

Evaluating Multilingual Features in Europeana

*Deriving best practices for digital cultural
heritage*

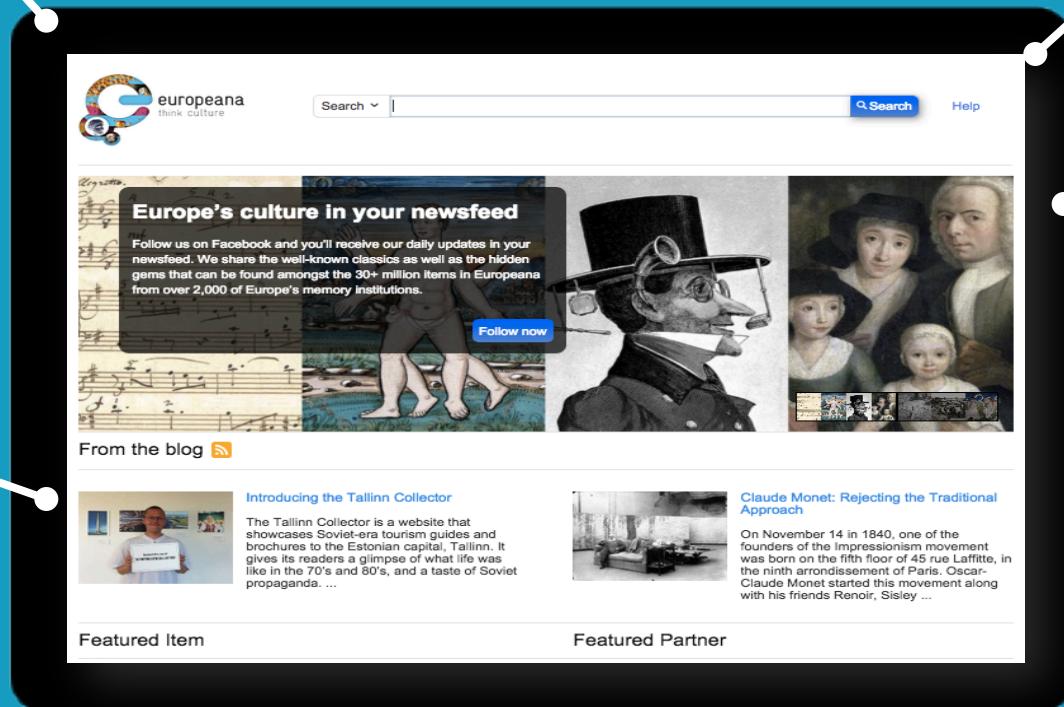
Juliane Stiller,
Humboldt-Universität zu Berlin

MultilingualWeb Workshop,
Riga, 29.04.2015

Europeana facts

42 millions
objects

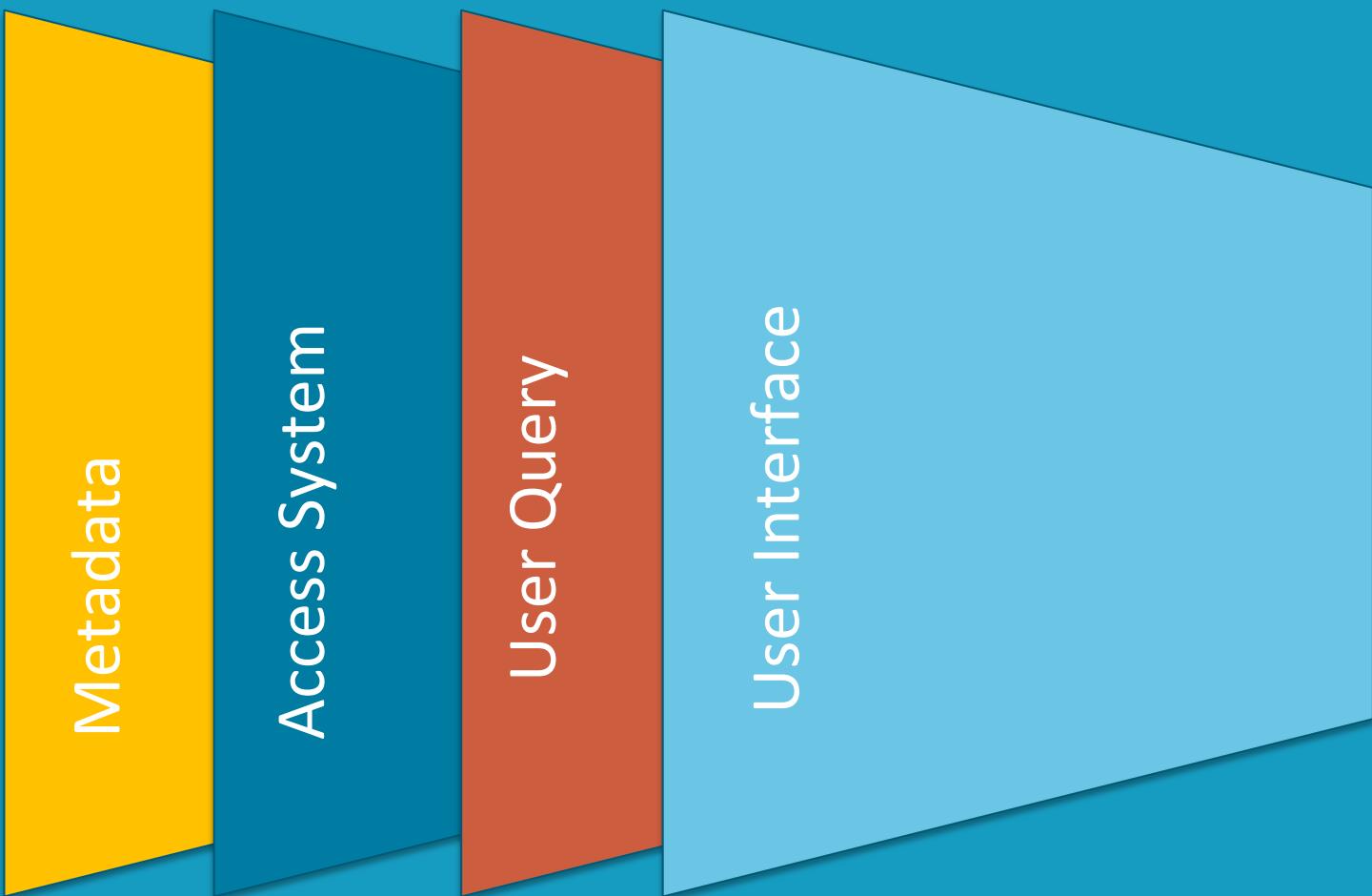
Text,
images,
video
sound



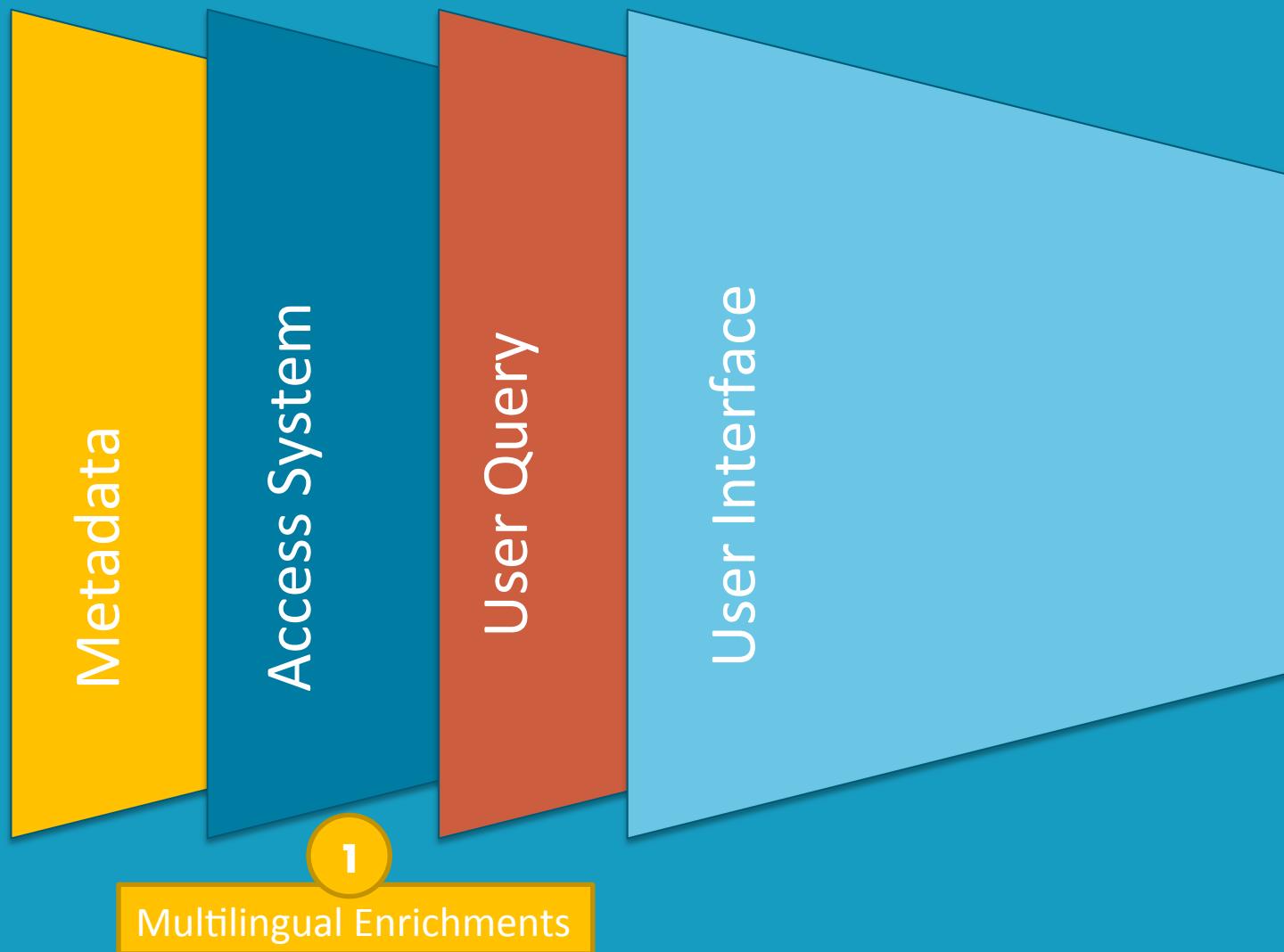
Multilingual
users, objects,
metadata

2300
institutions

Accessing digital cultural heritage



Crossing the language barrier



1. Multilingual enrichments



© Free access - no re-use

View item at
National Library of the Netherlands - Koninklijke Bibliotheek
[↗](#)

De koppelaarster

Description: Litt. : Christopher Braider: Refiguring the real. Picture and modernity in word and image. 1400-1700. Princeton. 1993

Creator: Johannes Vermeer ;

http://dbpedia.org/resource/Johannes_Vermeer

Contributor: Johannes Vermeer

Date of creation:

Type: StillImage

Format: schilderij, image/jpeg

Subject: koppelen

Auto-generated tags ▾

Who ▾

Agent Term: http://dbpedia.org/resource/Johannes_Vermeer

Agent Label: [jan vermeer] (de) ; [扬·弗美尔] (zh) ; [jan vermeer] (it) ; [johannes vermeer] (pt) ; [jan vermeer] (pl) ; [johannes vermeer] (sv) ; [johannes vermeer] (fr) ; [johannes vermeer] (en) ; [верmeer, ян] (ru) ; [johannes vermeer] (es) ; [johannes vermeer] (nl)

Number of enriched objects, their type and vocabularies



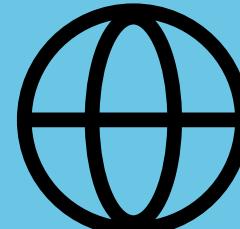
Time

Semium Time
10.2 Millions



Concept

GEMET, DBpedia
9.2 Millions



Locations

GeoNames
7 Millions



Agents

DBpedia
144,000

What happens if automatic solutions fail?



Devaluation of curated metadata
Loss of trust from providers, users & service re-users
Irrelevant search results



We need to evaluate features to derive best practices and improve services!



Picture from Poland by Staff Photographer.

Description:

Picture from Poland by Staff Photographer. A Polish soldier takes a last walk with his family before leaving for the front.

dc:subject:

<http://www.beeldbankkw02.nl/cgi-bin/miserver>
def:
Familieeven
Gezinnen
Kinderen
Militairen - Zie ook: Officieren, Soldaten
Mobilisatie - Zie ook: Demobilisatie
Oorlogvoorbereiding
Polen
Poolse Strijdkrachten

Semantically incorrect enrichment

dc:subject:
 Translate details

Select language ▾

Pow

Polen (Dutch) ≠ Polen (Basque)

- » Helpdesk
- » User directory
- » Roles
- » Organisations
- » NFP/Eionet IG
- » Mails to NFPs
- » SERIS
- » Workplan/planner
- » Meetings & events
- » Priority dataflows
- » AQ Portal

Find a person

Account services
I have
» lost my password

EIONET
European Environment Information and Observation Network

General Multilingual Environmental Thesaurus

Thematic Listings INSPIRE Data Thesaurus

pollen

Microspores of seed-producing plants. Each pollen grain contains a much-reduced male gametophyte. Pollen grains are transferred by wind, water, birds or other animals. (S)

Themes

- » biology

Groups

- » BIOSPHERE (organisms, ecosystems)

Related terms

Broader: [plant component](#)

Related: [aerobiology](#)

Other relations

Has close match: [UMTHES: Pollen](#)

Has exact match: [AGROVOC: Pollen](#)

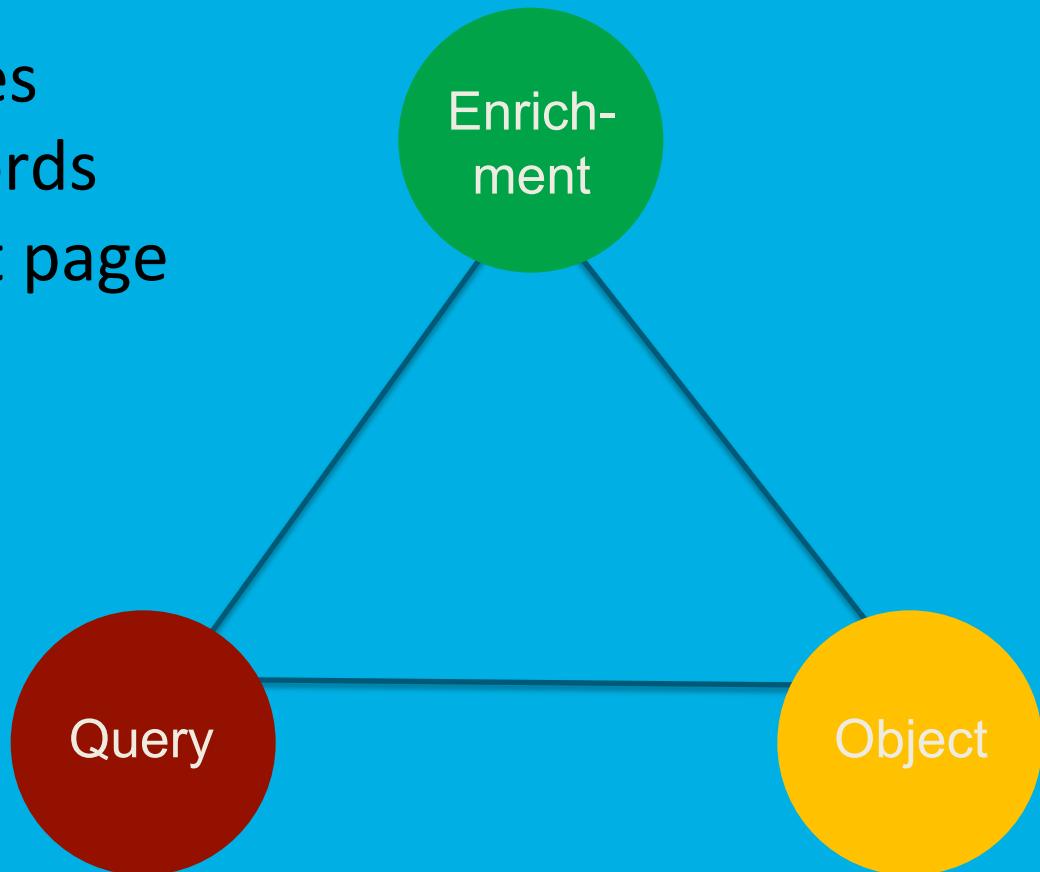
Translations

Arabic:	النَّاقِح
Basque:	polen

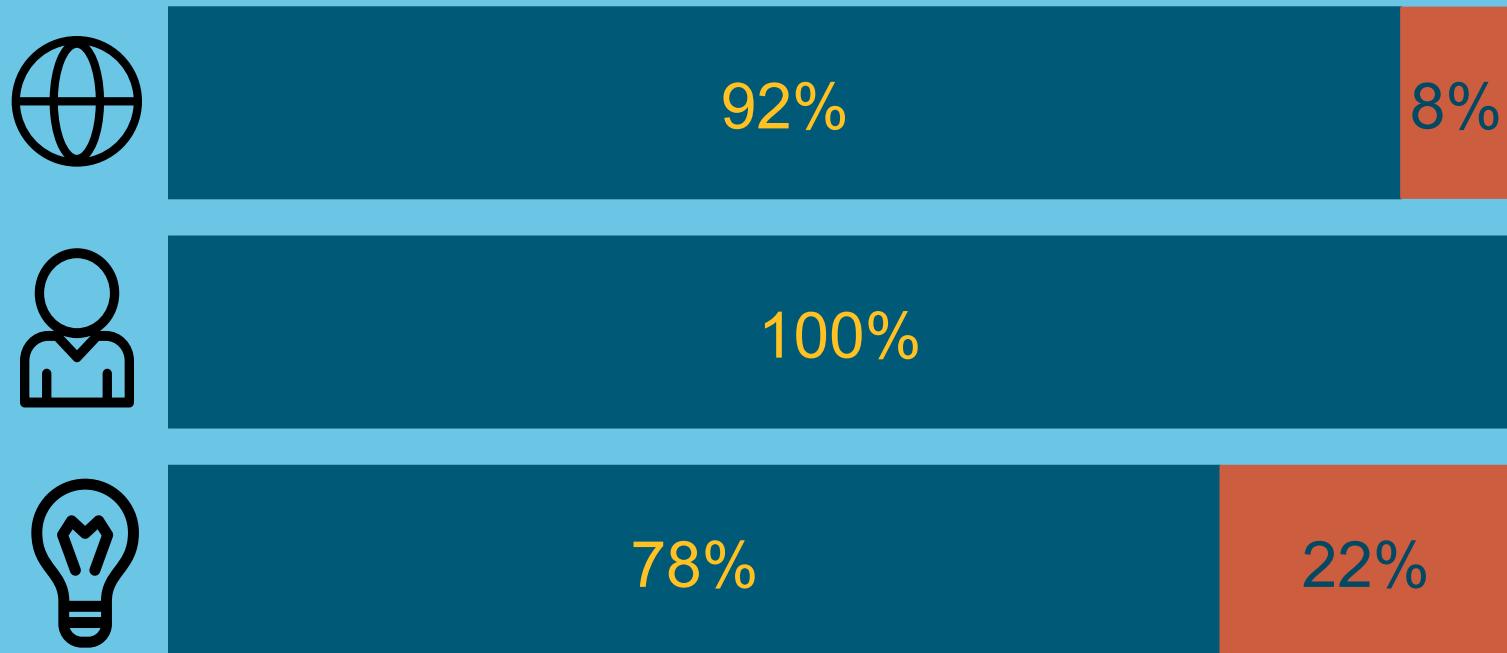
Afrikaans:	polen
Czech:	pyl
Danish:	pollen
Dutch:	stuifmeel
English:	pollen
English (US):	pollen
Estonian:	õielõlm
Finnish:	siitepöly
French:	pollen
German:	Pollen
Greek:	γύρη
Hungarian:	poller/virágpor
Irish:	páilín
Italian:	polline
Latvian:	zielputekšņi
Lithuanian:	žiedadulkės

Evaluation method – enrichments

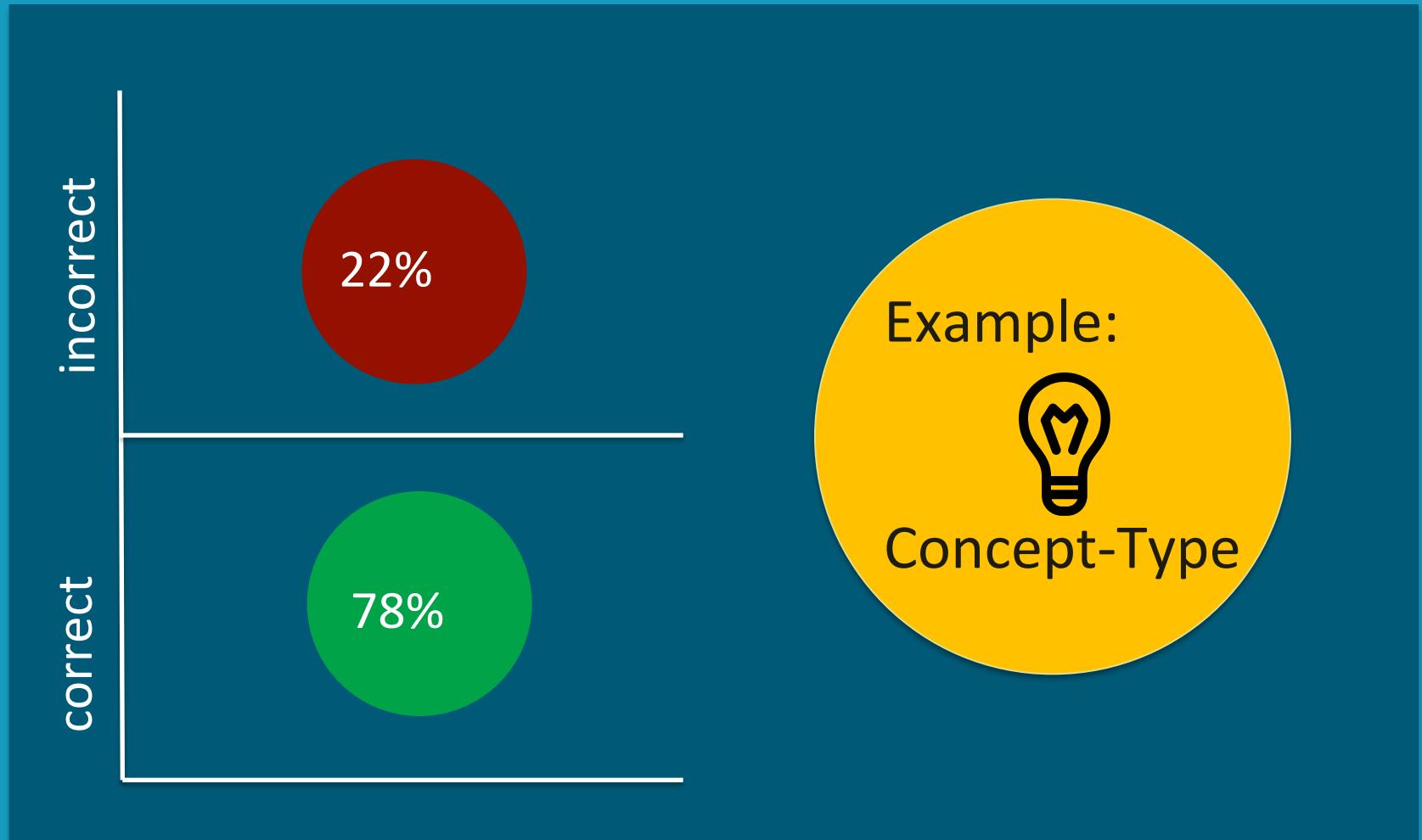
- 100 queries
- 1,121 records
- First result page



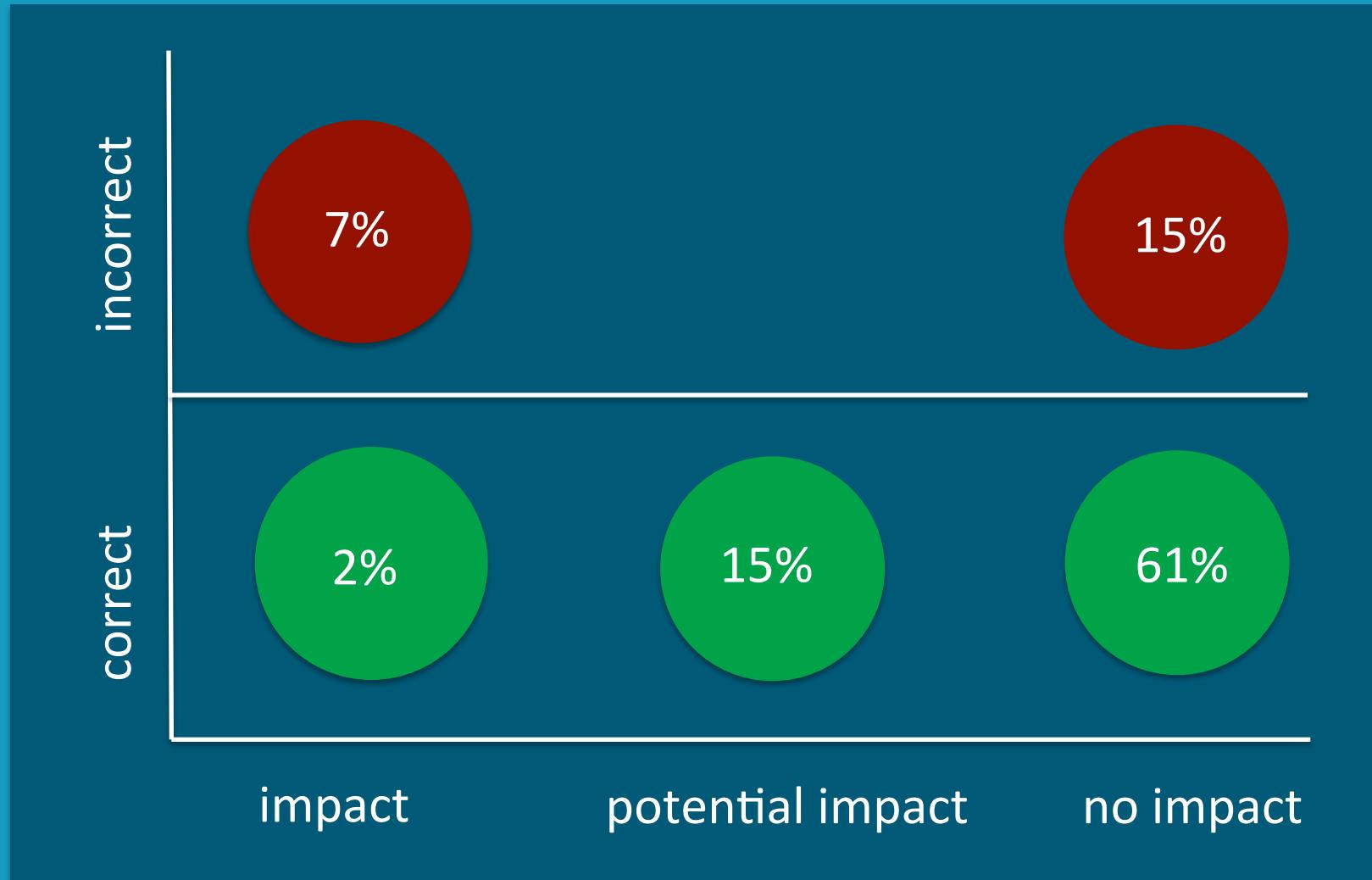
Correct and incorrect enrichments per type



Percentage of incorrect enrichments



Impact of enrichments on retrieval



Source: Stiller, J., Olensky, M., Petras, V.:

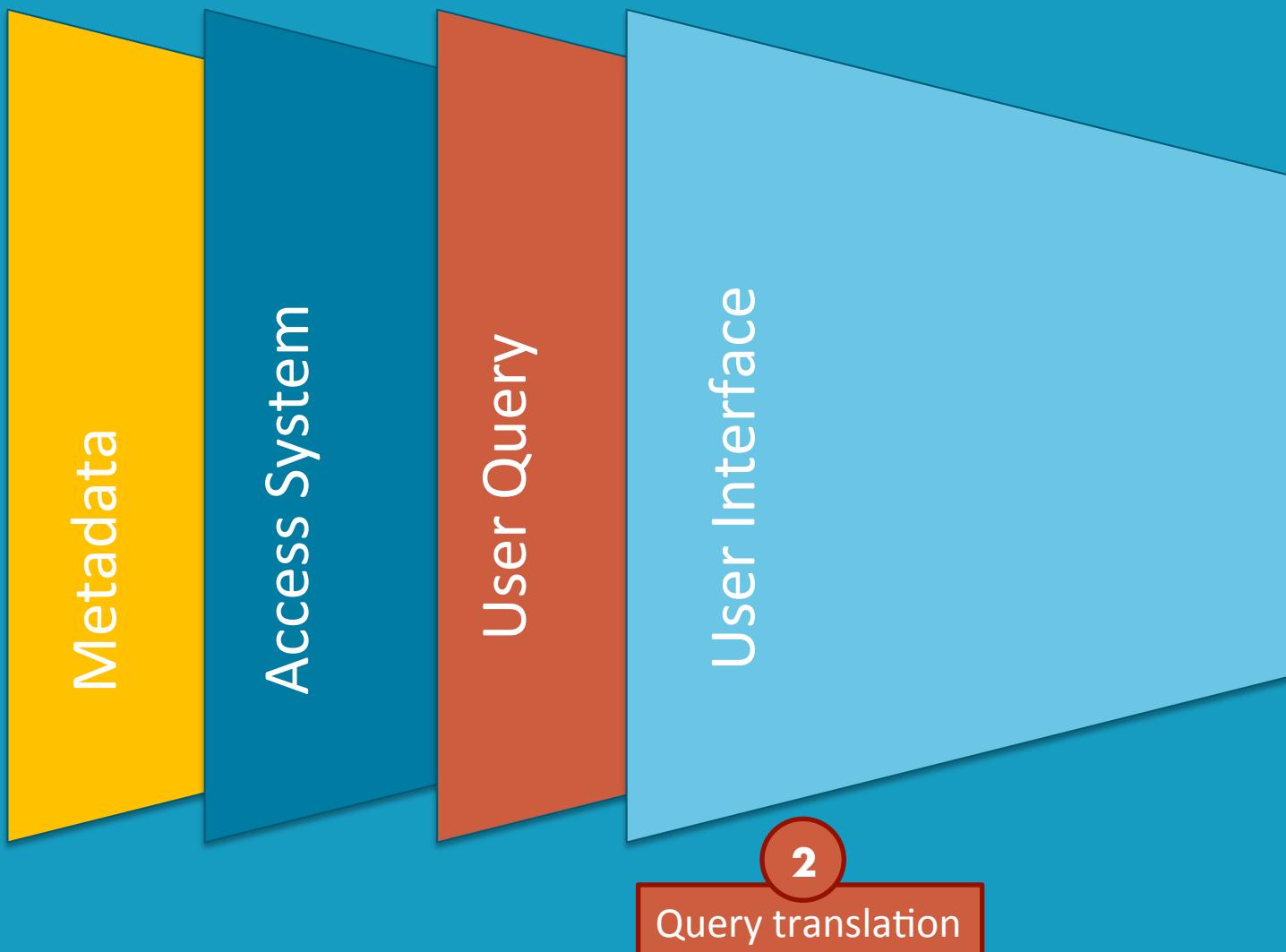
A Framework for the Evaluation of Automatic Metadata Enrichments. MTSR 2014

Method provided:

- Holistic view on the quality and impact of enrichments
- Query-dependent measures may deliver different results
- Measures are system-focused and independent of the user's point of view
- Focus on first result page -> it might be different for long tail

Task force on enrichment and evaluation:
*[http://pro.europeana.eu/europeanatech-task-forces/
evaluation-and-enrichments](http://pro.europeana.eu/europeanatech-task-forces/evaluation-and-enrichments)*

Crossing the language barrier



2. Query translation

User preferences

- | | |
|---|--|
| <input checked="" type="checkbox"/> English | <input type="checkbox"/> Español |
| <input type="checkbox"/> Basque | <input type="checkbox"/> Eesti |
| <input type="checkbox"/> Български | <input checked="" type="checkbox"/> Français |
| <input type="checkbox"/> Català | <input type="checkbox"/> Gaeilge |
| <input type="checkbox"/> Čeština | <input type="checkbox"/> Hrvatski |
| <input type="checkbox"/> Dansk | <input type="checkbox"/> Íslenska |
| <input checked="" type="checkbox"/> Deutsch | <input type="checkbox"/> Italiano |
| <input type="checkbox"/> Ελληνικά | <input type="checkbox"/> Lietuvių |

Wikipedia API



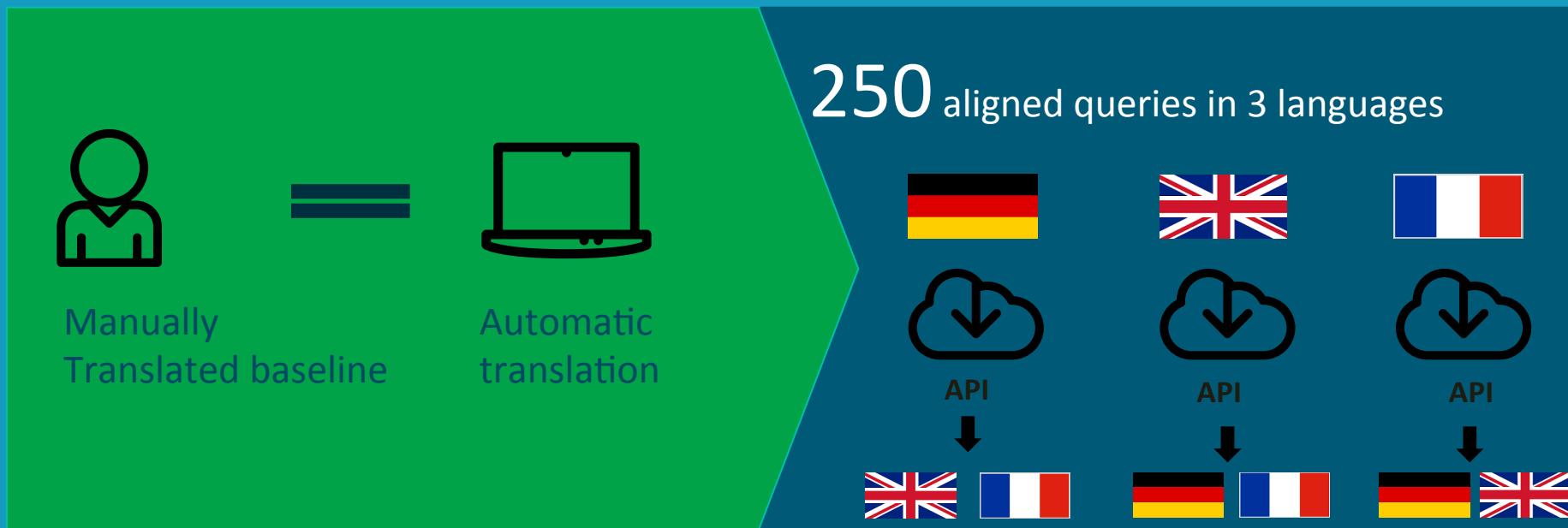
Query
construction



Mixed language
result list

The detailed and long version of the process is in Péter Király: Query Translation in Europeana. Code4lib, Issue 27, <http://journal.code4lib.org/articles/10285>

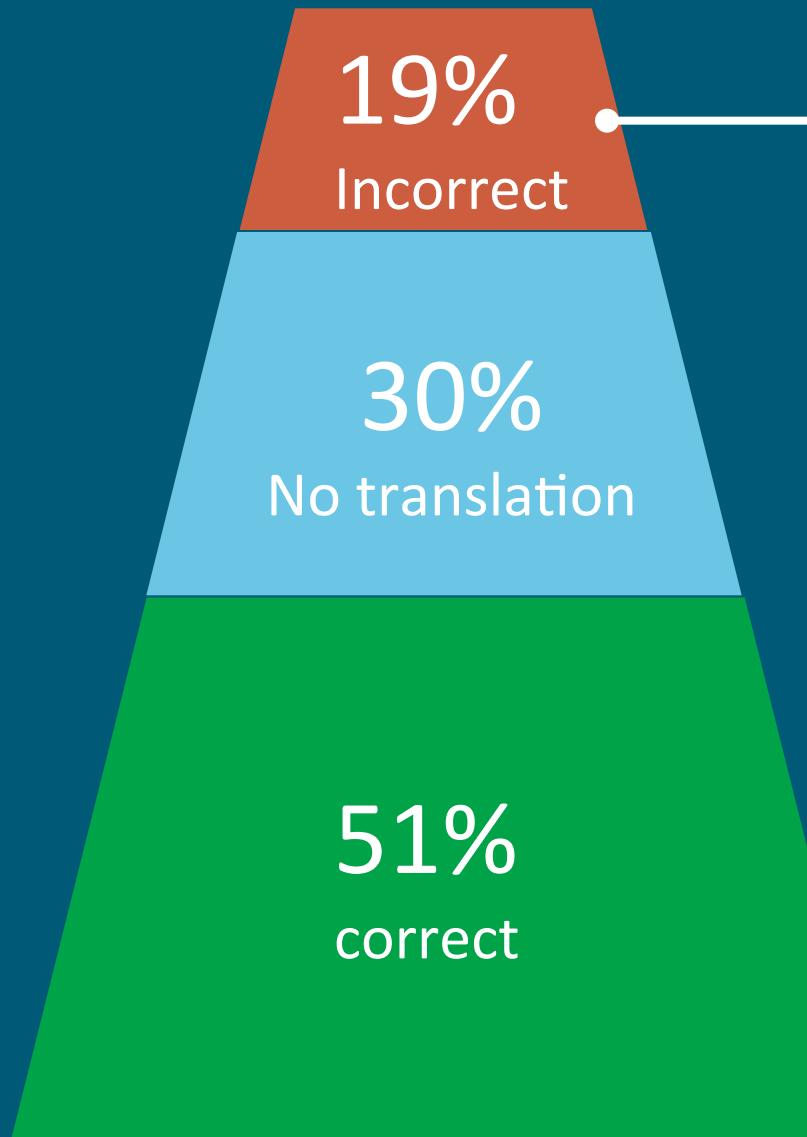
Evaluation method – query translation



Example

Baseline	From to	From to	From to
Stummfilm	-	Stummfilm	Stummfilm
Silent film	Silent film	-	Silent film
film muet	cinéma muet	cinéma muet	-

Results



Query Translation

Query	Translation
Unarmed	unarmed – Best of 25th Anniversary
Bulgaria	Bulgarien (Begriffserklärung)

For over **80%** of the queries,
the automatic solution is
suitable

Evaluations help

- revealing workflow problems and issues
- Giving insight on the real impact of automatic solutions
- Targeting efforts and adapt algorithms
- Deriving best practices

Best practices

- Should also incorporate evaluation methodologies



Email: juliane.stiller@ibi.hu-berlin.de

Acknowledgements: All icons designed by [freepik](#)