

DIRECTRICES PARA PROXECTOS DE DIXITALIZACIÓN EN GALICIA

Biblioteca de Galicia
Marzo 2011

I. Introducción

As presentes directrices básicas non pretenden ser un manual de dixitalización senón simplemente un instrumento para o traballo en común e a coordinación dos proxectos que se están a realizar ou se realicen no futuro dentro do ámbito galego. Trátase dun documento ademais adaptado ás actuais características e panorama das normativas e directrices nacionais e internacionais vixentes para estes procesos, que requirirá de posteriores e previstas actualizacións que permitan adaptar o seu contido a futuros desenvolvementos no ámbito de aplicación das mesmas

A existencia polo demais de pautas e directrices xa publicadas e adoptadas en ámbitos nacionais e internacionais, aconsella non establecer novas normativas que poidan diverxer dos ámbitos comúns xa comunmente adoptados e establecidos. Desta forma o presente documento segue e respecta as directrices xa consolidadas, especialmente aquelas ás que a nosa comunidade e especialmente a Biblioteca de Galicia xa leva tempo adaptándose, como son as “Directrices para proxectos de dixitalización de coleccións y fondos de dominio público, en particular para aqueles custodiados en bibliotecas y arquivos” redactadas por un grupo de expertos da IFLA e a tradución ó castelán da cal pode consultarse (en versión actualizada a 2 de xullo de 2010) en

http://travesia.mcu.es/portaln/jspui/bitstream/10421/3342/1/pautas_digitalizacion1.pdf

Así como as diferentes normativas ou directrices técnicas xurdidas da participación da Biblioteca Dixital de Galicia (Galiciana) no proxecto Europeana

<http://version1.europeana.eu/web/europeana-project/technicaldocuments/>

especialmente as seguintes recomendacións:

European Semantic Elements

http://version1.europeana.eu/c/document_library/get_file?uuid=a830cb84-9e71-41d6-9ca3-cc36415d16f8&groupId=10602

Europeana Data Model

http://version1.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605

No apéndice sinálanse outros textos consultados e tidos en conta para a redacción destas directrices ou considerados de interese.

A función pois desta publicación está, por un lado, en divulgar no noso ámbito unha serie de conceptos que consideramos teñen unha básica importancia cara ó deseño dunha política común, tanto técnica como documental de dixitalización do noso patrimonio bibliográfico e por outro, concretar ou desenvolver aqueles aspectos desta política que poidan ser específicos da nosa comunidade.

Obxectivos

En termos xerais, as funcións ou desexables obxectivos destas directrices serían:

- Normalizar no posible e necesario as características técnicas da dixitalización que faciliten o intercambio de recursos, possibiliten as buscas integradas e simplifique o acceso libre e gratuïto a través da web e para todo tipo de usuarios a aqueles temas de interese para Galicia que se atopen libres de dereitos ou que os teñan cedidos ou abertos polos seus detentadores.
- Coordinar os esforzos de forma que non se dupliquen traballos e inversións, rendabilizando no posible os altos gastos que supoñen os procesos de dixitalización e creación de bibliotecas dixitais.

Trátase en calquera caso de procesos tanto técnicos como normativos en constante evolución e desenvolvemento. O presente documento pretende establecer obxectivos a medio e longo prazo pero coa constancia de que estes deben ser fixados de forma que poidan ó mesmo tempo adaptarse a posibles cambios ou melloras xurdidas no campo que atinxe estes procesos e ós elementos constitutivos dos mesmos. Deberán así mesmo actualizarse cunha periodicidade suficiente.

2. Proceso técnico

O proceso integral de dixitalización

A dixitalización como proceso documental asociado á conservación e difusión do patrimonio é un proceso no que inciden tanto elementos de carácter informático, referentes especialmente ás características técnicas do proceso e resultados da dixitalización en si, como documentais, que afectan ós datos (metadatos na terminoloxía actualmente en uso) que axuntamos ás imaxes para facilitar a súa descrición e posterior recuperación polos usuarios das bibliotecas dixitais.

Consideracións previas

A dixitalización é un proceso que debe estar precedido dun estudio suficientemente fundamentado acerca da conveniencia, rendibilidade e posibilidades do proxecto, e que so en caso de que os resultados así o aconsellen debe asumirse o seu desenvolvemento. Algúns aspectos fundamentais a ter en conta son:

- O **custo** do proceso de dixitalización é alto e esixe un laborioso traballo de planificación, execución e mantemento. Debe terse en conta adecuadamente o uso e demanda que vai ter o resultado da dixitalización, e se a inversión a realizar correspóndese co interese presente ou futuro dos materiais a expoñer.
- Ademais da inversión precisa para poñer en marcha unha biblioteca dixital é importante prever a súa **sustentabilidade**, polo que debe contar cunha previsión de mantemento.
- É importante non **duplicar** esforzos coa dixitalización de materiais xa accesible a través da rede. É esencial nese senso coordinar os esforzos e decisións, especialmente no campo das institucións públicas, cara a iniciación dun proxecto.
- Os múltiples **beneficios** da dixitalización deben ser tidos en conta, especialmente os derivados da súa **difusión e acceso** fácil, xeneralizado e, en xeral, gratuito. A biblioteca dixital debe crearse cunha idea clara de servizo ós usuarios ós que vai dirixida. Debe facilitar o seu acceso a través de normativas, especialmente as que afectan o seu uso y accesibilidade, facilitando ademais a súa integración con outras ofertas similares na rede. Non se pode esquecer tampouco o papel educativo que pode xogar. Tamén hai que ter en conta o seu papel na **preservación**, especialmente no feito de limitar (non necesariamente eliminar) o uso do material orixinal. A dixitalización non é sen embargo a forma máis barata nin con máis garantías de preservación.
- Especialmente importante é o tema dos **dereitos de autor**. Non debemos esquecer que moitos dos materiais bibliográficos e documentais susceptibles de dixitalización e de difusión a través da rede, están suxeitos a dereitos de autor, de copia ou doutro tipo, que debemos respectar ou negociar de tratarse o caso.

No capítulo 1 das “Directrices para Proxectos de Dixitalización...” da IFLA establécense pautas para a toma de decisión cara os posibles proxectos e para a selección de materiais a dixitalizar.

http://travesia.mcu.es/portaln/b/jspui/bitstream/10421/3342/1/pautas_digitalizacion1.pdf

2.1 Características e requisitos técnicos da conversión das imaxes a formato electrónico

Características técnicas: consideracións xerais

As “Directrices para Proxectos de Dixitalización...” da IFLA definen a “*imaxe dixital*” como unha “fotografía electrónica” convertida nun conxunto de elementos pictóricos (píxeles) e ordenada de acordo cunha relación predefinida de columnas e filas.

Estas imaxes dixitais son o resultado da conversión, a través de escanares, das imaxes procedentes dos obxectos ou documentos orixinais, que son “capturadas” e convertidas en ficheiros a través dos softwares normalmente incluídos nos citados escanares.

Neste proceso poden utilizarse diferentes formatos de almacenamento, e distintos niveis de calidade, en función da resolución, profundidade ou compresión. Os elementos máis importantes a ter en conta son:

Formato: Os datos recollidos polo software do escáner, deben ser almacenados de acordo cun formato, dos que existen no mercado diferentes ofertas. Os máis coñecidos e utilizados son TIFF, JPEG, GIF, PDF, etc. Algúns deles están pensados así mesmo para compresión das imaxes.

O formato máis usualmente requirido para imaxes de preservación sen compresión nas bibliotecas dixitais é o TIFF (“Tagged Image File Format”), actualmente na súa versión 6. Neste formato as imaxes poden acompañarse con etiquetas que conteñen información sobre as mesmas. Para imaxes de acceso aconséllanse sistemas de compresión sen perda. A Biblioteca de Galicia está a solicitar JPEG (“Joint Photographic Experts Group”).

Actualmente se está a demandar tamén o JPEG2000, que permite traballar con niveis de compresión maiores e que está a ser aceptada como formato de preservación

Resolución: Indica o número de píxeles utilizados para representar a imaxe, medidos en proporción á unha superficie determinada. Utilízase a unidade ppp (píxeles por pulgada ou puntos por pulgada equivalentes a “ppi = píxeles per inch”, ou “dpi = dots per inch”, respectivamente). Cantos máis píxeles ou puntos se utilicen para definir unha pulgada da imaxe, máis resolución terá esta.

A maior resolución a imaxe terá máis peso e tamén máis custo de produción, polo que é importante considerar para cada material a resolución máis adecuada. Para prensa non ilustrada, por exemplo, a resolución poderá ser menor que para cartografía, que precisa de maior capacidade de soportar ampliacións.

Profundidade ou definición: Indica o número de bits utilizado para representar cada píxel. Cantos máis bits se utilicen maior será o número de combinacións posibles para representar cores ou escalas de grises. Un único bit so permitiría representar o branco e o negro, 8 bits permitirían ata 256 gamas de grises e 24 bits, ata 16,7 millóns de cores.

O mesmo que coa resolución, cada tipo de material require unha definición óptima que permita explotar ó máximo as súas posibilidades de información e o tempo non desaproveite recursos innecesarios. Non se precisa unha mesma calidade para poder ler unha páxina de prensa ou dun libro por Internet que para reproducir un cartel a cor, un debuxo, gravado ou unha obra pictórica, nin para contar con arquivos que permitan unha reprodución facsímil.

Tipo de cor: branco e negro, escala de grises ou cor, en relación tamén coa definición escollida.

Compresión: Co fin de reducir o peso das imaxes, facilitando o seu almacenamento e a súa transmisión, hai desenvolvementos diversos programas que permiten a compresión (con algunha perda ou sen perda de calidade) das imaxes dixitalizadas.

Como norma xeral, e en relación cos citados valores técnicos no proceso de dixitalización é conveniente a produción de dous diferentes tipos de arquivo:

- Un **arquivo de conservación ou preservación**, con maior calidade (formato sen compresión, maior resolución e profundidade).
- Un **arquivo para difusión a través de Internet**, que permita a súa máis fácil transmisión pola rede, con menor resolución e profundidade ata onde sexa factible, e en formatos comprimidos preferentemente.

Requisitos técnicos: especificacións da Biblioteca de Galicia

Segundo o tipo de material e soporte e da función dos mesmos variarán os requisitos ou directrices concretas a ter en conta. A continuación indícanse unha serie de pautas que son nestes momentos as que están a demandarse tanto pola Biblioteca de Galicia como por diversas institucións, tanto nacionais como de fora de España, especialmente aquelas de especial peso no campo das bibliotecas dixitais. Non se presentan como normativa de obrigado cumprimento, pero aconséllanse en liñas xerais en canto conducen á desexada estandarización, e contribúen á busca e integración de recursos. Os datos, e especialmente a resolución, refírense sempre en relación o formato do orixinal (100% do seu tamaño).

Nos anexos A e B inclúense textos e enlaces a diferentes requisitos acerca de estes aspectos da dixitalización.

Tipo de documento	Archivo de conservación TIFF 6		Archivo de difusión JPEG, PDF ¹ , GIF		
	Resolución	Definición	Resolución	Definición	Compresión
Material textual: libro antigo e moderno, publicacións seriadas	Mínimo 400 ppp	24 bits color	JPEG: 150 ppp GIF: 72 ppp	24 bits color	JPEG: baixa GIF: estándar
Manuscritos	Mínimo 400 ppp	24 bits color	JPEG: 150 ppp GIF: 72 ppp	24 bits color	JPEG: baixa GIF: estándar
Mapas	Mínimo 400 ppp Aconséllase 500 ppp	24 bits color	300 ppp	24 bits color	Compresión baixa
Material gráfico: Fotografías, Gravados, Carteis, Debuxos...	Mínimo 400 ppp Aconséllase 500 ppp	24 bits color	300 ppp	24 bits color	Compresión baixa
Microfilms	Mínimo 400 ppp	24 bits color	300 ppp	24 bits color	Compresión baixa
Música: partituras, partituras anotadas, manuscritos de música	Mínimo 400 ppp	24 bits color	300 ppp	24 bits color	Compresión baixa

¹ O PDF adopta as características do formato original.

2.2. OCR (Reconocimiento Óptico de Caracteres)

Con cada imaxe de texto deberase acompañar un arquivo asociado de texto lexible por ordenador, elaborado a través dun programa de OCR que deberá contar con seguintes elementos para a xeración dos arquivos OCR²:

1. Un arquivo de texto OCR independente por cada imaxe de páxina.
2. Cada arquivo de texto OCR corresponderase coa imaxe que representa.
3. Texto en set de caracteres UTF-8.
4. No texto OCR non se deben recuperar elementos gráficos.
5. O texto OCR ordenado columna por columna (ou sexa nun orde de lectura natural).
6. O arquivo de texto OCR con cadro delimitador da coordenada de datos a nivel de palabra.
7. O OCR deberá ser conforme co esquema ALTO XML .
8. Todas as imaxes de páxinas deberán estar acompañadas por un arquivo ALTO XML que conteña o texto recoñecido.

Estes ficheiros entregaranse segundo o esquema ALTO XML. Analyzed Layout and Text Object é un esquema XML que especifica os metadatos técnicos do deseño e contido de recursos textuais físicos, tales como páxinas dun libro ou periódico. Xeralmente usase como esquema que estende a sección de metadatos administrativos en Metadata Encoding and Transmission Schema (METS). Nembargantes, pódese usar ALTO independentemente de METS.

A vinculación entre os ficheiros de imaxe e texto especificarase na sección correspondente do ficheiro METS que debe acompañar a cada obra dixitalizada.

2.3. OS METADATOS

Tan importante como o proceso de conversión das imaxes a estruturas dixitais é a asignación de terminados datos ás mesmas que permitan tanto a súa descrición como a posterior recuperación a través de buscas.

A normalización das normas ou estándares aplicables ós obxectos dixitais, está a sufrir importantes cambios, nunha evolución que está a conducir a novas estruturas de datos que habitualmente coñécense como “metadatos”.

Un dos principais obxectivos do plan de dixitalización da Biblioteca de Galicia e á de integrar as coleccións, tanto propias como doutras bibliotecas dixitais galegas, de forma que a través dunha simple busca poida accederse os fondos en conxunto de todas elas. O mesmo tempo tamén se persigue que os instrumentos de busca dende o exterior de Galicia

² Segundo as pautas de: “The National Digital Newspaper program (NDNP). Technical Guidelines for Applicants” (2010)

poidan acceder fácil e libremente ós nosos recursos e de forma suficientemente normalizada como para ofrecer resultados axustados ó seu interese.

Os novos desenvolvementos no campo da descrición de obxectos dixitais permiten ademais avanzar por novos camiños como son os de integración de obxectos de diferente procedencia ou fin (arquivos, bibliotecas, museos ou proxectos educativos e institucionais online) e así mesmo tamén no imprescindible perfeccionamento dos sistemas de descrición e busca, desenvoltos ó carón da web semántica. Faise polo tanto imprescindible formular unha maior coordinación e mesmo unificación nos casos que sexa posible.

Actualmente son varios os estándares aplicables ós obxectos dixitais, non existindo un único que recolla todos os ámbitos aplicables á descrición destas imaxes. Por esa razón faise preciso aplicar diferentes esquemas, que polo demais actualmente intégranse a través de linguaxes como o XML en estruturas internacionalmente aceptadas como o de METS.

De entre tódolos esquemas e recomendacións relativos a metadatos existentes actualmente hai varias normas que se están a establecer como estándares máis ou menos aceptados de forma global no mundo das bibliotecas dixitais. A Biblioteca de Galicia nos seus proxectos de dixitalización está a adoptar estes estándares aceptados pola meirande parte das bibliotecas dixitais nacionais e internacionais e son estes os que se propoñen como pautas para proxectos a levar a cabo no resto de Galicia.

Os esquemas de metadatos

As actuais pautas da Biblioteca de Galicia establecen como modelo de metadatos a entrega de ficheiros XML estruturados segundo o esquema Metadata Encoding & Transmission Standard (METS), actualmente na súa versión 1.9. <http://www.loc.gov/standards/mets/METSPrimerRevised.pdf>

METS é un esquema de metadatos desenvolto por iniciativa da Digital Library Federation e mantido por la “Network Development and MARC Standards Office” da Biblioteca do Congreso (LC) norteamericana (<http://www.loc.gov/standards/mets/>). Utiliza como linguaxe o XML Schema do World Wide Web Consortium (W3C).

O METS aporta unha metaestrutura que contempla as diversas categorías de datos significativos para unha imaxe dixital e ó mesmo tempo permite a integración dentro do propio rexistro ou a través de enlaces a arquivos externos, de outros esquemas de metadatos.

Ademais das recomendacións xerais para a utilización do esquema METS establecidas pola “Network Development and MARC Standards Office” da LOC, a Biblioteca de Galicia sigue as recomendacións particulares realizadas polo Ministerio de Cultura para a inxesta e preservación de recursos dixitais³.

³ Documento en acceso web proximamente.

Os ficheiros METS de la Biblioteca de Galicia elabóranse conforme os seguintes perfíles rexistrados na “Network Development and MARC Standards Office” da LOC e aprobados pola METS Editorial Board:

Para Monografías: **Model Paged Text Object Profile**

<http://www.loc.gov/standards/mets/profiles/00000005.html>

Para Publicacións Periódicas: **METS Profile for Historical Newspapers**

<http://www.loc.gov/standards/mets/test/ndnp/00000010.html>

Requisitos mínimos dos ficheiros de metadatos establecidos pola Biblioteca de Galicia.

Seguindo as normativas anteriormente sinaladas, especialmente as definidas pola Subdirección Xeral de Coordinación Bibliotecaria do MCU, a Biblioteca de Galicia recomenda que os arquivos METS deben de ter, como mínimo, as seguintes seccións:

1. Etiqueta raíz (mets)
2. Cabeceira (metsHdr)
3. Sección de metadatos decriptivos (dmdSec)
4. Sección de metadatos administrativos (amdSec)
5. Sección de arquivos (fileSec)
6. Sección de mapa estrutural (structMap)

1. Etiqueta raíz

Nesta etiqueta, ademais das mencións de namespaces e URIs necesarias, hase de incluír un atributo LABEL que vai identificar o arquivo. O valor do devandito atributo será o título da obra (ou as primeiras palabras do título se este é moi longo). De ser o caso de que houbese varios mets co mesmo valor deste atributo LABEL (varios exemplares dunha obra ou partes dunha obra...), diferenciaranse os devanditos valores dalgún xeito, por exemplo engadindo un número de parte ou a sinatura do exemplar.

Exemplo de etiqueta raíz:

```
<mets ID="IDI" OBJID="1001287"
LABEL="Apertura de la Escuela Normal de la Provincia de Lugo"
PROFILE="http://www.loc.gov/standards/mets/profiles/00000005.html"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:marc="http://www.loc.gov/MARC21/slim"
xsi:schemaLocation=http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd
http://www.loc.gov/mix/ http://www.loc.gov/standards/mix/mix02/mix02.xsd
http://cosimo.stanford.edu/sdr/metsrights/ http://cosimo.stanford.edu/sdr/metsrights.xsd
http://www.loc.gov/standards/premis/v2 http://www.loc.gov/standards/premis/v2/premis-
v2-0.xsd">
```

2. Cabeceira (metsHdr)

Na cabeceira ha de haber dúas etiquetas < altRecordID > que nos servirán para asociar os grupos de imaxes aos rexistros correspondentes:

a) Etiqueta <altRecordID> con atributo TYPE="Nº control". O valor do atributo TYPE será o número de control *ou o código interno de identificación* asignado ó rexistro bibliográfico por la organización que o crea, utiliza ou distribúe. Por exemplo:

```
<mets:altRecordID TYPE="Nº control">PFB002201056</mets:altRecordID >
```

b) Etiqueta < altRecordID > con atributo TYPE="Institución e sinatura". O valor do atributo TYPE neste caso será o código de biblioteca (que nos rexistros debe aparecer no campo 852 \$a) e a sinatura do exemplar (que nos rexistros debe aparecer no campo 852 \$j). Ambos os dous valores deben estar separados por un espazo en branco. É importante que nos códigos de biblioteca non haxa espazos en branco entre as súas letras, iso produciría erros. Por exemplo:

```
<mets:altRecordID TYPE="Institución e signatura"> C(S)-CBG BL10784  
</mets:altRecordID>
```

Onde C(S)-CBG é o código da biblioteca (que aparece nos rexistros no 852 \$a) e BL10784 é a sinatura do exemplar (que aparece nos rexistros no 852 \$j). Cada exemplar debe ter a súa propia sinatura, estas non deben repetirse.

Nota: de entregar só rexistros bibliográficos, neles estará o mencionado campo 852. Se se entregaron rexistros independentes de exemplar (de fondos) neles é onde deberá estar o campo 852.

Para a asignación do código da biblioteca recoméndase a utilización da codificación establecida para o Catálogo Colectivo do Patrimonio Bibliográfico.

3. Sección de metadatos descritivos (dmdSec)

-Debe estar identificada por un atributo ID. Por exemplo:

```
<mets:dmdSec ID="DMD001">
```

-Como requisito mínimo débese incluír nesta sección de metadatos descritivos o rexistro bibliográfico en formato Marc21 XML.

<http://www.loc.gov/standards/marcxml///>

Pódese engadir ademais o rexistro de fondos cando se dispoña del (en formato Marc21 XML de fondos).

<http://www.loc.gov/marc/holdings/echdspa.html>

-Pódense incluír ademais metadatos noutros formatos. Para a participación en proxectos europeos é obrigatorio a implementación do formato ESE na súa versión 3.3.1 e en adiante o será EDM (Europeana Data Model).

Cando o ficheiro de metadatos correspóndese cunha publicación periódica recoméndase a utilización do esquema MODS <http://www.loc.gov/standards/mods/>

4. Sección de metadatos administrativos (amdSec)

-Nesta sección pódense incluír metadatos PREMIS para preservación. A obtención destes metadatos farase seguindo o Dicionario de Datos PREMIS Versión 2.0.

<http://www.bne.es/es/Micrositios/Guias/DiccionarioPremis/VERSION2.0/>

5. Sección de arquivos (fileSec)

Nesta sección relaciónanse e enumeran ordenadamente os arquivos que van ser ou puideran ser cargados.

- Debe haber unha etiqueta <fileGrp> para cada unha das versións do obxecto dixitalizado. Por exemplo, se se vai cargar unha soa imaxe ou un só grupo de imaxes haberá un só <fileGrp>, se se pretende cargar a obra en varios formatos (por exemplo jpg e pdf), haberá un <fileGrp> para cada formato de imaxe.
- Os <fileGrp> deberán ter o atributo USE cos seguintes valores en función dos diferentes formatos de imaxe do obxecto dixitalizado que representen:

Para as imaxes TIFF: <fileGrp USE= "archive">

Para as imaxes JPEG: <fileGrp USE= "reference">

Para as imaxes GIF: <fileGrp USE= "thumbnail">

Para as imaxes PDF: o seu <fileGrp> non precisa do atributo USE

Será requisito mínimo obrigatorio a existencia dun <fileGrp> co atributo USE con valor "reference".

- Aniñaranse dentro de cada <fileGrp> tantas etiquetas <file> como obxectos pertencen ao grupo.

Se se trata dun conxunto de imaxes jpg, haberá un <file> para cada imaxe.

```
<file ID="FID0" MIMETYPE="image/jpeg" SEQ="1" OWNERID="11148168"
DMDID="DMI">
```

```
  <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://galiciana.bibliotecadegalicia.xunta.es/i18n/catalogo_imagenes/imagen
_id.cmd?idImagen=11148168"/>
</file>
```

```
<file ID="FID1" MIMETYPE="image/jpeg" SEQ="2" OWNERID="11148169"
DMDID="DMI">
```

```
<FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://galiciana.bibliotecadegalicia.xunta.es/il8n/catalogo_imagenes/imagen
_id.cmd?idImagen=11148169"/>
</file>
```

De tratarse dun só pdf haberá un só <file>

```
<mets:file MIMETYPE=" application/pdf " ID="FPID0">
  <mets:FLocat LOCTYPE="URL" xlink:href=" ./alto/1918-1.pdf" />
</mets:file>
```

- Cada unha destas etiquetas <file> ha de estar definida cun atributo ID. Tamén se acompañará dun atributo MIMETYPE cun valor que indique o formato do obxecto (image/jpeg; application/pdf; image/tiff...).
- Aniñada en cada unha das etiquetas <file> ha de haber unha etiqueta <FLocat> cos seguintes atributos obrigatorios LOCTYPE (con valor "URL") e xlink:href.
- O valor do atributo xlink:href poderá ser:

a) **Unha dirección absoluta ao obxecto.** Por exemplo, se as imaxes atópanse nun disco duro externo, o valor deste atributo sería a dirección absoluta á imaxe correspondente.

P. ex: G://BdG/proxecto_dix_2010/img_0006.jpg
<D:\Monografias\references\7\imagenes>

b) Outra opción sería poñer o texto `file://($$SOPORTEI $$) /` antes do nome da carpeta contenedora que contén todos os grupos de imaxes (a carpeta co nome ou código da biblioteca que contén en resto de carpetas).

Por exemplo:

```
< file ID="FID5" MIMETYPE="image/jpeg" SEQ="1" DMDID="DMI " >
< FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="file://( $$SOPORTEI $$ ) /NOVA
GALICIA reference/Nova galicia 16-17.0005.jpg"/>
</File>
```

Onde NOVA GALICIA reference é o nome da carpeta contenedora.

6. Sección de mapa estrutural (structMap)

- Haberá tantos <structMap > como estruturas se vaian cargar. Se se vai cargar un grupo de imaxes jpg por un lado e un pdf por outro, haberá dous mapas estruturais.
- A etiqueta <structMap > ten de levar o atributo TYPE co valor "physical".

- Ha de haber unha primeira etiqueta <div> que identifique a obra co atributo TYPE (con valor "libro, mapa... "). Este primeiro <div> tamén ha de levar un atributo LABEL co mesmo valor que o atributo LABEL da etiqueta raíz.
- Tamén ha de levar un atributo ORDER, para indicar a orde do <div> dentro deste nivel de aniñamento.
- O resto de etiquetas <div> dependerán da estrutura da obra que se está a cargar. Deberán incluírse os niveis de <div> necesarios para expresar esta estrutura e en cada uns destes <div> deberán incluírse os atributos ORDER, para indicar o número de orde do obxecto (mediante este atributo ordénanse as imaxes) e LABEL, que é o texto que se vai visualizar asociado ao obxecto, por exemplo un número de páxina ou de capítulo...
- A cada obxecto particular (imaxe jpg, pdf, tiff...) corresponderalle unha etiqueta <fptr> co atributo obrigatorio FILEID. O valor deste atributo será igual ao valor do atributo ID da etiqueta <file> (da sección de arquivo) correspondente ao devandito obxecto dixital. Esta relación é indispensable para que a aplicación saiba que imaxe das enumeradas na sección de arquivo corresponde a cada parte da estrutura.
- Cada unha destas etiquetas <fptr> estará aniñada nunha etiqueta <div> onde se indicase mediante os atributos ORDER e LABEL, a orde do obxecto respecto aos seus irmáns e o texto que se visualizará asociado ao devandito obxecto (por exemplo, o número de páxina).

```

<mets:structMap ID="SMI" TYPE="PHYSICAL">
<mets:div ORDER="1" LABEL="NOVA GALICIA" TYPE="book">
<mets:div ORDER="1" LABEL="[CUBIERTA]" TYPE="page">
<mets:fptr FILEID="FMID1"/>
<mets:fptr FILEID="FRID20"/>
</mets:div>
<mets:div ORDER="2" LABEL=" " TYPE="page">
<mets:fptr FILEID="FMID2"/>
<mets:fptr FILEID="FRID21"/>
</mets:div>
<mets:div ORDER="3" LABEL=" " TYPE="page">
<mets:fptr FILEID="FMID3"/>
<mets:fptr FILEID="FRID22"/>
</mets:div>
<mets:div ORDER="4" LABEL=" " TYPE="page">
<mets:fptr FILEID="FMID4"/>
<mets:fptr FILEID="FRID23"/>
</mets:div>
...
</mets:structMap>

```

```
<mets:structMap ID="SM2" TYPE="PHYSICAL">
<mets:div ORDER="1" LABEL="NOVA GALICIA " TYPE="book">
<mets:div ORDER="1" LABEL="NOVA GALICIA " TYPE="page">
<mets:fptr FILEID="FRID38"/>
</mets:div>
</mets:div>
</mets:structMap>
```

Encarpetado das imáxes e dos ficheiros METS

A Biblioteca de Galicia recomenda que os resultados dos proxectos de dixitalización se encarpenen segundo o seguinte esquema:

- Crearase unha carpeta por cada formato de imaxe resultante da dixitalización (TIFF, JPEG, GIF, PDF).
- Dentro de cada una desas carpetas crearanse tantas carpetas como obras dixitalizadas obxecto de ese proxecto. Ditas carpetas identificaranse mediante o código da biblioteca e sinatura da obra. Cada unha delas conterá tódalas imaxes desa obra.
- Os arquivos METS deberán agruparse nunha carpeta independente.

Autoridades

De igual modo, entregarase un ficheiro XML estruturado segundo o esquema MADS no que se detallarán os rexistros de autoridades correspondentes a cada publicación (persoas, entidades e títulos), con todas as variantes e referencias que poidan desprenderse da catalogación completa da totalidade do exemplar dispoñible. Con estes ficheiros XML MADS constituirase o ficheiro de autoridades da Biblioteca de Galicia.

Repositorios dixitais: Protocolo OAI-PMH e Directrices DRIVER / OpenAire para provedores de contidos

Dende outubro do ano 2008 a Biblioteca de Galicia mantén en funcionamento “Galiciana”, repositorio OAI accesible na dirección <http://www.galiciana.bibliotecadegalicia.xunta.es>. Este repositorio permite recolectar rexistros mediante o Protocolo OAI-PMH versión 2.0. Os seus rexistros están codificados como mínimo en Dublin Core sen cualificar, aínda que Galiciana traballa con case tódolos esquemas de metadatos máis utilizados actualmente (OAI_MARC, OAI_DC, MARC 21, ESE e DIDL).

Hoxe en día o repositorio de Galiciana é provedor de datos nos principais proxectos de agregadores de contidos de bases de datos de coleccións dixitais a nivel estatal (HISPANA), europeo (EUROPEANA) e mundial (OAISTER). Tamén está presente nos principais rexistros internacionais de provedores de datos como “Registry of Open Access Repositories (ROAR)”. Todo elo é posible xa que Galiciana cumpre cas principais directrices e protocolos de interoperabilidade vixentes actualmente.

Protocolo OAI-PMH versión 2.0.

A formulación esencial dos arquivos abertos é permitir o acceso aos materiais na web por medio de repositorios que, interoperando uns con outros, permitan o intercambio de metadatos, o seu almacenamento e publicación. Perseguindo este obxectivo "*a Open Archives Initiative desenvolve e promove normas de interoperabilidade que teñen por obxecto facilitar a difusión eficaz dos contidos*".

O Protocolo OAI para a Recolección de Metadatos (OAI-PMH), define un mecanismo para a recolección de rexistros que conteñen os metadatos dos repositorios. O OAI-PMH ofrece aos Provedores de Datos unha opción técnica sinxela para poñer os seus metadatos a disposición de servizos baseados nos estándares abertos HTTP (Hypertext Transport Protocol) e XML (Extensible Markup Language). Os metadatos que son reunidos poden estar en calquera formato establecido por unha comunidade (ou por calquera conxunto específico de Provedores de Datos e Provedores de Servizos), con independencia de que establecesen o Dublin Core non cualificado para proporcionar un nivel básico de interoperabilidade.

Directrices DRIVER

DRIVER, "Dixital Repository Infrastructure Vision for European Research" (Visión de infraestrutura de repositorios dixitais para a investigación europea), é un proxecto realizado por un consorcio financiado pola Unión Europea que está a crear un marco de traballo tecnolóxico e organizativo para implementar unha capa paneuropea de datos, que permita o uso avanzado dos recursos de contido no ámbito da investigación e a educación superior. DRIVER desenvolve unha infraestrutura de servizos e unha infraestrutura de datos. Ambas as dúas están concibidas para orquestrar os recursos e os servizos existentes na rede de repositorios.

Directrices OpenAire 1.1

As Directrices OpenAIRE 1.0, orientarán os administradores de repositorios para definir e implementar as súas políticas locais de administración de datos cumprindo coas demandas da Comisión Europea sobre Acceso Aberto. Ademais, cumprirán cos requisitos técnicos da infraestrutura de OpenAIRE, deseñada para soportar e monitorizar a implementación do piloto 7PM OAI.

Ao implementar estas Directrices para administradores de repositorios, facilítase aos autores que depositan as súas publicacións no repositorio cumprir cos requisitos da CE sobre Acceso Aberto.

Para desenvolvedores de software de repositorios, as Directrices son unha guía para engadir en futuras versións do software, funcionalidades que axuden aos autores con financiamento da CE para as súas investigacións.

ANEXO A

Relación de directrices e normativas propuestas e consultadas

Guidelines for preparing your historical newspapers for online acces.

http://www.oclc.org/services/brochures/212553usf_contentdm_guidelines_historical_newspapers.pdf

CDL (California Digital Library) Guidelines for Digital Objects.

<http://www.cdlib.org/inside/diglib/guidelines/>

<http://www.cdlib.org/inside/diglib/guidelines/bpgimages/>

Estándares de digitalización. Elementos mínimos. Propuesta del Consorcio de Bibliotecas Universitarias de Catalunya (actualización octubre 2010).

http://www.recercat.net/bitstream/2072/97497/3/1010EstandardsRequerimentsMinims_Public spa.pdf

Framework of Guidance for Building Good Digital Collections. 3ª ed. 2007. (Edición electrónica 2008)

<http://www.niso.org/publications/rp/framework3.pdf>

“Directrices para proyectos de digitalización de colecciones y fondos de dominio publico, en particular para aquellos custodiados en bibliotecas y archivos”

<http://hdl.handle.net/10421/3342>

<http://travesia.mcu.es/portalnjb/jspui/bitstream/10421/3342/1/PAUTASDIGIT010052011.pdf>

Europeana

<http://version1.europeana.eu/web/europeana-project/technicaldocuments/>

European Semantic Elements

<http://hdl.handle.net/10421/2403>

http://version1.europeana.eu/c/document_library/get_file?uuid=a830cb84-9e71-41d6-9ca3-cc36415d16f8&groupId=10602

Europeana Data Model

<http://hdl.handle.net/10421/3861>

http://version1.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605

National Digital Newspaper Program

<http://www.neh.gov/grants/guidelines/ndnp.html>

The National Digital Newspaper Program (NDNP). Technical Guidelines for Applicants (2009).

http://www.loc.gov/ndnp/guidelines/archive/NDNP_201113TechNotes.pdf

Network Development and MARC Standards Office – Library of Congress

<http://www.loc.gov/standards/mets/>

Metadata Encoding & Transmission Standard (METS), actualmente na súa versión 1.9.

<http://www.loc.gov/standards/mets/METSPrimerRevised.pdf>

Technical Metadata for Optical Character Recognition: Analyzed Layout and Text Object (ALTO)

<http://www.loc.gov/standards/alto/>

Model Paged Text Object Profile

<http://www.loc.gov/standards/mets/profiles/00000005.html>

METS Profile for Historical Newspapers

<http://www.loc.gov/standards/mets/test/ndnp/00000010.html>

Formato Marc21 XML.

<http://www.loc.gov/standards/marcxml///>

Formato Marc21 XML de fondos.

<http://www.loc.gov/marc/holdings/echdspa.html>

Esquema MODS

<http://www.loc.gov/standards/mods/>

The Dublin Core Metadata Initiative (DCMI)

<http://dublincore.org/>

Diccionario de Datos PREMIS Versión 2.0.

<http://www.bne.es/es/Micrositios/Guias/DiccionarioPremis/VERSION2.0/>

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).
Protocol Version 2.0

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

OAI para principiantes - tutorial en línea del Open Archives Forum

<http://hdl.handle.net/10421/1823>

<http://travesia.mcu.es/portaln/jspui/html/10421/1823/intro.htm>

Open Archives Initiative - Object Reuse and Exchange (OAI-ORE)

<http://www.openarchives.org/ore/>

Directrices DRIVER 2.0. Directrices para proveedores de contenido - Exposición de recursos textuales con el protocolo OAIPMH [Noviembre 2008].

http://www.driver-support.eu/documents/DRIVER_2_0_Guidelines_Spanish.pdf

Directrices OpenAire 1.1

http://www.recolecta.net/buscador/documentos/OpenAIRE-Guidelines_v1-1_es.pdf

Digitisation guidelines – National Library of Australia

<http://www.nla.gov.au/digital/standards.html>

Manual de Autoridades. Biblioteca Nacional de España (2010)

[http://www.bne.es/es/Micrositios/Guias/Autoridades/resources/docs/Manual de Autoridades.pdf](http://www.bne.es/es/Micrositios/Guias/Autoridades/resources/docs/Manual_de_Autoridades.pdf)

Resource Description and Access (RDA) (oct. 2009)

<http://hdl.handle.net/10421/1561>

<http://www.rda-jsc.org/rda.html>