



# Liburutegi digitalak

## Formatuak eta estandarrak

Xabier Arregi

IXA taldea (EHU)

[xabier.arregi@ehu.es](mailto:xabier.arregi@ehu.es)

# Liburutegi digitalak = Askotariko objektuen biltegiak

- Liburutegi digitala
  - *Objektu* digital heterogeneoen biltegia
- Objektu horiek ondo deskribatzea komeni da...
  - Gutxienez paperezkoetan bezala,  
biltzeko, katalogatzeko, antolatzeko, bilatzeko, trukatzeko
  - Bai eta programa informatikoen bidez ere  
atzitzeko, nabigatzeko, ustiatzeko, zehar-bilaketa aurreratuak egiteko,  
informazioa erazteko, iragazteko, irakaskuntzan erabiltzeko...
  - Bildumak ere aintzat hartuta
    - Kolaborazio instituzionala bultzateko, instituzio arteko biltegi birtualak  
eratzeko, edukiak partekatzeko (2.0 eredua)...

# Nola deskribatu

- Metadatuak
  - Datuei buruzko datuak
- Helburuak beteko badira, ados jarri beharra
  - Formatu bateratua
  - Estandarrak
- Bi joera kontrajarri
  - Informazio konplexuaren beharra vs. estandar sinpleak

# Metadatuaren tipologia

1. Deskriptiboa
  - Objektuaren edukiei buruzko ezaugarri orokor nagusiak
2. Administratiboa
  - Kudeatzeko behar den informazioa
  - Besteak beste:
    - Metadatu teknikoak
    - Eskubideen kudeaketa (jabetza kontuak...)
    - Jatorria, bilakaera, babeskuntza
3. Egiturazkoa
  - Barne-egiturari buruzko zehaztasunak

# Oinarrizko formatua:

## XML

- XML, *eXtensible Markup Language*, dokumentuak *markatzeko* estandarra da.
  - Sintaxi orokorra definitzen du.
    - Etiketa sinpleak eta irakurgarriak erabiltzean datza
    - Etiketa horiek dokumentuetako edukiak deskribatzen dituzte, eta testuan bertan txertatzen dira
    - Markatzean jarraitu beharreko arau sintaktikoak definitzen dira: nola txertatu etiketak dokumentuetan, nola mugarritu *elementuak*, nola idatzi etiketak, zein izen jarri elementuei...

# Zer da XML

- Formatu estandarra eskaintzen du.
  - Malgua eta irekia:
    - edozein dokumentu marka daiteke
    - etiketa multzoa ez dago aldezturik mugatua:  
*eXtensible*
  - Zurruna:
    - oinarrizko arauak errespetatu behar dira, nahitaez
    - XML gramatika

# Adibidea

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<?xml-stylesheet href="pertsona.css" type="text/css" ?>
<pertsona>
  <izen-abizenak>
    <izena>Pier Paolo</izena>
    <abizena>Pasolini</abizena>
  </izen-abizenak>
  <lanbidea>zine-zuzendaria</lanbidea>
  <lanbidea>idazlea</lanbidea>
</pertsona>
```

# XMLren ezaugarriak

- Markatzen dena testua da. Markatze-etiketak testua dira. Dena da testua.
  - Edozein testu-editore aski da XML dokumentua irakurtzeko eta sortzeko.
- Dokumentu-mota jakin baterako onartzen den markatzea *eskema* baten bidez adierazten da.
  - Eskema-lengoaiak: DTD, XML Schema (W3C), RELAX NG...

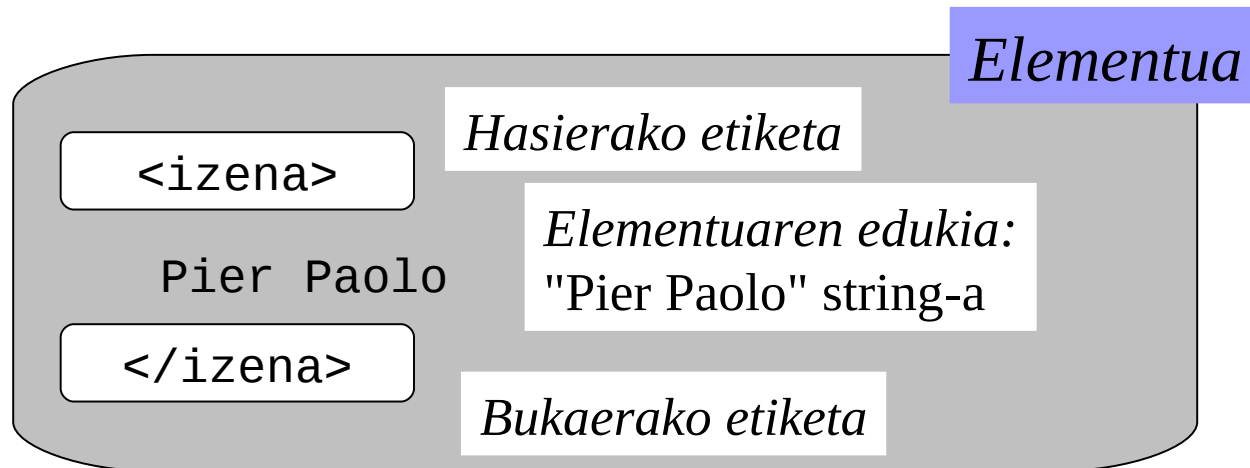


# XMLren ezaugarriak

- Dokumentu-motak erabilera zabala duenean, eta erakunde edo talderen batek adostua bada: *XML aplikazioa*.
  - Adibidez: XSL, XML Schema, RSS, MathML, MODS, Dublin Core...
- Testu soila denez, ez dago arazorik plataforma artean informazioa pasatzeko.
  - Dokumentuetako edukiak eta egiturazko ezaugarriak esplizituak dira.
- Horregatik, informazioa trukatu, gorde, kudeatu eta argitaratzeko erabiltzen da.

# XML dokumentuak eta elementuak

- XML dokumentuak prozesagarriak izan daitezzen, sintaxi-arau zorrotzak bete behar dituzte.
  - Elementuak dira oinarrizko osagaiak.
- Oinarrizko markatze-eredua. Elementuak, etiketak eta karaktere-datuak. Adibidea:

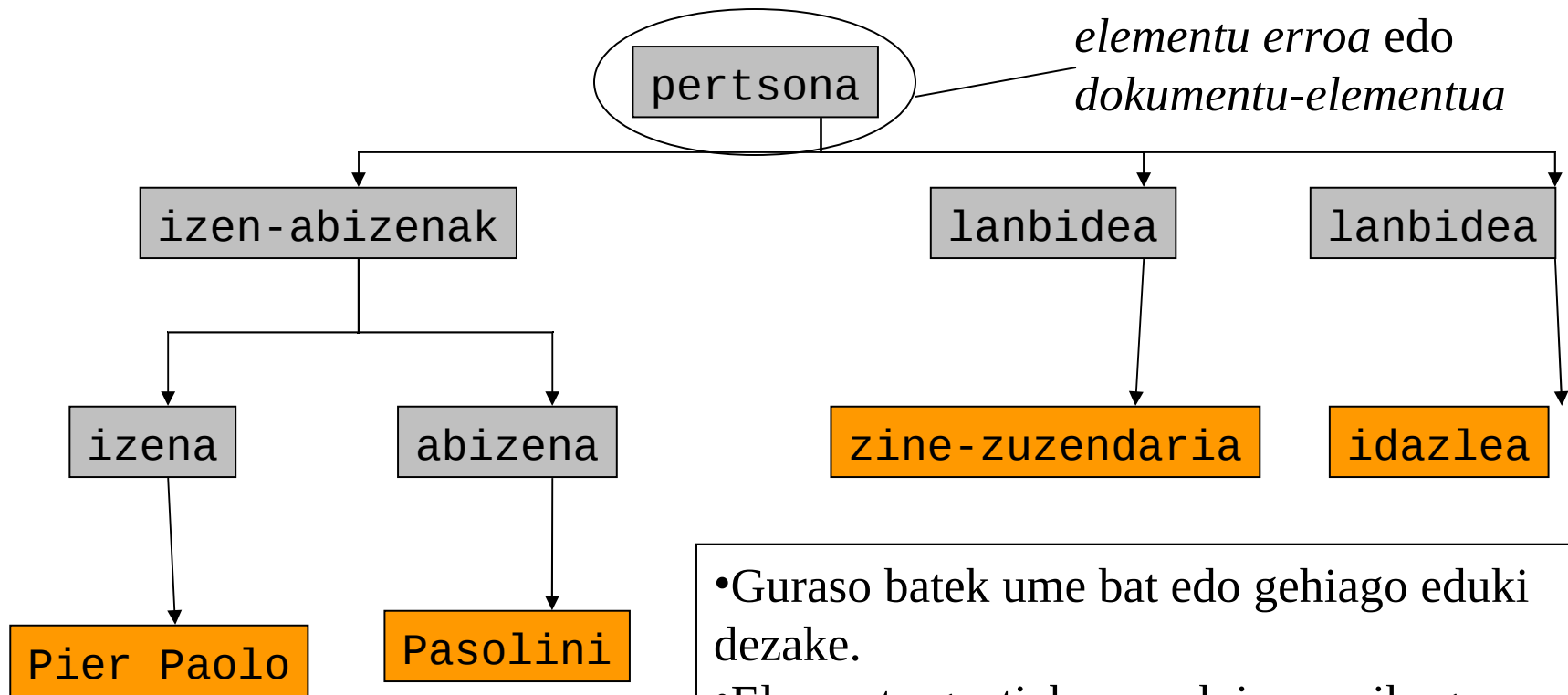


# XML zuhaitzak

- XML dokumentuak zuhaitz-erako datu-egituratzat har daitezke.
- Har dezagun honako adibide hau:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<pertsona>
  <izen-abizenak>
    <izena>Pier Paolo</izena>
    <abizena>Pasolini</abizena>
  </izen-abizenak>
  <lanbidea>zine-zuzendaria</lanbidea>
  <lanbidea>idazlea</lanbidea>
</pertsona>
```

# XML zuhaitzak



- Guraso batek ume bat edo gehiago eduki dezake.
- Elementu guztiek, erroak izan ezik, guraso bakarra dute.
- Erro bakarra eduki behar dute dokumentuek.

# Atributuak

- Elementuek atributuak eduki ditzakete.
- Atributua izen-balio bikotea da, elementuaren hasierako etiketan doana.
- Balioak, beti, komatxoan artean jartzen dira.

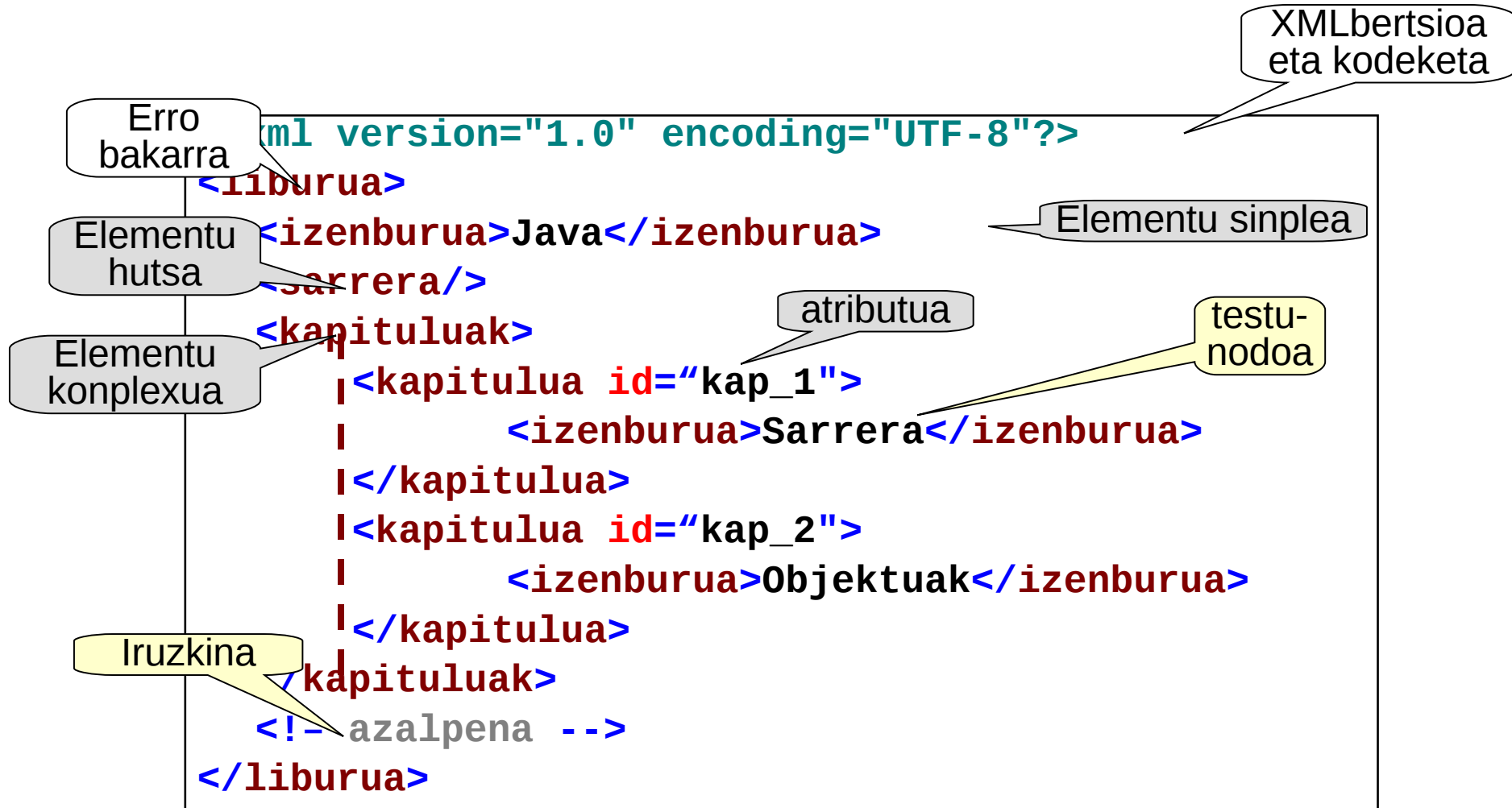
```
<pertsona id="G1900000" info-jatorria="wikipedia">  
  <izen-deiturak>  
    <izena>Pier Paolo</izena>  
    <deitura>Pasolini</deitura>  
  </izen-deiturak>  
  <jaioa>1922</jaioa>  
  <hila>1975</hila>  
</pertsona>
```

# Atributuak

```
<pertsona info-jatorria="wikipedia">
  <jaiotza-urtea>1922</jaiotza-urtea>
  <izen-abizenak>
    <izena>Pier Paolo</izena>
    <abizena>Pasolini</abizena>
  </izen-abizenak>
  <lanbidea mota="nagusia">zine-zuzendaria</lanbidea>
  <lanbidea mota="bigarrena">idazlea</lanbidea>
</pertsona>
```

- Noiz erabili elementu (ume)ak? Noiz atributuak?
  - Irizpide bat:
    - atributuak metainformaziorako erabili
    - elementu (ume)ak informaziorako
  - Kontuan hartzekoa:
    - elementu batek ez du onartzen atributu bat baino gehiago izen berarekin
    - atributuen balioek ez dute egiturarik onartzen.

# XML dokumentuak: laburpena



# XML dokumentu ongi eratua vs baliozkoa

- XML dokumentu bat ongi eratua (well-formed) bada:
  - XMLren sintaxia betetzen du
    - → dokumentua prozesagarria da.
- Baliozkoa (valid) bada:
  - Dokumentua ongi eratua dago, eta, gainera, edukiek eskema edo gramatika jakin baten erregelak betetzen dituzte: dokumentua ondo egituratuta dago.



# Formatu estandarrak

## Aurrekaria: MARC

- 1960ko hamarkadan
- MARC, MACHine-Readable Cataloging.
  - Ezagutza eta dokumentazio bibliografikoa errepresentatzeko estandarra.
  - Informazioa adierazteko marko bat, kontainer bat, ezartzen du.
  - Erregistro bakoitza eremutan antolatzen da: egilea, izenburua...
    - Eremuak azpieremutan banatzen ahal dira.
  - MARC formatuak:
    - Autoritate erregistroak
    - Erregistro bibliografikoak
    - Sailkatze-erregistroak
    - Komunitate-erregistroak
    - Antolatze-erregistroak

# MARC markoen edukiak: AACR(2)

- AACR: Anglo-American Cataloguing Rules
- Liburutegietako katalogoak sortzeko arauak definitzen dira.
- Nola osatu, zein informazio jarri MARC erregistroetako eremuetan.
  - Azken batean, nola deskribatu gaur egun bil daitekeen liburutegietako edozein material.

# MARC eremuak “tag”en bidez

- Eremu bakoitzari hiru digituko zenbaki bat egokitzen zaio. Zenbaki horri “tag” deitzen zaio.
- “tag”ek ondorenean datorren datuaren mota adierazten dute.

- Batzuetan *indikatzaileak* gehitzen zaizkie

- Erregistro baten adibidea:

245 14 \$a The emperor's new clothes /

\$c adapted from Hans Christian Andersen  
and illustrated by Janet Stevens.

# Bilakaera: MARC 21

- Jatorrizko MARC formatua 21. mendera egokitzeko diseinatu zen.
- ANSI Z39.2 estandarrean oinarritzen da, eta software-aplikazioei datu-trukerako aukera ematen die.
- Estandar hau inplementatu dute, adibidez, The British Library-k, Europear erakundeek eta Estatu Batuetako liburutegi garrantzitsuek.

# MARCXML

- MARCXML erregistroa

- Adibidea:

```
245 10|aArithmetic /|cCarl Sandburg ; illustrated as an anamorphic
adventure by Ted Rand.
```

```
<datafield tag="245" ind1="1" ind2="0">
<subfield code="a">Arithmetic /</subfield>
<subfield code="c">Carl Sandburg ; illustrated as an
anamorphic adventure by Ted Rand.</subfield>
</datafield>
```

# MARCEn ondotik

- Liburutegien mundutik: METS
  - METS: Metadata Encoding and Transmission Standard
- Multimedia mundutik: DIDL
  - DIDL: Digital Item Declaration Language
  - DIDL = MPEG-21 estandarraren 2. partea
- Fedora Digital Object Model
  - XMLn definitutako eskema: FOXML
- Hiruek framework orokorrak definitzen dituzte, non informazio deskriptiboa, administratiboa eta egiturazkoa sartzen den.

# METS

- XML eskema bat da.
- Adierazten du:
  - Liburutegiko objektu digitalen egitura hierarkikoa.
  - Objektu horiek osatzen dituzten fitxategien izenak eta kokapenak.
  - Metadatu asoziatuak.
- Metadatuena eta fitxategi-esteken kontainerra da.
  - Metadatuak banatu egiten dira, eta beren arteko estekak adierazi. Egiturazko mapa moduko bat eratzen da.
  - Oro har, modu hedagarrian eta modularrean antolatzen da informazioa.
- Malgua da, eta beste eskemetako elementuak biltzen dira.
- Eredu hau oso baliagarria da informazioa biltzeko eta transmititzeko.
- **Mets eskema**

# METS dokumentu baten zazpi atalak

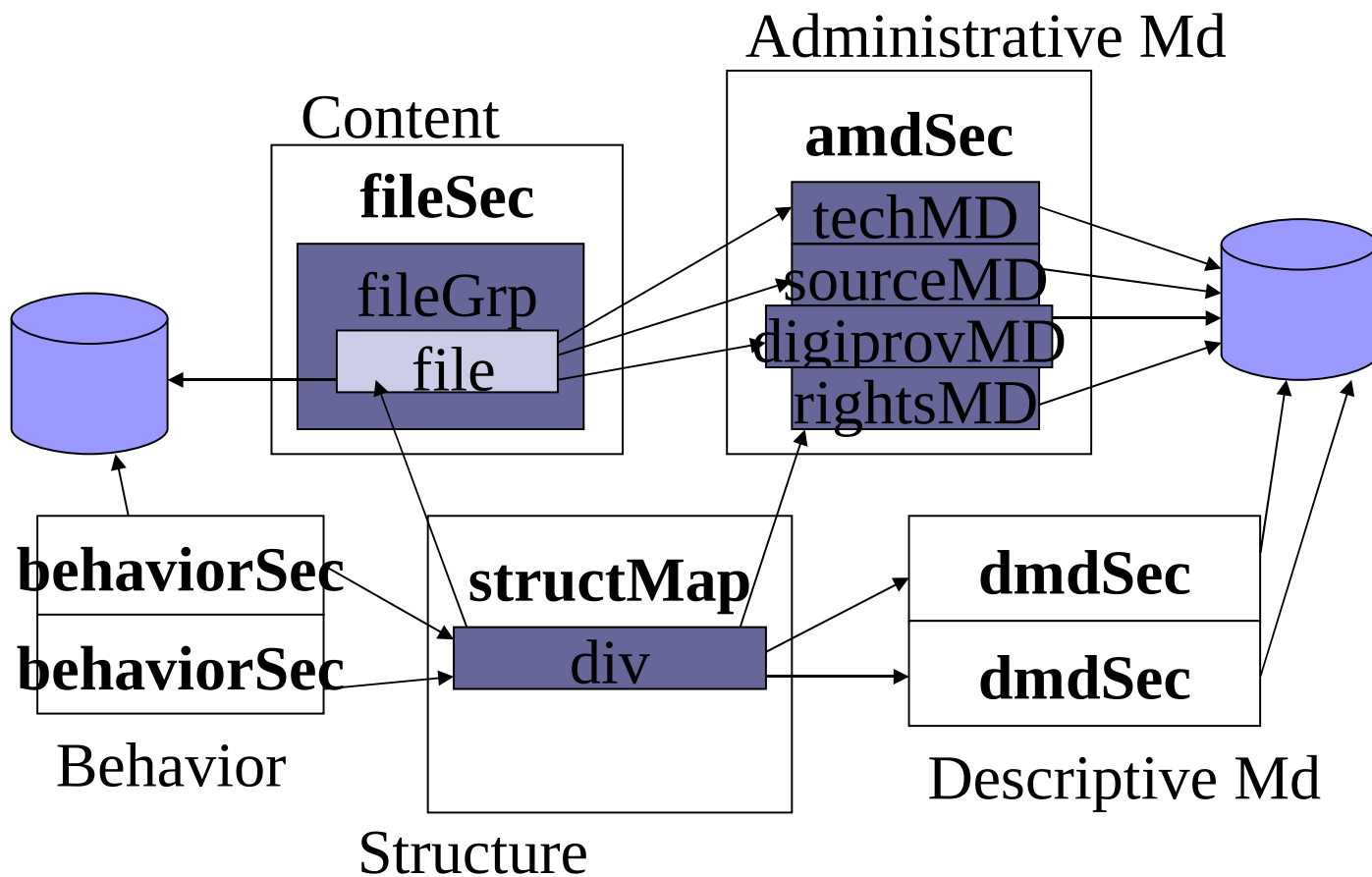
```
<mets>
```

```
  <metsHdr/>      -- Mets fitxategi berari buruz  
  <dmdSec/>       -- metadatu deskriptiboak  
  <amdSec/>       -- metadatu administratiboak  
  <fileSec/>      -- fitxategi-sekzioa  
  <structMap/>    -- mapa estrukturala  
  <structLink/>   -- estekak egituraren  
  barne  
  <behaviorSec/> -- exekutagarriak
```

```
</mets>
```



# METS egitura. Diagrama



# Metadatu deskriptiboak

- DC: Dublin Core
  - Oso erabilia edozein motatako baliabideak deskribatzeko.
  - Bi maila ditu: Simple eta Qualified.
    - Maila sinplean 15 oinarrizko elementu baino ez dira erabiltzen
  - Beste estandar garrantzitsuen oinarria:
    - OAI-PMH: Open Archives Initiative's Protocol for Metadata Harvesting.
- MODS: Metadata Object Description Schema
  - Elementuen multzo aberatsagoa eskaintzen du.
    - Zehaztasun handiagoa.
- Biek dituzte jarraitzaileak, eta zaila da esaten bietatik zein nagusituko den etorkizunean.

# Dublin Core

- Maila sinplean 15 elementu-mota baizik ez:
  - Adibidez: creator, subject, identifier
  - *Simple Dublin Core eskema*
- Maila kualifikatuan:
  - Beste hiru elementu (*audience, provenance* eta *rightsholder*)
  - Elementuen kualifikatzaileak (atributuak) gehitzen dira.
    - Adibidez: creator.author
- Zehaztasuna hobetzeak, hala ere, DC metadatuaren elkarreragingarritasuna zaildu egiten du.

# METS eta Dublin Core Adibidea

```
<mets:dmdsec id="dm3">  
  <mets:mdwrap mdtype="other">  
    <mets:xmldata>  
      <dc:dc>  
        <dc:title>[patrick breen diary]  
        </dc:title>  
        <dc:creator>breen, patrick  
        </dc:creator>  
      </dc:dc>  
    </mets:xmldata>  
  </mets:mdwrap>  
</mets:dmdsec>
```

# MODS

- Deskriptiboa da, ez du zehazten arau espezifikorik.
- Elementuen multzo egokia baliabide digitaletarako
- MARCetik eratorria
- MODS eskema

# MODS: goi-mailako elementuak

- Title Info
- Name
- Type of resource
- Genre
- Origin Info
- Language
- Physical description
- Abstract
- Table of contents
- Target audience
- Note
- Subject
- Classification
- Related item
- Identifier
- Location
- Access conditions
- Part
- Extension
- Record Info

# METS eta MODS

## Adibidea

```
<mets:dmdsec id="...">
  <mets:mdwrap mdtype="other">
    <mets:xmldata>
      <mods:mods>
        <mods:titleInfo>
          <mods:title>
            Bernstein conducts Beethoven and Mozart
          </mods:title>
        </mods:titleInfo>
        <mods:name>
          <mods:namePart>Bernstein, Leonard
        </mods:namePart>
        </mods:name>
      </mods:mods>
    </mets:xmldata>
  </mets:mdwrap>
</mets:dmdsec>
```

# Metadatu administratiboak

- Ikuspegi teknikoa
  - Irudiak, argazkiak
    - *MIX: Metadata for Images in XML*
    - *EXIF: Exchangeable Image File Format*
    - *IPTC Photo Metadata*
    - XMP (Adobe)
  - Testuak
    - *TEI: Text Encoding Initiative*
  - Audioa
    - Adibidez: AUDIOMD
  - Bideoa
    - VIDEOMD



# Metadatu administratiboak

- Eskubideen metadatuak
  - *METS Schema for Rights Declaration*
  - *XrML: eXtensible Rights Markup Language*
  - *ODRL: Open Digital Rights Language*
    - *Creative Commons agreement*

# Metadatu administratiboak

- Bilakaerari eta kontserbazioari buruzko metadatuak
  - *PREMIS: PREservation Metadata Implementation Strategies*
  - Eskema-konbinazio bat da:
    - PREMIS container
    - Object entity
    - Event entity
    - Agent entity
    - Rights entity
  - Gainjartzeak daude beste eskema batzuekin.

# Metadatuak integratzeko hurbilpena

- Izen-espazio desberdinak erabiltzeko aukera ematen du METS estandarrak.
- Horri esker, posible da METS dokumentu berean eskema lagungarriak erabiltzea.

# METS

## Estandarren kontainerra

Meta-datu deskriptiboak		<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• MODS</li> <li>• MARCXML</li> </ul>
Meta-datu administratiboak	Teknikoak	<ul style="list-style-type: none"> <li>• Irudiak: MIX, exif, iptc</li> <li>• Bideoa: VIDEO MD (PBCore)</li> <li>• Audioa: AUDIO MD</li> <li>• Testua: Eskema tekniko berezituak</li> </ul>
	Eskubideak	<ul style="list-style-type: none"> <li>• <i>METS Schema for Rights Declaration</i></li> <li>• XrML</li> <li>• ODRL</li> <li>• PREMIS (<i>rights object</i>)</li> </ul>
	Iturburua	<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• MODS</li> </ul>
	Jatorria	<ul style="list-style-type: none"> <li>• PREMIS (<i>events object</i>)</li> </ul>
Egiturazko meta-datuak		<ul style="list-style-type: none"> <li>• div elementuak</li> </ul>

## Logical (MODS)

```
<mods:mods ID="ver01">
  <mods:titleInfo>
    <mods:title>Original Work (vesion 1)</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="otherVersion" ID="ver02">
    <mods:titleInfo>
      <mods:title>Derivative Work 1</mods:title>
    </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="otherVersion" ID="ver03">
    <mods:titleInfo>
      <mods:title>Derivative Work 2</mods:title>
    </mods:titleInfo>
  </mods:relatedItem>
</mods:mods>
```

mods eta relatedItem type  
="otherVersion" elementuak

3 nodoen sekuentzia

## Physical (structMap)

```
<mets:structMap>
  <mets:div TYPE="photo:photoObject" DMDID="MODS1">
    <mets:div TYPE="photo:version" DMDID="ver01" >
      <mets:div TYPE="photo:image">
        <mets:fptr FILEID="FN10081"/>
      </mets:div>
    </mets:div>
    <mets:div TYPE="photo:version" DMDID="ver02">
      <mets:div TYPE="photo:image">
        <mets:fptr FILEID="FN10090"/>
      </mets:div>
    </mets:div>
    <mets:div TYPE="photo:version" DMDID="ver03">
      <mets:div TYPE="photo:image">
        <mets:fptr FILEID="FN1009F"/>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
```

div TYPE="photo:version" elementuak

3 nodoak sekuentzia logikoaren lotuta  
daude ID/IDREF atributuaren bidez

# METS eta MODS

## Adibidea

- <http://lcweb2.loc.gov/cocoon/ihas/loc.natlib>

# Integrazio-eredu honen zenbait arazo

- Eskema desberdinen arteko integrazioa bide egokitzat hartzen da liburutegi digitalen inguruneak antolatzeko, baina...
- Zenbaitetan, METS ereduko elementuak ez datoz (guztiz) bat eskema lagungarriekin.
  - Adibide garbiena PREMIS da:
    - PREMIS objektu bat ezin da osorik gorde
    - *agent* eskema anbigua da: *provenance/rights*
- Erredundantziak gertatzen dira:
  - PREMISeko *object* vs. MIX
  - METS structural map vs. PREMIS
- Metadatuak osatzeko software egokiaren premia!

# METS profila

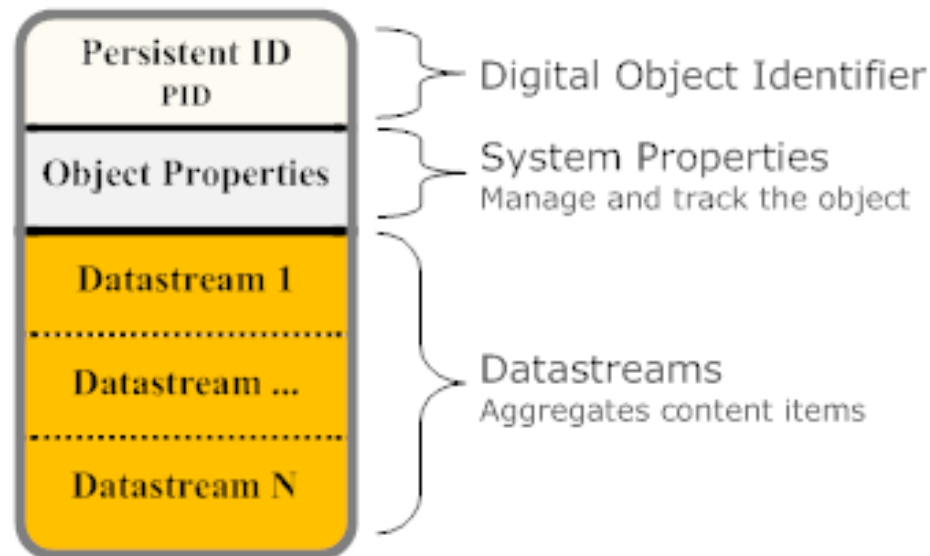
- XML dokumentu bat da.
- Helburu jakin bati erantzuten dion METS *zehatza*.
  - Testuinguru horretarako eskema.
- *Library of Congress*-en erabiltzen diren profilak, adibideak:
  - Sheet Music, Musical Score, Print Material (books, pamphlets, etc), Music Manuscript (score or sketches), Recorded Event (audio or video), PDF Document, Bibliographic Record, Photograph, Compact Disc, Collection



# Beste aukera bat

## *Fedora Digital Object Model*

- FOXML formatua:



- Bestelako formatuak ere irensten eta esportatzen ditu, adibidez: METS

# Aplikazio zehatzei orientatutako metadatuak

## ■ Irakaskuntza

- *IMS Learning Resource Meta-data Information Model*

  - Adibide mistoa: Dublin Core + IMS

- SCORM (e-learning)

## ■ Medikuntza

- TextMD (Technical Metadata for Text)

# Ondorioak eta etorkizuna

- Oraindik ez dago estandar simple eta bakarra!
- Sinpletasunaren eta zehaztasunaren arteko oreka zaila da.
- XMLren eta integrazioaren bidetik.
  - Eskema simple eta bakuna iristeraino?
- Deskribapen bibliografikoa eta autoritate-konbentzioak adostu beharra:
  - AACR -> RDA (Resource Description and Access)
- Beharrezko da software-garatzailleek estandar hauek erabiltzea, metadatu jabedunen ordezt.
  - Sekulako eragina izango du garatuko diren aplikazioetan.

# Ondorioak eta etorkizuna

## Kooperazio- eta sozializazio-tresnak

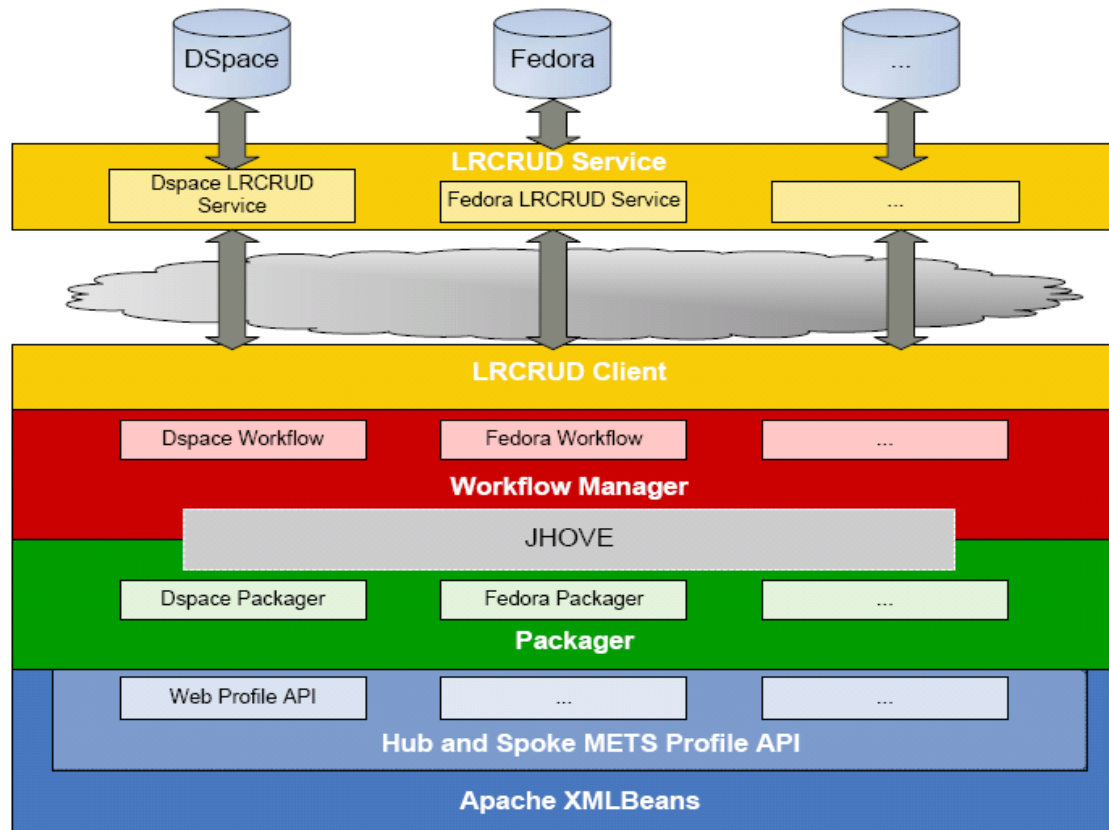
- Baliabideen bilketa eta zabalkunde kooperatiboa (elkarreraginkortasuna):
  - OAI: Open Archives Initiative
  - OAI-PMH: Protokolo bat da baliabideak biltzeko. Googlek, adibidez, erabiltzen du. Datu-emaileak behartuta daude Dublin Core formatuan datuak ematera
  - OAIster: Dublin Core sinplean kodetutako 14 milioi baliabide baino gehiago kontsultagarri.
- Sare sozialak, mobilak...
- Pertsonalizazioa:
  - Adibidez: Zotero

# Ondorioak eta etorkizuna

## Software-inguruneak

- Garrantzitsua da software-ingurune egokiak eta erabiltzaile-entzafaze erabilerrazak garatzea.
  - Bestela, ikusitako eskema estandarren erabilera gordina ez da praktikoa.
- Inguruneen adibideak:
  - *FedoraCommons*, *Greenstone* edo *Dspace*

# Ondorioak eta etorkizuna Software-arkitekturak



# Bibliografia

## Estandarrak

- <http://www.loc.gov/standards/marcxml>
- <http://www.loc.gov/standards/mets>
- <http://www.loc.gov/standards/mix>
- <http://www.loc.gov/standards/mods>
- <http://www.loc.gov/standards/premis/>
- <http://dublincore.org>

# Bibliografia

## Normalizazioa

- FRBR (*Functional Requirements for Bibliographical Records*): <http://www.frbr.org/>
- RDA (*Resource Description and Access*): <http://www.rda-jsc.org/rda.html>
- LOCeko autoritateak: <http://authorities.loc.gov/>



# Bibliografia

## Software-inguruneak eta kooperazioa

- Greenstone digital library software:  
<http://www.greenstone.org/>
- Fedora Commons Repository Software:  
<http://www.fedora-commons.org/>
- Dspace: <http://www.dspace.org/>
- OAI (Open Archives Initiative)  
<http://www.openarchives.org/>

# Bibliografia

## XML

- Arregi X., Artola X.. *XML: Oinarriak*.

<http://www.unibertsitatea.net/otarraa/ingeniaritza-eta-teknologia-1/informatika>

- Arregi X., Artola X.. *XML: DTD eta XML Schema*.

<http://www.unibertsitatea.net/otarraa/ingeniaritza-eta-teknologia-1/informatika>

# Bibliografia

- Gartner, R.. "Metadata for digital libraries: state of the art and future directions."  
[http://www.jisc.ac.uk/media/documents/techwatch/tsw\\_0801pdf.pdf](http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf)
- Carrasco, R.. Hitzaldia: "Interneteko liburutegi digitalak."  
<http://www.unibertsitatea.net/blogak/ixa/hitzaldia-interneteko-liburute>
- Cundiff, M.. Aurkezpena: "Using METS and MODS to Create an XML Standards-based Digital Library Application"  
[http://www.loc.gov/standards/mets/presentations/Digital\\_Future\\_Cun](http://www.loc.gov/standards/mets/presentations/Digital_Future_Cun)
- "Workbook on Digital Private Papers" (bereziki, 5. kapitulua: "Administrative and preservation metadata")  
<http://www.paradiam.ac.uk/workbook/>