Access IT Training

How to (and why) prepare a repository for Europeana

Europeana

- Portal which gives access to European cultural heritage
  - **http://europeana.eu**
- Information comes from:
  - Museums
  - Archives
  - Libraries
  - Audiovisual collections

Basic assumptions

- Europeana functional specification distinguishes five group of future users:
  - General User
  - School Child
  - Academic User (both students and teacher)
  - Expert Researcher
  - Professional user e.g. librarian, archivist, etc.

Basic assumptions

- Each group has different skills and needs
- Different objectives:
  - Looking for an answer to particular question
  - Looking for entertainment
- One thing all end-users have in common
  - They want access to the Europeana full content through search and browse

Basic assumptions

- Other user activities may include:
  - View and download a film footage
  - Copy and paste information for a paper they are writing
  - Create sets of preferred items
  - Study details of high resolution reproduction of cultural object
  - Upload a personal item
  - Enrich the description of materials through tagging

Basic assumptions

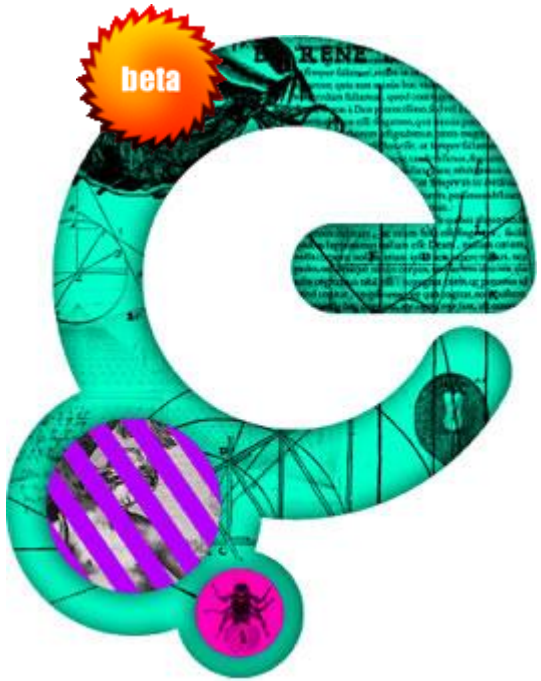
- Other user activities may include (2):
  - Search for information using simple/advanced/predefined queries
  - Sharing information with friends
  - Getting notifications about new objects from given thematic area
  - Browsing through time dimension - timeline

Basic assumptions

- External applications will be able to access Europeana through set of public APIs
  - e.g. content provider may get tags for their objects from Europeana
- Europeana will store
  - objects' metadata
  - thumbnail
  - link to content in original context

Europeana

- First prototype version was enabled on 20.11.2008
- Now Europeana gives access to over 5 million of digital objects distribute all over the Europe
- Europeana is a **metadata directory** access to the contents of the digital objects is made on the websites of their origin

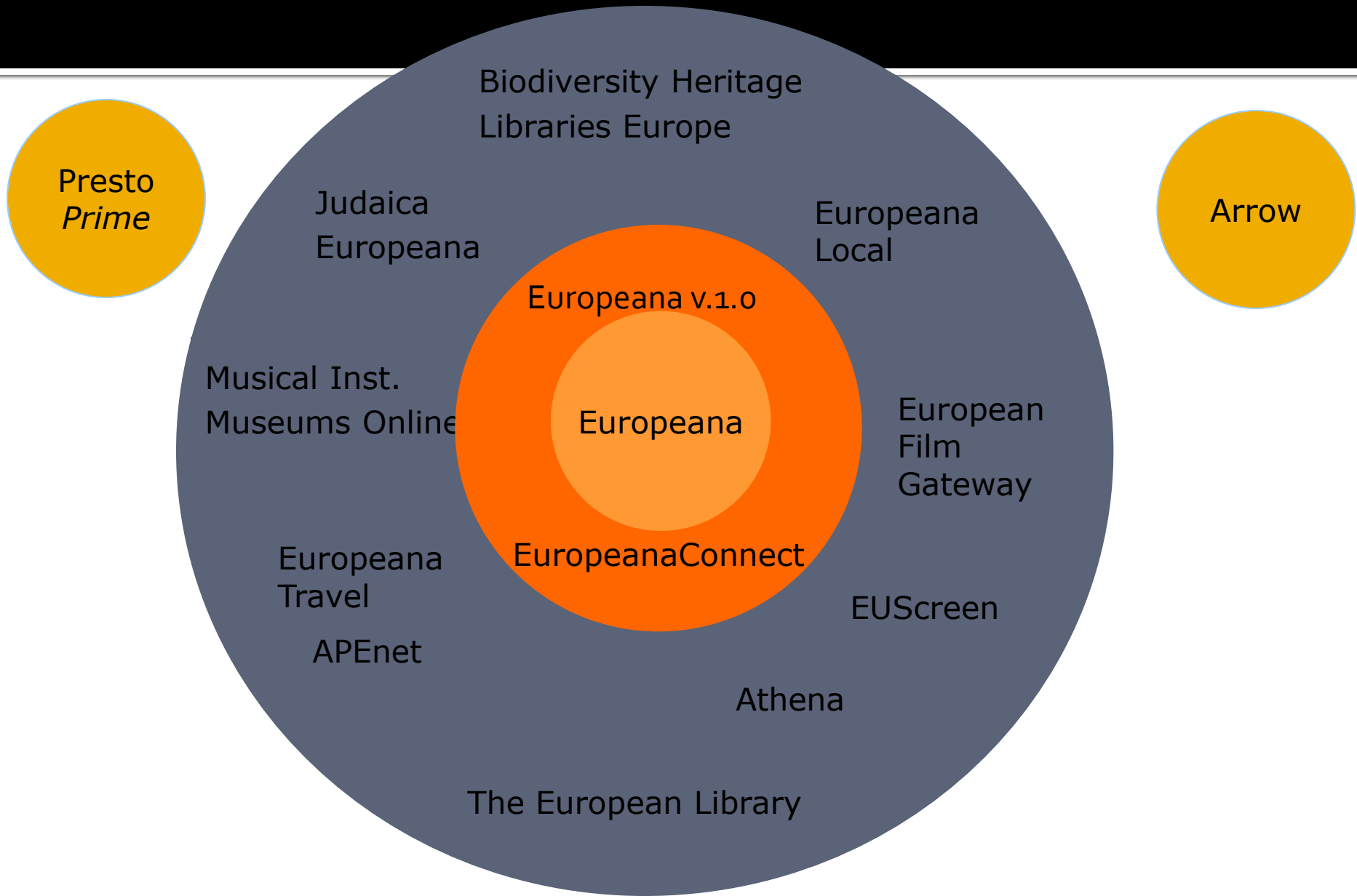


Europeana

- Main way of financial support for this initiative are projects co-funded by the European Commission
  - Previously under eContent*Plus* programme
  - Now CIP ICT-PSP
    - Theme 2: Digital Libraries
      1. European Digital Library – services
      2. European Digital Library – aggregating digital content in Europeana
      3. European Digital Library – digitising content for Europeana
      4. Open access to scientific information
      5. Use of cultural heritage material for education

europeana
pomyśl o kulturze

Europeana Group projects



Europeana

- Ongoing projects
  - Technical/organizational
    - Europena v1.0
      - Should result in a production-ready version of Europeana
    - Europeana Connect
      - Development of technologies necessary for the Europeana
    - PrestoPRIME
      - Long-term preservation of audiovisual materials

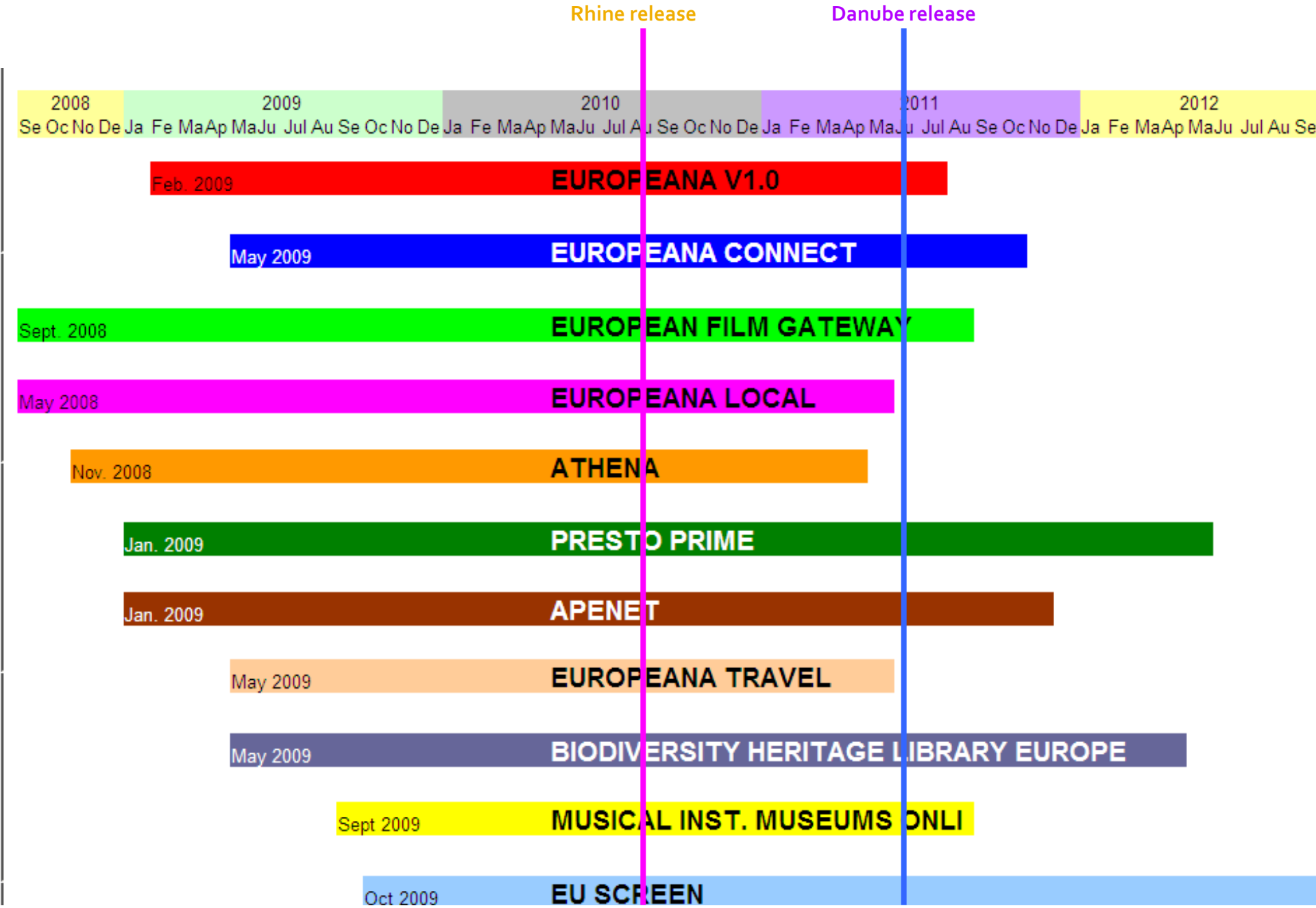
europeana
pomyśl o kulturze

Europeana

- Ongoing projects
  - Content providers
    - APEnet – national archives
    - ATHENA – museums (national level)
    - BHL – Europe – biodiversity heritage library
    - EUscreen – TV materials
    - Europeana Connect – audio materials
    - Europeana Local – materials from local and regional institutions
    - Europeana Travel – travel, tourism, …
    - Judaica Europeana – influence of Jewish culture on European cities
    - EFG – movies/cinema

europeana
pomyśl o kulturze

More information at: http://group.europeana.eu/


Rhine release
Danube release
2008
Se Oc No De
2009
Ja Fe Ma Ap Ma Ju Jul Au Se Oc No De
2010
Ja Fe Ma Ap Ma Ju Jul Au Se Oc No De
2011
Ja Fe Ma Ap Ma Ju Jul Au Se Oc No De
2012
Ja Fe Ma Ap Ma Ju Jul Au Se
Feb. 2009
EUROPEANA V1.0
May 2009
EUROPEANA CONNECT
Sept. 2008
EUROPEAN FILM GATEWAY
May 2008
EUROPEANA LOCAL
Nov. 2008
ATHENA
Jan. 2009
PRESTO PRIME
Jan. 2009
APENET
May 2009
EUROPEANA TRAVEL
May 2009
BIODIVERSITY HERITAGE LIBRARY EUROPE
Sept 2009
MUSICAL INST. MUSEUMS ONLI
Oct 2009
EU SCREEN

Upcomming releases

Europeana Version 1.0

- Full services and functionalities
- Greater content
  - Summer 2010
    - Rhine Release - 10 million items
  - 2011
    - Danube Release - expect to double content
  - 2012
    - 25 million items
    - Further growing content

Why bother with Europeana?

- Prestigious initiative
  - Endorsement from European Commission
  - Erasmus Award 2009
- Knowledge exchange with professional network
  - Metadata standards
  - Best practices
  - Technological innovation
- Popularity among users
  - User survey results:
    - Loyal user base (60% of respondents visiting the site more than 5 times);
    - Overall positive ratings for Europeana features and functions

Why bother with Europeana?

- Reaching out to users
  - Remain relevant
  - Put content where people are
  - Open up your marvelous collections
- Content remains within your organization
- Increase traffic to your site
  - User interest in viewing items in original context
  - 75% of Europeana user survey respondents thought it very useful to view the searched object in its original context.

Statistics - User survey

- Online User Survey **6-26 May 2009**
- **3,204** completed
- Replies from **54** countries - 53% of replies from five countries
- Almost everyone expects to visit the site again – **less than 1%** says they will not revisit
- Main route to Europeana is from a **paper or journal** (47.4%), second most popular is a link from **another web site** (21%)
- **Personal research** is dominant reason (72.9%)

Statistics - User survey

- Majority rate features and functions as "good" or "excellent". Around a third of all respondents only rate the general features and functions as "average".

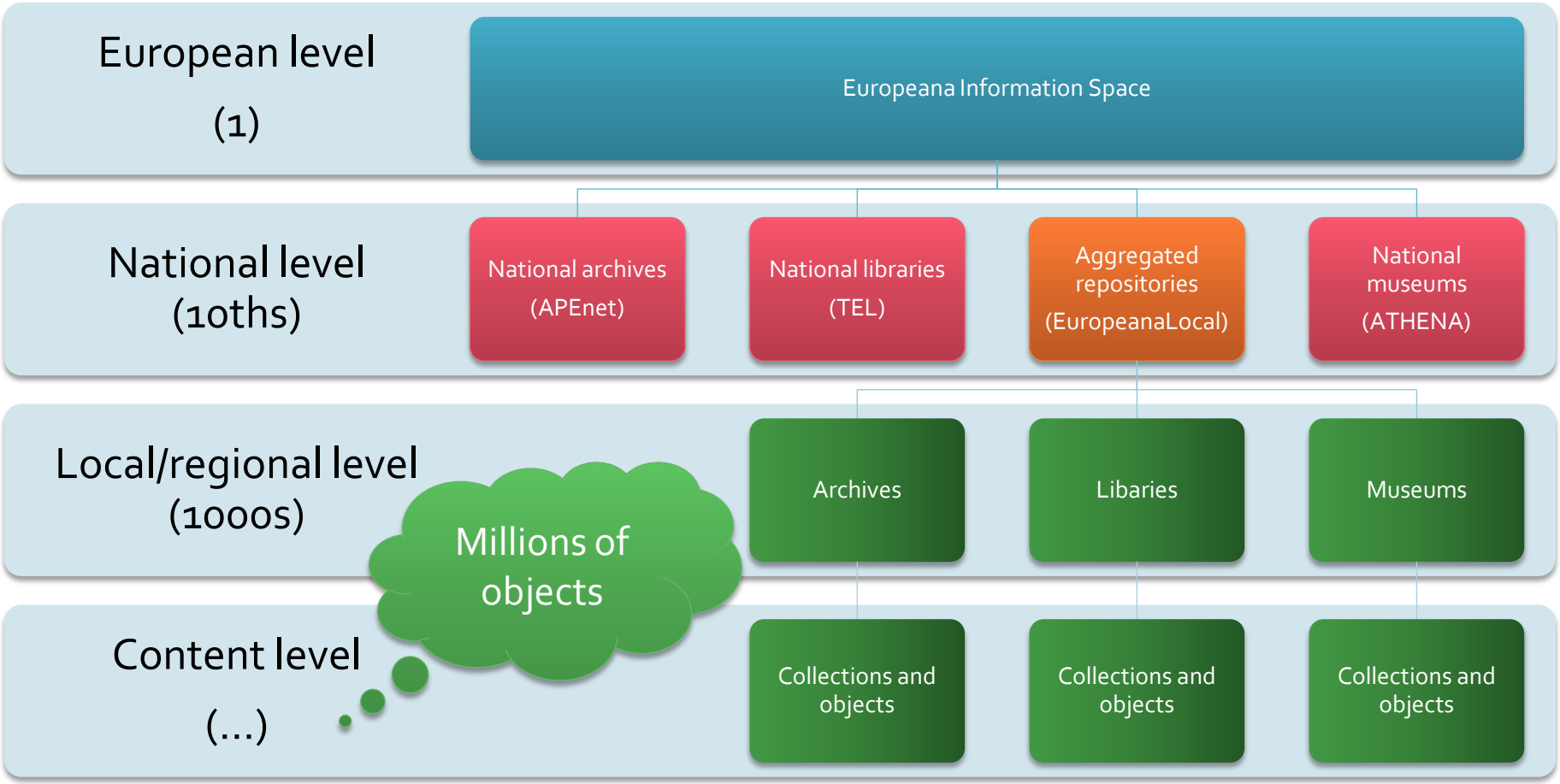
Content analysis

- Content at launch: **4.7 million items from every domain,** every EU member
- **3,500,000 images**: photos, paintings, drawings, postcards, posters
- **1,000,000 texts:** books, newspaper articles, manuscripts, letters
- **82,000 videos:** movies, documentaries, TV broadcasts, public information films
- **14,000 sounds:** cylinders, 78rpm discs, radio, field recordings

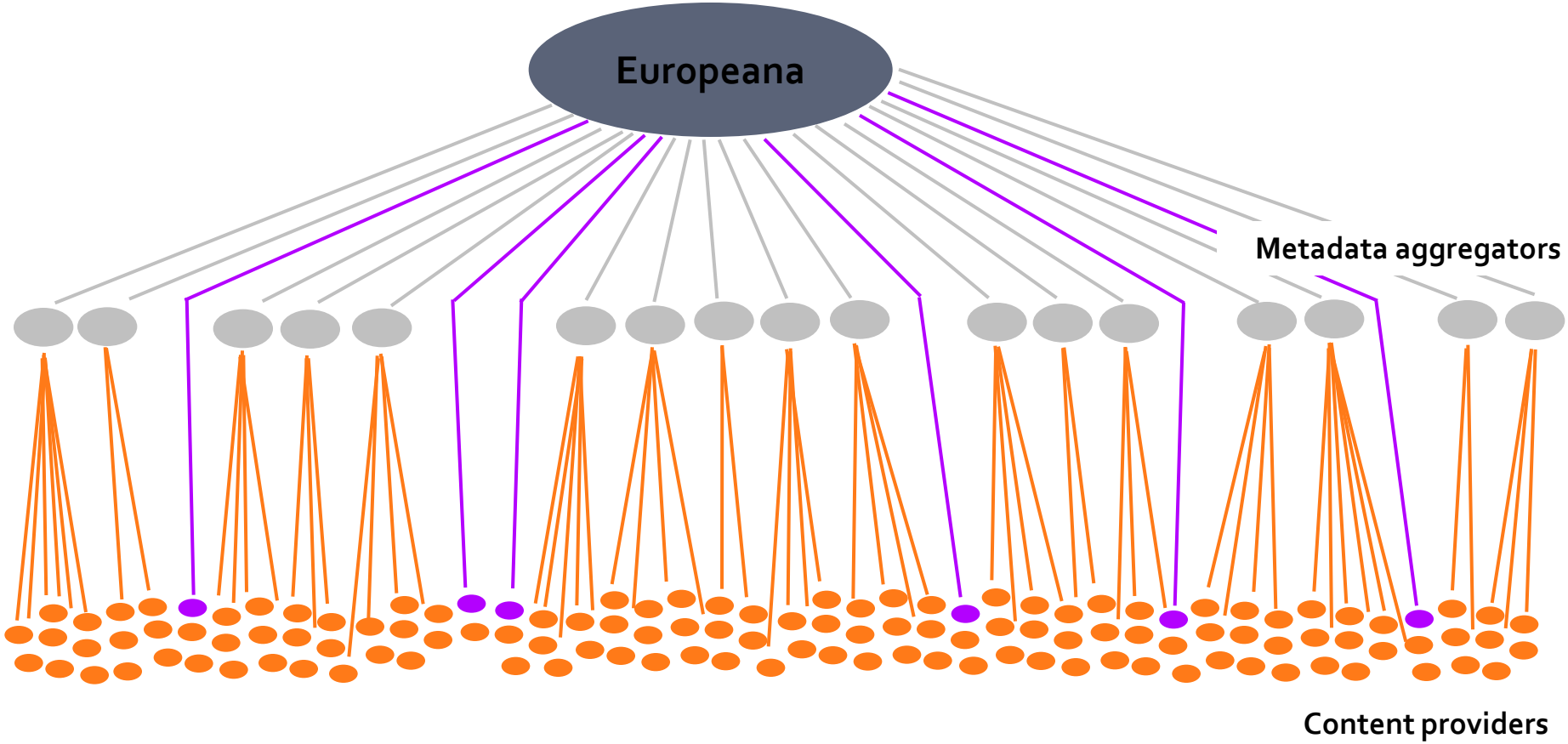
How to join Europeana?

- Choose a metadata aggregator
- Map your metadata to Europeana Semantic Elements schema
- Normalize the metadata
- Test the metadata with Europeana
- Publish the metadata in the „production" version of Europeana

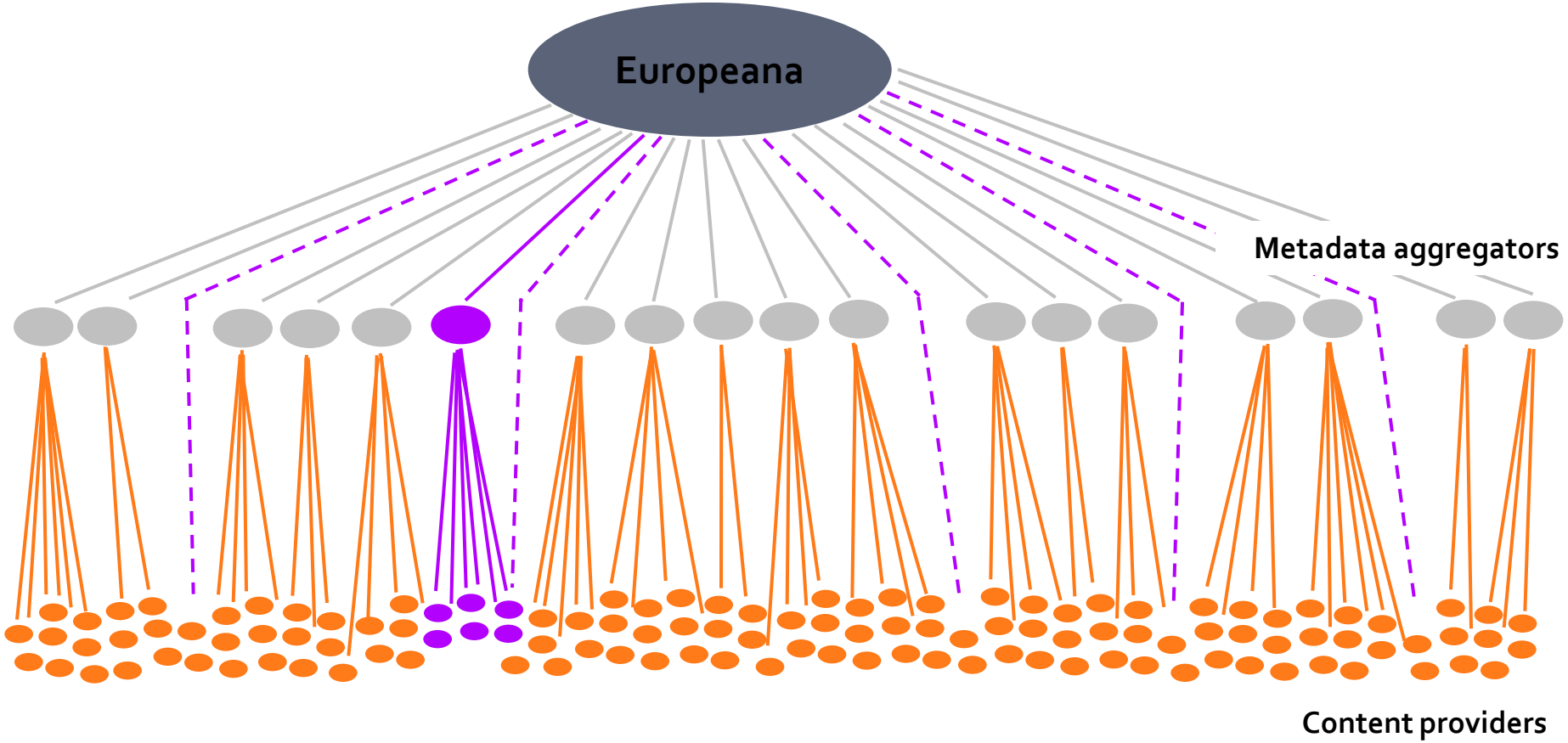
Local and regional resources in the European information space



Present Europeana model for metadata aggregation



Target Europeana model for metadata aggregation



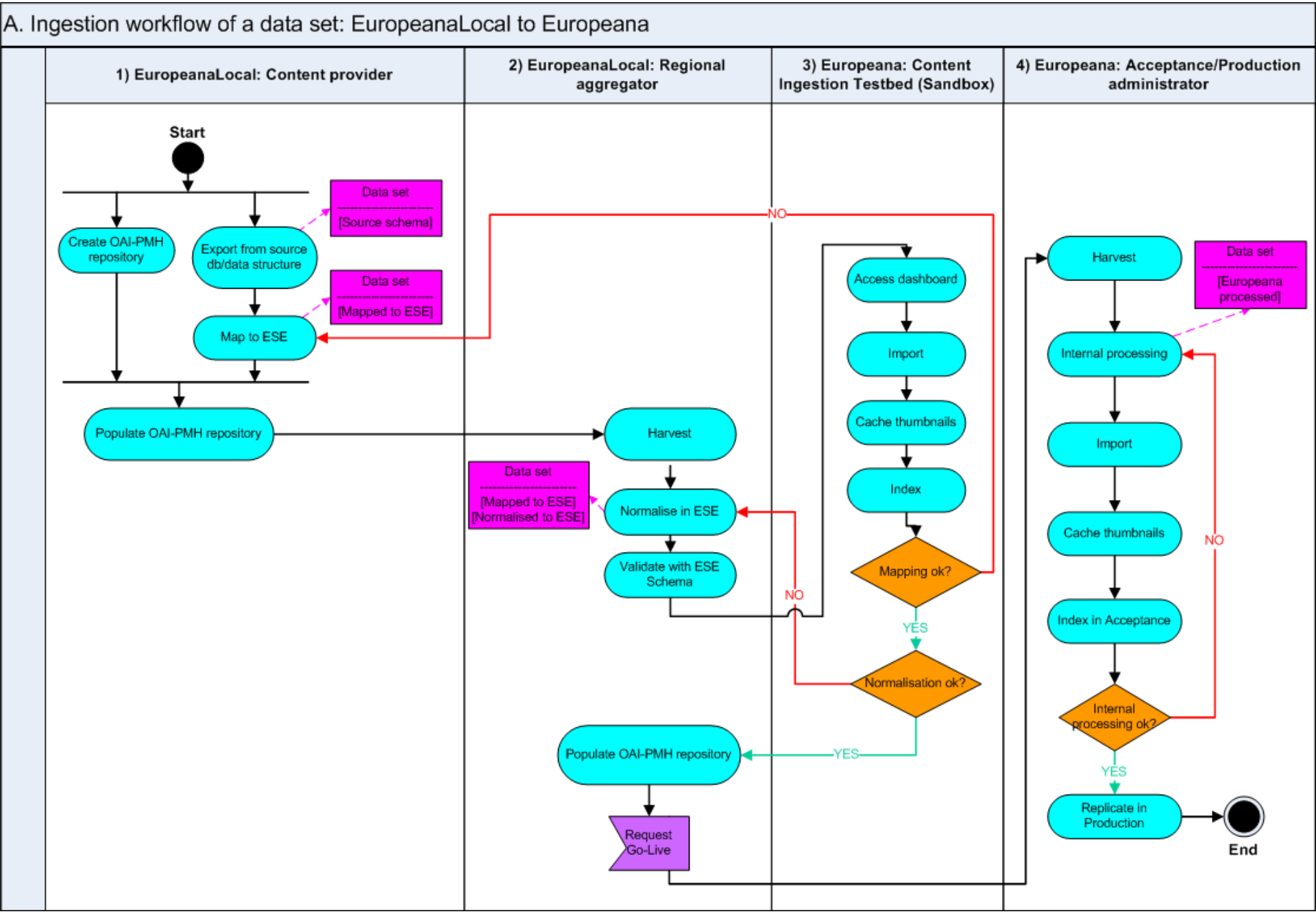
Metadata aggregators

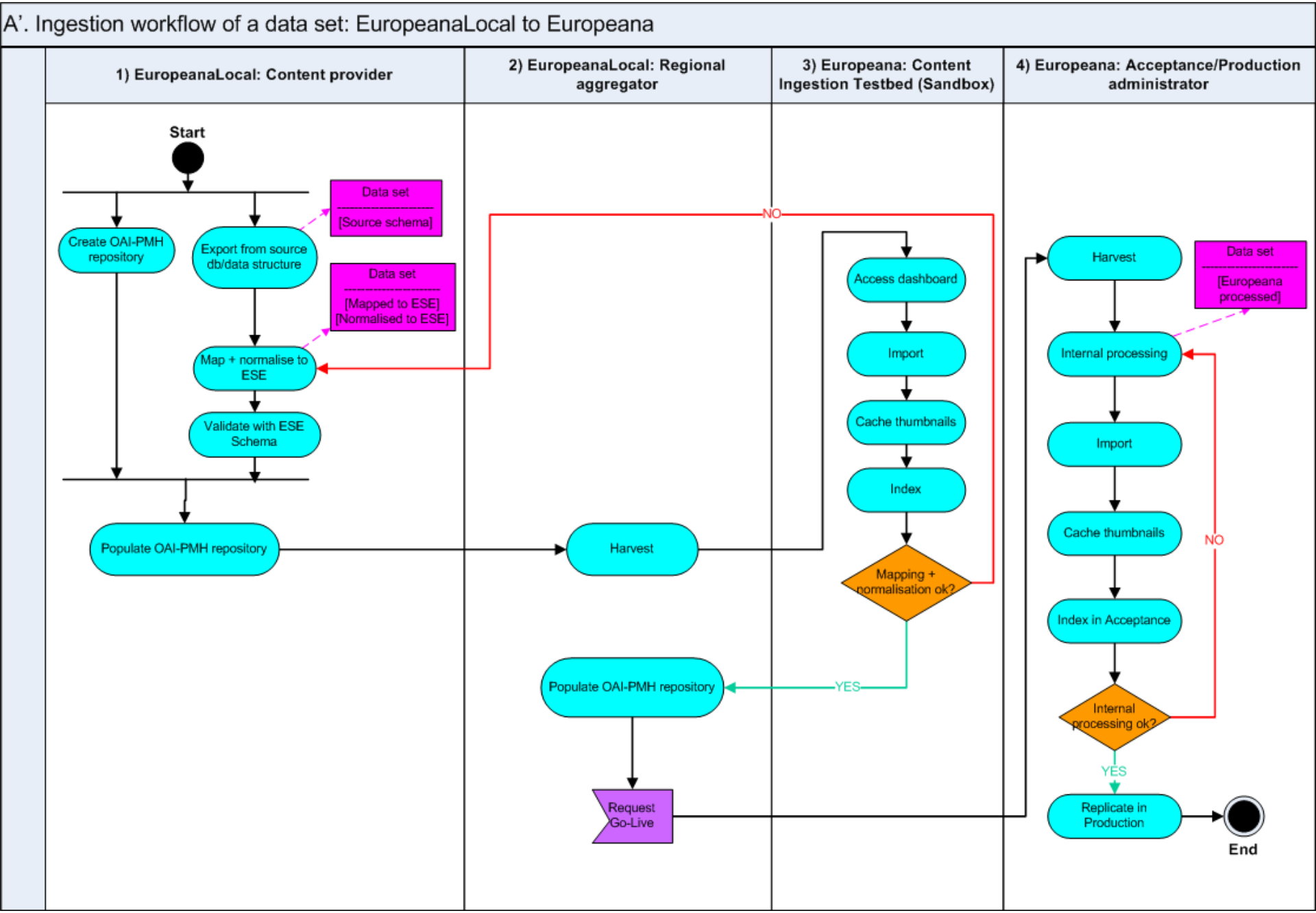
- According to the present version of Europeana Outline Functional Specification tasks for the aggregator are:
  1. To gather the information about content providers and their information systems
  2. To gather the metadata of objects that should be visible in Europeana
  3. To remove duplicates, clean-up the metadata, normalize it and enrich it
  4. To confirm the accessibility of digital objects
  5. To expose the aggregated metadata for Europeana via the OAI-PMH protocol

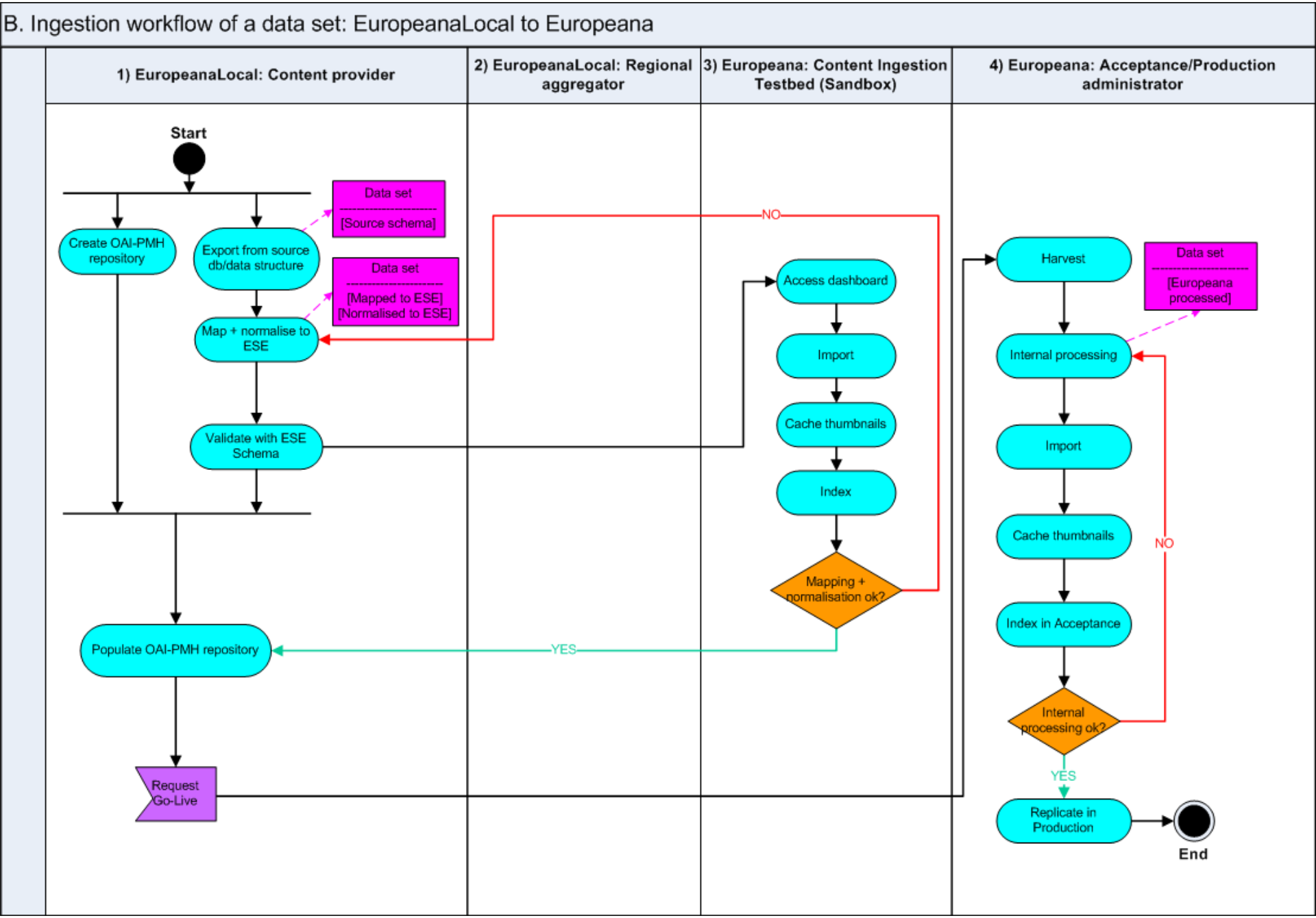
http://dev.europeana.eu/public_documents/EDLnet%20D2.5_Outline_Functional_Specifications20090301_version%201.7_consWithoutHistory_lossless.pdf

A. Ingestion workflow of a data set: EuropeanaLocal to Europeana

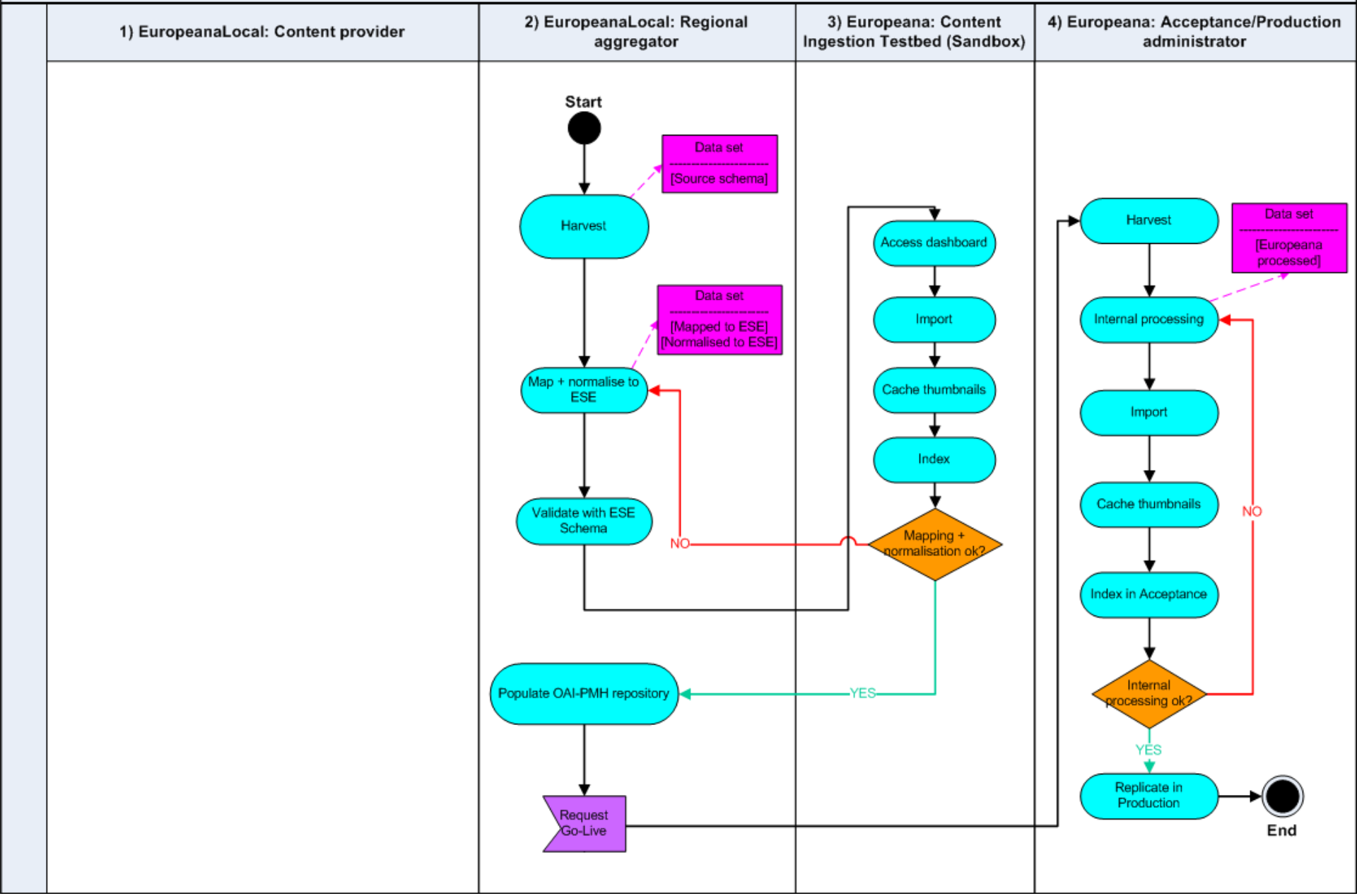
1) EuropeanaLocal: Content provider	2) EuropeanaLocal: Regional aggregator	3) Europeana: Content Ingestion Testbed (Sandbox)	4) Europeana: Acceptance/Production administrator
Start			
			End


A. Ingestion workflow of a data set: EuropeanaLocal to Europeana
1) EuropeanaLocal: Content provider
2) EuropeanaLocal: Regional aggregator
3) Europeana: Content Ingestion Testbed (Sandbox)
4) Europeana: Acceptance/Production administrator
Start
Create OAI-PMH repository
Export from source db/data structure
Data set
[Source schema]
Map to ESE
Data set
[Mapped to ESE]
Populate OAI-PMH repository
NO
Harvest
Data set
[Mapped to ESE]
[Normalised to ESE]
Normalise in ESE
Validate with ESE Schema
Access dashboard
Import
Cache thumbnails
Index
Mapping ok?
NO
YES
Normalisation ok?
NO
YES
Populate OAI-PMH repository
Request Go-Live
Harvest
Data set
[Europeana processed]
Internal processing
Import
Cache thumbnails
Index in Acceptance
Internal processing ok?
NO
YES
Replicate in Production
End


A'. Ingestion workflow of a data set: EuropeanaLocal to Europeana
1) EuropeanaLocal: Content provider
2) EuropeanaLocal: Regional aggregator
3) Europeana: Content Ingestion Testbed (Sandbox)
4) Europeana: Acceptance/Production administrator
Start
Create OAI-PMH repository
Export from source db/data structure
Data set
[Source schema]
Data set
[Mapped to ESE]
[Normalised to ESE]
Map + normalise to ESE
Validate with ESE Schema
Populate OAI-PMH repository
Harvest
Access dashboard
Import
Cache thumbnails
Index
Mapping + normalisation ok?
NO
YES
Populate OAI-PMH repository
Request Go-Live
Harvest
Data set
[Europeana processed]
Internal processing
Import
Cache thumbnails
Index in Acceptance
Internal processing ok?
NO
YES
Replicate in Production
End


B. Ingestion workflow of a data set: EuropeanaLocal to Europeana
1) EuropeanaLocal: Content provider
2) EuropeanaLocal: Regional aggregator
3) Europeana: Content Ingestion Testbed (Sandbox)
4) Europeana: Acceptance/Production administrator
Start
Create OAI-PMH repository
Export from source db/data structure
Data set
[Source schema]
Data set
[Mapped to ESE]
[Normalised to ESE]
Map + normalise to ESE
Validate with ESE Schema
NO
Access dashboard
Import
Cache thumbnails
Index
Mapping + normalisation ok?
YES
Populate OAI-PMH repository
Request Go-Live
Harvest
Data set
[Europeana processed]
Internal processing
Import
Cache thumbnails
Index in Acceptance
Internal processing ok?
NO
YES
Replicate in Production
End

C. Ingestion workflow of a data set: EuropeanaLocal to Europeana



Europeana Semantic Elements

- Metadata schema required by the Europeana
- Current version is 3.2.2, 18/01/2010
  - http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602
- Metadata Mapping & Normalisation Guidelines for the Europeana Prototype
  - Version 1.2.1, 18/01/2010
  - http://version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&groupId=10602

Europeana Semantic Elements (ESE)

- ESE ver. 3.2.2 consists of:
  A. 15 Dublin Core elements
     + 22 Dublin Core qualifiers / terms
  B. 11 Europeana-specific elements
- Majority of elements from group A should be harvested from aggregated digital library
- Some of these elements may be extracted/mapped from other elements
  - It depends on the metadata standards used in particular digital library
- Majority (all?) of elements from group B may be extracted from A group elements or is obvious

ESE - Dublin Core

- Title
  - Alternative
- Creator
- Subject
- Description
  - Table of Contents
- Publisher
- Contributor
- Date
  - Created
  - Issued
- Type
- Format
  - Extent
  - Medium
- Identifier
- Source
- Language
- Relation
  - isVersionOf; hasVersion;
  - isReplacedBy; replaces;
  - isRequiredBy; requires;
  - isPartOf; hasPart;
  - isReferencedBy; references;
  - isFormatOf; hasFormat;
  - conformsTo
  - *isShownBy; isShownAt (Europeana)*
- Coverage
  - Spatial
  - Temporal
- Rights
- Provenance (DC Terms)

ESE - Europeana-specific elements

- Please note that the DC and Europeana namespaces both have **Type** and **Language** elements
- When making mapping decision, providers are also asked to consider how their data will perform in response to **„who, what, where and when"** questions

ESE - Europeana-specific elements

- Elements whose values will be provided by Europeana
  - **User tag**
    - tag created by a user through the Europeana interface
  - **Language**
    - language assigned to the resource with reference to the Provider
  - **Year**
    - This is a 4 digit year in the Gregorian calendar (e.g. 1523), which is derived by Europeana from date values in the source metadata.
  - **Country**
    - Country name
  - **URI**
    - This is a record identifier for the object in the Europeana system.
  - **hasObject**
    - Indicates the availability of thumbnails of digital objects for the Europeana system to understand and process them.

ESE - Europeana-specific elements

- Elements gathered from providers
  - **Unstored**
    - Everything that was not mapped to other fields
  - **Object**
    - Link to miniature/sample of an object
  - **Provider**
    - Provider of this object (aggregator)
    - Name of institution should be placed in **dc:source**
  - **Type**
    - Object type, one of: Text, Image, Video, Sound

Mapping to ESE

- Required elements:
  - **europeana**: provider, type, isShownAt **or** isShownBy
- Strongly recommended elements:
  - **dc**: title, creator, contributor, date
  - **dcterms**: alternative, created, issued

Mapping to ESE

- Recommended elements:
  - **dc**: coverage, description, language, publisher, source, subject, type
  - **dcterms**: spatial, temporal, isPartOf
- Additional elements:
  - **dc**: format, identifier, rights, relation
  - **dcterms**: extent, medium, provenance, conformsTo, hasFormat, isFormatOf, hasVersion, isVersionOf, hasPart, isReferencedBy, references, isReplacedBy, replaces, isRequiredBy, requires, tableOfContents

IsShownAt, IsShownBy

- See examples from **Annex A** in "*Metadata Mapping and Normalisation Guidelines for Europeana Prototype*"

Mapping to ESE – general rules

- Map as many as possible of the original source elements to the available ESE elements
- If it is not possible to map the source element to an appropriate ESE element then leave it unmapped or consider using **europeana:unstored**

Mapping to ESE – general rules

- If possible use one of the more specific dcterms refinements
  - Remember that the semantic of the source term have to clearly correspond to the narrower term
- The persistent link to digital object and/or full information page should be given as a URL
  - These may need to be constructed from metadata values and information external to the metadata.

Mapping to ESE – general rules

- If it is difficult to decide which ESE element to map a source term to, consider how best to meet expectations of the user and the functionality of the system
- Where there are multiple values for the same element repeat the element for each instance of the value

Mapping to ESE – general rules

- To ensure that your data will be meaningful when displayed in the new context consider adding a prefix or suffix.
  - e.g. "100 x 200" could become "100cm x 200cm"
- Currently, the Europeana portal cannot use BC, BCE or BP dates
  - Such dates should be retained in the mapped metadata (e.g. dc:date) in order to be present for future development of the portal.

Mapping to ESE – Date

- Date should be machine readable
- Textual time periods will display in a result list but cannot be represented in the Timeline or Date facet and should also be provided as numeric dates
  - <localtimeperiod>**17th century**</localtimeperiod>
  - Transform and map also as <dc:date>**1601**</dc:date> and <dc:date>**1700**</dc:date>

Mapping to ESE – Language

- This element should be used to state the language of the digital object and should be repeated if the object has more than one language
  - If there is no language aspect to the object (for instance, a photograph) then the element should be ignored
  - The use of RFC 4646 is highly recommended
  - Best practice is to use ISO-639-1 or ISO-639-2

Mapping to ESE – Source

- Europeana recommends that the name of the content holder should be recorded using dc:source
- Thanks to this Europeana will show this information in the brief record display
- If multiple instances are to be provided containing different values it is suggested that they should be provided in a consistent order
  - Always put the name of the content holder first
  - <dc:source>The British Library</dc:source>
  - <dc:source>ISBN 1-86197-612-7</dc:source>

Polish Digital Libraries in Europeana

- Resorces from Polish digital libraries are available in Europeana since **11th December 2009**
- More than 340 000 objects at the moment
- How it was done?

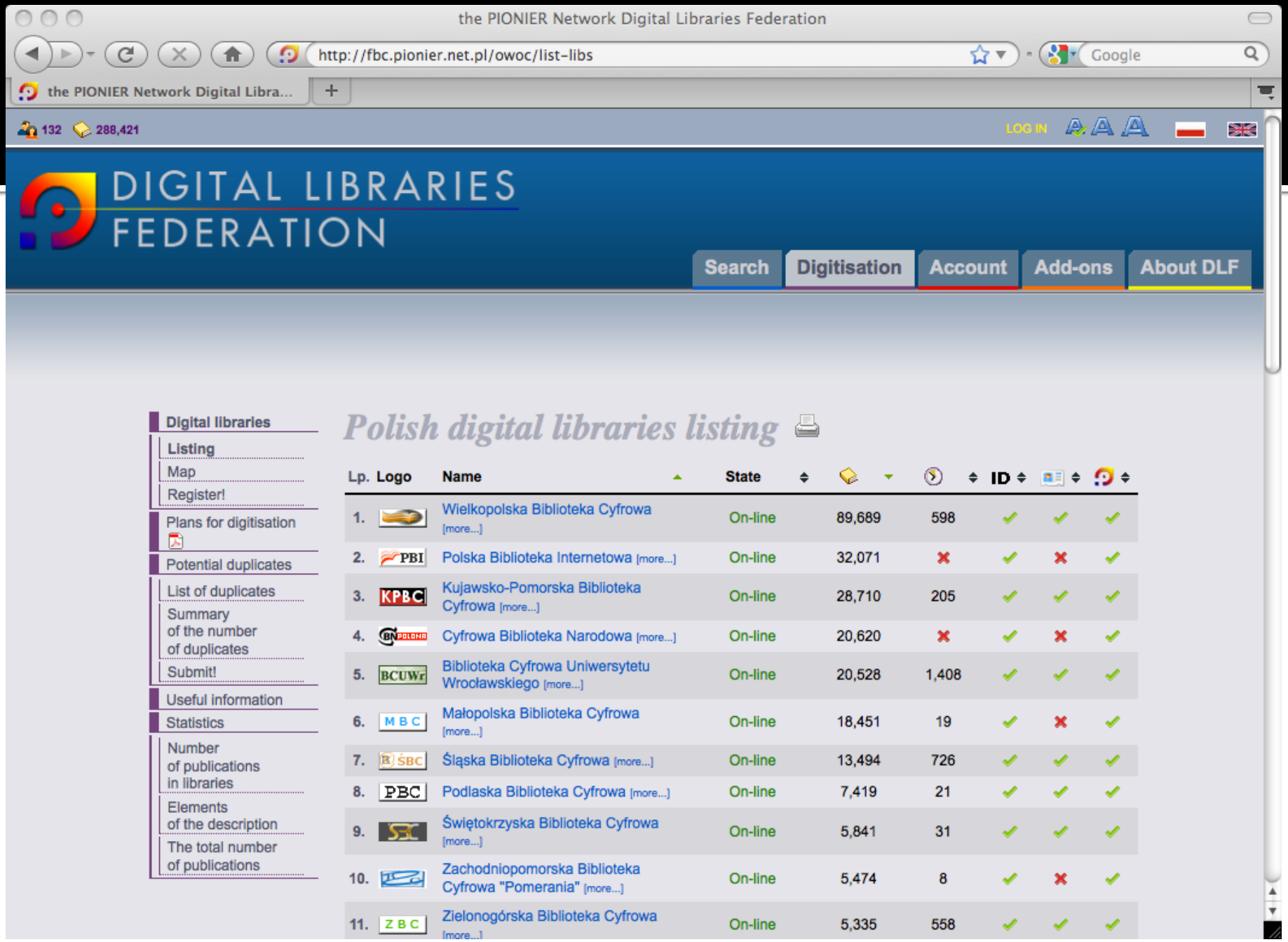
Metadata aggregators

- According to the present version of Europeana Outline Functional Specification tasks for the aggregator are:
  1. To gather the information about content providers and their information systems
  2. To gather the metadata of objects that should be visible in Europeana
  3. To remove duplicates, clean-up the metadata, normalize it and enrich
  4. To confirm the accessibility of digital objects
  5. To expose the aggregated metadata for Europeana via the OAI-PMH protocol

http://dev.europeana.eu/public_documents/EDLnet%20D2.5_Outline_Functional_Specifications20090301_version%201.7_consWithoutHistory_lossless.pdf

Digital Libraries Federation as a metadata aggregator for Europeana

- To collect information about providers and their delivery systems
  - Name and logo of a digital library, its website URL and the address of the OAI-PMH interface for digitized objects and objects planned for digitization
  - Detailed description with list of participating institutions
  - Sample objects
  - Basic statisitics


the PIONIER Network Digital Libraries Federation
http://fbc.pionier.net.pl/owoc/list-libs
Google
the PIONIER Network Digital Libra...
132
288,421
LOG IN
DIGITAL LIBRARIES FEDERATION
Search
Digitisation
Account
Add-ons
About DLF
Digital libraries
Listing
Map
Register!
Plans for digitisation
Potential duplicates
List of duplicates
Summary of the number of duplicates
Submit!
Useful information
Statistics
Number of publications in libraries
Elements of the description
The total number of publications
Polish digital libraries listing
Lp. Logo Name State ID
1. Wielkopolska Biblioteka Cyfrowa [more...] On-line 89,689 598
2. Polska Biblioteka Internetowa [more...] On-line 32,071
3. Kujawsko-Pomorska Biblioteka Cyfrowa [more...] On-line 28,710 205
4. Cyfrowa Biblioteka Narodowa [more...] On-line 20,620
5. Biblioteka Cyfrowa Uniwersytetu Wrocławskiego [more...] On-line 20,528 1,408
6. Małopolska Biblioteka Cyfrowa [more...] On-line 18,451 19
7. Śląska Biblioteka Cyfrowa [more...] On-line 13,494 726
8. Podlaska Biblioteka Cyfrowa [more...] On-line 7,419 21
9. Świętokrzyska Biblioteka Cyfrowa [more...] On-line 5,841 31
10. Zachodniopomorska Biblioteka Cyfrowa "Pomerania" [more...] On-line 5,474 8
11. Zielonogórska Biblioteka Cyfrowa [more...] On-line 5,335 558

Digital libraries
Listing
Map
Register!
Plans for digitisation
Potential duplicates
List of duplicates
Summary of the number of duplicates
Submit!
Useful information
Statistics
Number of publications in libraries
Elements of the description
The total number of publications

Digital libraries in Poland

Show digital library on the map:

-- select --

Additional map overlays: the PIONIER network route | the institutions creating digital libraries

Digital libraries
Listing
Map
Register!
Plans for digitisation
Potential duplicates
List of duplicates
Summary of the number of duplicates
Submit!
Useful information
Statistics
Number of publications in libraries
Elements of the description
The total number of publications

Digital libraries description

Choose digital library:

KPBC Kujawsko-Pomorska Biblioteka Cyfrowa (KPBC)

General information | Institutions | Recommended | OAI-PMH

State: *On-line*

Type: *Regional*

Web page: *http://kpbc.umk.pl/*
Contact email: *kpbc@umk.pl*
Start date: *Sep 1, 2005*
Description:

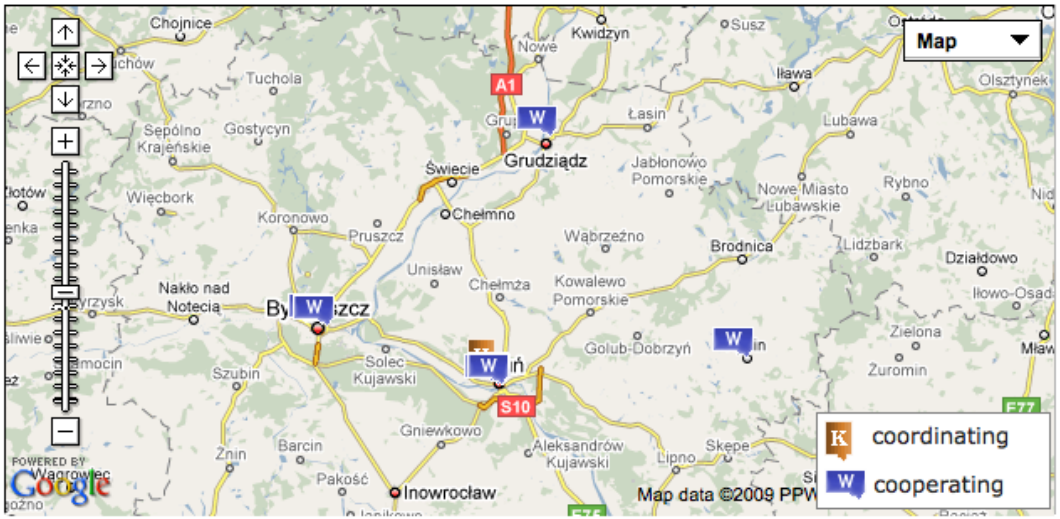
Kujawsko-Pomorska Biblioteka Cyfrowa (KPBC) jest projektem realizowanym przez instytucje współpracujące w ramach Konsorcjum Bibliotek Naukowych Regionu Kujawsko-Pomorskiego. I faza projektu (lata 2004-2006) była realizowana ze wsparciem funduszy europejskich, aktualnie biblioteki regionu rozwijają zasoby samodzielnie. Celem tworzenia regionalnej biblioteki cyfrowej jest wspieranie edukacji i nauki, turystyki oraz potencjału intelektualnego i innowacyjnego społeczeństwa. Bibliotekarze realizują ten zamiar przez umożliwienie wszystkim zainteresowanym dotarcia do zasobów wiedzy oraz cennych zabytków kultury piśmienniczej. Kujawsko Pomorska Biblioteka Cyfrowa ma przede wszystkim służyć naukowcom, studentom, uczniom oraz wszystkim zainteresowanym regionem kujawsko-pomorskiegim.

Map
Register!
Plans for digitisation
Potential duplicates
List of duplicates
Summary of the number of duplicates
Submit!
Useful information
Statistics
Number of publications in libraries
Elements of the description
The total number of publications

Choose digital library:

KPBC Kujawsko-Pomorska Biblioteka Cyfrowa (KPBC)

General information | Institutions | Recommended | OAI-PMH



Biblioteka Główna Uniwersytetu Mikołaja Kopernika w Toruniu
(coordinating)

http://www.bu.umk.pl/

dr Mirosław A. Supruniuk
Dyrektor
send... +48-56 611-4408

Biblioteka Główna Uniwersytetu Kazimierza Wielkiego w Bydgoszczy
(cooperating)

http://biblioteka.ukw.edu.pl/

dr Aldona Chlewicka
Dyrektor
send... 052 34 19 356

lic. Ewa Wójcik
Referent
send... 052 34 19 356

Map
Register!
Plans for digitisation
Potential duplicates
List of duplicates
Summary of the number of duplicates
Submit!
Useful information
Statistics
Number of publications in libraries
Elements of the description
The total number of publications

Choose digital library:

KPBC Kujawsko-Pomorska Biblioteka Cyfrowa (KPBC)

General information | Institutions | Recommended | OAI-PMH

Kujawsko-Pomorska Biblioteka Cyfrowa recommend:

Apokalypse - Heinrich von Hesler

Apokalipsa św. Jana - biblia. Kodeks 30 x 21,5 cm. W rękopisie znajduje się 35 przepięknych, złoconych miniatur. Było ich więcej, jednak na przestrzeni dziejów wyciętych lub wyrwanych zostało 12 kart z miniaturami. Fundatorem rękopisu był Luther von Braunschweig, wielki mistrz zakonu krzyżackiego

Ryciny Erika Dahlberga z dzieła Samuela Pufendorfa

Ryciny Erika Dahlberga z dzieła Samuela Pufendorfa "De rebus a Carolo Gustavo Sueciae Rege ...", pochodzącego ze zbioru Biblioteki Uniwersytetu Kazimierza Wielkiego w Bydgoszczy.

Stemmata genealogica praecipuarum in Prussia Familiarum Nobilium - Hennenberger, Johann

Rękopis z końca XVI w.

Digital Libraries Federation as a metadata aggregator for Europeana

- To gather the metadata of objects that should be visible in Europeana
  - Done with the OAI-PMH
    - In most cases we require the OAI-PMH interface
    - In really special cases we can do it in different way (e.g. Polish Internet Library)
  - Now we harvest only Dublin Core Simple
    - Works on new national metadata schema started in September 2009

Digital Libraries Federation as a metadata aggregator for Europeana

- **To remove duplicates**, clean-up the metadata, normalize it and enrich
  - Two types of duplication:
    - Duplicated metadata records describing the same digital object
    - Digital objects being a representation of the same physical object
      - Makes sense mostly in the context of libraries, where there may be several, practically identical editions of the same book
        - In museums and archives each object is unique
  - De-duplication in the DLF is based on the metadata comparison with some similarity threshold
    - Around 0.2% of aggregated objects makes the list of the „potential duplicates"
    - Similar mechanisms are used for the prevention of duplicated digitization

Digital Libraries Federation as a metadata aggregator for Europeana

- To remove duplicates, **clean-up the metadata, normalize it** and enrich
  - On the DLF level there are automatically built dictionaries on the basis of aggregated metadata
    - Separately for each metadata element
    - Separately for each metadata language
  - Differences between the metadata from various digital libraries have negative impact for the searching possibilities of the end-users
  - That is why the metadata normalization is so important
  - The basic analysis shows which elements are crucial and which should be easy to clean-up
    - The analysis was done in April 2009 on the metadata of 214 254 aggregated objects

Digital Libraries Federation as a metadata aggregator for Europeana

DC element	No. of unique values	Number of associations	Average no. of occurrences
format	39	209 789	5 379,2
language	195	210 529	1 079,6
type	822	211 816	257,7
rights	1 192	246 093	206,5
coverage	66	2 390	36,2
publisher	18 002	310 764	17,3
contributor	12 979	83 464	6,4
subject	78 440	438 871	5,6
relation	9 292	48 319	5,2
date	47 581	209 589	4,4
identifier	6 426	27 666	4,3
description	43 657	180 391	4,1
source	16 996	52 506	3,1
creator	21 908	67 503	3,1
title	210 745	227 039	1,1

Digital Libraries Federation as a metadata aggregator for Europeana

- Format
  - In 99% of descriptions: MIME type
    - e.g. text/html, image/x.djvu
- Language
  - In most cases: ISO 639-2 (pol, ger, lat, fre etc.)
  - Sometimes one value „pol, ger" instead of „pol", „ger"
- Rights
  - Name of the institution which holds the original object
- Type
  - …

Digital Libraries Federation as a metadata aggregator for Europeana

Values for „Type” (top 20)	Number of objects with the value	% of aggregated objects	% of aggr. obj. (after clean-up)
czasopismo	44 709	20,9%	33,8%
gazeta	32 921	15,4%	31,3%
gazety	23 119	10,8%	
Czasopismo	20 965	9,8%	
książka	12 503	5,8%	
Gazeta	11 098	5,2%	
pocztówka	5 768	2,7%	
czasopisma	4 962	2,3%	
text	4 452	2,1%	
grafika	3 863	1,8%	
fotografia	3 596	1,7%	
artykuł z czasopisma	3 164	1,5%	2,6%
artykuł	2 455	1,1%	
Czasopisma	1 710	0,8%	
dzienniki urzędowe	1 516	0,7%	
stary druk	1 222	0,6%	1,1%
starodruk	1 221	0,6%	
rysunek	1 094	0,5%	
rękopis	1 062	0,5%	
mapa	1 028	0,5%	
Sum		**85,1%**	**68,9%**

Digital Libraries Federation as a metadata aggregator for Europeana

- To remove duplicates, clean-up the metadata, normalize it and **enrich**
  - Basic enrichment can be the creation of the Europeana specific metadata elements from :
    - Other Dublin Core fields
    - Additional information
      - e.g. used DL software – standard link structure

Dates patterns analysis

- Dates patterns analysis
  - Basic measurement: length of DC:date value

Length	No. Of occurences	%
4	92 606	44,03%
10	82 182	39,07%
9	12 833	6,10%
6	5 133	2,44%
11	4 772	2,27%
5	2 420	1,15%
13	2 038	0,97%
7	1 975	0,94%
8	1 484	0,71%
16	866	0,41%

  - Top ten values covers 98,09% of all objects

Dates patterns analysis

- Dates patterns analysis
  - Looking for a pattern – step 1

Pattern	No. Of occurences	%
DDDD	92 402	43,93%
DDDD!DD!DD	81 162	38,59%
DDDD!DDDD	9 029	4,29%
!DDDD!	4 350	2,07%
!ca DDDD!	3 219	1,53%
!DDDD!DDDD!	2 208	1,05%
DDDD!	1 783	0,85%
DDDD!DD	1 354	0,64%
!ante DDDD!	924	0,44%
DDDD!D!DDDD	836	0,40%

  - Top ten patterns cover 93,79% of all objects

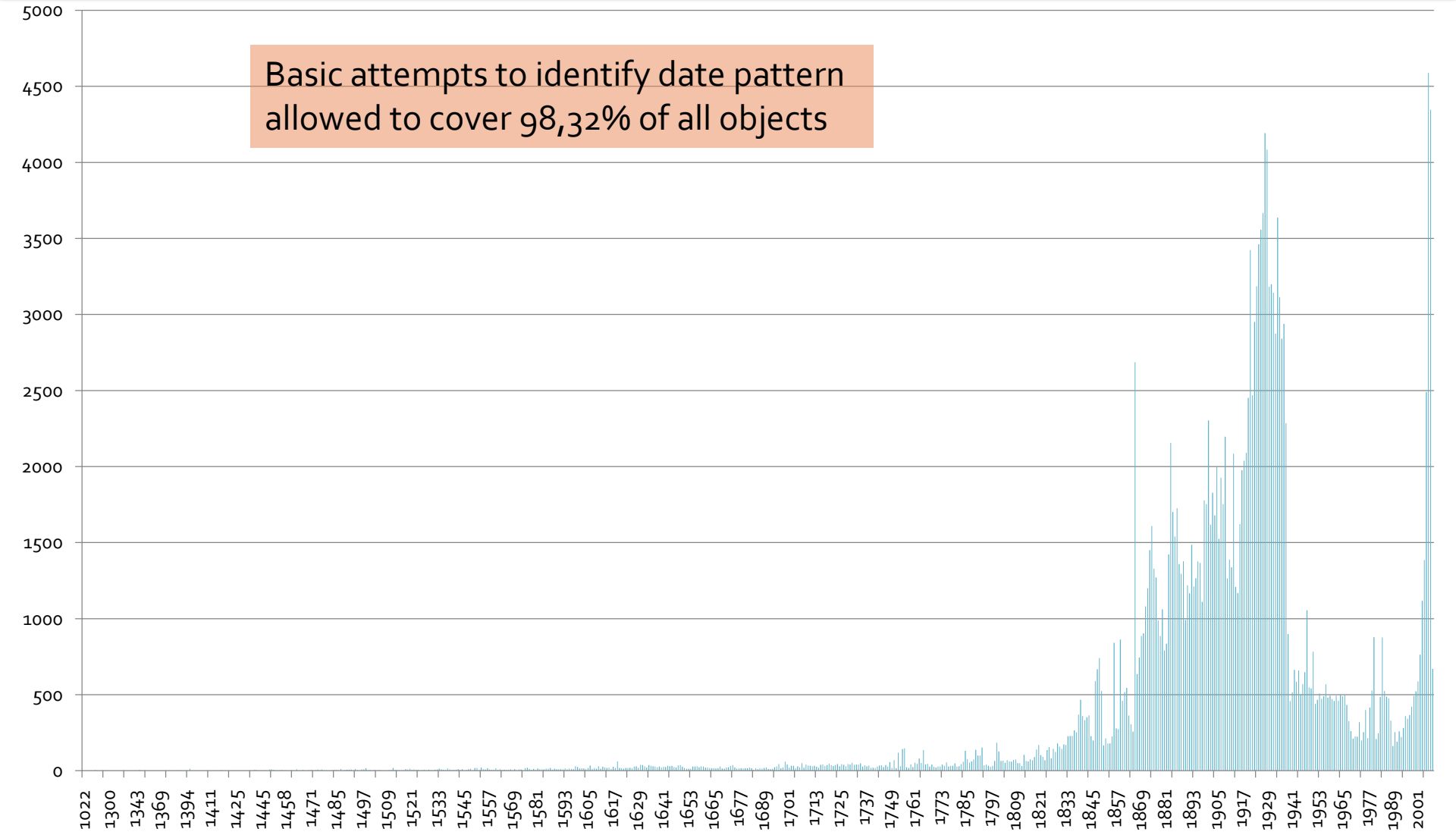
Dates patterns analysis

- Dates patterns analysis
  - Looking for a pattern – step 1

Pattern	No. of Occurences	%
DDDD	92 402	43,93%
DDDD.DD.DD	62 710	29,82%
DDDD-DD-DD	18 287	8,69%
DDDD-DDDD	8 935	4,25%
[DDDD]	4 327	2,06%
[ca DDDD]	3 208	1,53%
[DDDD-DDDD]	2 202	1,05%
[ante DDDD]	924	0,44%
DDDD.	906	0,43%
DDDD.DD	840	0,40%

  - Top ten dc:date patterns covering 92,59% of all objects

Timeline


Basic attempts to identify date pattern allowed to cover 98,32% of all objects
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1022
1300
1343
1369
1394
1411
1425
1445
1458
1471
1485
1497
1509
1521
1533
1545
1557
1569
1581
1593
1605
1617
1629
1641
1653
1665
1677
1689
1701
1713
1725
1737
1749
1761
1773
1785
1797
1809
1821
1833
1845
1857
1869
1881
1893
1905
1917
1929
1941
1953
1965
1977
1989
2001

Q&A