

Archivo digital y servicios sin papel



Estándar Técnico

Arquivo Electrónico do Patrimonio de Galicia

Acordos de envío – SIP, AIP y DIP



PROGRAMA
COOPERACIÓN TRANSFRONTERIZA
ESPAÑA ~ PORTUGAL
COOPERAÇÃO TRANSFRONTEIRIÇA
2 0 0 7 - 2 0 1 3



UNIÓN EUROPEA
FONDO EUROPEO DE
DESENVOLVEMENTO
REXIONAL
"Unha maneira de facer Europa"

CONTROL DE VERSIÓNS E DISTRIBUCIÓN DO DOCUMENTO

Nome	Acordos de envío – SIP, AIP e DIP	Versión	1.2
Código			
Elaborado por	Proxecto ARPAD	Data	15-10-2015
Validado por		Data	
Aprobado por		Data	

REXISTRO DE CAMBIOS

Versión	Causa da nova versión	Data de aprobación

Índice

1	Introdución	4
1.1	Propósito e alcance.....	4
2	Identificación, acordo de envío e SIP	4
2.1	Identificación	4
2.2	Acordo de envío.....	5
2.3	SIP (Submission Information Package) – Condicións comúns.....	5
2.3.1	Condicións específicas para a área de preservación Arquivo Electrónico Administrativo	6
2.3.2	Condicións específicas para as áreas cubertas polo entregable Dixitalización do Patrimonio Documental.....	6
2.4	AIP e DIP	7
3	Referencias de aplicación.....	7

1 Introducción

1.1 Propósito e alcance

O seguinte documento serve como especificación das condicións en que os obxectos dixitais serán preservados no sistema de preservación do Arquivo Dixital Integrado de Galicia.

As condicións aquí expresadas serán de cumprimento e rexistro obrigatorios, e quedarán reflectidas a través do formulario de rexistro que acompaña a este documento. Todos os formularios cumprimentados para cada unha das inxestións de obxectos dixitais preservados serán custodiados nun repositorio centralizado que garde a información referente ao sistema de preservación.

Os formularios cumprimentados gardaranse asinados, xa sexa manual ou dixitalmente, protexidos contra variacións, e custodiados en formato PDF/A, para a súa posterior consulta.

2 Identificación, acordo de envío e SIP

2.1 Identificación

A identificación da inxestión de cada obxecto dixital realizarase a través da cumprimentación do formulario de rexistro. Sendo obrigatorio que se cubran todos os campos especificados.

A continuación móstrase un exemplo orientativo de uso.

Entidade	Cardinalidade	Descrición
Identificador de proceso	Inxestión 2015-0023	Número correlativo ou identificador do proceso de inxestión.
Descrición, rango ou referencia de contidos	É.GA.36000.AFP_XXXXX É.GA.40320.AFP_XXXXX	a Referencia de códigos ou descrición do contido do paquete de inxestión.
Axente	Deputación de Pontevedra	Organización creadora
Actividade	Preservación bibliográfica	Función do axente
Regulación		Marco normativo de xeración
Relación	Proxecto Expte. 54/2011	Contexto de creación e uso
Data de elaboración	28/09/2012	Data de xeración do paquete de inxestión
Data de inserción	13/10/2016	Data de inserción (primeiro intento) do paquete de inxestión

Operador	Manuel Martínez González	Operador que realiza a inxestión
Área de preservación	Material textual manuscrito	Área de preservación asignada en sistema de preservación
Plan de preservación	Material textual manuscrito, versión 1	Plan de preservación activo na área de preservación en momento de inxestión
Esquema de metadatos	EAD3	Para catalogación dos obxectos dixitais

2.2 Acordo de envío

Todos os obxectos dixitais a ser inxestados no sistema de preservación deben cumprir as seguintes condicións:

1. Cada obxecto dixital (nativo ou dixitalizado) está contido nun cartafol, que pode conter ficheiros e subcarpetas con ficheiros .
2. Todos os ficheiros foron comprobados contra infección de virus.
3. Cúmrese o formato SIP especificado no seguinte apartado para a área de preservación na que se inxeste o obxecto.
4. O obxecto dixital foi verificado en orixe previamente, e seguíronse os procesos de xeración/extracción previstos para o proxecto no que se engloba.
5. Todos os ficheiros de cada obxecto dixital poden ser validados segundo a caracterización do formato que se derive da súa composición utilizando ferramentas comúns.

2.3 SIP (Submission Information Package) – Condicións comúns

A estrutura do SIP (*Submission Information Package*) para cada obxecto dixital a inxestar cumprirá as condicións comúns para todas as áreas de preservación. Que son:

1. Non existen carpetas baleiros.
2. Cada cartafol creado contén polo menos un ficheiro ou outro cartafol.
3. O tamaño máximo de cada ficheiro individual é de 1Gb.
4. Os caracteres de todos os nomes de ficheiros e carpetas son válidos nun esquema basee UTF-8, para o sistema de almacenamento provisto para o sistema.
5. O ficheiro de metadatos atópase situado nunha subcarpeta *\mets* que colga da cartafol raíz do obxecto.
6. O ficheiro de metadatos do obxecto dixital segue a estrutura do estándar METS.
7. O ficheiro de metadatos contén un campo de metadatos que contén un dato único que identifica o obxecto dixital, e que se define como *identificador*.
8. O *identificador* do obxecto é único para o universo de obxectos dixitais a preservar; a súa repetición no sistema de preservación implicaría unha versión igual ou diferente do obxecto.
9. O nome do ficheiro de metadatos é da forma *identificador.xml*.
10. O nome do cartafol do obxecto dixital é da forma *identificador* ou, se aplica para os ficheiros cubertos polo documento entregable *Dixitalización do Patrimonio Documental*, da forma establecida no punto 3.3 *Estrutura de directorios* para o terceiro nivel -> “Por regra xeral, a identificación incluírá o código do ficheiro de orixe e a signatura do documento. Ambos os elementos

estarán separados por un guión baixo. Nos casos nos que o arquivo non conte cunha signatura única para todos os seus fondos poderá engadirse algún elemento que facilite a identificación, por exemplo, o código de clasificación do arquivo ou a introdución de letras diferenciadoras dos diferentes tipos de instalación: c para caixas, l para libros, p para planeiros.”.

Así mesmo, a estrutura do SIP (*Submission Information Package*) para cada obxecto dixital a inxestar cumprirá as condicións específicas para a súa área de preservación.

2.3.1 **Condicións específicas para a área de preservación Arquivo Electrónico Administrativo**

1. O ficheiro de metadatos inclúe polo menos unha sección *dmdSec/mdWrap* cun esquema de metadatos ARPAD, para catalogación do obxecto.
2. Os metadatos asociados ao obxecto conteñen un campo *ARPAD.NODEREF* do esquema de metadatos ARPAD, correspondente ao EMGDE2.1, que se tomará como *identificador* do obxecto.
3. Os metadatos asociados ao obxecto conteñen polo menos os campos obrigatorios segundo o documento [*ARPAD – Arquivo Electrónico Administrativo*] *Definición metadatos*, de especificacións de metadatos da Xunta.

2.3.2 **Condicións específicas para as áreas cubertas polo entregable Dixitalización do Patrimonio Documental**

O documento entregable *Dixitalización do Patrimonio Documental* establece unha estrutura de obxecto igual para todo o material correspondente ás áreas de material: textual manuscrito, textual impreso, textual fotocopias, fotográfico, gráfico, e gráfico de gran tamaño.

Onde as diferenzas están no perfil METS seguido, que non afecta o plan de preservación a configurar no sistema de preservación, e aos directorios opcionais que resultan de xerar un PDF auxiliar e/ou realizar un OCR sobre o material referenciado.

Por ser comúns, establécense a continuación as condicións específicas para todas esas áreas.

1. O ficheiro de metadatos inclúe polo menos unha sección *dmdSec/mdWrap* cun esquema de metadatos EAD3, para catalogación do obxecto, segundo o perfil especificado no documento entregable *Dixitalización do Patrimonio Documental*.
2. O esquema de metadatos EAD3 contén un campo *<unitid label="Número de identificación do sistema">*, cuxo valor se tomará como *identificador* do obxecto.
3. Todos os ficheiros do obxecto están nomeados como *identificador_XXXXX.extensión*. Onde *XXXXX* indica un número correlativo de 5 díxitos que empeza en 00001 para cada cartafol, e *extensión* é a extensión propia do tipo de ficheiro que vai no cartafol.
4. Existe un cartafol *tiff*, que contén todos os ficheiros TIFF do obxecto.
5. Existe un cartafol *jpeg*, que contén todos os ficheiros JPEG do obxecto.
6. Existe unha cartafol miniaturas, que contén miniaturas en JPEG de todos os ficheiros JPEG do obxecto.
7. Pode existir unha cartafol *pdf*, que contén un ou máis ficheiros PDF/A.

8. Pode existir un cartafol *alto*, que contén todos os ficheiros de aplicación de OCR sobre cada imaxe en formato TIFF, e cuxos ficheiros teñen o mesmo nomee que estes con extensión xml.
9. Se existe o cartafol *alto* existe a cartafol *pdf*.
10. O tamaño máximo de cada ficheiro da cartafol miniaturas é de 1Mbyte.
11. Non existe ningunha outro cartafol que as indicadas nos puntos anteriores.

2.4 AIP e DIP

Como complemento á definición dos SIP, queda establecido que a forma dos AIP (Archival Information Package) e DIP (Disemmination Information Package) que se aplican aos obxectos dixitais preservados, baixo calquera área e plan de preservación, é igual á esixida para o SIP correspondente coa soa adición de dous ficheiros de metadatos sobre a estrutura procesada dos ficheiros, así como dos metadatos de cada obxecto dixital, segundo almacénase no sistema de preservación do Arquivo Dixital Integrado de Galicia.

Estes ficheiros noméanse como *LUOI_LibsafeMetadataXMLObject_X--X.xml* e *LUOI_LibsafeXMLObject_X—X.xml*. Onde *LUOI* é un identificador de obxecto en o sistema de preservación do Arquivo Dixital Integrado de Galicia, e *X—X* é unha marca única do proceso de inxestión de 23 caracteres.

3 Referencias de aplicación

Os seguintes son os documentos e webs de referencia que xustifican e determinan as condicións aquí detalladas, ao redor da aplicación do estándar OAIS e normativa aplicable aos proxectos tecnolóxicos da Xunta:

- a web do Library of Congress <http://www.digitalpreservation.gov/index.php>.
- a de uso dos seus estándares en <http://www.loc.gov/standards/>.
- a definición do formato PREMIS <http://www.loc.gov/standards/premis/>.
- a especificación de niveis de preservación dixital <http://www.digitalpreservation.gov/nds/activities/levels.html>.
- a recomendación de especificacións de formatos a usar segundo o tipo de obxecto a preservar (<http://www.loc.gov/preservation/resources/rfs/rfs%202015-2016.pdf>).
- as recomendacións do Hathi Trust (<https://www.hathitrust.org/partnership>) e a definición dos seus propios SIPs e acordo de envío.
- as normas e recomendacións técnicas do Esquema Nacional de Interoperabilidade (do Goberno de España) (http://administracionelectronica.gob.es/pae_Home/pae_Estratexias/pae_Interoperabilidad_de_Inicio.htm).
- documento [ARPAD – Arquivo Electrónico Administrativo] *Definición metadatos*, de especificacións de metadatos da Xunta.
- definición do EAD3 <https://github.com/SAA-SDT/EAD3>.
- documento entregable Dixitalización do Patrimonio Documental.