europeana

Evaluation of metadata enrichment practices in digital libraries: steps towards better data enrichments

Valentine Charles and Juliane Stiller | **SWIB 2015**

Quadruplettes à Buffalo
Agence de presse Meurisse
1909, National Library of France
France, Public Domain

Co-financed by the European Union
Connecting Europe Facility

Semantic enrichment, a solution for better quality data?

Automatic and manual enrichment are more and more commonly used in digital libraries to:

- "standardize data" by linking it to authority resources

- improve multilingual coverage in datasets

- contextualise resource

- And more

Semantic enrichment, a solution for better quality data?

Automatic and manual enrichment are more and more commonly used in digital libraries to:

- "standardize data" by linking it to authority resources

- improve multilingual coverage in datasets

- contextualise resource

- And more

This growth is also accompanied by the adoption of semantic web technology in the cultural heritage!

What do we mean by semantic enrichment ?

(Semantic) enrichment can be described:

- As a process that improves metadata about an object by adding statements about the object that this metadata describes

- As a process e.g. the application of an enrichment tool

- As a result e.g. the new metadata created at the end of the process

Enrichment strategy refers to all workflows components and the processes which determined these components.

What do we mean by semantic enrichment ?

(Semantic) enrichment can be described:

- As a process that improves metadata about an object by adding statements about the object that this metadata describes

- As a process e.g. the application of an enrichment tool

- As a result e.g. the new metadata created at the end of the process

Enrichment strategy refers to all workflows components and the processes which determined these components.

Co-referencing, alignment, contextualisation, annotation are types of enrichment !

Europeana Semantic Enrichment - Example

CC BY

View item at
The Wellcome Library

Share

W Cite on Wikipedia

Translate details

Select language ⬍

Powered by **Microsoft®** Translator

A man building a wig on to the head of a woman on a kind of

Description:
A man building a wig on to the head of a woman on a kind of scaffolding; another woman wearing a tall wig looks on.

Date:
18-th, 18th, 18th century; Part of: Second millenium AD; From: 01-01-1701 — To: 31-12-1800

Type:
Coloured etching; Still image

Subject:
Maximilien Rapine; Costume; Fashion; Peruke; Wig; Wig macker

Identifier:
http://wellcomeimages.org/ixbin/hixclient.exe?MIROPAC=V0019839

Is part of:
http://data.theeuropeanlibrary.org/Collection/a1000

Rights:
Wellcome Library, London; Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0
http://creativecommons.org/licenses/by/4.0/

Source:
V0019839

Data provider:
The Wellcome Library

Provider:
The European Library

Providing country:
United Kingdom

What ▾

Concept Term:
http://dbpedia.org/resource/Costume

Concept Label:
[Costume] (en); [Костюм] (ru)

Concept Term:
http://dbpedia.org/resource/Fashion

Concept Label:
[Mode] (de); [時裝] (zh); [Moda] (it); [Moda] (pt); [Moda (styl)] (pl); [Mode] (sv); [Mode (habillement)] (fr); [Fashion] (en); [Мода] (ru); [Moda] (es); [Mode] (nl); [フアツシヨン] (ja)

An incorrect enrichment is worse than no enrichment at all !

Colombes : championnats de France d'Athlétisme : rivière, le speaker
Agence de presse Meurisse
1921, National Library of France
France, Public Domain

Our starting point

Incorrect enrichments lead to

- Devaluation of curated metadata

- Loss of trust from providers

- Propagation of errors to different languages

- Irrelevant search results

- Bad user experiences

Our starting point

Incorrect enrichments lead to

- Devaluation of curated metadata

- Loss of trust from providers

- Propagation of errors to different languages

- Irrelevant search results

- Bad user experiences

Better understanding of impact of enrichments is needed !
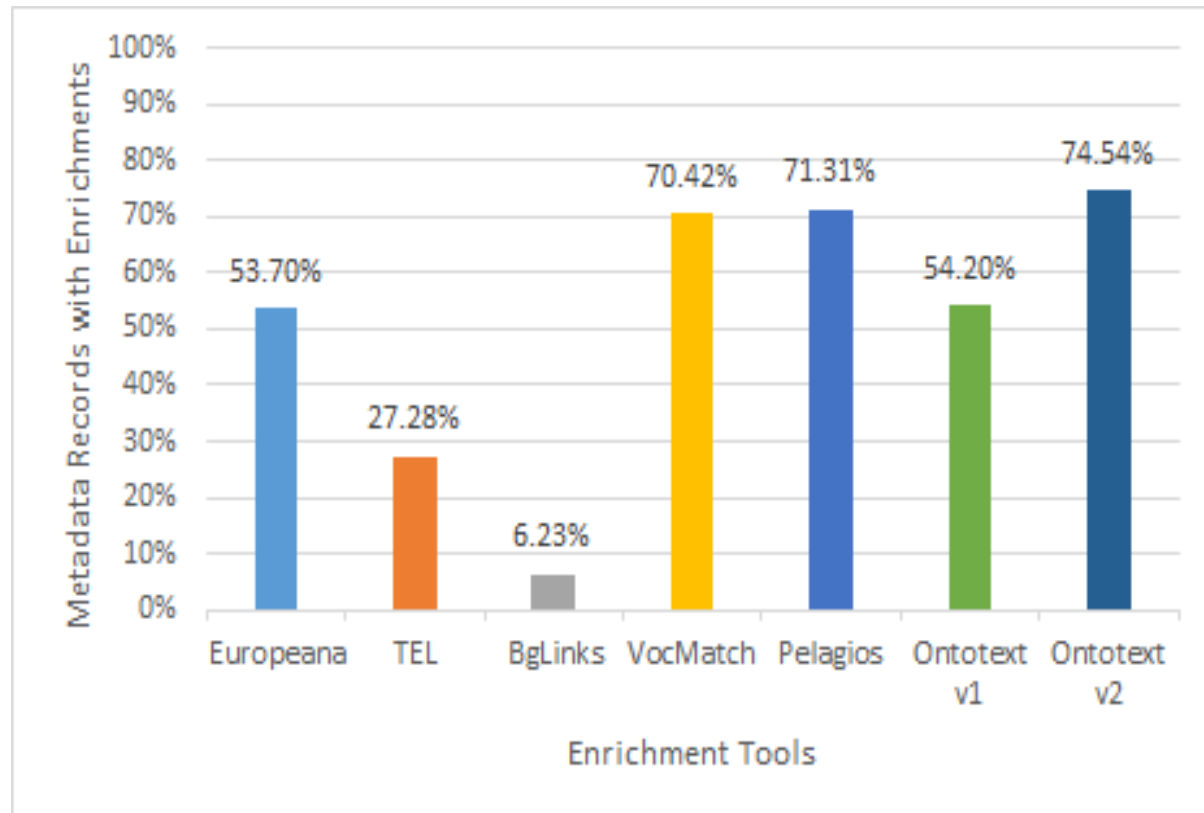
Comparative evaluation

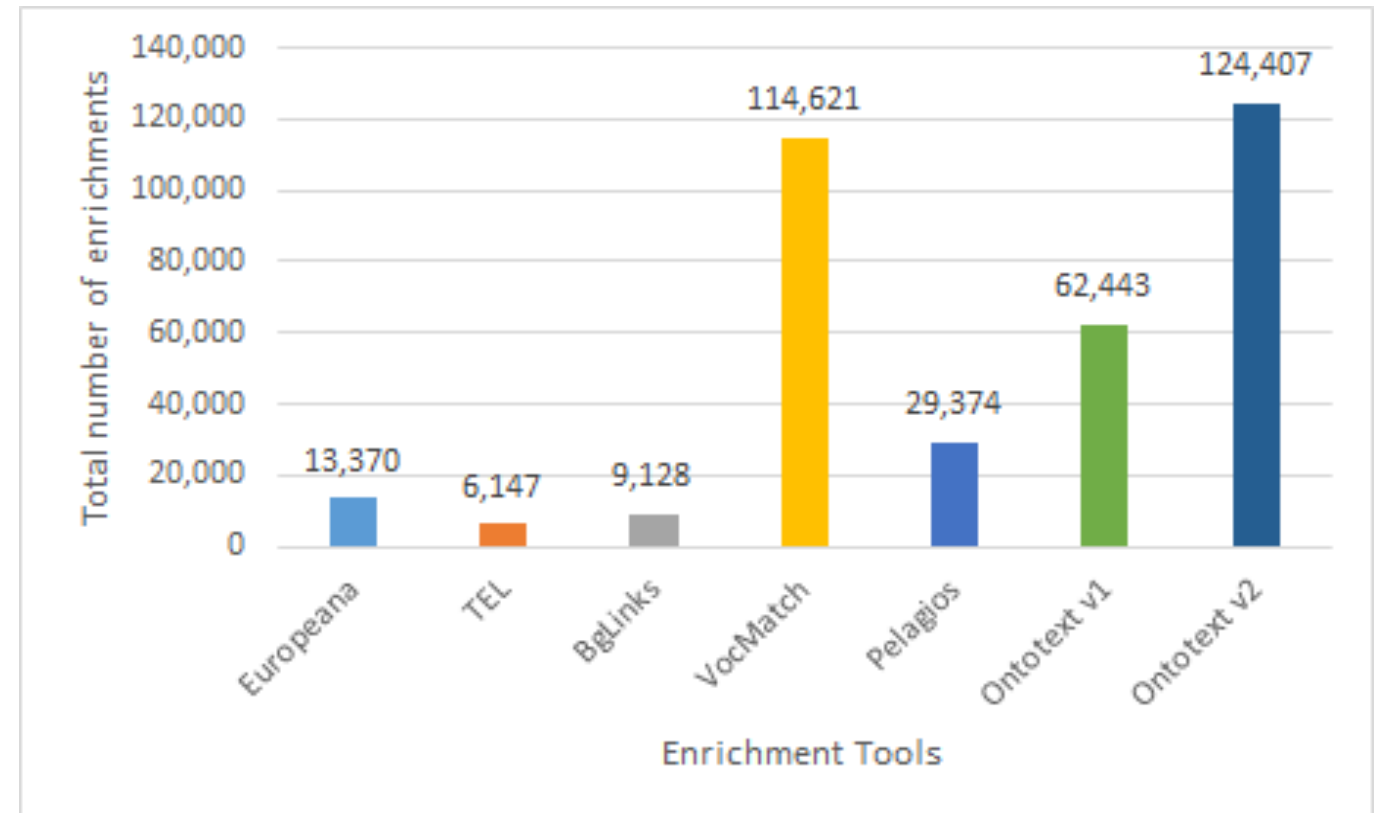A group of 23 experts in the Cultural Heritage domain run a quantitative and qualitative evaluation:

- An evaluation dataset contributed by The European Library

- Enriched by 7 different tools or tool settings

- 1757 enrichments were sampled and manually annotated

- Goal was to determine the quality of the enrichments

Comparative evaluation

Frequency of enriched records for each enrichment tool.

Number of enrichments by enrichment tool.

10 recommendations for better semantic enrichment services

1 Define your enrichment goals

What problems are you trying to solve?

What type of enrichment do you want to achieve?

On which criteria will you evaluate the success of the enrichment ?

What type of data do you want as input or output?

1   Define your enrichment goals

What problems are you trying to solve?

What type of enrichment do you want to achieve?

On which criteria will you evaluate the success of the enrichment ?

What type of data do you want as input or output?

Clear goals will help you to define your gold standard and guidelines for evaluating your enrichment

Choose the right method...

Choose enrichment methods and techniques that are adapted to the goals

- For instance a service providing an interface might be better if the enrichment is manual

http://cultuurlink.beeldengeluid.nl/app/#/

europeana

2 ...and the right service

Select a service that takes into account the semantics of your input data

- enrichment service or tool recognizes a specific metadata schema (schema/semantics-aware).

Select a service that provides output data in the desired representations

- Enrichment services should provide appropriate representations of enrichments following similar patterns and standard.

- They should also publish metadata for enrichment (e.g. provenance, confidence)

The selection of the target dataset that the enrichment service will use is also an important criteria for obtaining good quality enrichment.

3   Define the enrichment workflow

The workflow(s) supported by an enrichment service should be clearly identifiable so that an enricher:

- Can act upon the different components of the enrichment (source, target, and application scenario)

- Knows also the order in which to activate the steps.

Rules for the enrichment need to be defined and documented.

3 Define the enrichment workflow

The workflow(s) supported by an enrichment service should be clearly identifiable so that an enricher:

- Can act upon the different components of the enrichment (source, target, and application scenario)

- Knows also the order in which to activate the steps.

Rules for the enrichment need to be defined and documented.

Don' t forget user-initiated enrichment workflows !

4 Train the enricher

The enricher should have sufficient knowledge of the enrichment tool and on the parts of the data affected by the enrichment service

4 Train the enricher

The enricher should have sufficient knowledge of the enrichment tool and on the parts of the data affected by the enrichment service

This knowledge is crucial for the assessments of the enrichments results and the potential enrichment errors which might occur !

5 Test your enrichment workflow

Your enrichment service should be flexible enough to have adjustments made.

6 Assess the quality of your enrichment

The quality needs to be assessed against the goals set for the enrichment.

6 | Assess the quality of your enrichment

The quality needs to be assessed against the goals set for the enrichment.

Only if you know what the best possible outcome of the process is, you can evaluate how well the process worked !

7 Have an evaluation strategy and choose the right evaluation method

Assessing the quality of your enrichment will require an evaluation strategy taking into account:

- the correctness of the enrichments (intrinsic evaluation)

- Factors such as the search performance, user-experience and relevance (extrinsic evaluation)

Creation of Gold standards or annotated samples can be a good strategy but

- Make sure the diversity of the dataset represent the diversity of the data

- Make clear guidelines for raters and train them

7 Have an evaluation strategy and choose the right evaluation method

Assessing the quality of your enrichment will require an evaluation strategy taking into account:

- the correctness of the enrichments (intrinsic evaluation)

- Factors such as the search performance, user-experience and relevance (extrinsic evaluation)

Creation of Gold standards or annotated samples can be a good strategy but

- Make sure the diversity of the dataset represent the diversity of the data

- Make clear guidelines for raters and train them

Using a collaborative tool or computing inter-rater agreements can help the validation process.

Don't forget user-initiated enrichment workflows !

8 Apply user-initiated enrichment workflows

Don' t forget user-initiated enrichment workflows !

8 Apply user-initiated enrichment workflows

9 Document your enrichment process and learnings

Documentation will allow the improvements of future enrichment strategies and the maintenance of an enrichment process

10 Monitor your enrichment process and re-assess

An enrichment process needs constant re-evaluation

- Source data might have expanded with more information

- The target dataset might be richer (new terms, new languages, more granular)

- The workflow might have been improved

All the details in the report at http://bit.ly/Evaluation-Enrichment

Have a good read !

valentine.charles@europeana.eu

DISTANCE 8 MÈTRES

UNIQUE AU MONDE

Troupe Banola leur dernière création
Anonymous

1911, Circus Museum
Netherlands, CC BY-SA