esto no era viable ya que no utiliza una base de datos relacional para almacenar la información total del objeto sino que la almacena en un directorio aparte. Por lo que esta propuesta que se originó al inicio fue descartada.

Solución 2

La segunda solución consta de una copia diaria del directorio donde se almacenan los datos de la biblioteca digital y de esta forma copiarlo en un directorio específico de la BNU para que luego mediante un script se copiaran esos datos a la biblioteca digital de la BNU. Al principio nos pareció una buena idea y que solucionaba el tema de mantener la información actualizada en ambas instituciones pero luego vimos que resultaba un poco burocrático y tedioso el hecho de migrar diariamente la información de esta forma, dando paso a que se investigue en profundidad lo que vendría a ser la propuesta final con la cual exitosamente logramos el objetivo planteado.

Solución 3

Esta es la propuesta final con la cual se optó para cumplir exitosamente con el cometido de sincronizar ambos repositorios. Para ello se investigó como DSpace realiza la misma como se hace mención en el capítulo DSpace, este se basa en el protocolo OAI-PMH el cual promueve la difusión eficiente de contenidos en la web.

La elección de este protocolo se basa por la facilidad del entorno de interoperabilidad de archivos. El mecanismo para el cual OAI expone los metadatos en la web se detalla a continuación.

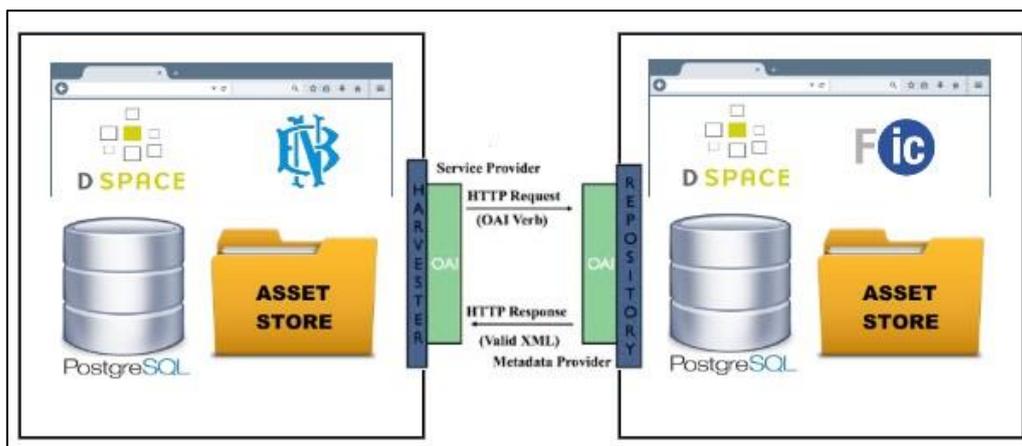


Figura 20- Esquema de sincronización OAI

En el esquema de la Figura 20 se puede ver la propuesta de sincronización que realizamos, la misma consiste en que la FIC expone los objetos y la BNU cosecha de esta.

Para esto vemos como el protocolo OAI expone la información proporcionada por la FIC, en este proceso se involucran dos actores uno es el llamado “proveedor de datos”, siendo en nuestro escenario la FIC como proveedor de los datos.

El mismo expone sus servicios para que luego los OAI servers localizan los metadatos de forma de utilizar esa información en los servicios de búsqueda de cada repositorio. Luego la BNU “cosecha” esta información. Esta propuesta fue la elegida como final y la más exitosa por su rapidez, eficacia y sencillez.

4.9. PROPUESTA AUTORES DEL URUGUAY

La Biblioteca Digital de Autores funciona como un conjunto de mini sitios (50 aprox.). Algunos de ellos conformaban el sitio archivos de prensa [53] y están hechos en HTML. Otros, la mayoría, están hechos como dokuwikis [54], permitiendo que colaboradores eventuales los construyan.

En principio no estaba previsto dentro de nuestro proyecto incluir el sitio mencionado, debido a que la migración masiva de este sitio no sería posible por la complejidad adicional que agrega al proyecto estando este requisito fuera del alcance del mismo.

Por pedido de la FIC que sería deseable incluir de alguna manera a Autores del Uruguay y que siguiera funcionando como una wiki colaborativa como hasta el momento. La posibilidad de que esta biblioteca digital se integrase al repositorio de FIC como una comunidad dentro del DSpace era una alternativa deseable ya que integra, centraliza y facilita la administración. A su vez, posibilita la contigüidad de sitios y materiales “hermanos”.

Por este motivo se realizó por nuestra parte la siguiente propuesta de solución para poder incorporar Autores del Uruguay a DSpace.

Se propone:

- Tener una Comunidad dentro del DSpace de FIC, incluye un flujo de trabajo diferente que simula una wiki colaborativa. Esto mediante usuarios con permisos para poder ingresar contenido al repositorio.
- En cuanto a los usuarios de la comunidad tendríamos dos opciones:
 - Cualquier usuario que se registre en el repositorio tendrá por defecto permisos para crear, subir y modificar ítems de esta comunidad.
 - Asignar perfil de Usuario Wiki, como cada comunidad puede tener su propio administrador de usuarios, podemos definir un administrador diferente, que otorgue los permisos de crear, subir y modificar los ítems, haciendo que esta sea independiente del resto del repositorio.

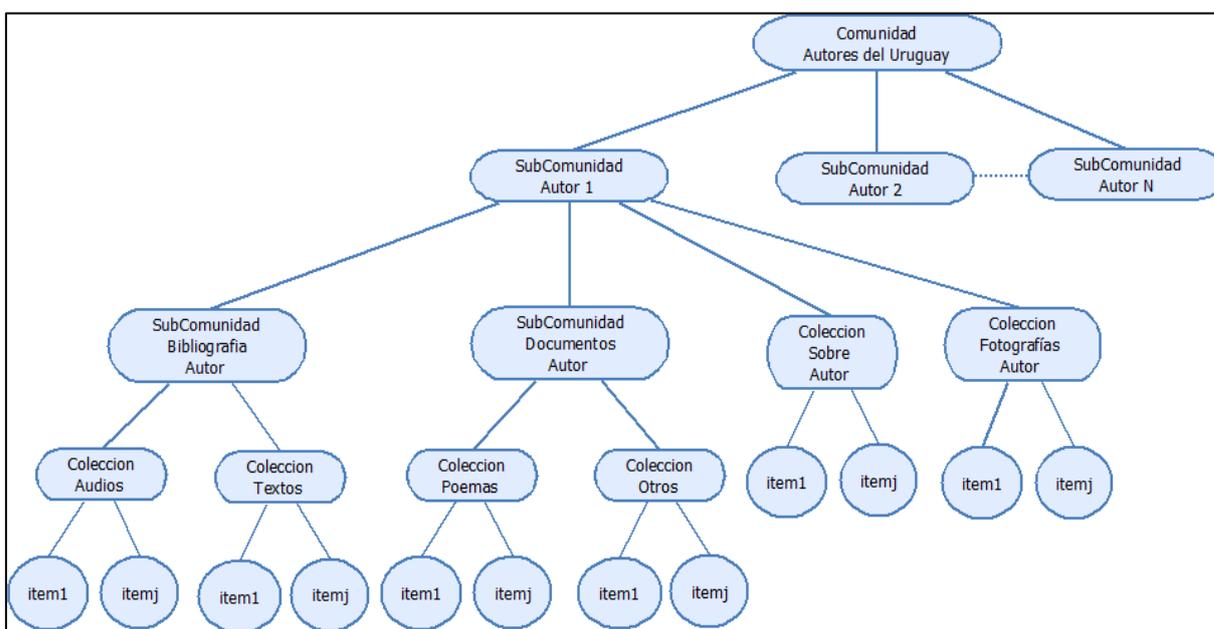


Figura 21-Propuesta de solución del árbol de Autores del Uruguay

Utilizar el flujo de trabajo de la siguiente manera:

- **Paso Inicial (Colaborador - Usuario Wiki)** - Se ingresa la información correspondiente a un ítem de una determinada colección.
- **Paso 1** - Confirmación de Datos ingresados y publicación.
- **Paso 2** - Iterativo Incremental de enriquecimiento de la información del ítem publicado, para simular una wiki en esta fase será ejecutado por cualquier persona que tenga permisos de edición del ítem o colección, siendo capaz de modificar la Metadata del ítem publicado, alimentando o validando la calidad de los metadatos ingresados por usuarios anteriores.

Para este paso de los 4 estados definidos, se pueden realizar cambios de estado a:

- En Desarrollo
- Verificación Pendiente
- Versión Final

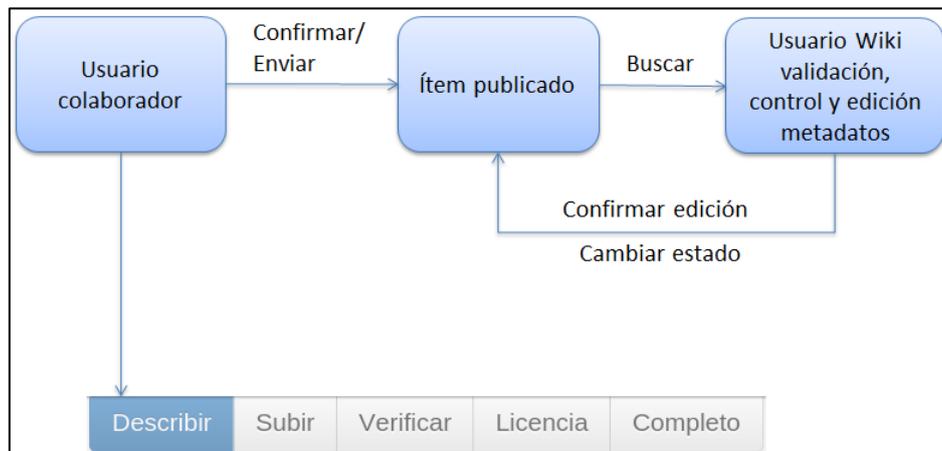


Figura 22- Flujo de trabajo propuesta

Sin embargo, las dificultades para la migración por motivos de la heterogeneidad de las fuentes y la imposibilidad de hacer un realojamiento manual de los archivos por falta de recursos humanos, indican una solución intermedia.

Esta solución sería la creación de una comunidad “Biblioteca Digital de Autores del Uruguay” en la biblioteca digital de FIC, en el que los nuevos autores a ser integrados, utilicen las herramientas de alojamiento del repositorio. Un procedimiento similar al utilizado para los sitios en Dokuwikis.

Se decidió realizar esta solución intermedia, donde los ítems no estarán alojados en el repositorio DSpace, sino que cada autor estará representado como una colección y cada colección tendrá un enlace a la página actual del mismo.

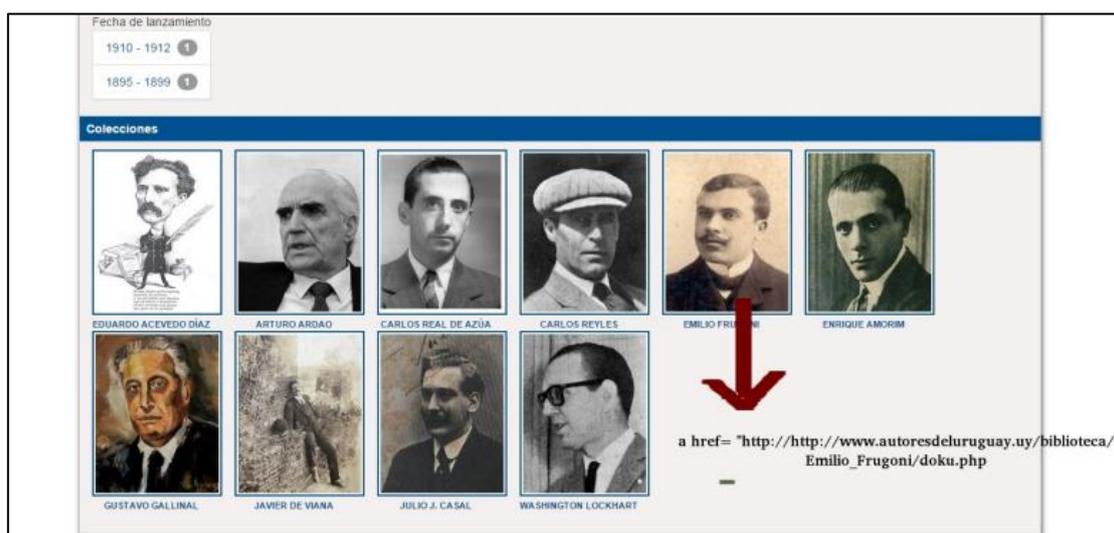


Figura 23 - Propuesta de FIC para Autores del Uruguay

4.10. PROPUESTA FRONTEND PROTOTIPO INICIAL

Debido a dudas e inquietudes por ambas partes, se realizó un primer prototipo de pruebas para poder mostrar las funcionalidades que permite el software DSpace. Además se realizaron modificaciones a la interfaz de usuario para que ambas partes validaran y nos dieran un feedback de si la interfaz cumplía con sus expectativas. Para estas modificaciones nos basamos en la Biblioteca Digital de la Biblioteca Nacional de La Plata, Argentina, llamada Trapalanda [11]. Ambas instituciones hicieron especial énfasis desde un principio de la clara interfaz de usuario que los desarrolladores habían logrado con el software DSpace, sobre todo como se presentaban gráficamente las comunidades, colecciones e ítems y en el uso de complementos. El trabajo realizado por los desarrolladores de Trapalanda, para llegar al producto actual fue de aproximadamente 3 años.

Realizamos la instalación de DSpace en un servidor de pruebas de la BNU, de modo que el prototipo quedará disponible para realizar pruebas.

En la presentación se mostraron las funcionalidades a las cuales puede acceder el administrador, se mostró cómo agregar/borrar colecciones e ítems, flujos de trabajo posibles, carga masiva, mapeo de ítems entre colecciones, entre otras.

Detallaremos la customización realizada para este prototipo.

- **Paleta de colores**
Se seleccionó una paleta de colores a modo de ejemplo
- **Portada**
Se posicionó el menú tope y se agregó un cabezal donde iría el logo de las instituciones.

Debajo del menú un texto introductorio y a la derecha del panel de búsquedas

Se modificó la barra de navegación (en la versión por defecto la barra horizontal verde), mostrando solo el texto y alineado a la derecha.

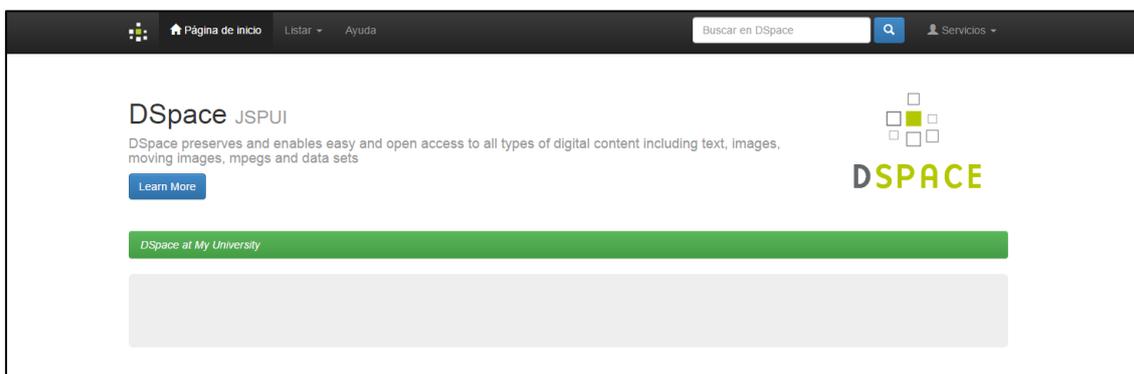


Figura 24 - Localización del menú y texto introductorio en la versión por defecto de DSpace



Figura 25- Localización del menú y texto introductorio en la versión del prototipo

Se modificó la visualización de las comunidades de vertical a horizontal así como la imagen a visualizar, en caso que no se tenga una imagen se agregó una por defecto cuyo texto es “Imagen no disponible”.

Además se ocultó la sección de “Otras opciones relacionadas” de la página principal.

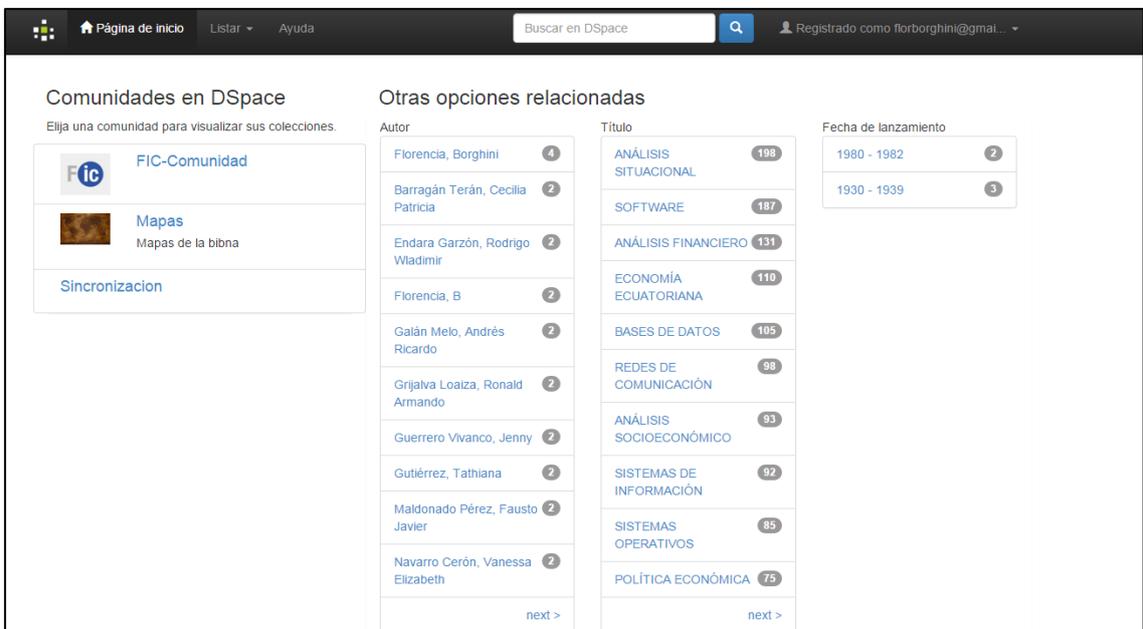


Figura 26- Comunidades en versión por defecto



Figura 27 - Comunidades y área de noticias en el prototipo

- **Pie de página**



Figura 28- Pie de página versión por defecto



Figura 29 - Pie de página prototipo

- **Subcomunidades y colecciones**

Para las subcomunidades y colecciones se eliminó el texto introductorio (este solo se muestra en la portada), se modificó la posición de la imagen y el texto de la subcomunidad y colección.

Además se modificó el panel de búsqueda agregando la búsqueda en todo DSpace que en la versión por defecto se encuentra en el menú.

El listado de subcomunidades o de las colecciones se muestra en forma vertical, al igual que en la página principal se colocó en forma horizontal y se agrandaron las imágenes. Se agregó también una imagen por defecto para las subcomunidades o colecciones que no tienen imagen, la misma contiene el texto de "Imagen no disponible".

Se sacó de esta página la sección de envíos recientes.



Figura 30- Página de una subcomunidad - colección en versión por defecto

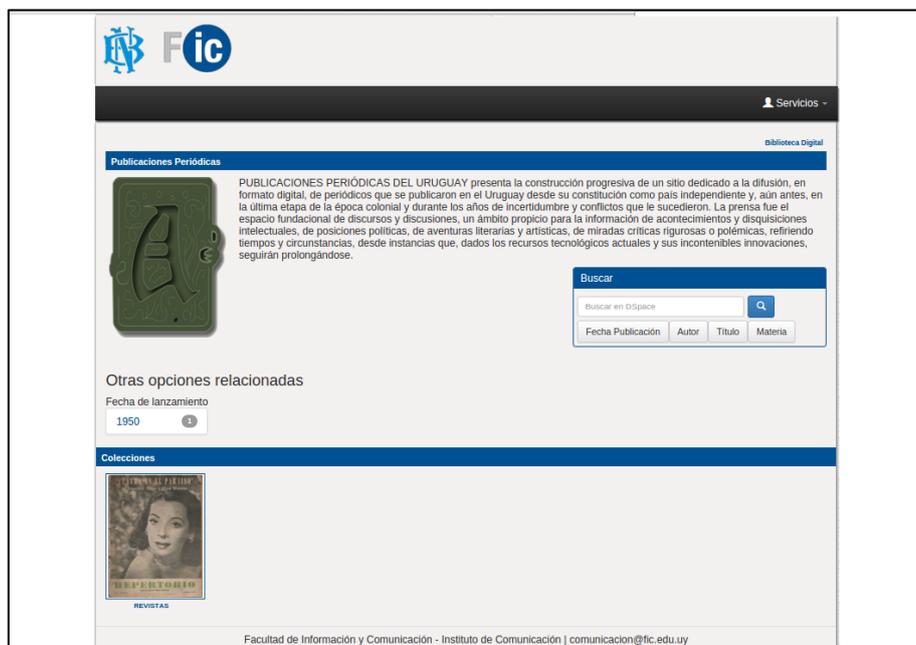


Figura 31 - Página de una subcomunidad en prototipo

- Ítems

Dentro de las colecciones, los ítems se despliegan en forma tabular en la versión por defecto de DSpace. Al igual que en las subcomunidades y colecciones, se modificaron los ítems de forma horizontal junto a su imagen.

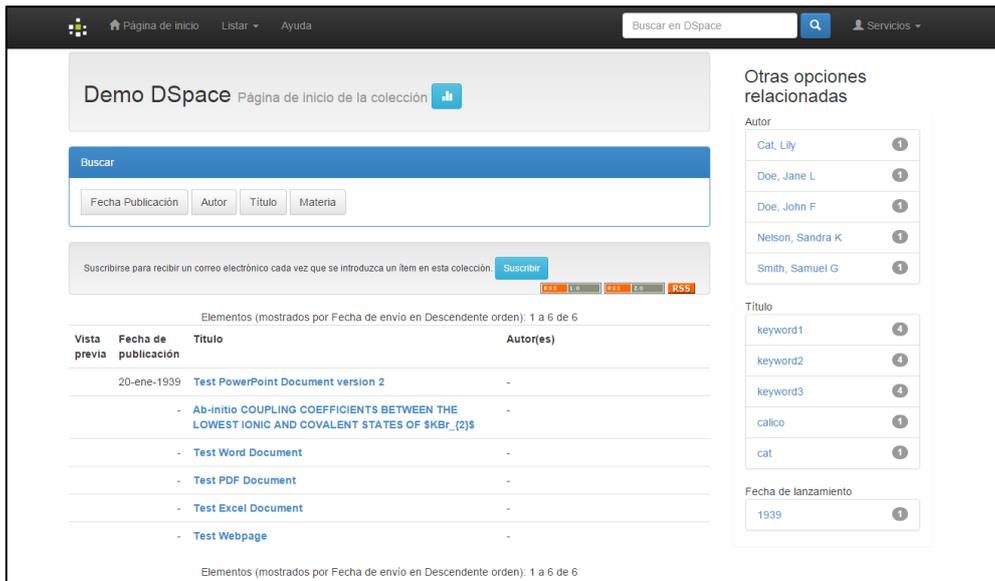


Figura 32 - Página de despliegue de ítems de una colección en versión por defecto

Para esta versión no se realizaron grandes cambios en el despliegue de ítems ni se agregaron complementos para la visualización, esto se planifico para ser realizado en la versión final.

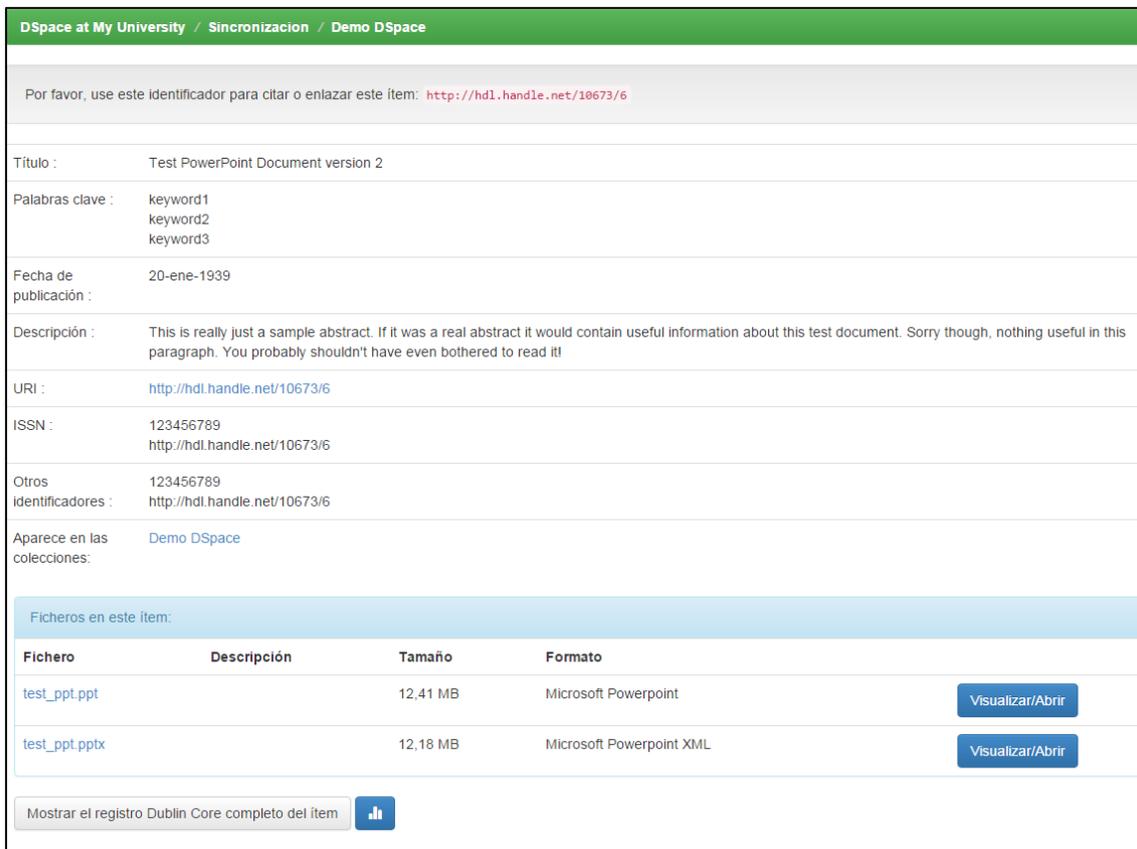


Figura 33 - Despliegue de un ítem en versión por defecto

Marcha



"MARCHA surge a la vida en momentos extraordinariamente graves. El fascismo, asesino y bestial, se ciernen sobre toda la humanidad y amenaza ahogarla en la noche más espantosa. Vivimos días decisivos. En la medida de nuestras posibilidades, lo decimos sin ambages, MARCHA será una trinchera contra el fascismo. Desde ella, al entrar en el combate, enviamos el cordial saludo a todos los luchadores antifascistas".

Buscar

Suscribirse para recibir un correo electrónico cada vez que se introduzca un ítem en esta colección.



Items



MARCHA, AÑO 1, N. 8

MARCHA, AÑO 1, N. 9

MARCHA NRO 1

Elementos (mostrados por Fecha de envío en Descendente orden): 1 a 3 de 3

Otras opciones relacionadas

Autor

- Baudelaire, Charles (1821-1867) 1
- Franco, Rafael (1896-1973) 1
- Giraudoux, Jean (1882-1944) 1
- Kipling, Rudyard (1865-1936) 1
- Labat, Juan Carlos. 1
- Montherlant, Henry de (1895-1972) 1
- Onetti, Juan Carlos (1909-1994) 1
- Priacel, Stefan. 1

Fecha de lanzamiento

- 1939 4

Figura 34 - Despliegue de ítems de una colección en prototipo

4.11. PROPUESTA FRONTEND BNU

Luego de unos días de prueba con el prototipo anterior surge la propuesta de mejora por parte de la BNU:

- **Paleta de colores**
Tomar la paleta utilizado actualmente por la página web de la biblioteca: fondo blanco, letras negras en tipografía Verdana, menú en color celeste.
- **Portada**
El nombre de la página sería “Biblioteca Nacional - Colecciones digitales” (BNU-CD).

En términos generales el cambio propuesto con respecto a la instalación de prueba es enviar el menú hacia el tope de la página y dejar el área de título abajo. Se propone además poblar el menú de más opciones e integrar el cuadro de búsqueda hacia la izquierda el menú.

El menú se posicionaría en la parte superior de la página contra el marco de la ventana

Menú:

Alineados a la izquierda:

- Botón para volver a la web de la Biblioteca, se mostraría el logo de la BNU
- Botón Inicio para regresar a la página principal de BNU-CD
- Botón Búsqueda para desplegar la página de búsqueda avanzada

Alineados a la derecha:

- Un formulario de búsqueda compuesto de:
 - Campo Búsqueda
 - Botón de Búsqueda, con un icono de lupa tal como viene por defecto en DSpace
 - Combo al lado del botón buscar para buscar dentro de la colección que se esté
 - Botón Somos que abre a un menú formado por las opciones que indicará cada organización
 - Contacto (formulario de contacto)
 - Información - Acerca de información sobre BNU-CD, cuyo contenido está a definir.
- Botón de inicio de sesión en la extrema derecha, mostrará solamente un ícono de usuario anónimo, la idea es hacerlo lo menos visible que se pueda.

Logo y Descripción:

Queda como está en la instalación de prueba, pero debajo del menú, que va al top de la página. El logo se mueve a la derecha

- Logo de la BNU a la derecha
- Nombre y descripción del sitio a la izquierda

Cuerpo de la página:

Queda igual que en la instalación de prueba, pero se agrega en la parte inferior, al lado de la colección recomendada, un contenedor para novedades del sitio.

Parte central: Todas las comunidades, cada una identificada por una imagen, tal como en el prototipo de prueba.

Parte inferior

Dos cuadros:

- Uno para la Colección recomendada (tal como está en el prototipo de prueba)
- Otro para novedades

Pie de página

Irían datos de dirección y teléfono de la BNU, posiblemente iría el icono del MEC y la FIC

- **Subcomunidades y colecciones**

La principal solicitud de cambio es eliminar las tablas de Otras opciones relacionadas.

- Título y descripción a la izquierda
- Logo a la derecha
- Los iconos correspondientes al contenido se deberían mostrar a pantalla completa en una matriz de 4 columnas y 3 filas

- **Ítems**

Visualizador del objeto en la parte superior y central de la página ocupando todo el cuerpo de la misma en ancho.

A la izquierda:

Despliegue de los metadatos básicos. Serían los que despliega Trapalanda, agregando los campos Table of contents y hasPart. Además sería deseable los campos sólo si tienen algún valor.

A la derecha:

Cuadro para la descarga del Archivo y Licencia

- **Formulario de búsqueda**

El formulario de búsqueda ocupa mucho espacio en la página por lo que el despliegue de los resultados queda demasiado abajo. Se propone:

- Ocultar los cuadros de resultados en tanto no haya búsqueda.
- Colapsar las opciones de filtro
- Enviar el paginado de resultados al fondo

Despliegue del resultado de la búsqueda:

La solicitud sería marcar gráficamente en el despliegue alguna clase de diferencia entre los resultados que son ítems, comunidades y colecciones.

Para eso se propone visualizar las colecciones y comunidades mediante el logo que las identifica alineados horizontalmente, en tanto que los ítems aparecerán tal como en el prototipo inicial. En el caso de colecciones incluiría en el tope el nombre de la subcomunidad a la que pertenece.

4.12. PROPUESTA FRONTEND FIC

La propuesta de cambios de la FIC luego de testear el prototipo son las siguientes:

- **Paleta de colores**

La paleta de colores definida es la siguiente



Figura 35 - Paleta de colores FIC

- **Tipografía**

- Rúbrica Anáforas: Dominican
- Cuerpo DSpace: Times New Roman

- **Portada**

El nombre de la página sería “Anáforas”.

En la parte superior izquierda se localiza el logo y rúbrica de Anáforas y a la derecha el logo de la FIC y de la universidad de la República, todo sobre fondo blanco.

Menú

El menú queda como en el prototipo.

Alineados a la izquierda:

- Inicio conduce al index de Anáforas

Alineado a la derecha:

- Proyecto
- Equipo
- Contacto
- Botón de inicio de sesión (solo el icono)

Descripción de la página y barra de búsqueda

Queda como está en el prototipo de prueba

- Descripción del sitio a la izquierda
- Barra de búsqueda a la derecha

Cuerpo de la página

Se mantiene el contenedor de últimas colecciones o noticias del prototipo.

En caso de ser viable agregar un contenedor de nuevos ingresos, donde automáticamente figure el nuevo material.

- **Subcomunidades y colecciones**
Dejar como en el prototipo
- **Ítems**
Como en el prototipo, fundamental que se generen los iconos automáticamente de estos.
- **Búsquedas**
Tener un índice cronológico y alfabético de similares características al que cuentan en la actualidad.

Índice general alfabético: **A B C D E F G H I J L M N O P R S T U V Z**

Índice cronológico de la primera edición: **1800-1849 1850-1899 1900-1949 1950-1999**

Figura 36 - Índice actual de Periódicas

CAPÍTULO 5

5. IMPLEMENTACIÓN

5.1. METADATOS

Para la implementación de los metadatos en la plataforma se trabajó conjuntamente con la BNU y FIC, quienes proporcionaron para cada colección los metadatos correspondientes.

Brindaremos un detalle del formato de metadatos utilizado y la incorporación de estos en DSpace.

Formato de los metadatos

Los campos de metadatos se caracterizan por tener tres atributos:

- Elemento, que define al metadato.
- Cualificador, que añade información al elemento.
- Nota de alcance o descripción, que únicamente sirve para explicar su función.

Para todo metadato el elemento siempre deberá estar definido. El cualificado no tiene por qué estarlo, ya que no siempre es necesario. De esta forma, se hará referencia al campo de la siguiente forma:

- Sin cualificador: <nombre del esquema>.<elemento> Ej: dc.type
- Con cualificador: <nombre del esquema>.<elemento>.<cualificador>
Ej: dc.date.created

En este caso un metadato en formato dublin core es de la siguiente forma

```
<dcvalue element="description" qualifier="note">
Florencio Sánchez periodista en Minas y Mercedes: colaboraciones en La Voz del
Pueblo: Los Criik -- Los soplados -- Alrededor de un garrotazo -- Pepito -- A
un... escritor -- Actualidad -- Una tertulia -- Adivina adivinador.
</dcvalue>
```

Este es un ejemplo de un metadato en esquema Dublin Core que es el que se utiliza en la plataforma DSpace, se observa que el elemento se define como description y el cualificador es note, este metadato agrega más información al elemento en cuestión.

5.1.1. METADATOS EN ÍTEMS DE COLECCIONES

Cuando se definieron qué tipo de metadatos describirían los ítems para cada colección se llevó a cabo la modificación del archivo en el cual se configura que metadato describe a cada ítem en determinada colección.

Para modificar estos metadatos es necesario modificar el archivo [dspace]/config/input-forms.xml, siendo [dspace] el directorio de instalación de DSpace, en el cual se define todos los campos necesarios y sus tipos, bajo etiquetas <field> de xml.

Aquí se construyen los forms que identifican a cada colección y sus respectivos metadatos asociados a los ítems en cuestión.

Para cada colección en donde queremos introducir metadatos específicos debemos declarar un formulario con el siguiente esquema:

```
<form-map>
  <name-map collection-handle="123456789/767" form-name="revistas" />
</form-map>
```

Dentro del tag se especifica que para la colección con determinado Id el formulario que describirá a la misma se identifica con el nombre revistas.

He aquí una descripción de los parámetros que intervienen.

- **collection - handle:** Indica el identificador de la colección en la plataforma.
- **form-name:** El nombre del formulario de envío que se identifica más adelante.

Una vez que se define al formulario que describe un ítem de determinada colección pasamos a la creación del mismo. Para ello se encierra en los tags <form name="revistas"></form>, dentro de estos tags se definen los metadatos.

Aquí se presenta un ejemplo de cómo queda estructurado

```
<form name="revistas">
  <page number="1">
    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>title</dc-element>
      <dc-qualifier></dc-qualifier>
      <repeatable>>false</repeatable>
      <label>Titulo</label>
      <input-type>onebox</input-type>
      <hint>Número de volumen o año de la publicación imagen.</hint>
      <required>Debe introducir un titulo.</required>
    </field>
  </page>
</form>
```

Para indicar el metadato se debe encerrar entre los tags<field></field>, a continuación se detalla que describe cada campo dentro del form.

- **dc-schema-** Indica en qué esquema se encuentra el metadato, en este caso dc significa Dublin Core.
- **dc-element** - Indica el elemento del metadato
- **dc-qualifier** - Si el metadato en cuestión tiene un cualificador se describe aquí sino caso contrario se deja en blanco el campo.

- **repeatable** - Si el metadato es repetible o no aquí solo hay dos posibles descripciones true o false.
- **label** - La etiqueta que describe al metadato.
- **input-type** - Esta etiqueta define el tipo de campo los cuales pueden ser de los siguientes tipos:
 - **date**: Para campos de tipo fecha.
 - **name**: Aquí se indican dos cuadros de texto, una para el nombre y otra para el apellido.
 - **onebox**: Un solo cuadro de texto.
 - **twobox**: Dos cuadros de texto en una misma línea.
 - **dropdown**: Combo desplegable. Para este tipo de campos la etiqueta debe ser de la siguiente forma:

```
<input-type value-pairs-name="nombre del combo"> dropdown
</input-type>
```
- **hint** - Un mensaje informativo para el usuario
- **required** - Si el metadato es obligatorio o no, en caso de no serlo se deja el campo vacío.

Por lo tanto para cada colección, se introdujo un <form> dentro del archivo input-form.xml con las características descritas anteriormente detallando los metadatos que nos proporcionaron las instituciones. De esta forma para cada ítem que se ingrese en las colecciones tendrá asociado los metadatos específicos. Para el caso de modificar metadatos de colecciones difiere bastante de lo mencionado en este apartado.

5.1.2. METADATOS EN COLECCIONES

En este apartado veremos cómo se realizó la implementación de metadatos a nivel de colecciones. Uno de los requisitos que nos proporcionaron las instituciones fue la modificación de metadatos a nivel de colecciones para ello se realizó cambios a nivel de código.

Dado que DSpace es un tanto rígido a la hora de realizar cambios en los metadatos a nivel de colección y comunidad realizó un estudio acerca de cómo sería posible modificar los mismos para satisfacer dichos requerimientos.

Para ingresar nuevos metadatos en las colecciones se realizó lo siguiente:

Primero modificamos la interfaz gráfica en la cual se editan los metadatos de colección, la misma se encuentra en:

[dspace]/dspace/dspace-jsui/src/main/webapp/tools/edit-collection.jsp

Allí se agregan los metadatos “nuevos”, se declara un campo para cada uno de ellos agregando el código a continuación:

```

<div class="row">
  <label class="col-md-3"for="name">Tema</label>
  <span class="col-md-9">
    <input class="form-control" type="text" name="tema" value="tema" />
  </span>
</div><br/>

```

Luego se declara una variable asociada a este nuevo metadato a agregar y se le asocia lo siguiente:

```
tema = collection.getMetadata("tema");
```

La función `getMetadata` se encuentra definida en el objeto `Collection.java` contiene la definición del metadato asociado a dicho campo, la implementación de la misma presenta a continuación

```

public String getMetadata (String field) {
    String[] MDValue = getMDValueByLegacyField(field);
    String value = getMetadataFirstValue(MDValue[0], MDValue[1], MDValue[2],
Item.ANY);
    return value == null ? "" : value;
}

```

En este método se retorna el valor del metadato agregado.

Se utilizan `getMDValueByLegacyField` y `getMetadataFirstValue` definidas en el `DspaceObject.java` las cuales más adelante especificaremos por su importancia en la plataforma.

Para realizar el nuevo seteo del metadato se debe recurrir a la clase `DspaceObject.java` la cual encapsula las propiedades de los metadatos, una de las funciones principales que realiza esto se denomina `getMDValueByLegacyField (field)` quien de acuerdo al valor ingresado retorna un array con el valor del metadato asociado, en este caso tenemos la siguiente implementación:

```

protected String[] getMDValueByLegacyField (String field) {
    switch (field) {
        case "introductory_text":
            return new String[]{MetadataSchema.DC_SCHEMA, "description",
null};
        case "short_description":
            return new String[]{MetadataSchema.DC_SCHEMA, "description",
"abstract"};
        case "tema":
            return new String[]{MetadataSchema.DC_SCHEMA, "subject", null};
    }
}

```

Se observa que según el campo `field` de metadatos se obtiene el array compuesto por el esquema, tipo y cualificador del metadato, tal es así el ejemplo de `tema`.

Otro método de gran importancia es `getMetadataFirstValue`, quien se encarga del seteo del valor del metadato he aquí su implementación

```
protected String getMetadataFirstValue (String schema, String element, String
qualifier, String language){
    Metadatum[] dcvalues = getMetadata(schema, element, qualifier, Item.ANY);
    if(dcvalues.length>0){
        return dcvalues[0].value;
    }
    return null;
}
```

Esta se encuentra en la clase DSpaceObject.java y recibe como parámetro el esquema del metadato, el elemento, el cualificador y el lenguaje de este. Luego invoca a la función getMetadata quien también se encuentra implementada en la misma clase. Se encarga de obtener el valor “real” del metadato al cual solicita.

```
public Metadatum[] getMetadata(String schema, String element, String
qualifier,
                                String lang) {
    List<Metadatum> values = new ArrayList<Metadatum>();
    for (Metadatum dcv : getMetadata())
    {
        if (match(schema, element, qualifier, lang, dcv))
        {
            Metadatum copy = new Metadatum();
            copy.element = dcv.element;
            copy.qualifier = dcv.qualifier;
            copy.value = dcv.value;
            copy.language = dcv.language;
            copy.schema = dcv.schema;
            copy.authority = dcv.authority;
            copy.confidence = dcv.confidence;
            values.add(copy);
        }
    }
}
```

La lógica de este método realiza lo siguiente, se invoca por ejemplo como getMetadata(“dc”, “subject”, null, es) donde se indica el esquema de metadata en este caso Dublin Core luego el elemento del metadato que para tema es subject y el cualificador si es que lo tiene seguido del lenguaje en el cual se encuentra. Si el metadato indicado se encuentra en la lista de los metadatos disponibles que maneja la plataforma entonces arma una copia del mismo obteniendo los valores necesarios de este.

Luego retorna este quién será utilizado para el valor de la interfaz gráfica en el formulario de esta forma se asocia un metadato nuevo a una colección, este fue el ejemplo para asociar el metadato dc.subject al campo tema. Lo mismo se realiza si se quieren agregar otros metadatos pero con diferentes valores.

Tanto la clase Collection.java como DSpaceObject.java se encuentran en:
[dspace]/dspace-api/src/main/java/org/dspace/content

5.2. SINCRONIZACIÓN

En el proceso que lleva a cabo la sincronización intervienen dos actores, el proveedor de datos quien en este caso es la FIC y el cosechador BNU.

Para que la sincronización sea posible se realizaron ciertas configuraciones tanto del lado del proveedor de datos como del cosechador.

El mecanismo de sincronización viene incorporado como un módulo externo a DSpace, el cual es inicializado desde el servidor Tomcat.

Si las configuraciones fueron realizadas de forma exitosa se puede corroborar que accediendo desde la URL [dspace]/oai se desplegará una interfaz del estilo de la siguiente figura:

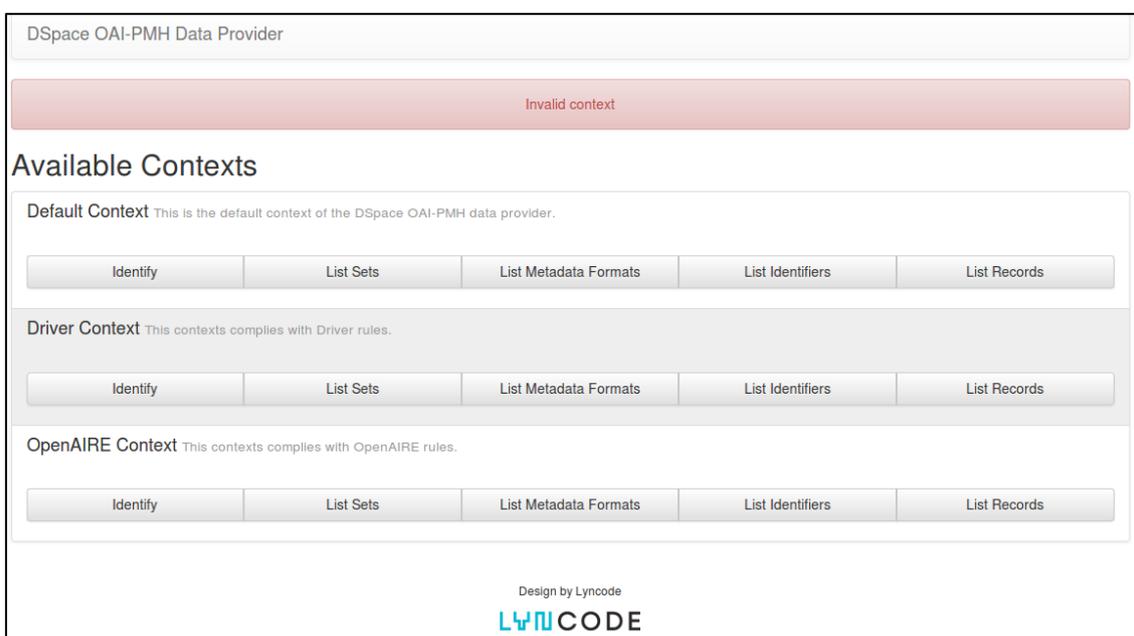


Figura 37 - Pagina de OAI desde DSpace

Si visualizamos esta interfaz se ha logrado exitosamente tener disponible el servicio de sincronización.

El siguiente paso a realizar es publicar los objetos en dicha interfaz para que puedan ser cosechados. Sin la realización de esto ningún cosechador tendría acceso a la información que se quiera recolectar, para ello se debe ejecutar el comando:

```
sudo [dspace]/bin/dspace oai import
```

Esto lo que hace es publicar en la interfaz todos los objetos digitales contenidos en la biblioteca digital para luego ser cosechados por otros repositorios o bibliotecas, en nuestro caso para ser cosechados por la BNU.

Cada vez que se ingresan nuevos ítems en la biblioteca digital, se debe ejecutar dicho comando a modo de que los objetos publicados queden actualizados y los que aún no, lo estén.

Una vez configurado el proveedor de datos (FIC) especificaremos que se necesita establecer del lado del cosechador (BNU) para realizar una cosecha exitosa.

Cuando se desea realizar una sincronización de una colección con otra fuente externa se debe especificar del lado del cosechador que esta colección será alimentada por el repositorio en cuestión. Como detalla en la imagen, la configuración se realiza al editar la colección y especificar los siguientes parámetros:

- **Content Source** - En este parámetro se debe indicar que la colección será cosechada desde una fuente externa como se muestra en la imagen.
- **OAI Provider** - La dirección en la cual será realizada la cosecha, en este caso se indica [URL del sitio]/oai/response.
- **OAI SET ID** - Indica el identificador de la colección a cosechar, el mismo se obtiene desde el proveedor de datos.
- **Metadata Format** - El esquema de metadatos en la cual realizamos la sincronización, el más utilizado es el Dublin Core.
- **Content Being Harvest** - Esto indica la forma en la cual queremos obtener los datos, hay tres tipos, el primero hace referencia a la obtención solamente de los metadatos del objeto, la segunda a los metadatos y los links referentes a los bitstreams de los objetos, por último la obtención de metadatos y bitstreams de los objetos.

En nuestro caso se opta por la opción tres, dado que a la BNU le es de interés tener los objetos digitales en su repositorio.

Figura 38 - Ingreso de colección a cosechar

Adicionalmente para las colecciones que quieran ser sincronizadas el cosechador deberá ejecutar el comando:

```
sudo [dspace]/bin/dspace Harvest -r -e email@email.com -c "123456789/6"
```

Se observa que se cosecha la colección de identificación "123456789/6", la bandera -c indica la cosecha de esa colección, la bandera -r que realice la cosecha y la bandera -e indica la e-persona que está autorizada a realizar esta. Esto último puede ser opcional pero para brindarle más nivel de seguridad siempre es bueno tener a cargo una persona autorizante de realizar dichas operaciones sobre las colecciones.

5.3. COMPLEMENTOS

En esta sección se detallan los complementos y las modificaciones realizadas para el correcto funcionamiento de los mismos.

5.3.1. COMPLEMENTOS PARA IMÁGENES:

Para las imágenes se utilizaron dos complementos, uno para visualizar la imagen dentro del navegador sin necesidad de salir de la página actual ni descargar (lightbox), y otro para realizar efecto zoom sobre la imagen, sea el caso de imágenes de gran tamaño como ser los mapas, donde se necesita ver en detalle el mismo y gracias al tamaño que posee no es viable.

En primer lugar se testeó con complementos de visualización de imágenes, como la galería de imágenes Prettyphoto [55]. Para el efecto zoom se utilizó el complemento Cloud-zoom [56], pero ambos complementos en conjunto no funcionaron

correctamente, por lo cual era necesario realizar modificaciones a ambos complementos.

Sugerido por los desarrolladores de Cloud-zoom y luego de una investigación previa se optó por utilizar Fancybox [57], siendo uno de los más utilizados y recomendados para el caso.

En el caso de los mapas, se tiene un tamaño aproximado de 25 MB, este permite visualizarlo en una escala más pequeña, ajustable al tamaño de la pantalla mediante la posición del mouse sobre la imagen. Esto produce el efecto de poder observar la imagen en forma de detalle, sin perder calidad de la misma.

Para la incorporación de complementos en DSpace, se agregaron las librerías Javascript a la ruta: [dspace]/dspace-jspui/src/main/webapp/static/js y las hojas de estilo a la ruta: [dspace]/dspace-jspui/src/main/webapp/static/css. Se agregan las referencias a estos en la página [dspace]/dspace-jspui/src/main/webapp/layout/header-default.jsp.

En el archivo: [dspace]/dspace-jspui/src/main/webapp/display-item.jsp se inicializa el complemento fancybox donde se le indica que cuando se muestre la imagen, se le agreguen las propiedades de Cloud-zoom para aplicar zoom a la imagen.

Por ultimo en la clase:

[dspace]/dspace-jspui/src/main/java/org/dspace/app/webui/jsptag/ItemTag.java se agregó la imagen a una escala más pequeña, tal que al hacer click sobre la misma se visualice en forma de popup. Para ello se tuvo que agregar lógica adicional en DSpace.

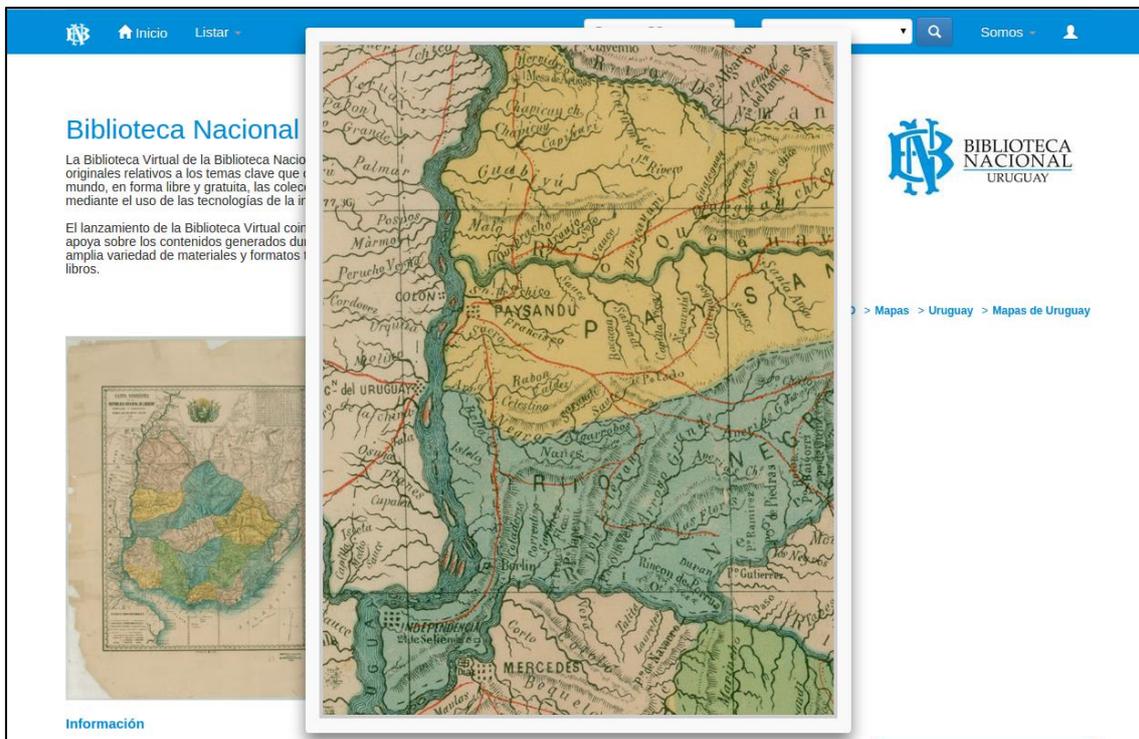


Figura 39 - Complemento Imagen y Zoom

5.3.2. COMPLEMENTO PARA PDF

Para los archivos en formato PDF, como también se manejan archivos de gran porte, fue necesario un complemento que descargara el PDF de forma progresiva, a medida avanzaban las páginas.

Se escogió el complemento PDF.js [58] el cual permite visualizar archivos PDF, embebiendo este en la página del ítem dentro de un contenedor, sin que el usuario tenga que utilizar complementos adicionales o programas externos.

PDF.js es un proyecto de software libre de la fundación Mozilla que permite utilizar un visor de archivos PDF para la web de forma genérica. La ventaja es que utiliza el mismo visor en todos los navegadores y por lo tanto nos aseguramos su correcto funcionamiento en la mayoría de los navegadores actuales.

Para la incorporación de este complemento a DSpace, se creó la carpeta PDFViewer bajo el directorio: [dspace]/dspace-jspui/src/main/webapp/static y se agregaron librerías Javascript y hojas de estilo. Se agregó en el directorio raíz el [dspace]/dspace-jspui/src/main/webapp el archivo viewer.html, visor que utiliza por defecto para los navegadores.

En el archivo [dspace]/dspace-jspui/src/main/webapp/display-item.jsp se inicializa el PDF al cual se le indica cual se cargar, tomando la URL de este definida en el archivo [dspace]/dspace-jspui/src/main/java/org/dspace/app/webui/jsptag/ItemTag.java.

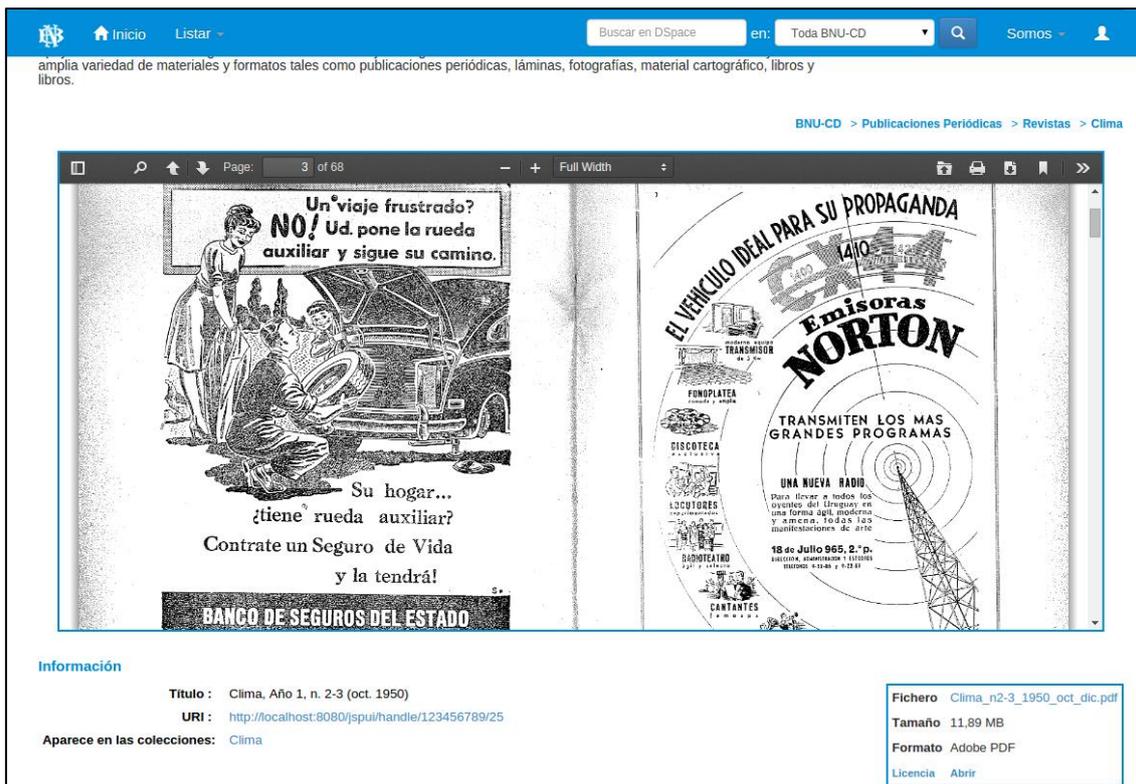


Figura 40 - Complemento PDF

5.3.3. COMPLEMENTO PARA VIDEO

Para los archivos de video se realizó las pruebas de dos complementos, video.js [59] y JWPlayer [60]. Si bien este último es pago, para el uso en la web tiene una versión gratuita.

Ambos están basados en HTML5 y permiten la reproducción de los formatos MP4, WebM y OGG dependiendo del navegador. El formato FLV no es soportado por HTML5 pero estos complementos lo soportan. El formato MPEG no tiene licencia libre por lo que no será posible reproducir este en video. Así como tampoco será soportado el formato de video AVI, propiedad de Microsoft, ya que no está soportado por HTML5 en ninguno de los navegadores, ni siquiera en Internet Explorer de Microsoft.

Navegador	WebM	Ogg	MP4
Firefox	Sí	Sí	No
Safari	No	No	Sí
Opera	Sí	Sí	No
Chrome	Sí	Sí	Sí
IE9	Parcial (requiere del componente WebM MF)	No	Sí

Tabla 7- Soporte de video HTML5 de los navegadores [61]

Luego de algunas pruebas realizadas e incorporar ambos complementos, se decidió utilizar el video.js por ser 100% código abierto. A lo largo de las pruebas que se

realizaron se encontró que el complemento JWPlayer en reiteradas ocasiones no se reproducía el video por lo que se debía refrescar la página, por tal motivo fue descartado la utilización del mismo.

Para la incorporación a DSpace, fue necesario agregar las librerías Javascript a la ruta [dspace]/dspace-jspui/src/main/webapp/static/js así como las hojas de estilo a la ruta [dspace]/dspace-jspui/src/main/webapp/static/css.

Luego en la clase:

[dspace]/dspace-jspui/src/main/java/org/dspace/app/webui/jsptag/ItemTag.java se agregó la lógica para embeber el video en la página.

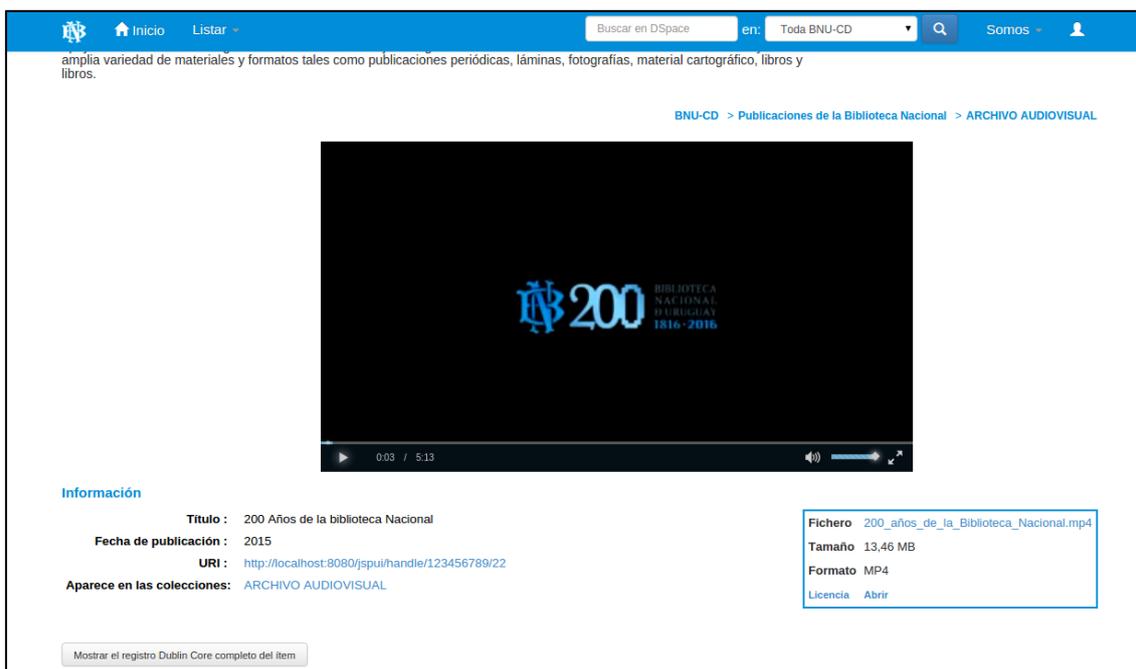


Figura 41- Complemento Video

Es necesario registrar los formatos AVI y FLV en el menú administrador opción "Registro de bitstreams", debido a que por defecto no los trae agregados. Los valores que deben ingresarse son los siguientes:

Nombre	Extensión	MIME Type
FLV	flv	video/x-flv
MP4	mp4	video/mp4

Tabla 8 - Formatos adicionales de video

5.3.4. COMPLEMENTO PARA AUDIO

Para los archivos de audio se probaron dos complementos, el SoundJS [62] y el JWPlayer mencionado en el apartado anterior cuya funcionalidad también incluye audio. Ambos están basados en el audio de HTML5. Luego de algunas pruebas, se optó por el JWPlayer. A pesar que algunos casos para video no funcionaba correctamente, sí lo hizo para archivos de audio.

Los archivos soportados por estos complementos son MP3, WAV y OGG en la tabla se detallan para qué navegador se encuentran soportados.

Navegador	MP3	Wav	Ogg
Firefox	Sí	Sí	Sí
Safari	Sí	Sí	No
Opera	No	Sí	Sí
Chrome	Sí	Sí	Sí
IE9	Sí	No	No

Tabla 9 - Soporte de audio HTML5 de los navegadores [61]

Para la incorporación de este complemento, se incorporaron las librerías Javascript a la ruta [dspace]/dspace-jspui/src/main/webapp/static/js y las hojas de estilo a la ruta [dspace]/dspace-jspui/src/main/webapp/static/css.

Por último se agregó en la clase:

[dspace]/dspace-jspui/src/main/java/org/dspace/app/webui/jsptag/ItemTag.java
la lógica para embeber el audio en la página.

The screenshot shows the 'Biblioteca Nacional - Colecciones digitales' page. At the top, there is a navigation bar with 'Inicio', 'Listar', and a search box. The main content area features the library's logo and a description of the digital collections. Below the description, there is a breadcrumb trail: 'BNU-CD > Publicaciones de la Biblioteca Nacional > Sonidos'. A central audio player is visible, showing a progress bar from 00:04 to 02:20. Below the player, the 'Información' section provides details for the selected item:

- Título:** Orquesta Sinfónica Metropolitana
- URI:** http://localhost:8080/jspui/handle/123456789/21
- Aparece en las colecciones:** Sonidos
- Fichero:** ORQUESTA_SINFÓNICA_METROPOLITANA_Montevideo_Uruguay_1.mp3
- Tamaño:** 5,47 MB
- Formato:** MP3
- Licencia:** Abrir

At the bottom left, there is a button labeled 'Mostrar el registro Dublin Core completo del ítem'.

Figura 42 - Complemento de audio

También es necesario el formato MP3 en el registro de bitstreams, debido a que por defecto no se encuentran agregados, el valor que debe ingresarse es el siguiente:

Nombre	Extension	MIME Type
MP3	mp3	audio/mpeg

Tabla 10 - Formatos adicionales de audio

5.3.5. COMPLEMENTO CARRUSEL

Para mostrar los últimos ítems agregados, se investigaron los complementos más utilizados que permitieran mostrar varias imágenes a la vez. Se optó por el complemento de carrusel jquery FlexSlider [63], el mismo permite modificar la velocidad y el tiempo de transición.

Dado que el carrusel que trae DSpace por defecto no muestra las miniaturas de los ítems, se agregó lógica al código para cumplir dicho objetivo.

Para la incorporación del complemento, se agregaron las librerías Javascript a la ruta [dspace]/dspace-jspui/src/main/webapp/static/js y las hojas de estilo a la ruta [dspace]/dspace-jspui/src/main/webapp/static/css.

Luego en la página [dspace]/dspace-jspui/src/main/webapp/home.jsp se corrigió la lógica existente de los últimos ítems ingresados, agregando las imágenes miniaturas de los ítems al carrusel.



Figura 43 - Carousel para últimos ítems agregados

5.3.6. PERSONALIZACIÓN DE LA INTERFAZ

Para hacer que ambas bibliotecas digitales sean amigables y usables para el usuario lo principal a considerar es la interfaz gráfica. Por tal motivo se realizaron varias modificaciones a la interfaz por defecto, de manera de dejarla lo más original y llamativa posible, cumpliendo con los comentarios realizados por ambas instituciones del prototipo mencionado anteriormente. Se detallan en el apartado siguiente los cambios realizados, separados por bibliotecas.

5.3.7. PERSONALIZACIÓN DE LA INTERFAZ BNU

- **Estilo e Imágenes**

Se incorporó una hoja de estilo para diferenciar los estilos que trae por defecto DSpace con los particulares impuestos por BNU. Fue agregada bajo [dspace]/dspace-jspui/src/main/webapp/static/css/bootstrap con el nombre de BNUcss.css

Para las imágenes estáticas como son los logos se creó el directorio BNU en [dspace]/dspace-jspui/src/main/webapp/image

- **Página principal**

Esta es una de las páginas en la cual se realizaron gran parte de los cambios asignados.

Menú

Por pedido de la BNU menú en la parte superior como lo trae la versión por defecto, pero se cambió el ancho y agregaron las opciones mencionadas en la propuesta de la BNU.

Para customizar la barra de menú se modificaron los siguientes archivos:

- Este archivo tiene la barra de menú, la cual acompaña al usuario no logueado durante toda la navegación [dspace]/dspace-jspui/src/main/webapp/layout/navbar-default.jsp
- Para usuarios logueados, la barra de menú se detalla en el siguiente archivo [dspace]/dspace-jspui/src/main/webapp/layout/navbar-admin.jsp
- Cuando un usuario logueado está agregando un ítem se despliega la barra de menú en la siguiente página [dspace]/dspace-jspui/src/main/webapp/layout/navbar-minimal.jsp

Además se modificaron las siguientes páginas correspondientes al cabezal de la aplicación:

- [dspace]/dspace-jspui/src/main/webapp/layout/header-default.jsp

- [dspace]/dspace-jspui/src/main/webapp/layout/header-submission.jsp
- [dspace]/dspace-jspui/src/main/webapp/layout/header-default.jsp

Barra de navegación

Esta barra es la que indica la ruta completa de la comunidad, subcomunidad o colección que estamos visualizando. Al hacer click sobre cualquiera de los niveles se redirige a la página de la comunidad, subcomunidad o colección padre seleccionada.

Se realizan los cambios en la página:

[dspace]/dspace-jspui/src/main/webapp/layout/navbar-default.jsp



Figura 44 - Barra de navegación

Título, Logo y Descripción

La descripción se configura en el archivo de idioma Message_es.properties con el nombre de propiedad `jsp.layout.header-default.brand.description`. Se le proporcionó el estilo y los colores del título y la descripción.

Cuerpo de la página

Para la customización del cuerpo de la página se realizaron los cambios en [dspace]/dspace-jspui/src/main/webapp/home.jsp.

Allí se modificó la visualización de las comunidades y se agregó el carrusel de imágenes con los últimos ítems incorporados. Además se agregó la sección customizable de colecciones recomendadas, noticias o la información que se desea desplegar. Esta última se puede modificar desde el administrador o desde la página ubicada en [dspace]/config/news-side_es.html y debe realizarlo una persona con algo de conocimientos en HTML, debido a que la estructura y el estilo se define en esta misma página con elementos HTML y CSS. De todas formas dejamos un template que pueden utilizar como base.

Pie de página

Se utilizó el mismo color de fondo que la barra de menú, se agregaron los logos pedidos de MEC y FIC así como la dirección y teléfonos de la BNU como fue pedido.

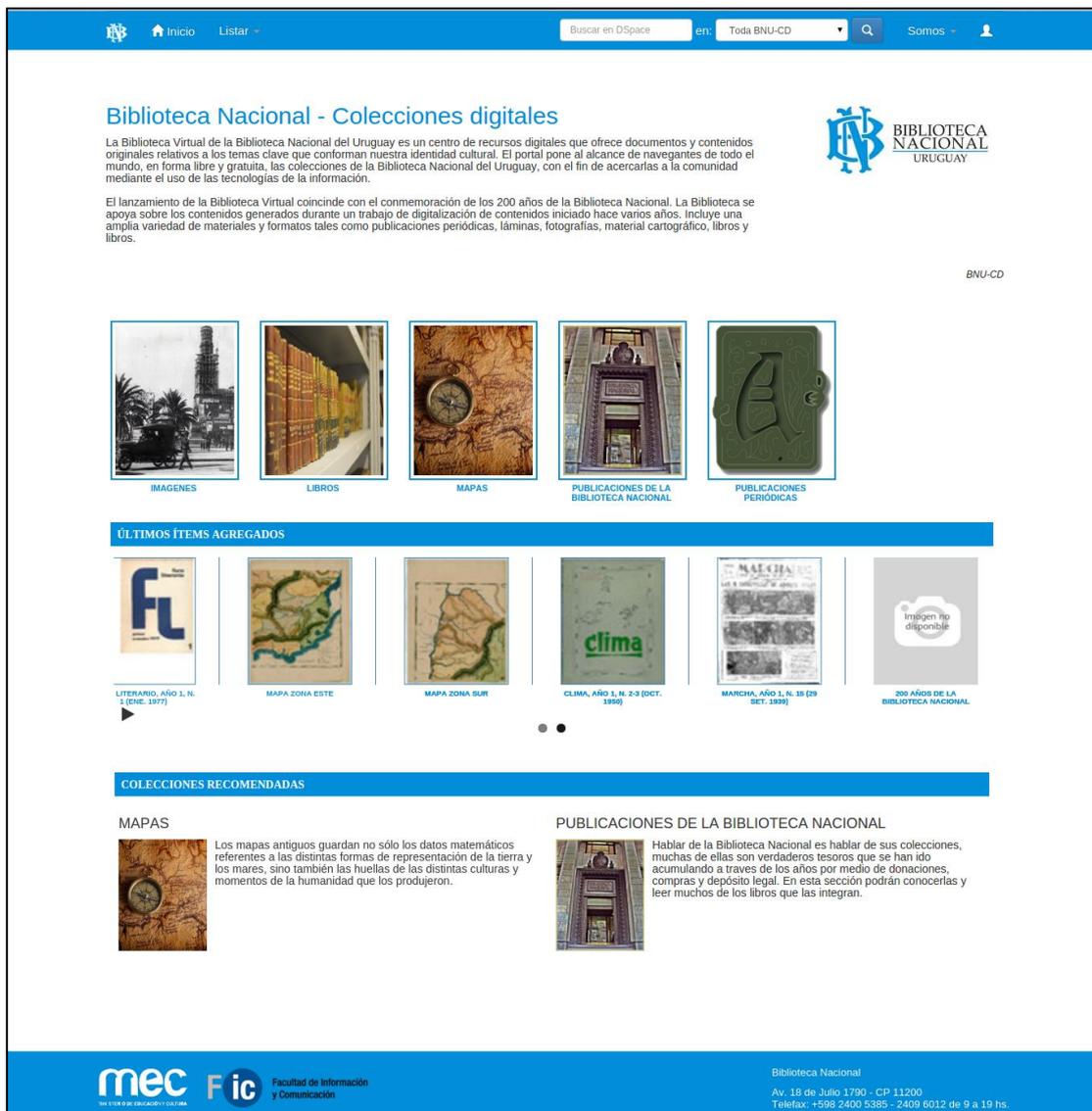


Figura 45 - Portada de la biblioteca digital de la BNU

- **Subcomunidades y colecciones**

Los cambios realizados a nivel de subcomunidad y colecciones son los mencionados en la solución del prototipo pero con las siguientes modificaciones:

Se elimina la tabla de Otras opciones relacionadas, el Título y la descripción a la izquierda y el logo a la derecha.

El pedido de mostrar la subcomunidades o colecciones en una matriz de 4 columnas y 3 filas no se realizó debido al espacio libre innecesario entre las imágenes.

Los cambios se realizaron en las páginas:

[dspace]/dspace-jspui/src/main/webapp/collection-home.jsp

y [dspace]/dspace-jspui/src/main/webapp/community-home.jsp

- **Ítems**

Se continuó las modificaciones del prototipo inicial, en el despliegue de los ítems de una colección. Se elimina la tabla de Otras opciones relacionadas, el Título y la descripción, se colocaron a la izquierda y el logo a la derecha.

Se realizaron modificaciones en la siguiente clase:

```
[dspace]/dspace-  
jspui/src/main/java/org/dspace/app/webui/jsptag/BrowseListTag.java
```

A nivel de ítem, se agregaron los complementos ya mencionados y se realizaron los cambios pedidos para desplegar los metadatos e información. Se agrega un link para la descarga de la licencia.

Para modificar la página donde se despliega el ítem se modificaron los archivos:
[dspace]/dspace-jspui/src/main/java/org/dspace/app/webui/jsptag/ItemTag.java y
[dspace]/dspace-jspui/src/main/webapp/display-item.jsp

- **Búsqueda**

Para ocultar la parte de filtros por pedido de la BNU se actualizó la página:

```
[dspace]/dspace-jspui/src/main/webapp/search/discovery.jsp
```

Para modificar cómo se despliegan las comunidades, subcomunidades y colecciones, se modificó el archivo:

```
[dspace]/dspace-  
jspui/src/main/java/org/dspace/app/webui/jsptag/CommunityListTag.java
```

También se modificó la página:

```
[dspace]/dspace-jspui/src/main/webapp/browse/full.jsp
```

- **Ingreso de un ítem (Usuario logueado)**

En el ingreso de un ítem, se debe ingresar los metadatos correspondientes a la colección, definidos en el archivo input-form.xml previamente mencionado. En la página de ingreso se realizaron algunos cambios, más que nada del tamaño que ocupan los campos.

Estos cambios se realizaron en:

```
[dspace]/dspace-jspui/src/main/webapp/submit/edit-metadata.jsp
```

- **Edición de metadatos**

Por pedido de ambas instituciones, se eliminó información no relevante de esta página de manera que sea más sencilla y limpia la edición de los metadatos de un ítem. Se posicionó en la parte superior de la tabla el agregado de metadatos, originalmente se posicionaba en la parte inferior.

Se modificó el archivo [dspace]/dspace-jspui/src/main/webapp/tools/edit-item-form.jsp

Elemento	Cualificador Valor	Lenguajidioma
	dc.bibliographicCitation	es
creator	González González, Adolfo.	es
creator	Guidi, Alfredo (1930-)	es
creator	Viera, Nelson A.	es
date	accessioned 2015-11-04T20:25:52Z	es
date	available 2015-11-04T20:25:52Z	es
date	issued 1978-07-01	es
identifier	uri http://localhost:8080/jspui/handle/123456789/20	es
description	provenance Submitted by Florencia Borghini (florborghini@gmail.com) on 2015-11-04T20:25:52Z	en
description	provenance Made available in DSpace on 2015-11-04T20:25:52Z (GMT). No. of bitstreams: 1	en
language	iso es	es
title	Foro literario, Año 2, n. 4 (jul. 1978)	es
type	Publicacion Periodica	es
description	version Version Inicial	es

Figura 46 – Formulario de edición de metadatos para BNU

5.3.8. PERSONALIZACIÓN DE LA INTERFAZ FIC

Las modificaciones con respecto al prototipo inicial para el repositorio de la FIC fueron pocos, se mantuvo prácticamente la estructura completa.

- **Estilo e Imágenes**

Se agregó una hoja de estilo para diferenciar los estilos que trae por defecto DSpace con los particulares de la FIC. La misma fue agregada bajo [dspace]/dspace-jspui/src/main/webapp/static/css/bootstrap con el nombre de fic-theme.css

Para las imágenes estáticas como los logos, se creó la carpeta con nombre FIC bajo el directorio [dspace]/dspace-jspui/src/main/webapp/image

- **Página principal**

Esta página al igual que la Biblioteca Nacional, es la que más cambios ha tenido, detallaremos los más relevantes.

Cabecal y Menú

Se mantuvo el cabezal original del prototipo, se agregó el logo de Anáforas a la izquierda con el título Anáforas con la rúbrica que nos proporcionó la FIC.

Se incorporaron los logos de la FIC y de la Universidad de la República a la derecha.

El menú se mantuvo como fue propuesto en el prototipo inicial, modificándose las opciones del mismo.

Las páginas modificadas son las mismas mencionadas en la personalización de BNU.

Barra de navegación

La barra de navegación se mantuvo como se propuso en el prototipo original, modificándose como se mencionó en la personalización de la BNU.

Título y Descripción

Se mantuvo la descripción de la página debajo del menú y a la derecha la barra de búsqueda. La descripción es configurada en el archivo de idioma `Message_es.properties` bajo el nombre de propiedad `jsp.layout.header-default.brand.description`. Allí se le dio el estilo y los colores del título junto con la descripción.

Cuerpo de la página

Las modificaciones realizadas al cuerpo son similares a las mencionadas para la BNU, las páginas y clases modificadas fueron las mismas.

Pie de página

Igual al prototipo original pero con datos de la FIC.

Figura 47 - Portada de Anáforas

- **Subcomunidades y colecciones**

Los cambios realizados a nivel de subcomunidad y colecciones son los mencionados en la solución del prototipo, pocas modificaciones con respecto al prototipo inicial.



anáforas




Inicio
Contacto
Equipo
Proyecto
Usuario

Anaforas

PUBLICACIONES PERIÓDICAS



PUBLICACIONES PERIÓDICAS DEL URUGUAY presenta la construcción progresiva de un sitio dedicado a la difusión, en formato digital, de periódicos que se publicaron en el Uruguay desde su constitución como país independiente y, aún antes, en la última etapa de la época colonial y durante los años de incertidumbre y conflictos que le sucedieron. La prensa fue el espacio fundacional de discursos y discusiones, un ámbito propicio para la información de acontecimientos y disquisiciones intelectuales, de posiciones políticas, de aventuras literarias y artísticas, de miradas críticas rigurosas o polémicas, refiriendo tiempos y circunstancias, desde instancias que, dados los recursos tecnológicos actuales y sus incontenibles innovaciones, seguirán prolongándose. Asimismo se podrán leer, en este sitio, las numerosas publicaciones organizadas o sostenidas por exilados e inmigrantes. Acontecimientos decisivos de la historia nacional se vieron reflejados en muchas de ellas (en italiano, francés, portugués e inglés, entre otras lenguas), por lo que son valioso testimonio tanto de la experiencia de una realidad extraña y nueva, la nuestra, como de los intereses y posiciones de los recién llegados ante sucesos locales y extranjeros. Efímeras o duraderas, influyentes o marginales, sus páginas sorprenden por la multiplicidad de sus planteos, muchos de ellos todavía vigentes; sorprenden, algunas veces, por la belleza de su diagramación; otras veces, por la calidad de sus contenidos y los elevados atributos de sus colaboradores, por la gracia y humor de sus ocurrencias, por la virulenta severidad de las diatribas pero también, y en casi todos los casos, por el lamentable desconocimiento general en el que hoy se encuentran. Rescatadas y difundidas, son huellas y reliquias de vicisitudes y visiones que podrán favorecer la consolidación de nuevas reflexiones, de invenciones, de iniciativas de la imaginación que, gracias a las redes y sus tecnologías, habilitan el acceso a un tiempo que pasó y aún nos concierne.

Buscar

Q

Fecha Publicación
Autor
Título
Materia

Otras opciones relacionadas

Autor	Título	Fecha de lanzamiento
Flores Mora, Manuel (1923-1985) 1	Cine - historia 1	1970 - 1976 1
Garce, Arturo (1931) 1	Cuentos ingleses - S. XX 1	1960 - 1969 3
Greene, Graham (1904-1991) 1	Filosofía griega 1	1950 - 1959 2
Otero Mendoza, Jorge. 1	Genética 1	1930 - 1939 3
Payró, Julio E. (1899-1971) 1	Pintura 1	
Platschek, Hans (1923-2000) 1	Poesía uruguaya - S. XX 1	
Somers, Armonía (1914 - 1994) 1		

COLECCIONES



BOLETINES



DIARIOS



OTROS IMPRESOS



PRIMEROS IMPRESOS PERIÓDICOS



REVISTAS



SEMANARIOS



ARTICULOS

Facultad de Información y Comunicación - Instituto de Comunicación | comunicacion@fic.edu.uy

Figura 48 - Subcomunidades y colecciones

- **Ítems**
Se mantuvo como el prototipo inicial, con el agregado de los complementos.

The screenshot displays a digital library interface. At the top, there is a header with the logo 'anáforas' and the 'Fic' logo (Facultad de Información y Comunicación) of the 'UNIVERSIDAD DE LA REPUBLICA URUGUAY'. A navigation bar includes links for 'Página de inicio', 'Contacto', 'Equipo', and 'Proyecto'. Below this, a breadcrumb trail reads 'Anaforas > Publicaciones Periódicas > Revistas > Sintaxis'. The main content area features a viewer window showing the cover of 'revista' for 'OCTUBRE 1976', issue number '3'. Below the viewer, the 'Registro completo de metadatos' section is visible, containing a table of DC fields and their values. To the right of the table is a file information box for a PDF document, including its filename, size (4.29 MB), and format (Adobe PDF), with a 'Visualizar/Abrir' button. At the bottom of the page, there is a footer with the text 'Facultad de Información y Comunicación - Instituto de Comunicación | comunicacion@fic.edu.uy'.

Registro completo de metadatos

Información

Campo DC	Valor	Lengua/Idioma
dc.creator	Garce, Arturo (1931)	es
dc.creator	Somers, Armonía (1914 - 1994)	es
dc.date.accessioned	2015-11-04T00:49:26Z	-
dc.date.available	2015-11-04T00:49:26Z	-
dc.date.issued	1976-10-01	es
dc.identifier.uri	http://localhost:8080/jspui/handle/123456789/35	-
dc.language.iso	es	es
dc.subject	Genética	-
dc.subject	Filosofía griega	-
dc.title	Revista Sintaxis, Año 1, n. 3 (oct. 1976)	es
dc.type	Publicacion Periodica	es
dc.description.version	Version Inicial	es

Aparece en las colecciones: Sintaxis

Mostrar el registro sencillo del ítem

Facultad de Información y Comunicación - Instituto de Comunicación | comunicacion@fic.edu.uy

Figura 49 - Ejemplo de un ítem con complemento PDF -

- **Búsqueda**

Se realizaron cambios con respecto al despliegue de comunidades y colecciones similares para los de la BNU.



Buscar

Buscar: por

Añadir filtros:
Usa los filtros para afinar la búsqueda.

Resultados por página: | Ordenar por: En orden: Autor/registro:

Resultados por colección:



CLIMA

Resultados por ítem:

Vista previa	Fecha de publicación	Título	Autor(es)
	1-jul-1950	Clima, Año 1, n. 1 (jul. 1950)	-
	1-oct-1950	Clima, Año 1, n. 2-3 (oct. 1950)	-

Resultados 1-3 de 3.

Otras opciones relacionadas

Autor
Flores Mora, Manuel (1923-1985) 1
Greene, Graham (1904-1991) 1
Otero Mendoza, Jorge. 1
Payró, Julio E. (1899-1971) 1
Platschek, Hans (1923-2000) 1

Título
Cine - historia 1
Cuentos ingleses - S. XX 1
Pintura 1
Poesía uruguaya - S. XX 1

Fecha de lanzamiento
1950 - 1959 2

Figura 50 - Búsqueda en Anáforas



Figura 51 - Búsqueda por autor

- **Autores del Uruguay**

Para la comunidad Autores del Uruguay, como ya comentamos se realizó una solución intermedia, donde las colecciones se re direccionan a los links de las paginas actuales de autores.

Como esta es una solución momentánea por las dificultades con la migración de Autores del Uruguay existentes en el sitio actual, se utilizaron los campos de metadatos ya existentes. Los campos de la colección utilizados para este fin son: Descripción corta, donde se debe ingresar la palabra "link" y en el campo Texto introductorio ingresar el link a la página actual.

Metadatos básicos de la Colección	
Nombre:	Eduardo Acevedo Diaz
Descripción corta	link
Texto introductorio (HTML):	http://www.autoresdeluruguay.uy/biblioteca/Eduardo_Acevedo_Diaz/doku.php

Figura 52 - Ejemplo de ingreso de link para Autores del Uruguay

Se modifica el archivo [dspace]/dspace-jspui/src/main/webapp/community-home.jsp

Esto proporciona flexibilidad para que en un futuro se trabajen las colecciones de Autores del Uruguay de manera tradicional sin necesidad de modificar código fuente.

Figura 53 - Colecciones de Autores del Uruguay

- Ingreso de un ítem (Usuario logueado)

Se modificó el formulario de forma similar a la BNU para que no ocupe tanto lugar

5.3.9. PROCESO DE CONVERSIÓN DE INFORMACIÓN

En la etapa de configuración y adaptación del software se convirtieron los registros correspondientes a la descripción de 22 títulos de publicaciones periódicas. Los registros originales fueron creados en el SGB de la Biblioteca Nacional, en forma de registros bibliográficos, con lo que se generaba la externalidad positiva de enriquecer el catálogo de la biblioteca con nueva información, para ser convertidos posteriormente a metadatos una vez que la biblioteca digital estuviera disponible.

Las etapas de conversión fueron:

1. Descargar los registros del SGB de la biblioteca, en un archivo MARC de comunicaciones
2. Insertar los registros en una base relacional
3. Unificar los registros correspondientes a un mismo número de una publicación

4. Modificar la puntuación a fin de adaptarla a la norma de metadatos desarrollada
5. Exportar los registros a archivos XML y generar los bundles de dSpace para su carga.

Además también contamos con un script de importación que crea la estructura de carpetas automáticamente para que la información sea transferida de forma exitosa a la plataforma.

Por cada ítem a subir se crea un directorio que contiene lo siguiente:

- license.txt - licencia del archivo
- archivo.extensión - el archivo digital a subir en la plataforma
- contents - contiene información acerca de la licencia del archivo así como también el nombre del archivo a importar.
- dublin_core.xml - archivo que contiene metadatos asociados al archivo

Estos datos se importaron exitosamente a la biblioteca digital de la FIC, dado que pertenecen a publicaciones periódicas y fueron sincronizados con la biblioteca digital de la BNU.

CAPÍTULO 6

6. PRUEBAS REALIZADAS

6.1. CARGA MASIVA

Para probar la carga masiva de datos vía DSpace lo primero que se investigó fue como esta plataforma incorpora los ítems a DSpace. Como ya hemos explicado en secciones anteriores este debe de seguir un esquema de directorio para que luego DSpace lo procese y lo importe.

Se realizó la importación de archivos de hasta 100MB de forma exitosa. Se realizaron estas pruebas ingresando como administrador desde la interfaz gráfica accediendo a la opción batch import desde el menú de administrador. Así mismo también se realizó la importación de estos desde una terminal ejecutando el comando:

```
[dspace]/bin/dspace import --add --eperson=admin@mail.com --collection=CollectionID --source=items_dir
```

No encontramos ninguna especificación donde dijera si DSpace tenía un máximo a cargar simultáneamente, pero en caso de que exista un máximo, se pueden realizar cargas de menor cantidad de archivos a la vez.

6.2. SINCRONIZACIÓN

Una vez que se realizaron las configuraciones pertinentes se comenzó con las pruebas de sincronización, para ello se tomó como primera instancia repositorios cuya puesta online tiene ya el proyecto OAI en producción, en principio se probó con el proyecto demo DSpace [64]. En este repositorio los ítems están disponibles para ser sincronizados con cualquier repositorio que lo desee. En principio se probó con unos 20 ítems luego se pasó a una prueba con más datos.

Como este proyecto maneja una gran cantidad de volumen de datos se buscó un repositorio en donde se pueda realizar la misma y que maneje grandes cantidades de archivos de mediano y gran porte. Utilizamos el repositorio [65] y con este repositorio se logró sincronizar más de 2500 objetos digitales con sus respectivos metadatos y bitstreams.

6.3. CARGA ARCHIVOS GRANDES

Se probaron las cargas de archivos de gran tamaño al repositorio. Se probaron cargas de PDFs de 100MB aproximadamente, los mismos son subidos sin demorar al repositorio y no tienen dificultades ni demoras al cargarse con el complemento utilizado. Además se probó subir un video de tamaño 1.5 GB aproximado y no hubo

ningún inconveniente, para imágenes se probaron con tamaño 50MB y para audio se probó con archivos de 40MB.

Existe una propiedad en el archivo de configuración (dspace.cfg), llamada upload.max, la cual indica el tamaño máximo el cual puede tener un archivo a subir. Si es necesario se puede modificar para que se acepten archivos de mayor tamaño.

CAPÍTULO 7

7. CONCLUSIONES Y TRABAJO A FUTURO

7.1. CONCLUSIONES

Consideramos que este proyecto ha sido y será de gran utilidad ya que se trabajó con dos instituciones referentes a nivel nacional como lo son la BNU y la FIC. Si bien el proyecto en principio no contaba con la realización de una biblioteca digital para la BNU se logró cumplir con los objetivos impuestos; la investigación realizada acerca de esta plataforma cuya usabilidad a nivel mundial la ubica en el primer lugar es y será de gran utilidad para posteriores trabajos que se realicen en base a la misma.

Destacamos la importancia a nivel cultural que poseen ambas bibliotecas digitales, y creemos que la difusión del valioso material que poseen ambas instituciones a través de esta digitalización y puesta en producción de las bibliotecas brindarán a la población un interés y mejor aprovechamiento del material. El aporte realizado por nuestra investigación en dicha plataforma sin dudas dará paso a que nuevas instituciones se interesen en esta para incorporarlo a futuro.

En general, estamos muy contentos con los resultados obtenidos ya que alcanzamos los objetivos que fueron propuestos cumpliendo con los requerimientos iniciales.

7.2. DIFICULTADES

En el transcurso de nuestro trabajo nos hemos encontrado con varias dificultades. Estas dificultades, en general, se han debido al poco conocimiento que teníamos de algunas herramientas a utilizar y en algunos casos, a la falta de documentación al respecto.

Para la instalación de la herramienta DSpace, tuvimos un proceso de ensayo-error debido a que siguiendo los pasos del manual oficial no se lograba la instalación con éxito. Luego de varias pruebas y búsquedas en los manuales, los cuales todos tenían diferentes procedimientos para la instalación, logramos crear nuestro propio manual con los pasos a seguir para la instalación.

Una vez que se tuvo a DSpace instalado y funcionando correctamente nos encontramos con un nuevo problema a la hora de hacer funcionar el correo electrónico para envío de mensajes. La configuración estaba correcta pero nos daba una excepción de autenticación (`AuthenticationFailedException`). Luego de investigar vimos que el problema no era del lado de DSpace sino que el servidor de correos bloqueaba la conexión dado que lo tomaba como una conexión no segura. Luego de permitir el acceso de aplicaciones no seguras logramos hacer funcionar correctamente el correo.

Para realizar las modificaciones en código, se utilizó la herramienta Netbeans. También encontramos dificultades para poder ejecutar correctamente el código fuente, luego de investigar encontramos algunas configuraciones particulares para poder ejecutar el proyecto.

Otro problema encontrado fue que en la versión que estábamos utilizando la búsqueda por fecha y autor no funcionaba correctamente, siempre retornaba que no se encontraban resultados, cuando sí los había.

Cuando comenzó el proyecto utilizamos la versión 5.2 estable al momento. Luego de realizar varias pruebas sin éxito, buscando en la comunidad DSpace, encontramos que el problema se debía a un bug que solo se encontró en la versión 5.2 el cual se corregía para la versión 5.3 liberada el 29 de Julio del 2015, también funcionaba correctamente en versiones anteriores.

Por lo tanto se tuvo que migrar todo el trabajo realizado en la versión 5.2 a la versión 5.3.

Otro error encontrado fue al intentar registrar nuevos bitstreams. El problema consistía cuando se agregaba un nuevo formato de archivo para que el sistema lo tomara como formato reconocido pero al intentar subir un ítem con ese tipo de formato producía un error.

Detectamos que esto se debía a que al registrar un formato bitstream, la opción para seleccionar el formato interno estaba oculto, por tal motivo no permitía subir ítems para el formato registrado. Se solucionó habilitando ese campo para elegir si el formato será interno o no.

ID	Tipo MIME	Nombre	Descripción larga	Nivel de soporte	¿Interno?	Extensiones	
1	application/octet-str	Unknown	Unknown data format	Desconocido	<input type="checkbox"/>	Extensiones	Actualizar
2	text/plain; charset=u	License	Item-specific license agreee	Conocido	<input type="checkbox"/>	Extensiones	Actualizar Eliminar...
3	text/html; charset=ut	CC License	Item-specific Creative Com	Conocido	<input type="checkbox"/>	Extensiones	Actualizar Eliminar...

Figura 54 - Registro de bitstream en versión por defecto

ID	Tipo MIME	Nombre	Descripción larga	Nivel de soporte	¿Interno?	Extensiones	
1	application/octet-str	Unknown	Unknown data format	Desconocido	<input type="checkbox"/>	Extensiones	Actualizar
2	text/plain; charset=t	License	Item-specific license agreec	Conocido	<input checked="" type="checkbox"/>	Extensiones	Actualizar Eliminar...
3	text/html; charset=u	CC License	Item-specific Creative Comi	Conocido	<input checked="" type="checkbox"/>	Extensiones	Actualizar Eliminar...
4	application/pdf	Adobe PDF	Adobe Portable Document I	Conocido	<input type="checkbox"/>	pdf	Actualizar Eliminar...
5							

Figura 55 - Registro de bitstream corregido

Otras dificultades encontradas se presentaron a la hora de encontrar los complementos correctos, esto fue mencionado anteriormente en la sección complementos.

Cuando se comenzó con la investigación acerca de la sincronización entre repositorios una de las mayores dificultades encontradas fue la falta de documentación al respecto. Cabe destacar que la poca información habilitada inclusive la que se encuentra en los manuales oficiales de las diferentes versiones que hay de DSpace es muy básica y

confusa para quien por primera vez se enfrenta con tal desafío. Para las primeras pruebas se comenzó con una especie de ensayo-error, a medida que nos encontrábamos con nuevas dificultades recurrimos a los grupos de ayuda en los foros para exponer dudas que DSpace ofrece donde algún experto puede guiar un poco el problema planteado.

Adicionalmente una vez lograda la correcta configuración para sincronizar distintas repositorios, nos encontramos con un nuevo obstáculo a la hora de sincronizar una colección. El problema consistía que a la hora de realizar la cosecha de una colección, por vez primera esto funciona de forma correcta, pero si se intenta volver a cosechar la misma colección, para actualizar los ítems de la misma, nos generaba un error.

Después de varios intentos fallidos y de realizar varias búsquedas en los foros de ayuda, encontramos que el problema era con la fecha de los servidores. El manejo de fechas en servidores es bastante confuso y de hecho siempre se expone a tener errores por configuración de hora, lenguaje etc. Se corrigió este para que funcione correctamente en cada sincronización que se realice. Este tipo de problemas llevo tiempo y varias pruebas hasta detectar que es lo que estaba impidiendo la cosecha, es un problema conocido pero nadie publicó una posible solución o al menos no se encontró nada referente a la solución del problema.

7.3. APORTES

7.3.1. BIBLIOTECAS DIGITALES

Mediante las bibliotecas digitales resultantes de este proyecto, se permitirá el acceso libre, universal y gratuito de contenidos originales relativos al patrimonio documental nacional contribuyendo a la identidad cultural del país. Pone al alcance de todo el mundo, las colecciones de la Biblioteca Nacional del Uruguay y las colecciones de la Facultad de Información y Comunicación, con el fin de acercarlas a la comunidad mediante el uso de las tecnologías de la información.

Mediante estas bibliotecas se ofrece el acceso a recursos importantes para la investigación, la enseñanza y el aprendizaje, para todo tipo de usuarios, tanto principiantes como expertos.

7.3.2. JORNADA BIBLIOTECA DIGITAL UNIVERSITARIA

Durante los días 5 y 6 de noviembre se realizó la 13ª Jornada sobre la Biblioteca Digital Universitaria y 1ª del Mercosur. Estas jornadas se realizan habitualmente en distintas ciudades de la República Argentina, en esta oportunidad se realizó en Montevideo teniendo como anfitrionas a la Biblioteca de Facultad de Ciencias y la Biblioteca Nacional.

Dado que nuestro proyecto se basaba en la investigación y en la realización de Bibliotecas Digitales para instituciones referentes de nuestro país, nos pareció oportuno presentarnos a participar de la misma y exponer nuestro proyecto de grado.

La exposición fue una gran experiencia para nosotros y a la vez un gran desafío puesto que varias personas de diferentes países se harían presentes en la exposición y desconocíamos qué inquietudes podrían llegar a preguntar.

Colmó nuestras expectativas al máximo dado el grado de interés que se vivió por parte de los presentes, al exponer la biblioteca digital de ambas instituciones así como también hacer referencia e hincapié en la investigación de DSpace que para nuestra sorpresa no muchos tenían el conocimiento de la misma.

7.3.3. URUGUAY CRECE CONTIGO

A través de uno de los colaboradores de la Biblioteca Nacional, Camilo Freire, nos contactaron del proyecto Uruguay Crece Contigo, para realizar un Repositorio Digital basado en el software DSpace.

Dado el conocimiento adquirido a lo largo del proyecto gracias a la investigación que llevamos a cabo, fuimos contratados con un cargo docente en la FIC para realizar la definición de jerarquías, metadatos y la customización del software DSpace.

7.4. TRABAJO A FUTURO

Existen numerosos aspectos en los que se puede trabajar en pro de mejorar las soluciones propuestas. A continuación detallaremos algunos de las mejoras que pueden ser útiles para las soluciones.

7.4.1. AGREGAR COMPLEMENTOS PARA OTROS FORMATOS DE ARCHIVOS

Por ejemplo para archivos Microsoft Word y Power Point no se agregaron complementos, esto podría ser necesario a futuro, no fueron agregados debido a que no fue pedido ni se relevaron documentos con estos formatos, pero podría ser una mejora para la aplicación.

7.4.2. CALIDAD

Con respecto a la calidad, no realizamos ningún agregado de controles excepto los propuestos por DSpace. Consideramos que con lo que trae DSpace era suficiente, pero existen varios controles de calidad adicionales que pueden ser interesantes de aplicar, pero debido al tiempo que contábamos no veíamos viable realizarlos.

Por ejemplo, unas de las posibles mejoras a incorporar podría ser la verificación de autores con otras bases de datos y otros sitios.

7.4.3. MEJORAR INGRESO DE METADATOS

El ingreso de metadatos en DSpace a veces puede ser un poco tedioso. Por ejemplo para los campos repetibles, DSpace permite agregarlos de a uno y por separado. Una posibilidad que fue mencionada por las instituciones y debido a la complejidad del cambio y por falta de tiempo no fue realizada, es poder ingresar en un solo campo todos los valores del campo repetible, separados por algún separador definido. De esta forma agilizaría el ingreso.

7.4.4. EXTRACCIÓN AUTOMÁTICA DE METADATOS

En el estudio [66] sugiere una modificación al flujo que trae DSpace y mediante un complemento de terceros para extraer metadatos de archivos, llenar automáticamente algunos de los metadatos, realizando la tarea de ingreso de metadatos menos tediosa.

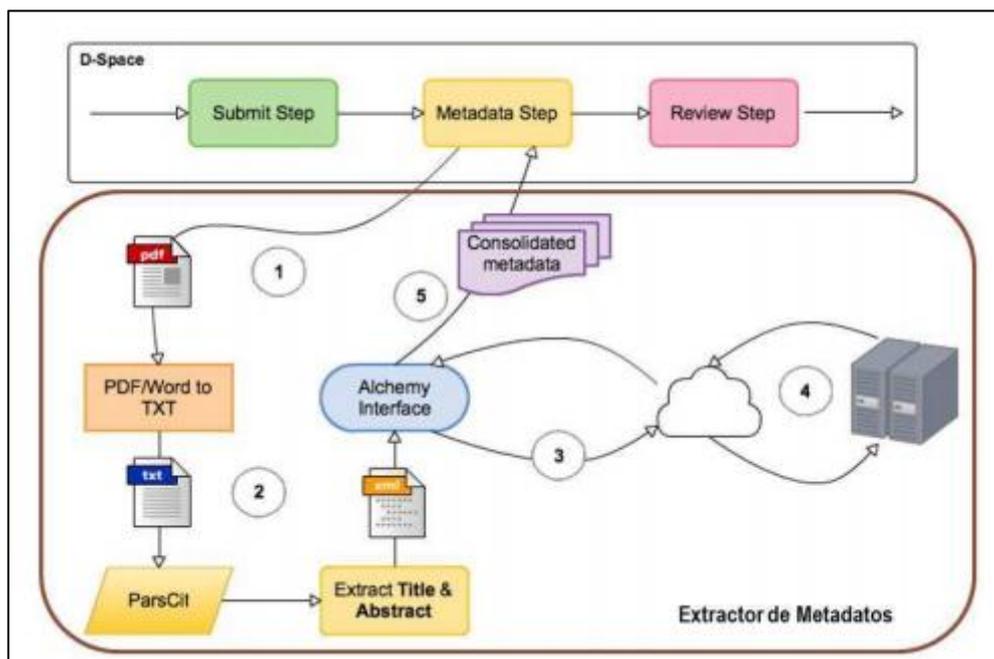


Figura 56 - Arquitectura propuesta

La propuesta consiste en el siguiente cambio:

- Cambiar el flujo para que primero el usuario ingrese los archivos en vez de la metadata
- Luego se utilizará un módulo intermedio donde se realiza la conversión a archivo de texto del archivo que cargó el usuario (en formato pdf, doc, ppt, etc.) y extrae el contenido en formato de texto (.txt).

- El módulo de estructuración del documento recibe como entrada el archivo en texto plano, organiza y da estructura al mismo generando un archivo XML en el cual se puede identificar título, autor, resumen y palabras claves. Esto mediante la herramienta ParsCit [67].
- Luego se envía el archivo en formato XML al servicio Alchemy. Alchemy envía la información al servidor extrayendo además el idioma y las palabras claves. El mismo retorna la respuesta en formato JSON.
- Se integran los resultados consolidando la respuesta del servidor y enviando a DSpace un archivo XML con los metadatos extraídos automáticamente.
- Por último se sigue con el flujo habitual, dándole la posibilidad al usuario de modificar los metadatos extraídos automáticamente o agregar nuevos.

Esta propuesta nos pareció muy interesante y nos parece una posible mejora pero de gran complejidad.

GLOSARIO

BNU Biblioteca Nacional de Uruguay

FIC Facultad de Información y Comunicación

UdelaR Universidad de la República

DC Dublin Core

MARC Machine Readable Cataloging

Metadatos Literalmente, los metadatos son datos sobre datos. Los metadatos son datos asociados a objetos o sistemas de información para fines de descripción, administración, uso, preservación, etc. Existen varios tipos de metadatos: descriptivos (que sirven para identificación y localización), administrativos (creación, derechos, control de acceso, etc.) y estructurales (que relacionan los objetos).

Cualificador Elemento que refina la semántica de los elementos de forma que pueden ser útiles para la recuperación de documentos.

OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting

Digitalización Reproducción digital por medio de escáneres, cámaras fotográficas y similares, de libros, periódicos, documentos, manuscritos y otros materiales afines.

Plugin Complemento para una aplicación que se relaciona con otra para aportar una función nueva y generalmente muy específica

Omeka Plataforma de software libre utilizada para la publicación de colecciones digitales u otros. La gran diferencia con otros programas pensados para construir repositorios digitales, es que Omeka no sólo es capaz de almacenar y mostrar colecciones digitales, sino que también permite crear exposiciones virtuales en torno a éstas.

DSpace Software de código abierto utilizado para gestionar repositorios de archivos (textuales, video, etc.) organizándolos en comunidades, asignándoles metadatos y permitiendo su difusión.

OpenAIRE Infraestructura de acceso abierto a la investigación en Europa, creada y desarrollada dentro de los proyectos del 7^º Programa Marco OpenAIRE y OpenAIRE Plus desde el año 2009.

Acceso Abierto, Open Access Acceso abierto significa la disponibilidad libre en Internet de la literatura de carácter académico o científico, permitiendo a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o enlazar los textos completos de los artículos científicos, y usarlos con cualquier otro propósito legítimo, sin otras barreras financieras, legales o técnicas más que suponga Internet en sí misma.

Interoperabilidad Es la capacidad de los repositorios para comunicarse e intercambiar procesos y datos. Es posible gracias al protocolo OAI-PMH.

Dokuwikis Software de código abierto que gestiona webs colaborativas de tipo wiki. Está enfocado para ser usado por grupos de desarrolladores, grupos de trabajo en general y pequeñas compañías.

Customizar Se trata de una adaptación del término inglés “*customize*”, que refiere a modificar y adaptar el software de acuerdo a los requerimientos establecidos por parte de los interesados.

Bundle Agrupación de archivos que pertenecen a los ítems, separan los diversos tipos de ficheros de modo que DSpace pueda tratarlos de forma diferenciada.

Bitstream Son los archivos contenidos en los bundles. Cada archivo subido a DSpace, o creado por DSpace se considera un bitstream.

Thumbnails o miniaturas Versiones de imágenes, usadas para ayudar a su organización y reconocimiento.

Framework Define en términos generales, un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar.

REFERENCIAS

- [1] CSE-BN, «Convenio entre la Universidad de la Republica - Comisión Sectorial de Enseñanza y Biblioteca Nacional» 2014. [En línea]. www.universidad.edu.uy/renderResource/index/resourceId/34170/.../1. [Último acceso: Octubre 2015].
- [2] «Publicaciones Periodicas» [En línea]. <http://www.periodicas.edu.uy/>. [Último acceso: Noviembre 2015].
- [3] «Omeka» [En línea]. <http://omeka.org/>. [Último acceso: Junio 2015].
- [4] «Dublin Core» [En línea]. <http://dublincore.org/>. [Último acceso: Junio 2015].
- [5] «DSpace» [En línea]. <http://www.dspace.org/>. [Último acceso: Noviembre 2015].
- [6] «Autores del Uruguay» [En línea]. <http://www.archivodeprensa.edu.uy/doku.php>. [Último acceso: Noviembre 2015].
- [7] «Library of Congress: "Marc 21 Format for bibliographic data"» [En línea]. <https://www.loc.gov/marc/bibliographic/>. [Último acceso: Noviembre 2015].
- [8] «Aleph» [En línea]. <http://www.exlibrisgroup.com/category/Aleph>. [Último acceso: Octubre 2015].
- [9] «Gallica» [En línea]. <http://gallica.bnf.fr/>. [Último acceso: Noviembre 2015].
- [10] «Biblioteca Nacional de España» [En línea]. <http://www.bne.es/es/Inicio/index.html>. [Último acceso: Noviembre 2015].
- [11] «Biblioteca Digital Trapalanda» [En línea]. <http://trapalanda.bn.gov.ar/jspui/handle/123456789/1>. [Último acceso: Noviembre 2015].
- [12] «Biblioteca Digital de Brasil» [En línea]. <https://bndigital.bn.br/>. [Último acceso: Noviembre 2015].
- [13] «Memoria Chilena» [En línea]. <http://www.memoriachilena.cl/602/w3-channel.html>. [Último acceso: Noviembre 2015].
- [14] «Library of Congress: "MARC Standards"» [En línea]. <https://www.loc.gov/marc/marcginf.html>. [Último acceso: Noviembre 2015].
- [15] «ISO 2709:2008» [En línea]. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=41319. [Último acceso: Noviembre 2015].
- [16] «Library of Congress: "MARC 21 XML Schema"» [En línea]. <https://www.loc.gov/standards/marcxml/>. [Último acceso: Noviembre 2015].
- [17] «Library of Congress: "Metadata Object Description Schema "» [En línea]. <http://www.loc.gov/standards/mods/>. [Último acceso: Noviembre 2015].

- [18] «Library of Congress: “MARCXML : MARC 21 XML Schema”» [En línea].
<http://www.loc.gov/standards/marcxml/>. [Último acceso: Noviembre 2015].
- [19] «Dublin Core Metadata Initiative» [En línea]. <http://dublincore.org/>. [Último acceso: Octubre 2015].
- [20] «Dublin Core - Metadata Element Set» [En línea]. <http://dublincore.org/documents/dces/>. [Último acceso: Julio 2015].
- [21] «Dublin Core Qualifiers» [En línea]. <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>. [Último acceso: Julio 2015].
- [22] R. Crow, «The Case for Institutional Repositories: A SPARC Position Paper. The Scholarly Publishing,» *Association of Research Libraries*, nº 223, 2002.
- [23] H. Van de Sompel, S. Payette, J. Erickson, C. Lagoze y S. Warner, «Rethinking Scholarly Communication,» vol. 10, nº 9, 2004.
- [24] Bustos y Gonzalez, Directrices para la creación de repositorios institucionales, Babel , 2008.
- [25] M. Lesk, Practical Digital Libraries, San Francisco: Morgan Kaufmann Publishers, 1997.
- [26] C. L. Borgman, «What are digital libraries? Competing visions,» vol. 35, Mayo 1999.
- [27] L. Candela, D. Castelli, P. Pagano, C. Thanos y Otros, «Setting the Foundations of Digital Libraries,» vol. 13, nº 3/4, 2007.
- [28] I. Witten, D. Bainbridge y D. Nichols, How to Build a Digital Library, Second Edition, New Zealand: Morgan Kaufmann, 2009.
- [29] V. Paulo, «Diseño de un modelo de repertorio institucional - EUBCA,» 2011.
- [30] «Fedora Commons» [En línea]. <http://www.fedora-commons.org/>. [Último acceso: Setiembre 2015].
- [31] «Greenstone» [En línea]. http://www.greenstone.org/index_es. [Último acceso: Setiembre 2015].
- [32] «EPRINTS» [En línea]. www.eprints.org/. [Último acceso: 09 2015].
- [33] «Archimede» Noviembre 2015. [En línea]. <http://www.bibl.ulaval.ca/archimede/index.en.html>.
- [34] «Open Repository» [En línea]. <http://www.openrepository.com/>. [Último acceso: Octubre 2015].
- [35] «Bepress» Digital Commons, [En línea]. <http://www.bepress.com/>. [Último acceso: Noviembre 2015].
- [36] «CONTENTdm» [En línea]. <http://www.oclc.org/en-US/contentdm.html>. [Último acceso: Noviembre 2015].
- [37] «OpenAIRE Guidelines» [En línea]. <https://guidelines.readthedocs.org/en/latest/>. [Último acceso: Setiembre 2015].
- [38] «Plataforams de software para Repositorios Institucionales» [En línea].
<https://repositorioinstitucional.wikispaces.com/Plataformas+de+software+para+Repositorios+Institucionales>.

[Último acceso: Setiembre 2015].

- [39] «Comparativa de Repositorios» [En línea]. <http://www.opendoar.org/find.php?format=charts>. [Último acceso: Octubre 2015].
- [40] F. Zubiri y A. Mereles, «Repositorios Institucionales» 2012. [En línea]. <https://www.colibri.udelar.edu.uy/bitstream/123456789/3023/1/tg-zubiri.pdf>. [Último acceso: Octubre 2015].
- [41] J. R. Pérez Agüera y R. Melero , «Plataforma digital de revistas científicas electrónicas españolas. Relación con el movimiento open access,» vol. 8, nº 170(74), 2005.
- [42] «Open Archives» [En línea]. <HTTPS://WWW.OPENARCHIVES.ORG>. [Último acceso: Noviembre 2015].
- [43] T. D. D. Team, «DSpace 5.x Documentation,» 2015.
- [44] «POSTGRESQL» [En línea]. <http://www.postgresql.org/>. [Último acceso: Agosto 2015].
- [45] «Oracle Database» [En línea]. <http://www.oracle.com/index.html>. [Último acceso: Agosto 2015].
- [46] «Cocoon Apache» [En línea]. <http://cocoon.apache.org/>. [Último acceso: Octubre 2015].
- [47] «XMLUI o JPSPUI» [En línea]. <http://www.arvo.es/dspace/xmlui-o-jspui/>. [Último acceso: Agosto 2015].
- [48] «Apache Tomcat» [En línea]. <http://tomcat.apache.org/>. [Último acceso: Noviembre 2015].
- [49] «SOLR» [En línea]. <http://lucene.apache.org/>. [Último acceso: Octubre 2015].
- [50] «Conociendo MARC bibliográfico : catalogación legible en máquina» [En línea]. <https://www.loc.gov/marc/umbspa/>. [Último acceso: Octubre 2015].
- [51] «Atmire» [En línea]. <http://atmire.com/website/>. [Último acceso: Octubre 2015].
- [52] «Dura Space» [En línea]. <http://www.duraspace.org/>. [Último acceso: Octubre 2015].
- [53] «Archivos de Prensa» Agosto 2015. [En línea]. www.archivosdeprensa.edu.uy.
- [54] «DokuWiki» [En línea]. <https://www.dokuwiki.org/dokuwiki#>. [Último acceso: Noviembre 2015].
- [55] «Pretty Photo» [En línea]. 1. Plugin Pretty Photo. [En línea] <http://www.no-margin-for-errors.com/projects/prettyphoto->. [Último acceso: Octubre 2015].
- [56] «Cloud-Zoom» [En línea]. <http://www.jqueryscript.net/zoom/Cloud-Zoom-Image-Zoom-jQuery-Plugin-2.html>. [Último acceso: Octubre 2015].
- [57] «Fancybox» [En línea]. <http://fancybox.net/>. [Último acceso: Octubre 2015].
- [58] «PDF.js» [En línea]. <https://mozilla.github.io/pdf.js/>. [Último acceso: Octubre 2015].
- [59] «Video.js» [En línea]. <https://github.com/videojs>. [Último acceso: Octubre 2015].

- [60] «JW-Player» [En línea]. <https://www.jwplayer.com/>. [Último acceso: Octubre 2015].
- [61] «Formatos admitidos de audio y video en HTML5» [En línea].
https://developer.mozilla.org/es/docs/Web/HTML/Formatos_admitidos_de_audio_y_video_en_html5.
- [62] «Sound JS» [En línea]. <http://www.createjs.com/soundjs>. [Último acceso: Octubre 2015].
- [63] «FlexSlider» [En línea]. <http://flexslider.woothemes.com/>. [Último acceso: Setiembre 2015].
- [64] «Demo DSpace» [En línea]. <http://demo.dspace.org/oai/request>. [Último acceso: Noviembre 2015].
- [65] «Repositorio de Universidad de Ecuador» [En línea]. <http://repositorio.puce.edu.ec/oai/request>. [Último acceso: Octubre 2015].
- [66] A. Casalia, C. Deco, C. Bender y S. Fontanarrosa, «Extracción Automática de Metadatos de Objetos Digitales Educativos,» de *Extracción Automática de Metadatos de Objetos Digitales*, Rosario, Argentina, 2014.
- [67] «ParsCit» [En línea]. <http://aye.comp.nus.edu.sg/parsCit/>. [Último acceso: Octubre 2015].
- [68] «Información y Documentación : formatos para el intercambio de la información : ISO,» *Revista española de Documentación científica*, vol. 28, nº Nro. 4, 2005.
- [69] J. Texier, «Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación,» Venezuela, 2013.

ANEXO MANUAL DE INSTALACIÓN DSPACE 5.X EN UBUNTU

14.02

A continuación detallaremos los pasos para instalación de DSpace en Ubuntu, aclaramos que el manual fue realizado por nosotros de varios manuales y que luego de varias pruebas logramos realizar la instalación con éxito.

Pre-requisitos:

1. Actualizar repositorios (lista de paquetes), para cuando se realiza la búsqueda y descarga sea más rápida:

```
sudo apt-get update
```

2. Instalar la jdk de java:

```
sudo apt-get install openjdk-7-jdk
```

3. Agregar variables de entorno (En este paso nos adelantamos y ya agregamos las variables de entorno para tomcat, maven y ant que serán instaladas más adelante)3.

Abrir el archivo `/etc/enviromen`: `sudo gedit /etc/enviromen`

Agregar lo siguiente al archivo (la variable PATH ya existe agregar lo que falta)

```
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr:%JAVA_HOME%/bin:%ANT_HOME%/bin:%CATALINA_HOME%:%CATALINA_BASE%"
```

```
JAVA_HOME="/usr/lib/jvm/java-1.7.0-openjdk-amd64"
```

```
JRE_HOME="$JAVA_HOME/jre"
```

```
MAVEN_OPTS="-Xms512m -Xmx1024m"
```

```
JAVA_OPTS="-Xmx2048M -Xms1024M -Dfile.encoding=UTF-8"
```

```
ANT_HOME="/usr/share/ant"
```

```
CATALINA_HOME="/var/lib/tomcat7"
```

```
CATALINA_BASE="/var/lib/tomcat7"
```

```
CATALINA_OPTS="-Xms1025m -Xmx1024m -XX:MaxPermSize=512m"
```

Luego de guardar el archivo ejecutar para ver los cambios:

```
source /etc/environment
```

4. Instalar tomcat 7

```
sudo apt-get install tomcat7
```

```
sudo apt-get install tomcat7-admin
```

Opcionales:

```
sudo apt-get install tomcat7-examples
```

```
sudo apt-get install tomcat7-docs
```

5. Se puede cambiar el puerto donde corre el tomcat, modificar en el archivo `/etc/tomcat7/server.xml` cambiar en el tag Connector el puerto 8080 a otro puerto (Ej.: 8181). Además dejarlo de la siguiente manera:

```
<Connector port="8181" maxThreads="150" minSpareThreads="25"
maxSpareThreads="75" enableLookups="false" redirectPort="8443"
acceptCount="100" connectionTimeout="20000" disableUploadTimeout="true"
URIEncoding="UTF-8"/>
```

6. Agregar los usuarios de tomcat:

En archivo `/etc/tomcat7/tomcat-users.xml` agregar usuario "admin" (para poder entrar como administrador) y el usuario "dspace".

Ej.:

```
<role rolename="admin-gui"/>
```

```
<role rolename="manager-gui"/>
```

```
<role rolename="manager-script"/>
```

```
<role rolename="manager-jmx"/>
```

```
<role rolename="manager-status"/>
```

```
<role rolename="dspace"/>
```

```
<user username="admin" password="admin" roles="admin-gui,manager-
gui,manager-script,manager-jmx,manager-status"/>
```

```
<user username="dspace" password="dspace" roles="dspace"/>
```

```
</tomcat-users>
```

7. Para que se impacten los cambios reiniciar el tomcat:

```
sudo service tomcat7 restart
```

8. Instalar Maven

```
sudo apt-get install maven
```

9. Instalar Ant

```
sudo apt-get install ant
```

10. Instalar base de datos postgresql:

```
sudo apt-get install postgresql-9.3
```

11. En el archivo `/etc/postgresql/9.3/main/postgresql.conf` descomentar `listen_addresses = 'localhost'`

12. En el archivo `/etc/postgresql/9.3/main/pg_hba.conf` agregar esta línea:

```
local all dspace md5
```

Además descomentar todo lo que tenga peer

13. Reiniciar la base

```
sudo /etc/init.d/postgresql restart
```

14. Crear el usuario dspace y crear la base dspace de la siguiente manera:

```
sudo su postgres
```

```
createuser -U postgres -d -A -P dspace (Poner una password)
```

```
createdb -U dspace -E UNICODE dspace (Poner una password)
```

```
exit
```

15. Crear el usuario DSpace y asignarle una contraseña

```
sudo useradd -m dspace
```

```
sudo passwd xxxx
```

16. Crear un directorio en donde poner DSpace (Ej: `sudo mkdir /opt/dspace5`), en los pasos siguientes para hacernos referencia a este directorio lo indicaremos como `[dspace]`

17. Descargar la versión 5.X de DSpace de <http://sourceforge.net/projects/dspace/>, descomprimir y copiar en `[dspace]`

También se puede instalar git y descargarlo desde ahí.

18. Abrir el archivo `[dspace]/build.properties` y setear en `dspace.install.dir` la ruta de donde quedo instalado

```
Ej: dspace.install.dir= [dspace]
```

19. Parados en `[dspace]` ejecutar el siguiente comando:

```
sudo mvn package
```

20. Ejecutar el siguiente comando parados en `[dspace]/dspace/target/dspace-installer`

```
sudo ant fresh_install
```

21. Cambiar permisos de la carpeta para el usuario tomcat:

```
sudo chown tomcat7:tomcat7 [dspace] -R
```

22. Reiniciar el tomcat:

```
sudo service tomcat7 restart
```

23. Hay que referenciar las aplicaciones de dspace para poder verlas desde el tomcat, en el archivo /etc/tomcat7/server.xml agregar antes de que termine </host>, agregar los módulos que se van a utilizar.

```
<!-- Define a new context path for all DSpace web apps -->  
  
<Context path="/xmlui" docBase="/directorio_instalacion/webapps/xmlui"  
allowLinking="true"/>  
  
<Context path="/sword" docBase="/directorio_instalacion/webapps/sword"  
allowLinking="true"/>  
  
<Context path="/oai" docBase="/directorio_instalacion/webapps/oai"  
allowLinking="true"/>  
  
<Context path="/jspui" docBase="/directorio_instalacion/webapps/jspui"  
allowLinking="true"/>  
  
<Context path="/lni" docBase="/directorio_instalacion/webapps/lni"  
allowLinking="true"/>  
  
<Context path="/solr" docBase="/directorio_instalacion/webapps/solr"  
allowLinking="true"/>
```

24. Asignar más memoria al Tomcat, en el archivo /etc/default/tomcat7, modificar la variable JAVA_OPTS

25. Reiniciar el tomcat

```
sudo service tomcat7 restart
```

ANEXO CONFIGURACIONES REALIZADAS

Estructura de directorio del código antes de la instalación

- [dspace] - Directorio de instalación
 - LICENSE - Licencia de código dspace
 - README - Información
 - build.properties - Información básica necesaria para la compilación e instalación de dspace
 - dspace - Directorio que contiene toda la compilación y configuración de DSpace
 - bin - Aplicaciones de línea de comando y scripts de perl para correr los comandos desde la línea de comando
 - config - Archivos de configuración
 - etc - Archivos administrativos
 - modules - Módulos web
 - solr - Configuración solr para índices
 - src - Configuración Maven
 - target - Este directorio se crea después de compilar, es el directorio que utiliza Maven para la instalación
 - dspace-api - Módulo Java Api
 - dspace-jspui - Código fuente del módulo JSPUI
 - dspace-lni - Código fuente del módulo LNI
 - dspace-oai - Código fuente del módulo OAI-PMH
 - dspace-rdf - Código fuente del módulo RDF
 - dspace-rest - Código fuente del módulo de Rest API
 - dspace-services - Código fuente del módulo servicios
 - dspace-sword - Código fuente del módulo SWORD
 - dspace-swordv2 - Código fuente del módulo SWORD v2
 - dspace-xmlui - Código fuente del módulo Manakin
 - dspace-xmlui-mirage - Theme Mirage 2 para XMLUI
 - pom.xml - Proyecto padre

Configuración

Luego de haber instalado DSpace lo primero que se tuvo que hacer es realizar las configuraciones básicas en el archivo de configuración de DSpace llamado `dspace.cfg`, el mismo se encuentra dentro del directorio instalación en la carpeta "Config" que se crea luego de haber instalado DSpace.

A continuación detallaremos las modificaciones realizadas a los archivos de configuración:

Información básica

- Directorio de Instalación
 - `dspace.dir = [dspace]`

- Nombre del Repositorio (sin puerto)
 - `dspace.hostname`
- URL base (con puerto)
 - `dspace.baseUrl`
- URL de DSpace por defecto
 - `dspace.url`
- Nombre del Repositorio
 - `dspace.name`

Configuración de Base de Datos

- URL de conexión a la BD
 - `dspace.url`
- Driver de conexión a la BD
 - `db.driver`
- Usuario y contraseña de la BD
 - `db.username`
 - `db.password`
- Lenguaje por defecto para los valores de los metadatos
 - `default.language = es`

Configuración de Mail

No pondremos valores reales por seguridad

- Servidor SMTP
 - `mail.server`
- Usuario y contraseña
 - `mail.server.username`
 - `mail.server.password`
- Puerto alternativo del SMTP (por defecto 25)
 - `mail.server.port`
- Dirección de correo electrónico desde donde se envían mails
 - `mail.from.address`
- Dirección de realimentación
 - `feedback.recipient`
- Dirección de correo del administrador del sitio
 - `mail.admin`
- Dirección de correo para el envío de errores y alertas
 - `alert.recipient`
- Dirección de correo para notificación de nuevos usuarios
 - `registration.notify`
- Propiedades extra para java mail.
 - `mail.extraproperties = mail.smtp.socketFactory.port=465, mail.smtp.socketFactory.class=javax.net.ssl.SSLSocketFactory, \ mail.smtp.socketFactory.fallback=false`

Localización de archivos

- Directorio donde se alojarán los documentos (Store 0)
 - `assetstore.dir`
- Directorios extra de alojamiento
 - `assetstore.dir.1`
 - `assetstore.dir.2`

- Cantidad de directorios de alojamiento (por defecto cero y corresponde al `assestore.dir`)
 - `assetstore.incoming`

Configuración de logueo

- Directorio para alojar los logs de la aplicación
 - `log.dir`

Prefijo Canónico

Este prefijo es el utilizado por dspace para generar el `dc.identifier.uri`. Existe la posibilidad de registrarse al servicio CNRI's para obtener una URL para utilizar. El servicio tiene costo por lo que no se utilizó y se configuró la URL de cada sitio.

- URL del prefijo Canónico
 - `handle.canonical.prefix`

Propiedades de la interfaz web

- Metadatos a mostrar en la vista simple del ítem
 - `webui.itemdisplay.default`

Configurar miniaturas (Thumbnails)

Las miniaturas son versiones reducidas de imágenes, usadas para ayudar al reconocimiento de ítems. DSpace permite por defecto visualizar miniaturas de los archivos de los ítems de tipo imagen `jpg`, `png` y `gif`. Incluye por defecto un filtro para la creación de éstos formatos de imágenes, éste es: "JPEGFilter" El resto de formatos (por ejemplo `pdf`, `odt`, `doc`, archivos de audio y video) no serán visibles con los cambios descritos a continuación.

Por defecto las imágenes miniaturas tanto en los ítems como en las búsquedas se encuentran deshabilitadas. Estas imágenes miniatura son generadas a través de los filtros de medios que será explicado más abajo. Para activar las miniaturas se deberá modificar los siguientes atributos en el archivo de configuración `dspace.cfg`:

- Habilitar las Imágenes en miniatura en las búsquedas
 - `webui.browse.thumbnail.show = true`
- Tamaño maximo de la miniatura en las búsquedas
 - `webui.browse.thumbnail.maxheight = 250`
 - `webui.browse.thumbnail.maxwidth = 180`
- Habilitar las Imágenes en miniatura en los ítems
 - `webui.item.thumbnail.show = true`
- Tamaño maximo de la miniatura en los ítems
 - `thumbnail.maxwidth = 250`
 - `thumbnail.maxheight = 180`

Estas líneas vienen descomentadas por defecto y con unos valores predefinidos, salvo la primera de ellas que hay que descomentarla y poner el valor a `true`.

Para documentos PDF se requiere realizar ciertas configuraciones e instalaciones que mencionaremos a continuación.

Filtro de Medios

DSpace puede aplicar filtros y transformaciones a los archivos/bitstreams creando nuevo contenido.

Los filtros incluyen la extracción de texto completo y la creación de miniaturas para los ítems.

Para generar las miniaturas mencionadas en el punto anterior y generar el texto para la búsqueda por texto completo, es necesario ejecutar una herramienta para la búsqueda y generación de las miniaturas de todos los archivos que dispongamos en DSpace.

El comando a ejecutar es el siguiente: `[dspace]/bin/dspace filter-media` y tiene varias opciones de ejecución, alguna de las más relevantes:

- -f Modo Forzado: Aplicar filtros para todos los bitstreams, incluso si ya han sido filtrados previamente
- -n Sin indexamiento: Elimina la creación de índices para búsquedas por texto completo
- -p Modo Complemento: Se aplica solo el complemento de filtro indicado

Los filtros disponibles son los siguientes y deben ser setados en el archivo de configuración:

- HTML Text Extractor - Extrae el texto completo de los documentos HTML para la indexación de texto completo
- JPEG Thumbnail - Crea miniaturas para los archivos de imagen con extensiones GIF, JPEG and PNG
- Branded Preview JPEG - Crea una imagen de previsualización de marca para los archivos GIF, JPEG y PNG
- PDF Text Extractor - Extrae el texto completo de documentos Adobe PDF (basados en texto y OCRred) para la indexación por texto completo. Utiliza la herramienta Apache PDFBox
- XPDF Text Extractor - Extrae el texto completo de documentos Adobe PDF (basados en texto y OCRred) para la indexación por texto completo. Utiliza la herramienta de línea de comandos XPDF para Unix
- Word Text Extractor - Extrae el texto completo de documentos Microsoft Word o documentos de texto plano para la indexación por texto completo.
- PowerPoint Text Extractor - Extrae el texto completo de ppts y notas de Microsoft PowerPoint and PowerPoint XML para indexación por texto completo. Utiliza la herramienta Apache POI.

Para generar la indexación por texto completo se utiliza el comando:

```
[dspace]/bin/dspace filter-media
```

Para la creación de miniaturas de gran calidad para PDF e imágenes, hay que instalar los complementos ImageMagick y Ghostscript en el servidor y luego habilitarlos en el archivo de configuración `dspace.cfg` de la siguiente manera:

Descomentar: en la propiedad `filter.plugins` agregar los complementos ImageMagick Image Thumbnail e ImageMagick PDF Thumbnail y comentar el complemento JPEG Thumbnail, debe quedar:

```
filter.plugins = PDF Text Extractor, HTML Text Extractor, Word Text Extractor, \
\
    PowerPoint Text Extractor, \
    Branded Preview JPEG, \
    ImageMagick Image Thumbnail, ImageMagick PDF Thumbnail
```

Además agregar en la propiedad

```
plugin.named.org.dspace.app.mediafilter.FormatFilter estos dos complementos
plugin.named.org.dspace.app.mediafilter.FormatFilter = \
org.dspace.app.mediafilter.PDFFilter = PDF Text Extractor, \
org.dspace.app.mediafilter.HTMLFilter = HTML Text Extractor, \
org.dspace.app.mediafilter.WordFilter = Word Text Extractor, \
org.dspace.app.mediafilter.PowerPointFilter = PowerPoint Text Extractor, \
org.dspace.app.mediafilter.BrandedPreviewJPEGFilter = Branded Preview JPEG, \
org.dspace.app.mediafilter.ImageMagickImageThumbnailFilter = ImageMagick
Image Thumbnail, \
org.dspace.app.mediafilter.ImageMagickPdfThumbnailFilter = ImageMagick PDF
Thumbnail
```

Descomentar la línea `filter.org.dspace.app.mediafilter.publicPermission` y agregar las propiedades `ImageMagickImageThumbnailFilter`, `ImageMagickPdfThumbnailFilter`
`filter.org.dspace.app.mediafilter.publicPermission = BrandedPreviewJPEGFilter,`
`ImageMagickImageThumbnailFilter, ImageMagickPdfThumbnailFilter`

Luego se debe ejecutar desde la línea de comando

Para la generación de miniaturas de imágenes:

```
[dspace]/bin/dspace filter-media -v -f -p "ImageMagick Image Thumbnail"
```

Para la generación de miniaturas de PDFs:

```
[dspace]/bin/dspace filter-media -v -f -p "ImageMagick PDF Thumbnail"
```

Para los archivos de video y sonido no hay complementos disponibles por defecto para la creación de miniaturas. La única forma de asignarles una miniatura es mediante la importación masiva, agregando la imagen miniatura e indicando en el archivo Contents que a la imagen se le asigne el bundle THUMBNAIL.

Estas tareas fueron agendadas a un cronjob para que se ejecute cada cierto tiempo, si no, no se podrá visualizar ninguna miniatura de los ítems añadidos posteriormente ni tampoco poder indexar por texto completo.

Configuración de Idioma

La versión estándar de DSpace tiene su propia traducción al español de todos los textos del repositorio realizada por un equipo ajeno a DSpace que se encuentra en la librería 'dspace-api-lang-5.3..jar' en la ruta `[dspace]/lib`.

Una vez estudiada la traducción, nos dimos cuenta de que contaba con numerosas traducciones erróneas y en algunos textos no se contaba con traducción al español y se desplegaba en el idioma por defecto (inglés).

Para poder realizar las traducciones, se descargó el archivo correspondiente al idioma español `messages_es.properties` y se colocó en el directorio `[dspace]/dspace-api/src/main/resources`. Para que este archivo se tome se debe especificar en el archivo de configuración el uso del idioma español como se especificó en la sección anterior de Configuración.

Estadísticas de DSpace

DSpace ofrece la posibilidad de mostrar estadísticas del acceso al repositorio como ser visualización de las páginas, búsquedas realizadas, autenticaciones de usuarios y errores ocurridos.

Hay dos tipos de estadísticas por Google Analytics y estadísticas SOLR.

Las estadísticas SOLR están activadas por defecto y en el archivo de configuración `dspace.cfg` mediante la propiedad `report.public` se indica si solo el administrador puede verlas o están accesibles públicamente.

Agregado de Filtro

Para el agregado de un filtro en la búsqueda se realizó las siguientes configuraciones:

En el archivo `[dspace]/config/spring/api/discovery.xml` se agregó el índice Versión con su metadato asociado en el cual para este campo elegimos `dc.description.version`

Para ello en el archivo se agregaron las líneas

```
<bean id="searchFilterVersion"
class="org.dspace.discovery.configuration.DiscoverySearchFilterFacet">
<property name="indexFieldName" value="version"/>
  <property name="metadataFields">
    <list>
      <value>dc.description.version</value>
    </list>
  </property>
</bean>
```

Y para agregar el campo `searchFilterVersion` a la búsqueda con filtro, se referencia al mismo de la forma (dentro de los beans con id `homepageConfiguration` y `defaultConfiguration`):

```
<property name="searchFilters">
  <list>
    <ref bean="searchFilterTitle" />
    <ref bean="searchFilterAuthor" />
    <ref bean="searchFilterSubject" />
    <ref bean="searchFilterIssued" />
    <ref bean="searchFilterVersion" />
  </list>
</property>
```

Luego de agregar este nuevo índice en el archivo debemos reindexar para que el motor de búsqueda tomé este nuevo cambio con el siguiente comando:

```
[dspace]/bin/dspace index-discovery -b
```

Versionado

Esta funcionalidad viene conjunto a DSpace pero no activada. Permite al usuario realizar más de una versión del mismo ítem, para ello debemos ingresar en el archivo [dspace]/config/modules/versioning.cfg

```
#####Versioning#####  
#Con esta opción en true se activa el versionado  
enabled=true  
#Esta opción en true activa el historial de un ítem, por ejemplo se puede ver  
cuántas versiones  
#de ese ítem existe y quien las realizó  
item.history.view.admin=true
```

Servidor OAI (FIC)

Para realizar la configuración de la sincronización se debe configurar en el archivo [dspace]/config/modules/oai.cfg

En este archivo se agrega la línea

```
bitstream.baseUrl = [url]/xmlui
```

Luego en una terminal corremos el siguiente comando para comenzar a exponer los ítems en el repositorio OAI

```
[dspace]/bin/dspace oai import
```

Este comando se agregó en un cron job para ejecutar diariamente.

Cliente OAI (BNU)

Para cada colección donde se desea realizar la cosecha se debe configurar los campos

- **OAI Provider** - Url del repositorio desde donde se cosecha la colección.
- **OAI Set Id** - Id de colección desde donde se realiza la cosecha
- **Metadata Format** - indica la opción de preferencia para obtener los items.

Luego correr en terminal lo siguiente

```
[dspace]/bin/dspace Harvest -r -e admin@gmail.com -c "idColeccion"
```

Estas tareas tanto fueron agendadas a un cronjob para que se ejecute todas las noches del lado de cada servidor, caso contrario no se tendrán los datos

ANEXO – REUNIONES

Reunión Inicial - 28/04/2015

Lugar: Facultad de Ingeniería - INCO

Participantes: Leonardo Casep, Florencia Borghini, Reyna Der Boghosian, Lisa Block de Behar, Regina Motz, Graciela Gargiulo, Arturo Rodríguez Peixoto, Rodrigo Echaniz, Camilo Freire, Maximiliano Basile, Federico Bello

Objetivo: plataforma adecuada para mostrar (exhibir) objetos digitales documentales, geolocalización, manuscritos, imágenes, etc.

Necesidades detectadas:

Definir un estándar de Metadatos que sea extensible, reutilizando los formatos ya definidos para el proceso de digitalización e ingreso de la información que se realiza actualmente.

Se entiende que la Biblioteca Nacional ya posee desarrollo de su Biblioteca en DSpace, por lo que se entiende que se debe realizar un repositorio para la FIC donde se debe sincronizar la información de publicaciones periódicas (en ambos repositorios la misma información).

Diseño del Repositorio FIC DSpace que almacenará en un formato único a definir, toda la información que se encuentra disponible actualmente en formatos y orígenes heterogéneos.

Migración de la base de datos OMEKA actualmente utilizada para almacenar la información digital a un repositorio DSpace destino de la misma.

Tener disponible online una plataforma adecuada para exhibir objetos digitales documentales, geolocalización, manuscritos, imágenes, etc. Asociar los datos que está generando la biblioteca nacional en Formato MARC21 para vincularlos con los documentos digitalizados que generan los becarios de la FIC

Debe manejar los archivos pesados lo cual es un cuello de botella por el ancho de banda que tiene la FIC.

Otros comentarios:

Gran énfasis en las funcionalidades de la Biblioteca Trapalanda.

Fuentes de datos

Publicaciones Periódicas: Consta de periódicos y materiales de apoyo como: artículos, libros y diccionarios. En su mayoría son periódicos y diarios. Estos datos son digitalizados por los becarios dentro de la Biblioteca Nacional y almacenados en

formato PDF en un directorio específico. Estos documentos ya digitalizados son posteriormente almacenados en una base de datos Omeka en formato Dublin Core.

Biblioteca Nacional : Actualmente los bibliotecólogos se encuentran generando los metadatos faltantes que serán vinculados a los documentos digitalizados por la FIC, los mismos son generados en formato MARC 21, por lo cual deberán ser transformados a formato Dublin Core para complementar los documentos digitalizados.

Autores del Uruguay: Los datos están en una wiki, donde quedan guardados en páginas HTML las cuales no son editables (no se incorpora actualmente este último pero podría ser incluido en futuro, se puede ver la opción de que se ingresen los datos en una wiki semántica, de modo de evitar tener que hacer un procesamiento).

Reunión - 30/04/2015

Lugar: Biblioteca Nacional

Participantes: Leonardo Casep, Reyna Der Boghosian, Rodrigo Echaniz, Graciela Gargiulo, Maximiliano Basile

Objetivo: Se observa el proceso que realizan a diario para subir archivos digitales a la plataforma OMEKA así como también que metadatos básicos utilizan. Se relevan los requerimientos de estos y que dificultades tienen a diario con esta forma de trabajo.

Reunión - 05/05/2015

Lugar: Facultad de Información y Comunicación

Participantes: Florencia Borghini, Reyna Der Boghosian, Lisa Block de Behar, Graciela Gargiulo, Arturo Rodríguez Peixoto, Camilo Freire, Maximiliano Basile, Federico Bello, Yonathan Benelli

Se aclara que la Biblioteca Nacional no tiene un repositorio, sino que se necesitaría para ingresar información proveniente de la Biblioteca y con las colecciones de Publicaciones Periódicas sincronizadas con el repositorio de FIC.

En primera instancia se comenzará con los Textos y luego se extenderá a los demás archivos (mapas, audio, etc.) Fuerte énfasis en permitir el ingreso de múltiples formatos.

Se habla la posibilidad de no duplicar la información de Publicaciones Periódicas en ambos repositorios, pero por convenio es necesario que la información esté duplicada.

Destacan nuevamente el repositorio Trapalanda, destacando su organización de información clara, buen buscador, fácil de usar, grafico, organización transparente. Además también destacan la organización jerárquica de este repositorio y del repositorio Chileno

Se comenta del proyecto colibrí del cual destacan el indexado.

La biblioteca se compromete a proporcionar la conversión del catálogo al metadato.

Dada la experiencia por parte de algunos integrantes de la BNU con el repositorio DSpace se destacó está desarrollado en Java, se va a necesitar complemento de visualización y optimización de archivos, tiene una gran comunidad de desarrolladores.

Se debe soportar la carga de archivos de más de 100MB, esto por defecto DSpace lo muestra en una nueva pantalla, se desea que mediante un complemento, como ser PDFJS, lo abra en la misma página, y que no cargue el PDF completo sino que progresivamente para evitar demoras por el tamaño.

Reunión 14/05/2015

Lugar: Biblioteca Nacional Informática

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Federico Bello, Camilo Freire

Se pide definir el flujo de trabajo y permisos, además de la estructura jerárquica La conversión de MARC21 a Metadatos no tendríamos que hacerlo nosotros, solo la importación masiva.

DSpace para la migración masiva utiliza un formato de carpetas, donde en cada una se debe poner un XML con los metadatos en formato DC, el archivo y la licencia. Mediante un script es posible generar toda esta estructura a partir de los archivos, donde cada archivo debe tener el nombre correcto.

En Omeka casi no se tienen metadatos ingresados, solo algunos mínimos como ser título y en algunos casos fecha.

Reunión 21/05/2015

Lugar: SECIU(Proyecto Colibri)

Participantes: Florencia Borghini, Reyna Der Boghosian, Federico Bello, Camilo Freire, Integrantes del Proyecto Colibri

Objetivo: Reunión para obtener información de su experiencia con la plataforma DSpace

Algunos detalles revelados de su experiencia con el repositorio DSpace:

- Customización con JSPUI, XMLUI lo encontraron más complicado y menos entendible.
- Se tiene definidos metadatos por tipo de material.
- Usan DC cualificado con campos propios, MODS les parece muy complejo

- Investigaron sobre preservación de metadatos pero les falta profundizar.
- Soporta versionado de ítems
- Es posible mapear objetos de una colección para poder hacer exposiciones por épocas por ejemplo. Se puede hacer 1 a 1 o hacer los insert en la base.
- No tuvieron problemas con los tamaños de los archivos, ya que son archivos de no muy gran tamaño.
- DSpace utiliza el framework Sprint
- Es posible sincronizar mediante OAI realizar la cosecha de metadatos y archivos pero no lo tuvieron que utilizar
- SWORD- Permite subir archivos desde otra aplicación sin pasar por Dspace.

Reunión 22/05/2015

Lugar: INCO - Facultad de Ingeniería

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Regina Motz

Objetivo: Reunión para definir cómo seguir

Investigar sobre Calidad de metadatos, metadatos excluyentes entre sí.

Extracción automática de metadatos de objetos digitales

Restricciones y tratamiento de los metadatos

Porque utilizar Dublin Core y no MARC21

Ver validaciones que se hacen actualmente, ver tiempo que les lleva para presentar mejoras y definir indicadores.

Reunión 09/05/2015

Lugar: Biblioteca Nacional Informática

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Federico Bello, Camilo Freire

Objetivo: Ver los procesos actuales de la Biblioteca Nacional

Actualmente el trabajo que se está realizando por parte de la Biblioteca Nacional es la Catalogación de metadatos.

Utilizan el software ExLibris de Aleph para el ingreso de los metadatos en formato MAR21. La catalogación se seguirá realizando en paralelo al Repositorio Digital, por lo que en conclusión no se tiene un proceso con el cual comparar mejoras, debido a que no hay proceso a sustituir.

Reunión 23/07/2015

Lugar: Biblioteca Nacional

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Federico Bello, Camilo Freire, Juan Maldini

Objetivo: Procedimiento de instalación de DSpace a pedido de Juan Maldini y mostrar posibles flujos de trabajo a proponer en la siguiente reunión

Se hizo la instalación de DSpace en el servidor de la Biblioteca Nacional, además el servidor se utilizará para mostrar el prototipo DSpace

Surgen algunos cambios del flujo a través de comentarios del proceso de trabajo de la FIC y Biblioteca Nacional

Reunión 30/07/2015

Lugar: Biblioteca Nacional

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Federico Bello, Camilo Freire, Lisa Block de Behar, Regina Motz, Arturo Rodríguez Peixoto, Maximiliano Basile, Yonathan Benelli

Objetivo: Presentar el flujo de trabajo en DSpace y sugerir flujos a utilizar

Se realizó una presentación donde se mostró

- 1- Diferentes alternativas de flujo de trabajo
- 2 - Control Metadatos
- 3 - Prototipo DSpace para eventuales pruebas de la plataforma.

Surge la idea de tener un campo para indicar la versión del ítem subido, por ejemplo un campo estado que tome los valores: Revisado o Versión Final

Por día se digitaliza aproximadamente 45 archivos, ver la posibilidad de migrar masivamente dado que no es viable subir de a uno los archivos.

Ambas instituciones se comprometieron a enviar documentos que especifiquen el esquema jerárquico de colecciones así como también los metadatos obligatorios para la descripción de los objetos digitales.

Por nuestra parte nos comprometimos a realizar un prototipo para pruebas así como también con una propuesta de interfaz gráfica.

Reunión 17/08/2015

Lugar: INCO - Facultad de Ingeniería

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Regina Motz

Objetivo: Reunión para mostrar avances

Se muestra la metadata definida, flujo de trabajo, el proceso a seguir y la propuesta de Autores del Uruguay.

Punto más importante es la sincronización entre los repositorios. Realizar una muestra de la sincronización para sacar dudas.

Datos redundantes en la sincronización

Autores del Uruguay proponer solución, ver el rol del editor. Control histórico de cambios. Qué rol cumplen los usuarios colaboradores para ellos.

Reunión 03/09/2015

Lugar: Biblioteca Nacional

Participantes: Florencia Borghini, Reyna Der Boghosian, Leonardo Casep, Federico Bello, Camilo Freire, Lisa Block de Behar, Graciela Gargiulo, Arturo Rodríguez Peixoto, Maximiliano Basile, Yonathan Benelli

Objetivo: Presentar el prototipo DSpace y realizar propuesta de Autores del Uruguay

- Se mostró un prototipo de DSpace con algunos cambios de interfaz de usuario donde mostramos varias funcionalidades (funcionalidades del administrador, agregar ítems a colecciones con y sin flujo de trabajo, crear/borrar colecciones, carga masiva, dar de baja momentánea a ítems y borrarlos)
- Se realizó una propuesta de autores del Uruguay, adjuntamos la misma para ajustes si es necesario.
- Se pidió agregar estadísticas de ingreso (como Google Analytics)

En paralelo con las pruebas de la plataforma, se pide que cada parte defina la interfaz de usuario del repositorio digital, esto es definir colores, textos, imágenes, links a otras páginas, etc.

Se entrega documento con capturas de la plataforma, indicando lo que deben definir.