

iArxiu: Estructura i creació de Paquets d'Informació de Transferència (PIT) utilitzant el model METS

Índex

Índex	2
Informació general	3
Els participants	3
Glossari	4
1. Introducció	5
1.1 El concepte de paquet d'informació	5
1.2 Paquets d'informació de transferència	6
2. Estructura general METS	7
3. Creació d'un METS	9
3.1 File Section	9
3.1.1 Incrustats en Base64.....	9
3.1.2 Referència a fitxers	10
3.2 Structural Map Section	10
3.3 Descriptive Metadata Section	11
3.4 Administrative Metadata Section	12
3.4.1 Metadades PREMIS	13
4. Vocabularis de metadades per iArxiu	16
4.1 Vocabulari d'expedient	16
4.2 Vocabulari de document integrat en expedient	18
4.3 Vocabulari de document independent	19
4.4 Vocabulari de signatura	20
4.5 Altres vocabularis	20
5. Exemple PIT segons el model METS	22
6. Implementació tècnica	24
7. Annexes	25
7.1 Taula d'il·lustracions	25

Informació general

Estat formal	Elaborat per:	Aprovat per:
	Equip del projecte iArxiu de CATCert	Grup d'Innovació Tecnològica de la Subdirecció General d'Arxius i Gestió Documental
Data de creació	17/04/2008	
Control de versions	Data:	17/04/2008
	Descripció:	Creació de la versió definitiva del document
Nivell accés informació	pública	
Títol	iArxiu: Estructura i creació de Paquets d'Informació de Transferència (PIT) utilitzant el model METS	
Fitxer	iArxiu v2 0 - Estructura_PIT_ METS_final_20080417.doc	
Drets d'autor	Aquesta obra està subjecta a una llicència Reconeixement-No comercial-Sense obres derivades 2.5 Espanya de Creative Commons. Per veure'n una còpia, visiteu http://creativecommons.org/licenses/by-nc-nd/2.5/es/ o envieu una carta a Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.	

Els participants

El present document ha estat elaborat per l'equip de treball del projecte iArxiu de CATCert, comptant amb l'assessorament i la col·laboració de la Subdirecció General d'Arxius i Gestió Documental del Departament de Cultura i Mitjans de Comunicació de la Generalitat de Catalunya, a través del Grup d'Innovació Tecnològica, que té per missió oferir pautes i criteris per a la gestió de la documentació electrònica en el marc de les polítiques d'e-Governance que impulsa el govern de la Generalitat de Catalunya.

Per part de l'equip de treball del projecte iArxiu les persones que han participat en la redacció del document han estat les següents:

- Marta Cruellas Ibarz
- Joan Mir Rubio
- Raimon Nualart Mercadé
- Ralf Irion (HP)

Els integrants del Grup d'Innovació Tecnològica que han participat en tasques d'assessorament en la redacció del document i en la validació del seus continguts han estat les següents:

- Ferran Agelet, Ajuntament de Barcelona
- Ignacio Alamillo, Secretaria de Telecomunicacions i Societat de la Informació
- Gemma Bayó, Ajuntament de Barcelona
- Lluís-Esteve Casellas, Ajuntament de Girona

- Lluís Cermeno, Subdirecció General d'Arxius i Gestió Documental
- Anna Lorente, Agència Catalana de l'Aigua
- Betlem Martínez, Departament d'Economia i Finances
- Remei Perpinyà, Universitat Autònoma de Barcelona (UAB)
- Julio Quílez, Departament de Justícia, Arxius Judicials de la DGRAJ
- Tomàs Roy, Centre de Telecomunicacions i Tecnologies de la Informació (CTTI)
- Jaume Sardà, Subdirecció General d'Arxius i Gestió Documental
- Jordi Serra, Departament de la Vicepresidència

A tots ells CATCert els vol donar les gràcies per la seva col·laboració i participació en el projecte iArxiu, i en especial en la validació d'aquest document.

Glossari

CATCert: Agència Catalana de Certificació

CMS: Cryptographic Message Syntax. Format estàndard que s'utilitza per a signar electrònicament, autenticar a les parts, xifrar o generar resum de continguts (dades)

METS: Metadata Encoding & Transmission Standard

OAIS: Open archival information systems

PIA: Paquet d'informació d'Arxiu

PIC: Paquet d'informació de Consulta

PIT: Paquet d'informació de Transferència

PREMIS: PREservation Metadata Implementation Strategies)

PRONOM: Registre on-line sobre formats tecnològics desenvolupat per l'Arxiu Nacional del Regne Unit (TNA)

URI: Uniform Resource Identifier. Cadena de caràcters utilitzada per identificar un recurs o nom.

XSD: XML Schema Definition

1. Introducció

1.1 El concepte de paquet d'informació

Per entendre millor el funcionament del servei d'iArxiu i la gestió de la conservació dels documents a llarg termini cal explicar, encara que sigui breument, què entenem per documentació electrònica. En concret, ens estem referint tant a un document electrònic, sigui estructurat o no, com a agrupacions de documents o expedients electrònics, conservats com informació i prova per una organització en el desenvolupament de les seves activitats. Així mateix, un element inseparable i essencial per al tractament i la preservació dels documents electrònics és el de les metadades que descriuen el context, el contingut i l'estructura dels documents, i ens permeten garantir la seva autenticitat, integritat i disponibilitat al llarg del seu cicle de vida. Per tant, quan parlem de documents ens referim tant als documents com objectes d'informació (contingut), com a les metadades que ens ajuden a interpretar i gestionar correctament aquests documents.

Un requisit imprescindible per garantir la preservació de la documentació electrònica és definir, modelar i implementar una forma d'estructurar i agrupar tots els components que són objecte de conservació en un contenidor neutral. Si unim en un mateix contenidor els documents o objectes digitals (fitxers) i les seves metadades, es crea una entitat que pot existir en qualsevol entorn informàtic, ja que inclou tota la informació necessària per a accedir i representar els documents. El format XML és el més adient per a representar i emmagatzemar els documents i les metadades. Seguint la terminologia OAIS¹ existeixen tres tipus d'entitats, anomenades contenidors o paquets d'informació, segons la fase del cicle de vida del document:

1. Paquet d'informació de transferència (PIT).
2. Paquet d'informació d'arxiu (PIA).
3. Paquet d'informació de consulta (PIC).

En primer lloc, ens trobem els documents transferits pels productors de la documentació, que inclouen els continguts a preservar i totes aquelles metadades que faciliten la seva recuperació, el seu tractament i la seva preservació. Els anomenem *paquets d'informació de transferència* (PIT). Quan aquests ingressen a l'arxiu se'ls aplica un conjunt de controls per verificar la seva integritat, autenticitat i idoneïtat amb els requeriments de conservació. Se'ls afegixen altres metadades (per exemple, les relacionades amb la informació de representació) i, si es necessari, i així ho estableix la política de preservació, es pot procedir a la migració a un format més estàndard i estable. A aquesta normalització del PIT se l'anomena *paquet d'informació d'arxiu* (PIA). El PIA s'emmagatzemarà al repositori d'iArxiu el temps que s'estableixi. Quan un usuari vulgui consultar algun document emmagatzemat, el sistema li permetrà obtenir una còpia del PIA més coneguda com a *paquet d'informació de consulta* (PIC).

Davant l'absència d'un estàndard normalitzat d'estructuració i empaquetament de documents electrònics en format XML, ens hem decidit a utilitzar l'esquema METS² que és un model amb una àmplia acceptació, les característiques del qual s'adeqüen molt a les necessitats d'iArxiu per a la

¹ Per a més informació consulteu: *ISO 14721:2003 - Space data and information transfer systems. Open archival information systems. Reference model.*

Per a més informació sobre el model OAIS desenvolupat pel Consultative Comite for Space Data Systems de la NASA i que posteriorment es va convertir en la norma ISO 14721:2003 consulteu a: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

² L'esquema METS és un model per a codificar metadades descriptives, administratives i estructurals, d'objectes digitals emmagatzemats en un repositori, en format XML. Consulteu: <http://www.loc.gov/standards/mets/>

fase II. Per aquesta raó, s'ha optat per fer servir aquest com a format d'ingrés i consulta de paquets d'informació. Els motius que ens han dut a escollir METS són els següents:

- Permet estructurar els documents d'una forma flexible que s'adapta a les diferents maneres d'agrupar la documentació (document o expedient).
- Admet diferents diccionaris o vocabularis de metadades.
- Permet el creixement del paquet a mesura que s'apliquen processos de migració als fitxers.
- Permet especificar els fitxers, tant de forma incrustada en base64 com vinculats externament.
- Permet generar un paquet d'informació autònom i documentat en format XML, independent del sistema, i desproveït de les dependències tecnològiques de l'aplicació que el va generar. Inclou:
 - a. Els continguts a preservar.
 - b. Metadades descriptives, administratives i estructurals.

1.2 Paquets d'informació de transferència

Abans de procedir a la transferència de documents entre el sistema productor i iArxiu, el productor ha d'agrupar la documentació (fitxers més metadades) que vol transferir en *paquets d'informació de transferència (PIT)*. L'objectiu del present document és definir el protocol de creació i estructura de PIT seguint el model METS, quan la transferència a iArxiu sigui via webservice. La idea és representar en format XML el contingut dels diferents expedients/documentos i que sigui compatible amb el model de PIA d'iArxiu

L'objectiu del present document és desenvolupar una guia que faciliti als productors de la documentació (gestors documentals, gestors d'arxiu, etc.) la creació d'aquests paquets d'informació de transferència.

El document no inclou les diferents operacions de webservice amb iArxiu: operació de transferència, de consulta estat de transferència o de consulta. Tampoc s'inclouen els requeriments o característiques tècniques del client Webservice entre l'aplicació i iArxiu. Tots aquests punts es tracten en un altre document.

2. Estructura general METS

L'estructura d'un fitxer METS es compon de diverses seccions, de les quals farem servir les següents:

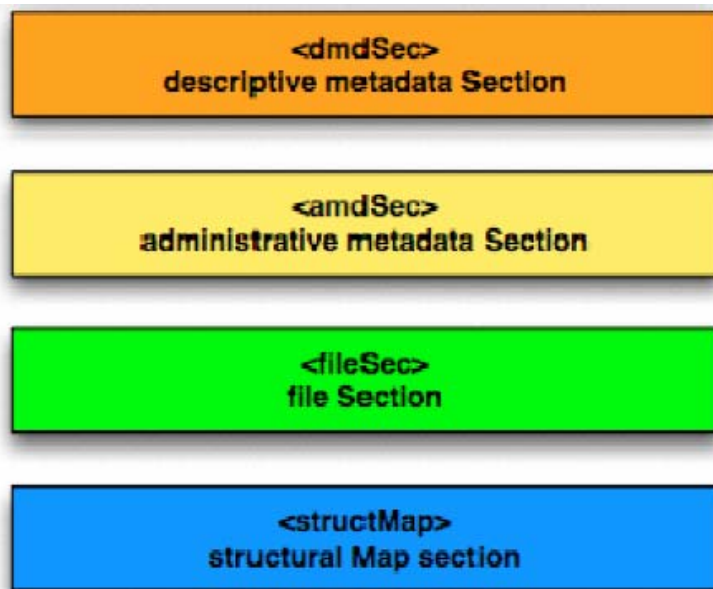


Figura 1: Estructura METS en blocs

Aquesta estructura s'adequa molt bé a les nostres necessitats pels següents motius:

- Permet crear una estructura jeràrquica d'elements (amb carpetes) dins del paquet.
- Permet associar metadades administratives i descriptives, tant a elements estructurals com a fitxers binaris.
- Permet especificar fitxers binaris, tant de forma incrustada com vinculada externament.

A continuació es detallen les seccions:

- Descriptive Metadata Section (<dmdSec>)

Aquesta secció conté tots els vocabularis descriptius que formen part del paquet, independentment del nivell (expedient, document) en què es trobin. Cadascun d'aquests vocabularis porta un identificador únic. Les metadades descriptives les formen uns vocabularis que tenen l'objectiu de descriure i identificar els paquets d'informació i serveixen per a la localització i recuperació de la informació, així com per a entendre el context del paquet.

- Administrative Metadata Section (<amdSec>)

Aquesta secció conté tots els vocabularis administratius que formen part del paquet, independentment del nivell (expedient, document, binari) en què es trobin. Cadascun d'aquests vocabularis porta un identificador únic. Les metadades administratives també les formen uns vocabularis, però no tenen la funció de descriure el paquet, sinó d'administrar-lo per garantir un tractament correcte al llarg del seu cicle de vida (preservació,

representació i ressegellat). Són metadades que faciliten la gestió i el processament dels paquets a llarg termini, la qual cosa cobreix una extensa gama d'informació com ara informació tècnica (formats i especificacions tecnològiques) necessària per mostrar i representar els objectes, metadades de provenença i metadades de cicle de vida i d'accés.

- File Section (<fileSec>)

Aquesta secció conté tots els fitxers binaris que formen part del paquet. Cadascun d'aquests fitxers binaris porta un identificador únic i pot estar referenciat amb una URI o estar incrustat en base64.

- Structural Map Section (<structMap>)

Aquesta secció ens permet estructurar els continguts de forma jeràrquica, tot permetent la creació de carpetes per agrupar els diferents fitxers binaris.

En aquesta mateixa secció també es vinculen els vocabularis de metadades descrits a les seccions de metadades amb les carpetes o fitxers binaris.

3. Creació d'un METS

Per construir un fitxer METS, s'ha de crear un fitxer XML amb les seccions descrites al punt anterior. Tot i que al fitxer METS (XML) creat s'han d'organitzar les diferents seccions en l'ordre indicat, començarem a definir la secció de fitxers (<fileSec>).

Als fragments XML que apareixen als punts següents apareixen certs espais de noms (namespaces) que, en el moment de redactar aquest document, són provisionals. Els espais de noms definitius es definiran durant l'anàlisi tècnica de l'iArxiu.

3.1 File Section

3.1.1 Incrustats en Base64

En aquesta secció s'han d'enumerar els diferents binaris que volem incloure al paquet. Suposem, per exemple, que el paquet es compona d'un binari en format Word i d'una signatura CMS. A més, tots dos s'informen en base64 (per simplificar, s'ha reduït molt la llargada de les codificacions b64, tal i com es pot observar a continuació):

```

<METS:fileSec>
<METS:fileGrp ID="bin_doc" >
  <METS:file ID="bin_doc1.0" MIMETYPE="application/msword" CHECKSUM="A454BF4E59BCD65343" CHECKSUMTYPE="SHA-
1">
  <METS:FContent>
    <METS:binData>Z2Rnc2Znc2dnZg==</METS:binData>
  </METS:FContent>
</METS:file>
</METS:fileGrp>
<METS:fileGrp ID="bin_cms" >
  <METS:file ID="bin_cms1.0" MIMETYPE="application/pkcs7-mime" CHECKSUM="A454BF4E59B22222" CHECKSUMTYPE="SHA-
1">
  <METS:FContent>
    <METS:binData>YmJiYmJiYg==</METS:binData>
  </METS:FContent>
</METS:file>
</METS:fileGrp>
</METS:fileSec>

```

Tal i com es pot observar, s'han d'indicar obligatòriament:

- Identificadors dels fitxers (*bin_doc* i *bin_cms* a l'exemple): aquests són lliures i es pot seguir qualsevol terminologia, sempre que no es repeteixin dins del mateix METS.
- MimeTypes: MimeTypes segons qualsevol registre de MimeTypes oficial, com per exemple el de PRONOM³ o iana⁴ ("application/msword" i "application/pkcs7-mime" a l'exemple).
- Resum del fitxer (checksum) i algoritme (SHA-1 a l'exemple). Els possibles valors per a l'algoritme són *SHA-1* i *MD5*.

³ Per a més informació consulteu a: <http://www.nationalarchives.gov.uk/pronom/>

⁴ Per a més informació consulteu a: <http://www.iana.org/assignments/media-types/>

- Contingut binari en base64.

3.1.2 Referència a fitxers

En cas de no voler incrustar els fitxers en Base64 dins del METS, es poden enviar els binaris per separat i indicar una referència des del fitxer METS de la següent forma:

```
</METS:fileGrp>
<METS:fileGrp ID="bin_cms_file" >
  <METS:file ID="bin_cms_file1.0" MIMETYPE="application/pkcs7-mime" CHECKSUM="A454BF4E59B22222"
CHECKSUMTYPE="SHA-1">
  <METS:FLocat xlink:href="dir_exp/doc_doc1/sig55.cms" LOCTYPE="URL"/>
</METS:file>
</METS:fileGrp>
```

L'atribut *xlink:href* ha de contenir la ruta del fitxer relativa a l'arrel del paquet, que és on es troba el fitxer METS.XML.

Per fer l'enviament del paquet a iArxiu, s'ha de crear un fitxer ZIP que contingui el fitxer METS.XML i tots els binaris referenciats.

3.2 Structural Map Section

En aquesta secció s'ordenen els binaris dins d'una estructura d'expedient. En l'exemple de sota crearem un expedient que només conté un document que agrupa el fitxer Word i la signatura associada:

```
<METS:structMap>
  <METS:div LABEL="expedient" DMDID="DMD_EXP">
    <METS:div LABEL="document_1" DMDID="DMD_DOC1">
      <METS:div LABEL="documento.doc">
        <METS:fptr FILEID="bin_doc"/>
      </METS:div>
      <METS:div LABEL="firma.p7b" ADMID="AMD_SIG">
        <METS:fptr FILEID="bin_cms"/>
      </METS:div>
    </METS:div>
  </METS:div>
</METS:structMap>
```

En aquest fragment tenim 3 nivells de jerarquització:

- expedient (entitat lògica)
 - o document_1 (entitat lògica)
 - fitxers binaris continguts (file)

Per als casos de documents simples que no pertanyen a un expedient, no existeix el nivell d'expedient, sent el primer nivell el de document.

S'han de informar obligatòriament:

- A l'últim nivell s'han d'informar les referències als binaris definits a la secció <fileSec> amb l'atribut FILEID (*bin_doc* i *bin_cms*)
- Amb l'atribut DMDID s'han d'identificar els vocabularis de metadades descriptives. En aquesta part només s'indiquen els identificadors del vocabulari. El contingut de les metadades descriptives s'indica a la secció <dmdSec>. Anàlogament a les metadades descriptives, l'atribut AMDID conté els identificadors dels vocabularis administratius, informats a la secció <amdSec>. Es pot observar que, tant a nivell d'expedient com a nivell de document, existeixen metadades descriptives (DMD_EXP i DMD_DOC1), mentre que a nivell de binari les metadades són administratives (metadades de la signatura).
- Amb els atributs LABEL es donen noms als diferents documents i s'indiquen noms de fitxers (si s'indiquen en Base64).

Nota: Els identificadors de vocabularis definitius dependran de la configuració definitiva de la plataforma iArxiu un cop aquesta estigui en producció.

3.3 Descriptive Metadata Section

Com ja s'ha esmentat anteriorment, en aquesta secció s'informen els valors dels vocabularis descriptius. Al nostre exemple, hi ha dos vocabularis descriptius: DMD_EXP i DMD_DOC1:

```
<METS:dmdSec ID="DMD_EXP">
  <METS:mdWrap MDTYPE="OTHER" MIMETYPE="text/xml" LABEL="Metadades descriptives: vocabulari doc1">
    <METS:xmlData>
      <exp:expedient xmlns:exp="http://www.iarxiu.cat/voc/exp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/exp expedient.xsd">
        <exp:codi_referencia>String</exp:codi_referencia>
        <exp:numero_expedient>String</exp:numero_expedient>
        <exp:codi_classificacio>String</exp:codi_classificacio>
        <exp:titol_serie_documental>String</exp:titol_serie_documental>
        <exp:nivell_descripcio>Unitat documental composta</exp:nivell_descripcio>
        <exp:titol>String</exp:titol>
        <exp:data_obertura>1967-08-13</exp:data_obertura>
        <exp:data_tancament>1967-08-13</exp:data_tancament>
        <exp:nom_productor>String</exp:nom_productor>
        <exp:unitat_productora>String</exp:unitat_productora>
        <exp:descripcio>String</exp:descripcio>
        <exp:classificacio_seguretat_acces>Accés públic</exp:classificacio_seguretat_acces>
        <exp:sensibilitat_dades_LOPD>Nivell baix</exp:sensibilitat_dades_LOPD>
      </exp:expedient>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>

<METS:dmdSec ID="DMD_DOC1">
  <METS:mdWrap MDTYPE="OTHER" MIMETYPE="text/xml" LABEL="Metadades descriptives: vocabulari de expedient">
    <METS:xmlData>
      <doc:document xmlns:doc="http://www.iarxiu.cat/voc/docExp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/docExp documentExp.xsd">
        <doc:codi_referencia>String</doc:codi_referencia>
        <doc:numero_document>String</doc:numero_document>
        <doc:titol>String</doc:titol>
        <doc:data_creacio>1967-08-13</doc:data_creacio>
        <doc:nivell_descripcio>Unitat documental simple</doc:nivell_descripcio>
      </doc:document>
    </METS:xmlData>
  </METS:mdWrap>
</METS:dmdSec>
```

```
<doc:suport>String</doc:suport>  
<doc:nom_productor>String</doc:nom_productor>  
<doc:unitat_productora>String</doc:unitat_productora>  
<doc:descripcio>String</doc:descripcio>  
</doc:document>  
</METS:xmlData>  
</METS:mdWrap>  
</METS:dmdSec>
```

Es pot observar que hi ha tants grups de metadades administratives com vocabularis diferents apareixen al paquet. En particular, cadascun d'aquests grups conté:

- Identificador de vocabulari (atribut ID). Els identificadors de vocabulari definitius dependran de la configuració definitiva de la plataforma iArxiu un cop estigui en producció.
- Descripció del vocabulari.
- Contingut del vocabulari (format XML).

Aquests vocabularis es poden emplenar tenint en compte un determinat esquema XSD, prèviament donat d'alta a la plataforma, o bé poden ser lliures. En aquest segon cas, els vocabularis no s'indexaran i no es podran fer cerques de paquets per aquestes metadades.

Quan el vocabulari segueix un determinat esquema XSD prèviament donat d'alta, la plataforma farà les validacions estructurals i de contingut corresponents.

Si a la plantilla existeixen també limitacions sobre aquest vocabulari, també es tindran en compte.

Al punt 4 es detallen els esquemes de metadades que els clients d'iArxiu han d'emplenar en cas d'enviar expedients i documents seguint el model de vocabulari de metadades proposat per iArxiu.

3.4 Administrative Metadata Section

Finalment, les metadades administratives apareixen en aquesta secció. En el nostre exemple, en tenim només una: AMD_SIG, que es correspon amb el vocabulari que descriu la signatura.

```
<METS:amdSec ID="AMD_SIG">  
<METS:techMD CREATED="2007-05-29T15:40:17.353Z" ID="AMD_SIG_1.0">  
<METS:mdWrap LABEL="Signature data" MDTYPE="OTHER" MIMETYPE="text/xml" >  
<METS:xmlData>  
<sig:signatura xmlns:sig="http://www.iarxiu.cat/voc/sig" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation="http://www.iarxiu.cat/voc/sig signatura.xsd">  
<sig:identificador>String</sig:identificador>  
<sig:Identificador_document>bin_doc</sig:Identificador_document>  
<sig:tipus_signatura> Independent (detached)</sig:tipus_signatura>  
<sig:format_signatura>CMS</sig:format_signatura>  
</sig:signatura>  
</METS:xmlData>  
</METS:mdWrap>  
</METS:techMD>  
</METS:amdSec>
```

Com es pot observar les metadades administratives funcionen de forma anàloga a les descriptives.

És important remarcar que la metadada *identificador_document* de les metadades de signatura conté "bin_doc", que és el document al que pertany aquesta signatura.

També són obligatòries les metadades *tipus_signatura* i *format_signatura*, que permeten a iArxiu saber com tractar i verificar la signatura.

A mode de resum, podem descriure un fitxer METS de la següent forma:

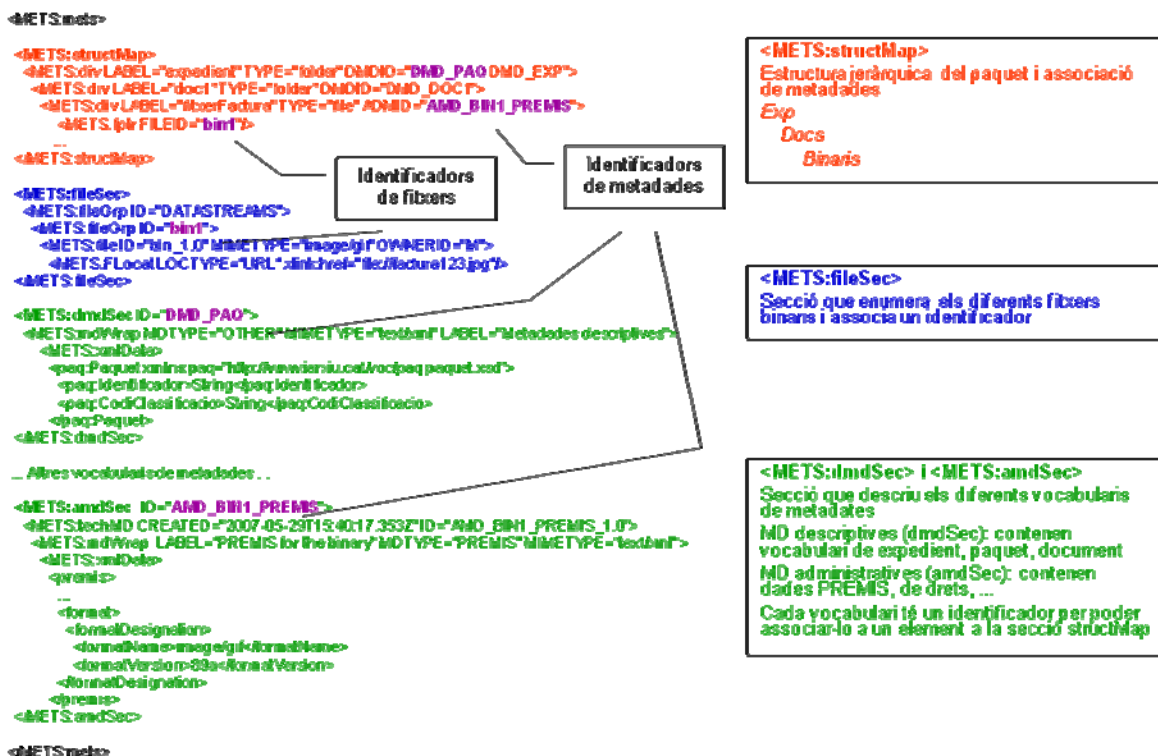


Figura 2: Associacions de metadades i estructura

3.4.1 Metadades PREMIS

PREMIS (PREservation Metadata Implementation Strategies) és un vocabulari estàndard que ens permet emmagatzemar la informació de representació de forma estructurada per als diferents fitxers binaris que formen part del paquet.

iArxiu farà servir els esquemes PREMIS-v1-1.xsd⁵ i Object-v1-1.xsd⁶ de PREMIS per emplenar les dades relatives a identificació i format. D'aquesta forma assurem una correcta migració, representació i preservació de continguts.⁷

⁵ <http://www.loc.gov/standards/premis/v1/PREMIS-v1-1.xsd>

⁶ <http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd>

⁷ En el moment de la redacció del present document el grup de treball de PREMIS està treballant en una nova versió dels esquemes o vocabularis LSD. Quan estiguin publicats iArxiu els farà servir com a model.

Com que iArxiu té la capacitat d'obtenir informacions de format de forma automàtica, **no és obligatori que el client informi aquest vocabulari.**

En aquells casos en què el client no informa aquest vocabulari, iArxiu el crearà de forma automàtica a partir de les informacions de *MIMETYPE* de *<fileSec>*, i d'introspeccions al fitxer per obtenir altres dades necessàries. En aquest cas, l'atribut *ADMID* del *<structMap>* no apareix.

Les metadades que pot informar el client, o que informarà la plataforma, per introspecció, són:

- **Identificació d'objecte (<objectIdentifier>)**

Com que cada vocabulari ja forma part d'un paquet, aquest element conté l'identificador del binari que està descrivint (per exemple: bin_doc).

D'aquesta forma es faciliten els traspassos, doncs no s'han de modificar aquests identificadors.

- **Format (<format>)**

El format d'objecte es compon dels següents components:

- formatName.
- formatVersion.
- formatRegistryName.
- formatRegistryKey.

L'element "formatName" es correspondrà amb el *ContentType* del binari, mentre que l'element "formatVersion" es correspon amb la versió d'aquest format. L'usuari podrà informar tots dos, però si no informa el segon, iArxiu l'obtindrà fent introspecció al fitxer.

"formatRegistryName" i "formatRegistryKey" contenen *PRONOM* i l'identificador intern *PRONOM*.

- **Aplicació creadora (creatingApplication)**

Si es coneix l'aplicació que es va fer servir per crear el binari, es pot informar en aquest element.

A continuació es mostra un exemple de vocabulari PREMIS:

```
<METS:amdSec ID="AMD_BIN1_PREMIS">
<METS:techMD CREATED="2007-05-29T15:40:17.353Z" ID="AMD_BIN1_PREMIS_1.0">
<METS:mdWrap LABEL="PREMIS for the binary" MDTYPE="PREMIS" MIMETYPE="text/xml" >
<METS:xmlData>
<premis:premis>
<premis:object>
<premis:objectIdentifier>
<premis:objectIdentifierType>iArxiu</premis:objectIdentifierType>
<!-- <premis:objectIdentifierValue>http://www.iarxiu.cat/iArxiu:400/DS_DATA</premis:objectIdentifierValue> -->
<premis:objectIdentifierValue>bin_doc</premis:objectIdentifierValue>
</premis:objectIdentifier>
<premis:objectCategory>file</premis:objectCategory>
<premis:objectCharacteristics>
```

000_20060208

Està prevista la seva publicació cap a mitjans del 2008. Per a més informació consulteu:
<http://www.loc.gov/standards/premis/schemas.html>

```
<premis:format>  
  <premis:formatDesignation>  
    <premis:formatName>application/msword</premis:formatName>  
    <premis:formatVersion>5.0</premis:formatVersion>  
  </premis:formatDesignation>  
  <premis:formatRegistry>  
    <premis:formatRegistryName>PRONOM</premis:formatRegistryName>  
    <premis:formatRegistryKey>fmt/4</premis:formatRegistryKey>  
  </premis:formatRegistry>  
</premis:format>  
</premis:objectCharacteristics>  
</premis:object>  
</premis:premis>  
</METS:xmlData>  
</METS:mdWrap>  
</METS:techMD>  
</METS:amdSec>
```

4. Vocabularis de metadades per iArxiu

Els vocabularis de metadades que els usuaris clients d'iArxiu han d'emplenar per ingressar un paquet d'informació de transferència (PIT) varien en funció de si es tracta d'un expedient o d'un document sol:

- Per a expedients (unitats documentals compostes), els vocabularis són els següents:
 - Vocabulari d'expedient (*Esquema_expedient.xsd*).
 - Vocabulari de documents integrats en expedients (*Esquema_documentExp.xsd*).
- Per a documents individuals (unitat de document simple), el vocabulari és el següent:
 - Vocabulari de documents que no formen part d'un expedient (*Esquema_documentSol.xsd*).

En funció de si s'envia un expedient o un document és obligatori informar al paquet, com a mínim, el vocabulari d'expedient o el de documentSol, respectivament. És necessari informar als paquets, com a mínim, un d'aquests vocabularis per tal de garantir la preservació i assegurar que es podran cercar els documents.

Adicionalment, si es volen arxivar les signatures dels documents, hi ha un vocabulari per descriure-les (*Esquema_signatura.xsd*).

Pel que fa als binaris, aquests tenen com a possible vocabulari a emplenar el de PREMIS com a metadades administratives, però aquest vocabulari és opcional. Les úniques dades a emplenar són les descrites al punt 3.1: Identificador, MimeType, resum de fitxer i algorisme de resum.

Als següents punts s'inclouen les representacions gràfiques dels esquemes mencionats anteriorment. Als fitxers XSD que s'adjunten amb aquest document es poden observar els tipus de dades dels diferents elements, i si són obligatoris o no.

4.1 Vocabulari d'expedient

L'esquema que descriu un expedient és el següent:

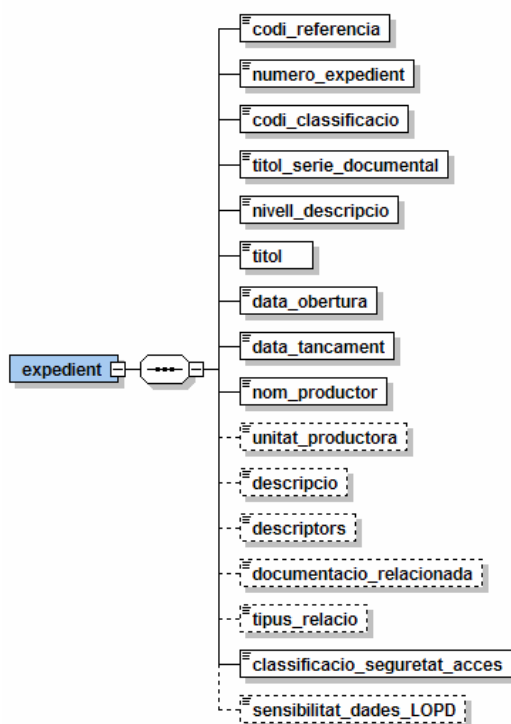


Figura 3: Vocabulari d'expedient

A continuació es mostra un exemple concret de metadades descriptives fent servir aquest vocabulari:

```
- <METS:dmdSec ID="DMD_EXP">
- <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="voc:generic:exp" MIMETYPE="text/xml">
- <METS:xmlData>
- <exp:expedient xmlns:exp="http://www.iarxiu.cat/voc/exp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/exp expedient.xsd">
<exp:codi_referencia>CAT/AACC/D1161_N-SCD_185_08</exp:codi_referencia>
<exp:numero_expedient>SCD/0185/08</exp:numero_expedient>
<exp:codi_classificacio>D1161</exp:codi_classificacio>
<exp:titol_serie_documental>Expedients d'emissió de certificats digitals personals</exp:titol_serie_documental>
<exp:nivell_descripcio>Unitat documental composta</exp:nivell_descripcio>
<exp:titol>Ajuntament de Castellví de Rosanes. Emissió 1 certificats digitals classe 1 CPISR+CPX</exp:titol>
<exp:data_obertura>2008-01-13</exp:data_obertura>
<exp:data_tancament>2008-03-13</exp:data_tancament>
<exp:nom_productor>Agència Catalana de Certificació</exp:nom_productor>
<exp:unitat_productora>Àrea de qualitat i procediments</exp:unitat_productora>
<exp:descripcio>Titulars: Lluís Tomàs, Pere Tomeu</exp:descripcio>
<exp:descriptors>String</exp:descriptors>
<exp:documentacio_relacionada>0233/01/07</exp:documentacio_relacionada>
<exp:tipus_relacio>Expedient híbrid</exp:tipus_relacio>
<exp:classificacio_seguretat_acces>Accés restringit</exp:classificacio_seguretat_acces>
<exp:sensibilitat_dades_LOPD>Nivell baix</exp:sensibilitat_dades_LOPD>
</exp:expedient>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
```

4.2 Vocabulari de document integrat en expedient

L'esquema que descriu un document que forma part d'un expedient és:

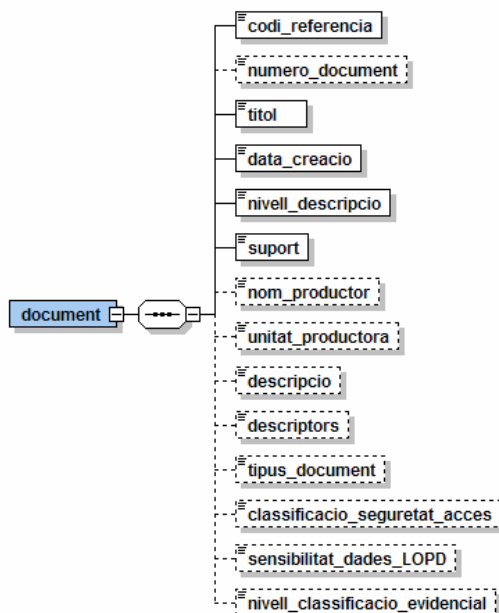


Figura 4: Vocabulari document integrat en expedient

A continuació es mostra un exemple concret de metadades descriptives fent servir aquest vocabulari:

```

- <METS:dmdSec ID="DMD_DOC_EXP">
- <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="vocsgeneric:docExp" MIMETYPE="text/xml">
- <METS:xmlData>
- <doc:document xmlns:doc="http://www.iarxiu.cat/voc/docExp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.iarxiu.cat/voc/docExp documentExp.xsd">
  <doc:codi_referencia>CAT/AACC/D1161_N-SCD_185_08</doc:codi_referencia>
  <doc:numero_document>01</doc:numero_document>
  <doc:titol>sol·licitud SCD/0185/2008</doc:titol>
  <doc:data_creacio>2008-01-13 10:13:04</doc:data_creacio>
  <doc:nivell_descripcio>Unitat documental simple</doc:nivell_descripcio>
  <doc:suport>Electrònic</doc:suport>
  <doc:nom_productor>Agència Catalana de Certificació</doc:nom_productor>
  <doc:unitat_productora>Àrea de qualitat i procediments</doc:unitat_productora>
  <doc:descripcio>String</doc:descripcio>
  <doc:descriptors>String</doc:descriptors>
  <doc:tipus_document>Sol·licitud</doc:tipus_document>
  <doc:classificacio_seguretat_acces>Accés restringit</doc:classificacio_seguretat_acces>
  <doc:sensibilitat_dades_LOPD>Nivell baix</doc:sensibilitat_dades_LOPD>
  <doc:nivell_classificacio_evidencial>Sense evidència (identificació al·legada)</doc:nivell_classificacio_evidencial>
</doc:document>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
  
```

4.3 Vocabulari de document independent

L'esquema que descriu un document independent (que no forma part d'un expedient) és el següent:

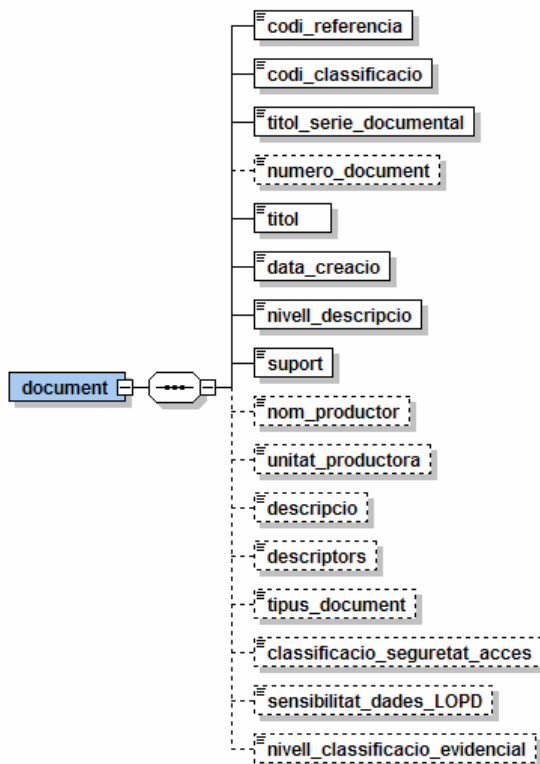


Figura 5: Vocabulari unitat documental simple

A continuació es mostra un exemple concret de metadades descriptives fent servir aquest vocabulari:

```

- <METS:dmdSec ID="DMD_DOC">
- <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="voc:generic:docSol" MIMETYPE="text/xml">
- <METS:xmlData>
- <doc:document xmlns:doc="http://www.iarxiu.cat/voc/docSol" xmlns:xsi="http://www.w3.org/2001/XMLSchema-Instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/docSol documentsol.xsd">
<doc:codi_referencia>CAT/AAAB/G0800 N-Conveni_UB</doc:codi_referencia>
<doc:codi_classificacio>G0800</doc:codi_classificacio>
<doc:titol_serie_documental>Convenis</doc:titol_serie_documental>
<doc:numero_document>01</doc:numero_document>
<doc:titol>Conveni de col·laboració amb la Universitat de Barcelona per a l'organització del congrés XX</doc:titol>
<doc:data_creacio>2007-06-12</doc:data_creacio>
<doc:nivell_descripcio>Unitat documental simple</doc:nivell_descripcio>
<doc:suport>Electrònic</doc:suport>
<doc:nom_productor>Ajuntament de Barcelona</doc:nom_productor>
<doc:unitat_productora>Alcaldia</doc:unitat_productora>
<doc:descripcio>String</doc:descripcio>
<doc:descriptors>String</doc:descriptors>
<doc:tipus_document>Conveni</doc:tipus_document>
<doc:classificacio_seguretat_acces>Accés restringit</doc:classificacio_seguretat_acces>
<doc:sensibilitat_dades_LOPD>Nivell baix</doc:sensibilitat_dades_LOPD>
<doc:nivell_classificacio_evidencial>Evidència de signatura electrònica</doc:nivell_classificacio_evidencial>
</doc:document>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>

```

000_20060208

- No es permeten atributs als elements

Per fer servir aquest tipus de vocabularis, l'administrador de la plataforma iArxiu els haurà d'haver donat d'alta prèviament.

La plataforma també permet fer servir vocabularis lliures que no segueixen cap esquema XSD concret, però aquests no seran indexats i, per tant, no es podran fer cerques per a aquests elements a l'hora de fer consultes.

5. Exemple PIT segons el model METS

A continuació es mostra un fitxer METS emplenat amb totes les seccions i vocabularis descrits als punts anteriors. En concret, es tracta d'un PIT que inclou un document i una signatura, a més d'incloure metadades a nivell d'expedient, de document i de signatura.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <METS:mets LABEL="Exemple expedient bàsic" PROFILE="Expedient" xmlns:METS="http://www.loc.gov/METS/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:premis="http://www.loc.gov/standards/premis/v1" xmlns:pac="iarxiu:vocabularis:package" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd">
- <!--

Metadades descriptives: vocabulari expedient
Aquest vocabulari és propi de dades d'un expedient

-->
- <METS:dmdSec ID="DMD_EXP">
- <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="voc:generic:exp" MIMETYPE="text/xml">
- <METS:xmlData>
- <exp:expedient xmlns:exp="http://www.iarxiu.cat/voc/exp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/exp expedient.xsd">
<exp:codi_referencia>CAT/AACC/D1161_N-SCD_185_08</exp:codi_referencia>
<exp:numero_expedient>SCD/0185/08</exp:numero_expedient>
<exp:codi_classificacio>D1161</exp:codi_classificacio>
<exp:titol_serie_documental>Expedients d'emissió de certificats digitals personals</exp:titol_serie_documental>
<exp:nivell_descripcio>Unitat documental composta</exp:nivell_descripcio>
<exp:titol>Ajuntament de Castellví de Rosanes. Emissió 1 certificats digitals classe 1 CPISR+CPX</exp:titol>
<exp:data_obertura>2008-01-13</exp:data_obertura>
<exp:data_tancament>2008-03-13</exp:data_tancament>
<exp:nom_productor>Agència Catalana de Certificació</exp:nom_productor>
<exp:unitat_productora>Àrea de qualitat i procediments</exp:unitat_productora>
<exp:descripcio>Titulars: Lluís Tomàs, Pere Tomeu</exp:descripcio>
<exp:descriptors>String</exp:descriptors>
<exp:documentacio_relacionada>0233/01/07</exp:documentacio_relacionada>
<exp:tipus_relacio>Expedient híbrid</exp:tipus_relacio>
<exp:classificacio_seguretat_acces>Accés restringit</exp:classificacio_seguretat_acces>
<exp:sensibilitat_dades_LOPD>Nivell baix</exp:sensibilitat_dades_LOPD>
</exp:expedient>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
- <!--

Metadades descriptives: vocabulari docl
Aquestes metadades descriuen el document que hi ha al expedient.
Amb més documents, n'hi haurà més blocs com aquest

-->
- <METS:dmdSec ID="DMD_DOC_EXP">
- <METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="voc:generic:docExp" MIMETYPE="text/xml">
- <METS:xmlData>
- <doc:document xmlns:doc="http://www.iarxiu.cat/voc/docExp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iarxiu.cat/voc/docExp documentExp.xsd">
<doc:codi_referencia>CAT/AACC/D1161_N-SCD_185_08</doc:codi_referencia>
<doc:numero_document>01</doc:numero_document>
<doc:titol>sol·licitud SCD/0185/2008</doc:titol>
<doc:data_creacio>2008-01-13 10:13:04</doc:data_creacio>
<doc:nivell_descripcio>Unitat documental simple</doc:nivell_descripcio>
<doc:suport>Electrònic</doc:suport>
<doc:nom_productor>Agència Catalana de Certificació</doc:nom_productor>
<doc:unitat_productora>Àrea de qualitat i procediments</doc:unitat_productora>
<doc:descripcio>String</doc:descripcio>
<doc:descriptors>String</doc:descriptors>
<doc:tipus_document>Sol·licitud</doc:tipus_document>
<doc:classificacio_seguretat_acces>Accés restringit</doc:classificacio_seguretat_acces>
<doc:sensibilitat_dades_LOPD>Nivell baix</doc:sensibilitat_dades_LOPD>
<doc:nivell_classificacio_evidencial>Sense evidència (identificació al·legada)</doc:nivell_classificacio_evidencial>
</doc:document>
</METS:xmlData>
</METS:mdWrap>
</METS:dmdSec>
- <!--
```

```

Metadades administratives: signatura
-->
-<METS:amdSec ID="AMD_SIG">
-<METS:techMD ID="AMD_SIG_1.0">
  -<METS:mdWrap MDTYPE="OTHER" OTHERMDTYPE="voc:generic:sig" MIMETYPE="text/xml">
    -<METS:xmlData>
      -<sig:signatura xmlns:sig="http://www.iarxiu.cat/voc/sig" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.iarxiu.cat/voc/sig signatura.xsd">
          <sig:identificador>010254545465787878788879879</sig:identificador>
          <sig:Identificador_document>BIN_1</sig:Identificador_document>
          <sig:tipus_signatura>Independent (detached)</sig:tipus_signatura>
          <sig:format_signatura>CMS</sig:format_signatura>
        </sig:signatura>
      </METS:xmlData>
    </METS:mdWrap>
  </METS:techMD>
</METS:amdSec>
-<!--

```

Secció de fitxers: En aquesta secció s'enumeren els diferent fitxers binaris que intervenen al paquet. No han de seguir cap estructura, doncs l'estructura es dona a la secció structMap, que apuntarà a cadascun d'aquests fitxers. Només és una llista d'identificadors i ubicacions.

```

-->
-<METS:fileSec>
-<METS:fileGrp ID="BIN_1">
  -<METS:file ID="BIN_1.0" MIMETYPE="application/msword" CHECKSUM="A454BF4E59BCD65343" CHECKSUMTYPE="SHA-1">
    -<METS:FContent>
      <METS:binData>Z2Rnc2Znc2dnZg==</METS:binData>
    </METS:FContent>
  </METS:file>
</METS:fileGrp>
-<METS:fileGrp ID="BIN_2">
  -<METS:file ID="BIN_2.0" MIMETYPE="application/pkcs7-mime" CHECKSUM="A454BF4E59B22222" CHECKSUMTYPE="SHA-1">
    -<METS:FContent>
      <METS:binData>YmJiYmJiYg==</METS:binData>
    </METS:FContent>
  </METS:file>
</METS:fileGrp>
</METS:fileSec>
-<!--

```

Secció de estructura del PIT. En aquesta secció indiquen la estructura jeràrquica del paquet. En aquest cas és senzilla, doncs només reflexa un expedient amb un document, que conté un binari. En aquest secció també es relacionen els elements (exp, doc, binari) amb les seves metadades (enumerades a dalt).

```

-->
-<METS:structMap>
  -<METS:div LABEL="expedient" DMDID="DMD_EXP">
    -<METS:div LABEL="document_1" DMDID="DMD_DOC_EXP">
      -<METS:div LABEL="documento.doc">
        <METS:fptr FILEID="BIN_1" />
      </METS:div>
    -<METS:div LABEL="firma.p7b" ADMID="AMD_SIG">
      <METS:fptr FILEID="BIN_2" />
    </METS:div>
  </METS:div>
</METS:structMap>
</METS:mets>

```

Per veure el fitxer METS consulteu el document adjunt: *Exemple_generic_PIT_METS.XML*.

6. Implementació tècnica

Per a la creació d'un fitxer METS des de Java plantejarem dues opcions:

- Mitjançant tecnologia XMLBeans: XMLBeans és una llibreria genèrica i lliure que ofereix la funcionalitat de generació d'XMLs a partir d'esquemes XSD. Per a més informació sobre el projecte consulteu a: <http://xmlbeans.apache.org/>
- Mitjançant la llibreria específica de generació de fitxers METS: creada per la Universitat de Harvard. La descàrrega es pot fer des de la següent pàgina web: <http://hul.harvard.edu/mets/>

7. Annexes

7.1 Taula d'il·lustracions

Figura 1: Estructura METS en blocs	7
Figura 2: Associacions de Metadades i estructura.....	13
Figura 3: Vocabulari d'expedient	17
Figura 4: Vocabulari document integrat en expedient	18
Figura 5: Vocabulari unitat documental simple	19
Figura 6: Vocabulari de signatura	20