Europeana Cloud and Metis

Cécile Devarenne

Operations Officer

Aggregation Workshop

Belgrade, 28th January 2015

Content

- Europeana Cloud: from a project to a product
- Europeana and the cloud technologies
- Europeana ingestion tools: current status and concerns
- Europeana ingestion tools: lessons learnt
- Metis and the will to share experience and technologies
- Building strong Metis requirements
- Metis: how to?
- Connection between Metis and Europeana Cloud
- Metis: first plans for releases

→ **Europeana Cloud: from project to product**

Europeana Cloud project

- Goals are:
    - Store/Aggregate data and content; part of the aggregated data is submitted to Europeana
    - Set up a cloud technical infrastructure for storage of content and metadata
    - Set up a framework for the stored content and metadata to deal with access rights
- Main focus is:
    - Store, share data and collaborate on technologies

Europeana Cloud project

- Goals are:
  - Store/Aggregate data and content; part of the aggregated data is submitted to Europeana
  - Set up a cloud technical infrastructure for storage of content and metadata
  - Set up a framework for the stored content and metadata to deal with access rights
- Main focus is:
  - Store, share data and collaborate on technologies

Why the Europeana Cloud project?

- The current aggregation model would gain from sharing expertise: we have common knowledge on what processing data means!
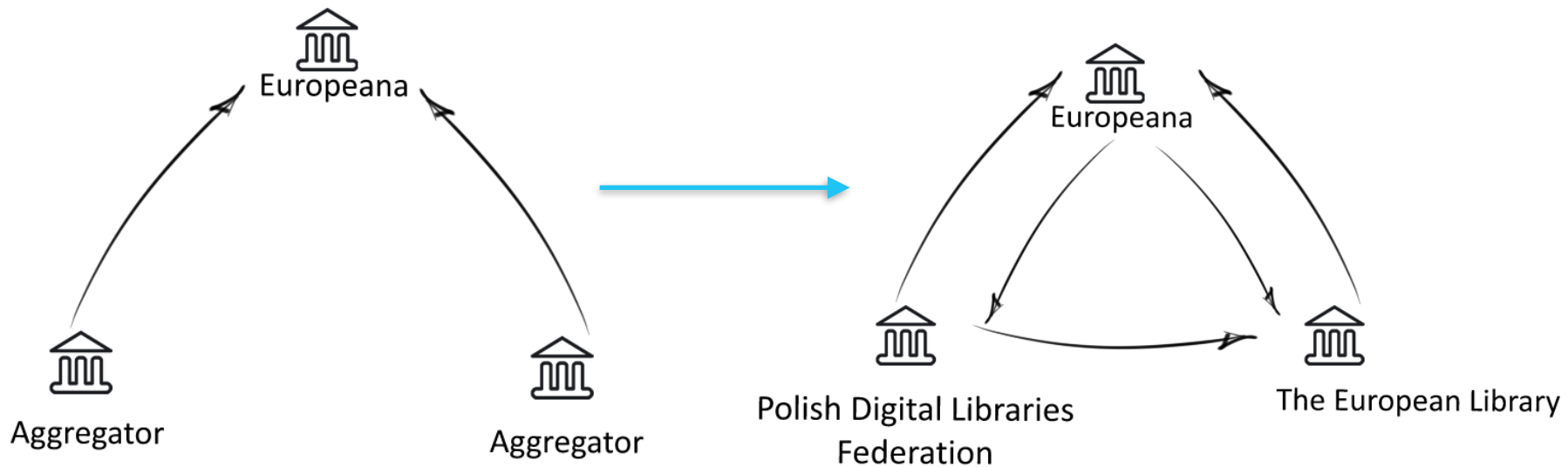
    Common interests

- The Cloud project aims at providing smaller institutions technologies to be able to share their data and content.

    A « dropbox » for metadata and cultural heritage content

    Data is exposed and can be shared using standard protocols (API as well as OAI-PMH etc)

Europeana Cloud: improving the aggregator model

Europeana Cloud: from project to product

- Sustainability and improvements:
  - Provide cloud generic services that will benefit to data/content providers and aggregators (on going thinking): sharing and giving access rights for re-use of the stored data
  - On top of the storage layer, support the development of products that enable aggregation and publication of data
- The first clients to the Europeana Cloud: Polish Digital Libraries Federation, Europeana and The European Library. All 3 are thinking about ways to re-develop and improve their aggregation set of tools

Europeana and the cloud infrastructures

- Europeana is already investing in the cloud technologies: migration of the full current stack to a new hosting provider using cloud technologies

- Europeana is willing to invest resources in making the ingestion tools work against Europeana Cloud's technical infrastructure: Metis

→ **Europeana current set of ingestion tools: issues and concerns, lessons learnt**

Europeana ingestion tools: UIM

- Back to the beginning ;-)

  Our current set of ingestion tools (see ingestion process presentation) is stable and enables Europeana to aggregate and publish your data successfully on the Europeana portal and API

- But it does not fulfill all our and your needs!

Europeana ingestion tools: UIM, only used by Europeana ingestion team

- UIM is complex to use
    - its original focus was internal only
    - the minimal user interface and the way the tool was developed makes evolutions difficult to implement
- Aggregation domain specific operations on data are disconnected from Europeana data processes: 2 or more distinct sets of data processing workflows
- Lack of tools to support aggregators and data providers for submission of quality data to Europeana

Europeana ingestion tools: UIM, only used by Europeana ingestion team

Result in:

- Data providers:

Europeana's workflow seems complex and lacks transparency, submission process is long and not smooth enough

- Operations officers:

Manual feedback and all operations on data are handled internally, leaving little room for increased focus on quality

Europeana ingestion tools: UIM, not fully satisfactory for our goals

- A lot of resources spent for maintenance

  Stand-alone tools are maintained by external teams, UIM and its plugins are maintained internally but expertise is needed

- Scalability issue

  Our current tools cannot hardly enable us to process collections with more than 2 million records, which is not satisfactory for our growing amount of aggregated data

- « Black box » also for the tech community

  Lack of modularity and granularity of the existing plugins leaves little room for collaboration on new data services with external researchers and development teams (e.g. generic service for data cleaning)

→ **Future plans: the ideas behind Europeana Metis**

Europeana Metis: identifying common areas and share experiences

- Operations performed on data are common between Europeana and European aggregators:
  - Mapping data, transforming data following a crosswalk
  - Cleaning data
  - Enriching data
  - Exposing/Sharing/Publishing data
- Some of the functionalities of our ingestion tools can be offered to providers as services
- Part of the development of these services can be shared: aligning internal development with projects and external developments

Europeana Metis: the vision

New set of ingestion tools for Europeana:

- Focus on quality!
    - implement new generic data flows to improve data quality on the Europeana dataset:
        - automated cleaning/normalization
        - improved enrichment
    - storage of the provenance information for each record
- Build a set of services/tools for data ingestion that rely on Europeana Cloud for storage
- Use standard technologies so that our services can be re-used by the tech community (and vice-versa :-)

Europeana Metis: the vision

Data submission tool for Europeana providers

- Open up progressively: first make the process of submission of data to Europeana smoother (upload of data, validation against EDM schema)

- On longer term, give external users the ability to perform operations on data prior to submission (mapping of data, enrichments)

- Expose via APIs the processed data which is stored on the cloud to support publication processes

Europeana Metis: the vision

Advantages of opening up:

- The process is more efficient and transparent to everyone contributing

- Aggregators/projects (and in the future data providers with technical expertise) are able to use Metis to prepare and check their data for submission to Europeana

- Part of the operations we do internally can be performed directly by providers (e.g. automated validation)

- What the providers could do directly in Metis does not need to be repeated by the Europeana operations officers (e.g. uploading data in several places vs unified storage)

Europeana Metis: the vision

- Longer term: Europeana believes that we can contribute to building a shared data management infrastructure

Europeana Metis, from vision to implementation

- Maintaining our current set of tools in order to pursue with aggregating data
- Building user requirements for Metis:
  - Internally: aggregation team
  - In consultation with potential future users; trying to reach aggregators and projects first:
    - Series of aggregators interviews organized by the Europeana Cloud team last year
    - Aggregator survey (results in February)
    - Workshops: gathering ideas
- Thinking about a product
  - Stable generic and modular set of tools
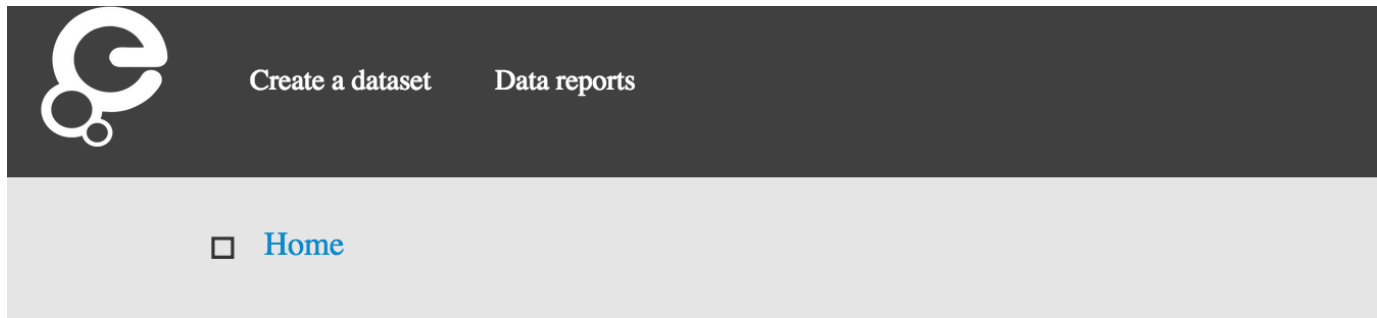 - Progression in opening up

Europeana Metis, first plans for releases

- Release 1 (2015):
  - Europeana operations officers are shifting from UIM to Metis
  - A set of providers can validate and submit data to Europeana
- Release 2:
  - Europeana operations officers are using Metis
  - More providers can validate and submit data to Europeana
  - A set of providers can also use Metis to transform, clean and submit their data to Europeana
- Release 3, additional features:
  - More Metis data services can be opened to external users
  - Data is exposed at different stages for re-use, publication of data is facilitated on various platforms

Europeana Metis, first attempts for wireframes

- Consultation for user design is going to take place (spring 2015)

Europeana Ingestion Tools

The Europeana Ingestion Tools are for content providers to upload and enrich their data to Europeana.

Europeana Metis, first attempts for wireframes

□ Home > Create a dataset

Create a dataset

Fill in the following fields to create your dataset.

Choose organisation

BNF

Name dataset

TIP: what makes a good name?

Europeana Metis, first attempts for wireframes

Data report

Advanced view options

Date		Organisation		Dataset name		Status		Actions
9-7-14		BNF		2020785_1989_17-03-2014.zip		☐ Saved		Edit Harvest
9-7-14		BNF		09308_20140203_154828.zip		☐ Harvested		Edit Map to Mint
4-7-14		BNF		09310_20140203_154244.zip		☐ Mapped		Edit Transform
4-7-14		BNF		09309_20140203_131955.zip		☐ Transforming		Cancel
4-7-14		Rijksmuseum		itemdata-20140306.xml		☐ Transforming		Cancel
2-7-14		BNF		storydata-20140326.xml		☐ Failed		Find out why Retry
29-6-14		BNF		storydata-20140326(1).zip		☐ Transformed		Edit Validate

Questions and input welcome!

Thank you!

Cécile Devarenne

cecile.devarenne@europeana.eu or content@europeana.eu