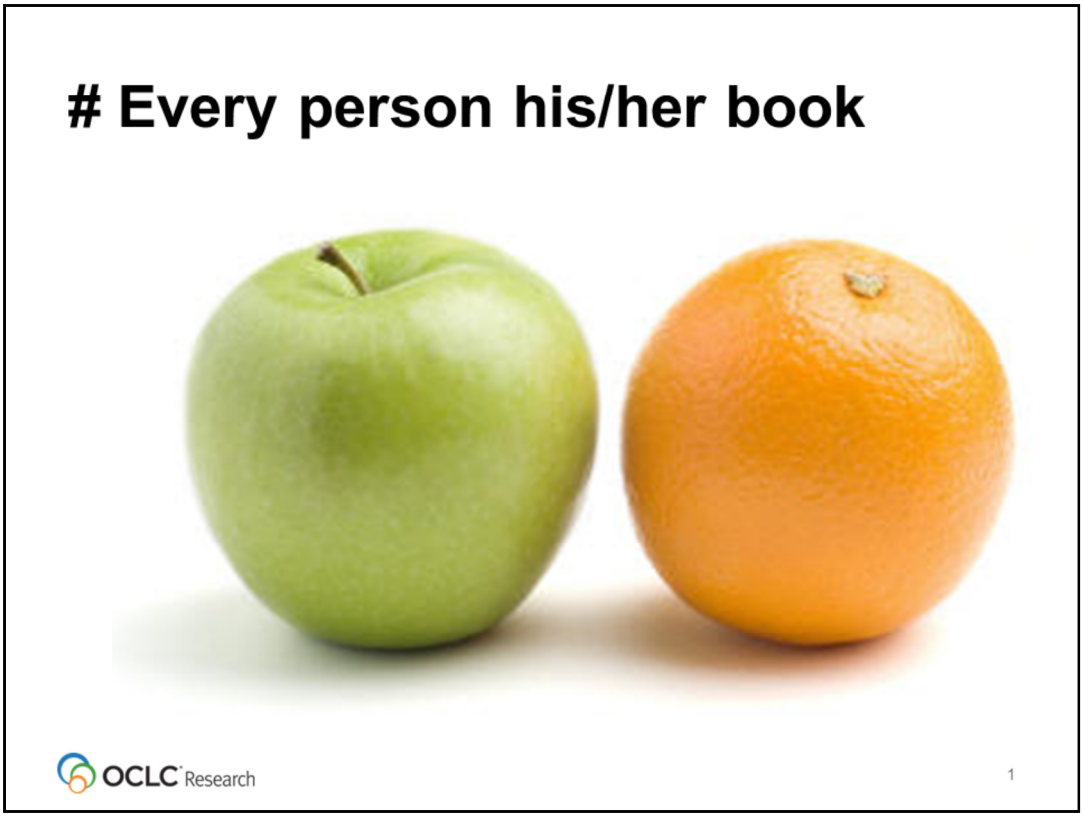
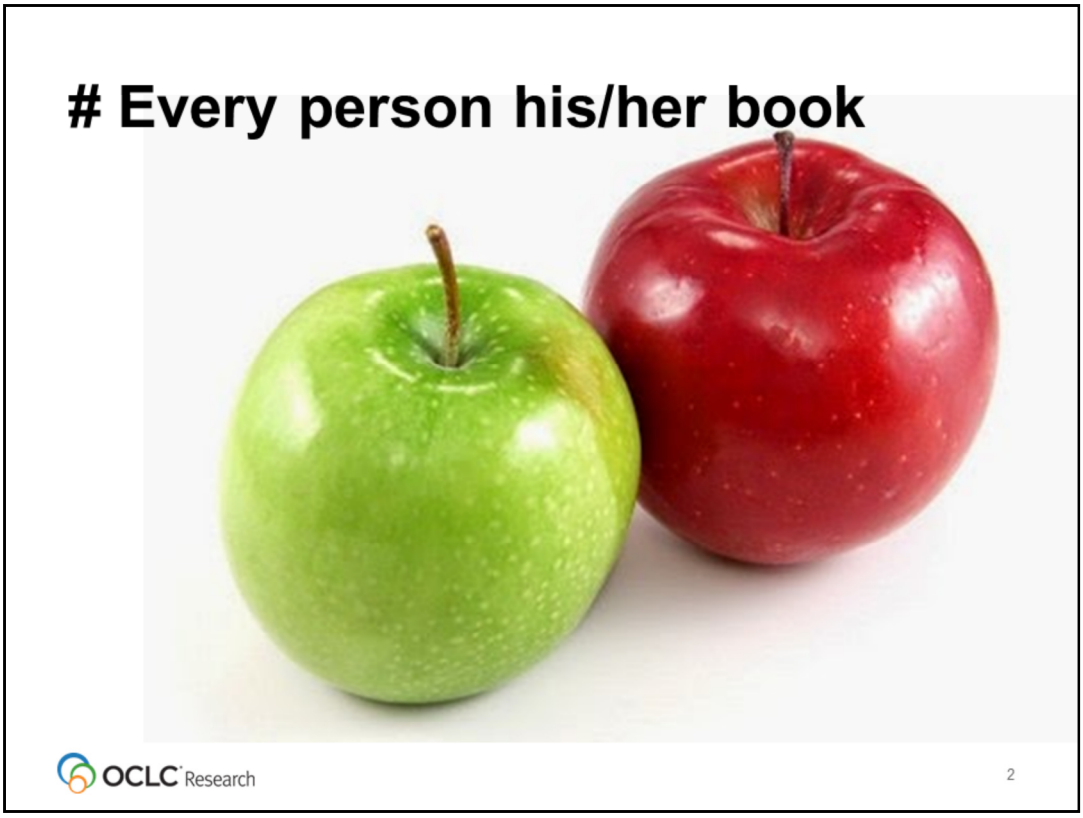
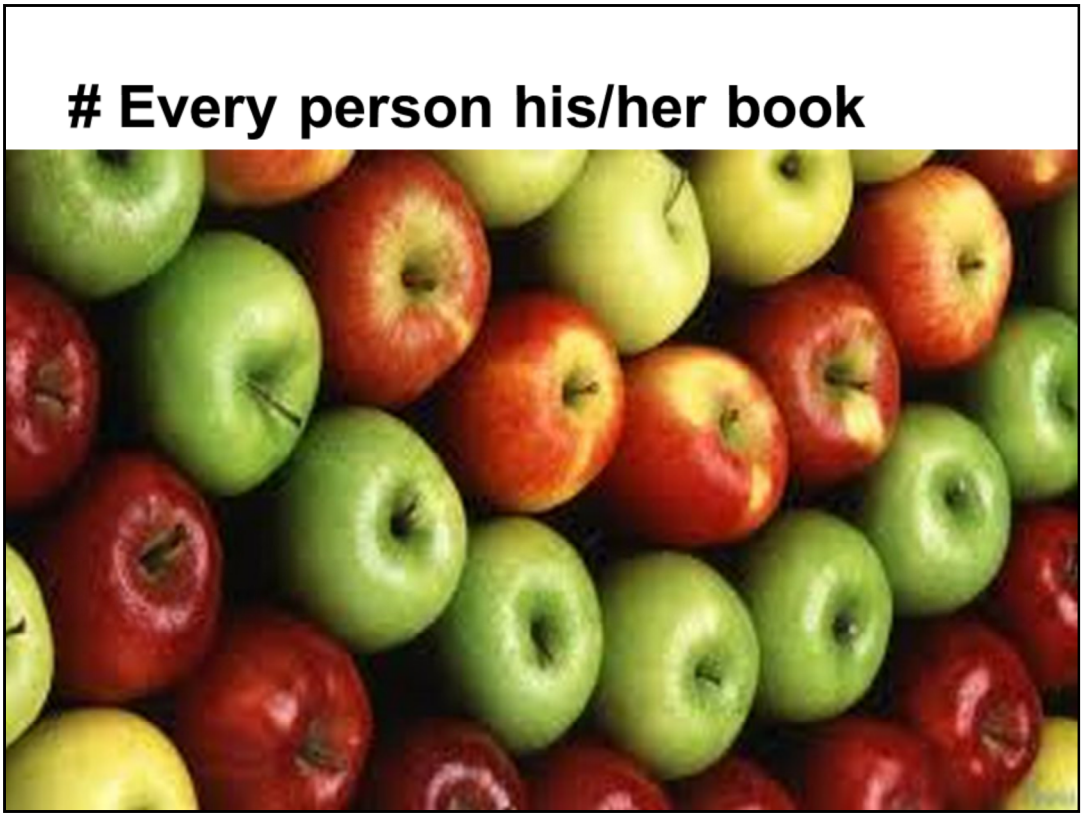


TITIA
OK, so this is where we continue from…

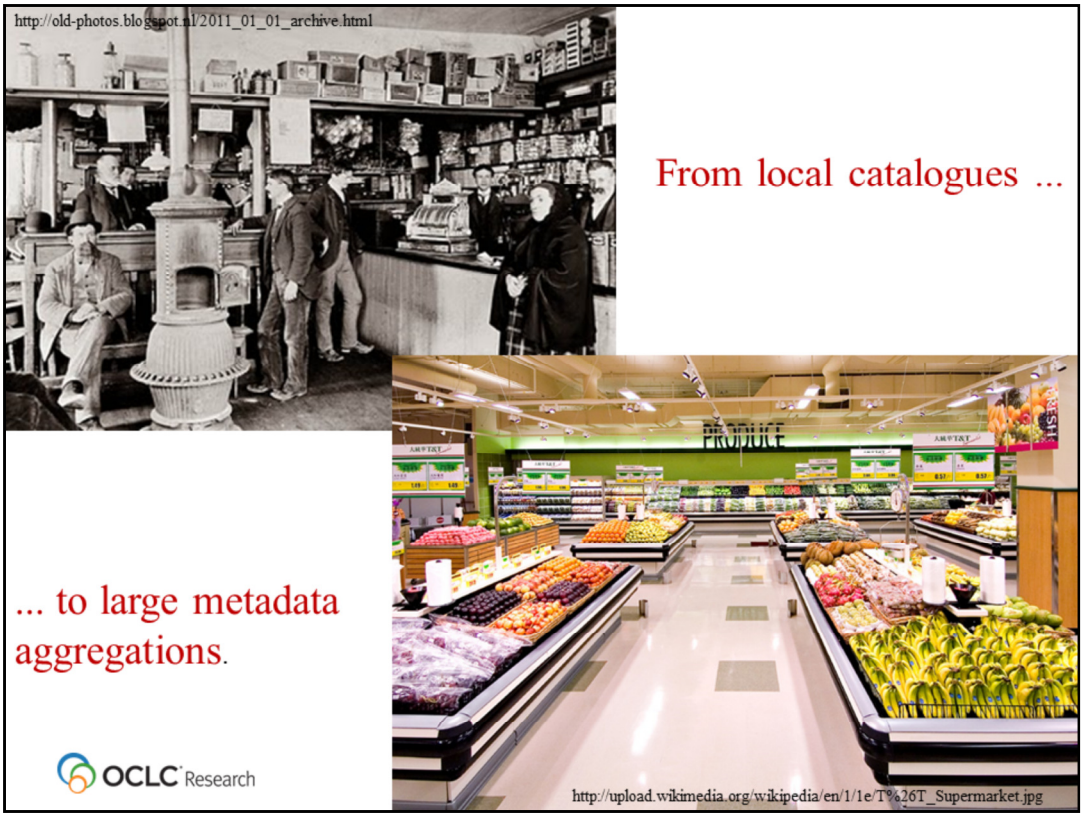


TITIA
In fact, we had an OCLC Research activity that was symbolized by a similar picture – apples that look similar but are not the same – in this case not symbolizing different readers with different needs, but metadata that looks similar but is not. As Lynn pointed out, this Ranganathan Law is about the ability to "pinpoint the information readers need".

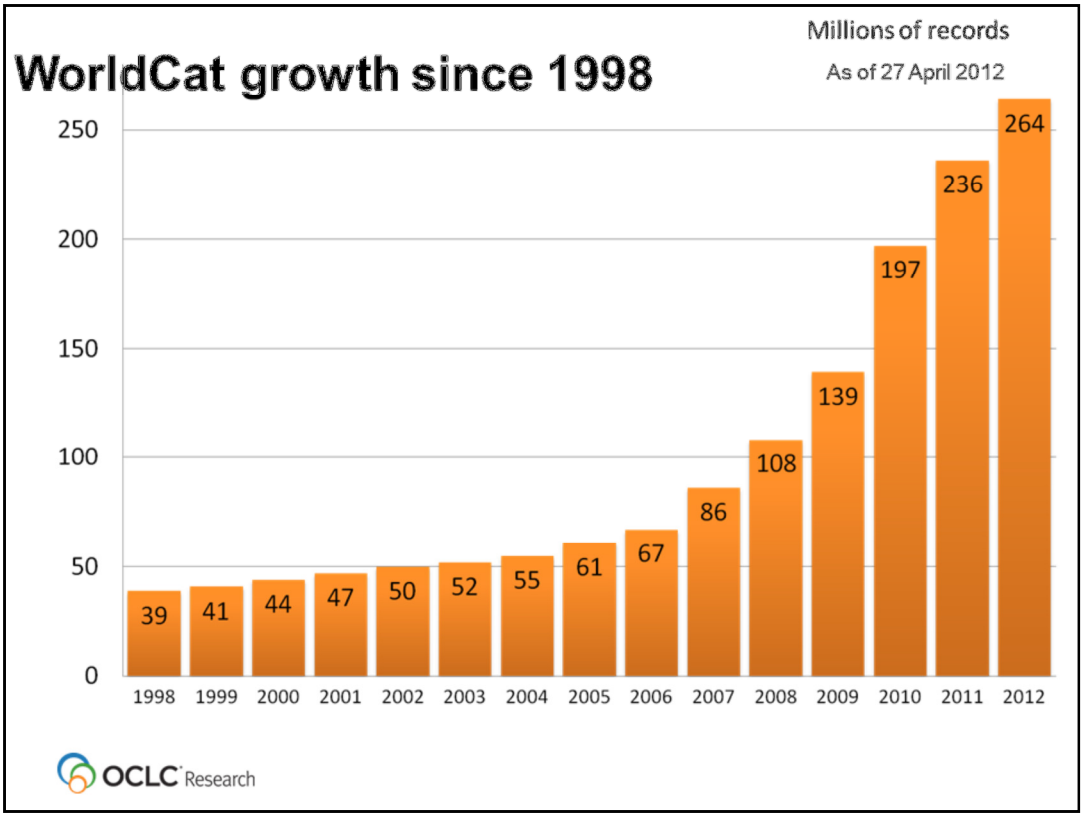


TITIA

In the digital environment this law is about the unlimited growth of content and the challenge for the reader to find what he/she is looking for. Metadata is the key to unlock this wealth of information, but the problem is that our library metadata aggregations are not helping the user find the information he/she wants.

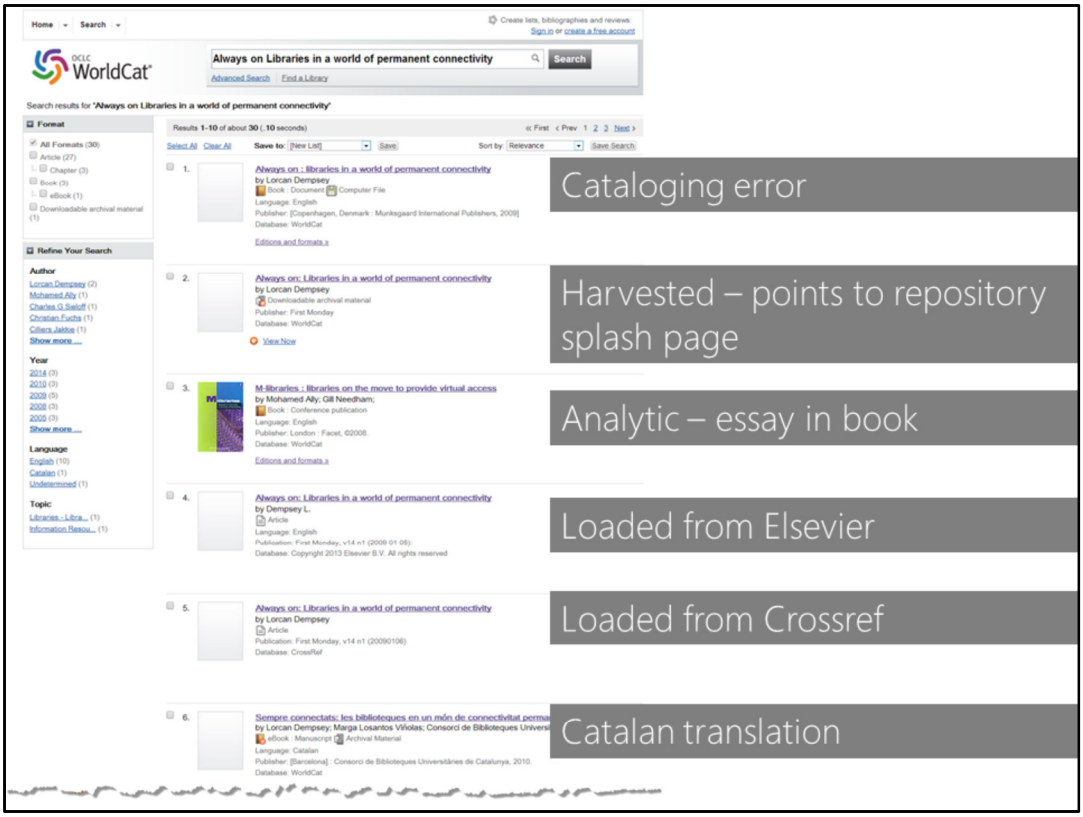


TITIA
The reason is that we have evolved from local catalogues to large metadata aggregations.



TITIA

The WC aggregation is a good example.


Always on Libraries in a world of permanent connectivity
Search
Cataloging error
Harvested – points to repository splash page
Analytic – essay in book
Loaded from Elsevier
Loaded from Crossref
Catalan translation

Aggregating data

In aggregations:

- data lose their local context
- data get lost in the bigger context

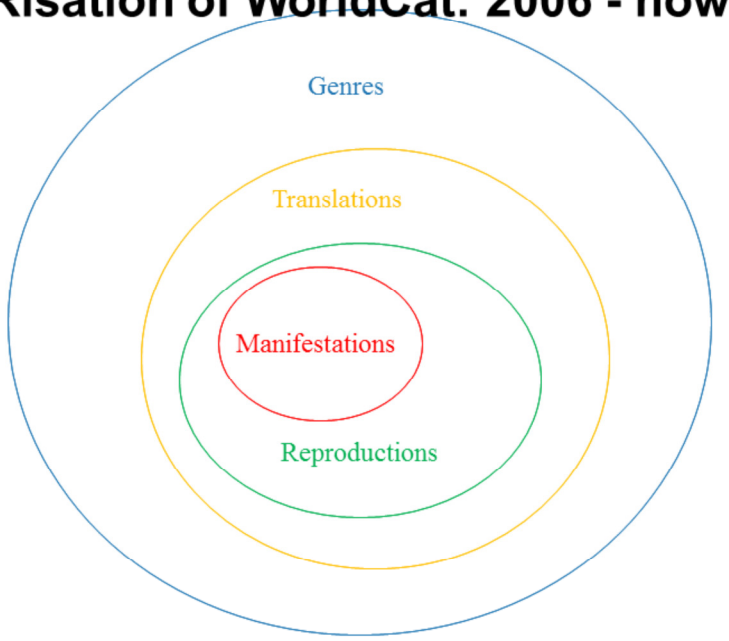
Making sense of data at the aggregate level:

- FRBR
- GLIMIR
- VIAF
- FAST
- Mining for entities/names

Applying techniques & algorithms to other datasets

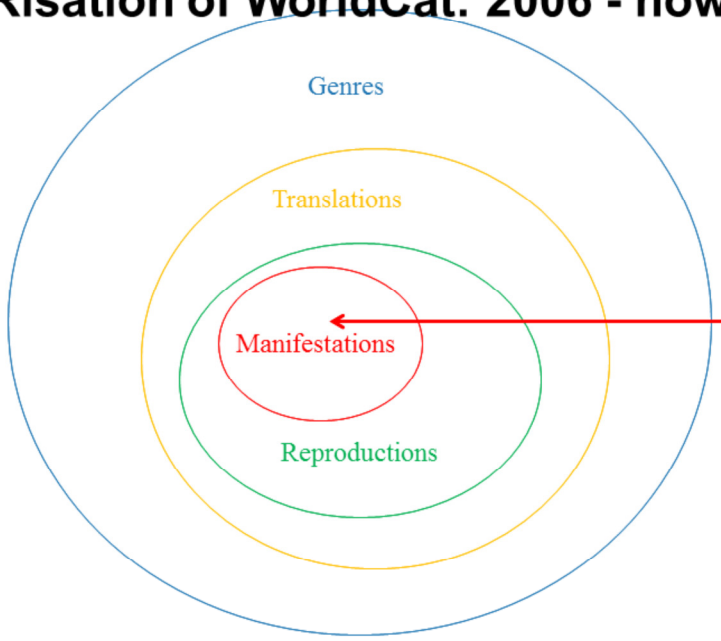
OCLC Research

FRBRisation of WorldCat: 2006 - now



OCLC Research

FRBRisation of WorldCat: 2006 - now



GLIMIR: Clustering records which differ in language and cataloguing rules

OCLC Research

VIAF

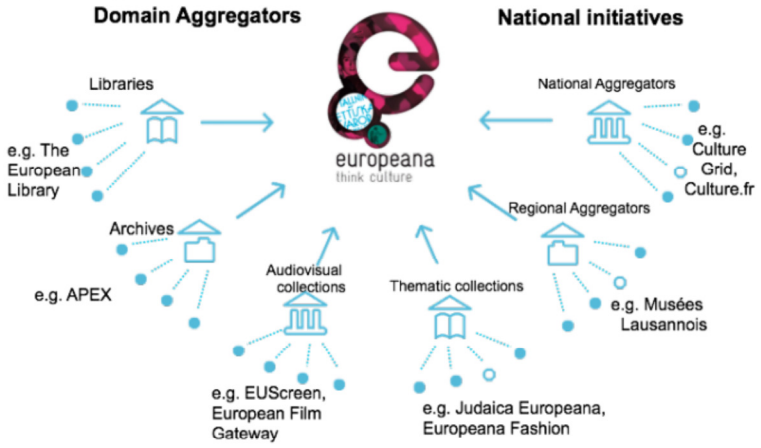
What is VIAF?
Virtual International Authority File

- Merge of 24+ national level authority files
- Cooperative program run by OCLC
- Initiated by LoC, DNB, BnF and OCLC
- 29 million authority records
- 112 million bibliographic records

- Migrated from an OCLC Research project to an OCLC service in 2012
- VIAF is available as linked data

OCLC Research

Applying techniques to other datasets



Hunting for Semantic Clusters: Hierarchical Structuring of Cultural Heritage Objects within Large Aggregations

17th International conference on Theory and Practice of Digital Libraries (TPDL), 22-26 September 2013, Valletta (Malta)

OCLC Research

Duplicates

© Rights reserved - Free access

View item at
Jewish Museum London

Pair of silver rimmonim

Description: A pair of late eighteenth century Italian silver rimmonim.; Repository/Location: Jewish Museum London
Geographic coverage: Italy [Place] [Production]
Type: rimmonim
Format: 406 mm; silver
Identifier: local JML 58095809 [Metadata]
Rights: Digital image: Jewish Museum London
Source: Jewish Museum London

© Rights reserved - Free access

View item at
Jewish Museum London

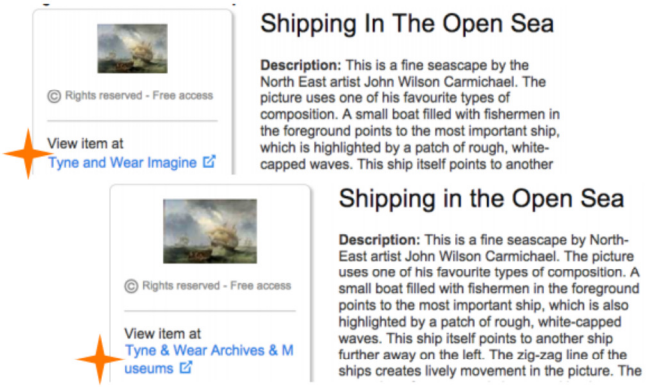
Pair of silver rimmonim

Description: A pair of late eighteenth century Italian silver rimmonim.; Repository/Location: Jewish Museum London
Geographic coverage: Italy [Place] [Production]
Type: rimmonim
Format: 406 mm; silver
Identifier: local JML 58095809 [Metadata]
Rights: Digital image: Jewish Museum London
Source: Jewish Museum London

Data provider: Jewish Museum London
Provider: Judaica Europeana
Providing country: United Kingdom

OCLC Research

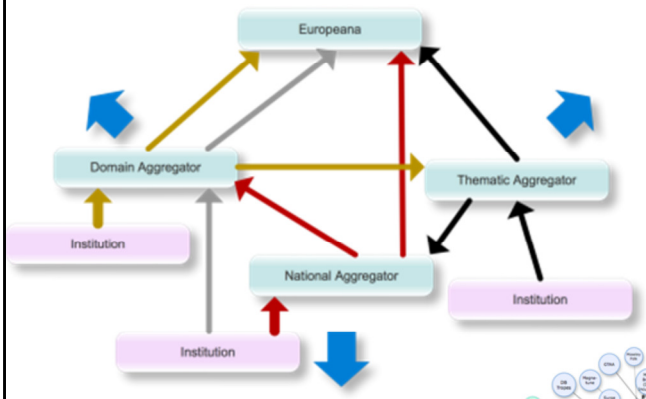
Duplicates? Same object: different providers



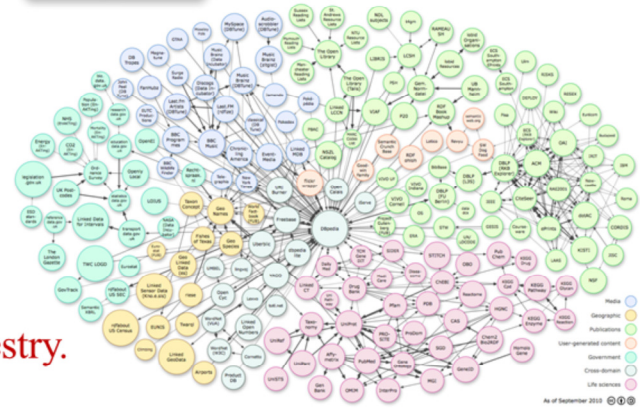
age digitised three times

OCLC Research

http://www.theeuropeanlibrary.org/confluence/display/wiki/A+Library+domain+aggregator



From an infrastructure of aggregators ...

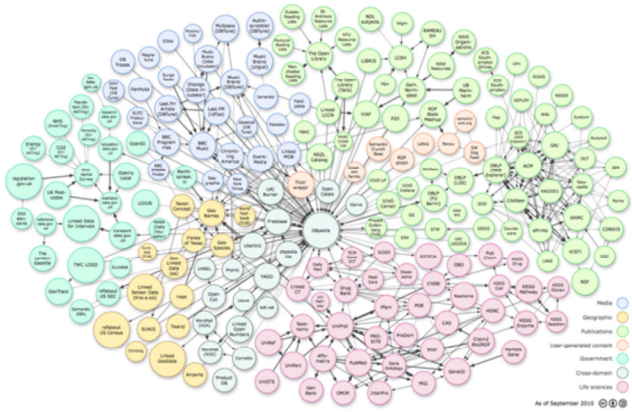


... to the web of data tapestry.

OCLC Research

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

Getting ready to move to the global aggregation: the semantic web



Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

OCLC Research

We are, in fact, still in the middle of developing the proper logistics for our supermarkets. We are still developing metadata harvesting workflows, metadata harmonisation and enrichment procedures, etc. We are setting up large digital stacks to be able to supply from our digitised and digital collections. We have clearly not yet realised the promise of smooth d2d logistics within our own controlled aggregations. And already, developments on the Web are raging past us, weaving d2d logistics into the tapestry of the Web.

Data quality: OCLC

- Who = author => VIAF
- What = work => clustering works (FRBRization and Glimirization of WorldCat)
- Where = location => Registry of Libraries; Find a Library Near You

OCLC Research

Continuous quality improvement

- Looking at new data modeling solutions: Europeana (EDM), LoC (BibFrame), etc.
- Libraries are structuring systems – continuously organising and reorganising the same data.

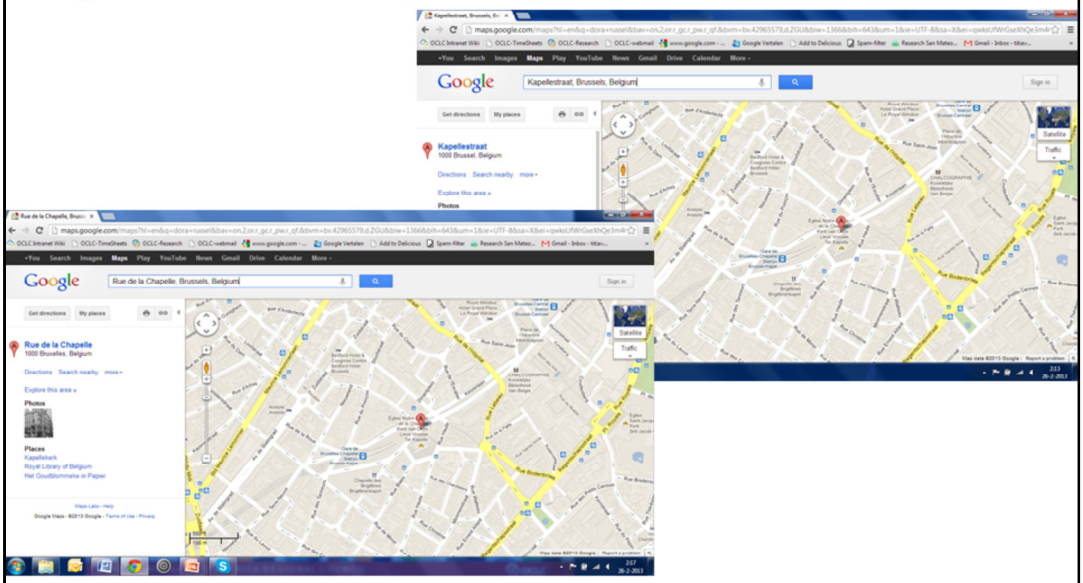
OCLC Research

Continuous quality improvement

- Wikipedia and Google are learning systems – continuously improving the data and relying on crowd sourcing for correcting and improving the quality of the data

OCLC Research

Quality of the system is only as good as the quality of the data



OCLC Research