EEXCESS

SEVENTH FRAMEWORK PROGRAMME

Bringing Your Content to the User, not the User to Your Content – A lightweight approach towards integrating external content via the EEXCESS framework

Martin Höffernig, Werner Bailer
JOANNEUM RESEARCH
SWIB 2015, Hamburg, 2015-11-23

JOANNEUM RESEARCH · Collections Trust · Archäologie und Museum, Bildungs-, Kultur- und Sportdirektion Kanton Basel-Landschaft · MENDELEY · INSA LYON

wissenmedia · UNIVERSITÄT PASSAU · KNOW Center · ZBW Leibniz-Informationszentrum Wirtschaft, Leibniz Information Centre for Economics · bit media, member of the bit group

Outline (1)

- Introduction to EEXCESS
- Tools for content injection
  - Install & try Chrome plugin
- Integrating a new data provider
  - Introduction to the data model
  - PartnerWizard
  - Integrate data provider with a web-based tool

EEXCESS

Outline (2)

- Refining data mapping
  - Introduction to mapping tool
  - Review and update mappings
  - Test and check mappings
- Metadata quality assessment
  - Checking input and mapping quality

EEXCESS

Logistics

- Wifi
  - SSID: SWIB*
  - Password: berners-lee
- Coffee break 15.30-16.00
- Short breaks in each of the blocks before & after (flexible timing)

EEXCESS

Materials

Links, examples etc.

http://eexcess-dev.joanneum.at/swib15.html

Accounts: see handout

Slides: will be made available on EEXCESS website

EEXCESS

EEXCESS - Enhancing Europe's eXchange in Cultural Educational and Scientific resourceS

- EU FP7 project (Feb. 2013-Jul. 2016)
- 10 partners
  - technical partners
  - scientific partners
  - cultural institutions

EEXCESS


JOANNEUM
RESEARCH
DIGITAL


UNIVERSITÄT
PASSAU


KNOW
Center


INSA
LYON


ZBW
Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics


bit media
member of bit group


Archälogie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft


Collections
Trust


MENDELEY


wissenmedia

EEXCESS

EEXCESS

SEVENTH FRAMEWORK PROGRAMME

Overview

JOANNEUM RESEARCH

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

INSA LYON

wissenmedia

UNIVERSITÄT PASSAU

KNOW Center

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

bit media
member of the bit group

Motivation

- Vast amounts of digital cultural and scientific resources available

europeana
think culture

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

- Still memory organisations (i.e. library, museums, archives) face challenges in disseminating their content
- Two reasons, addressed by EEXCESS:
  - Todays content dissemination processes are optimised for mainstream content
  - Long tail content needs contextualisation

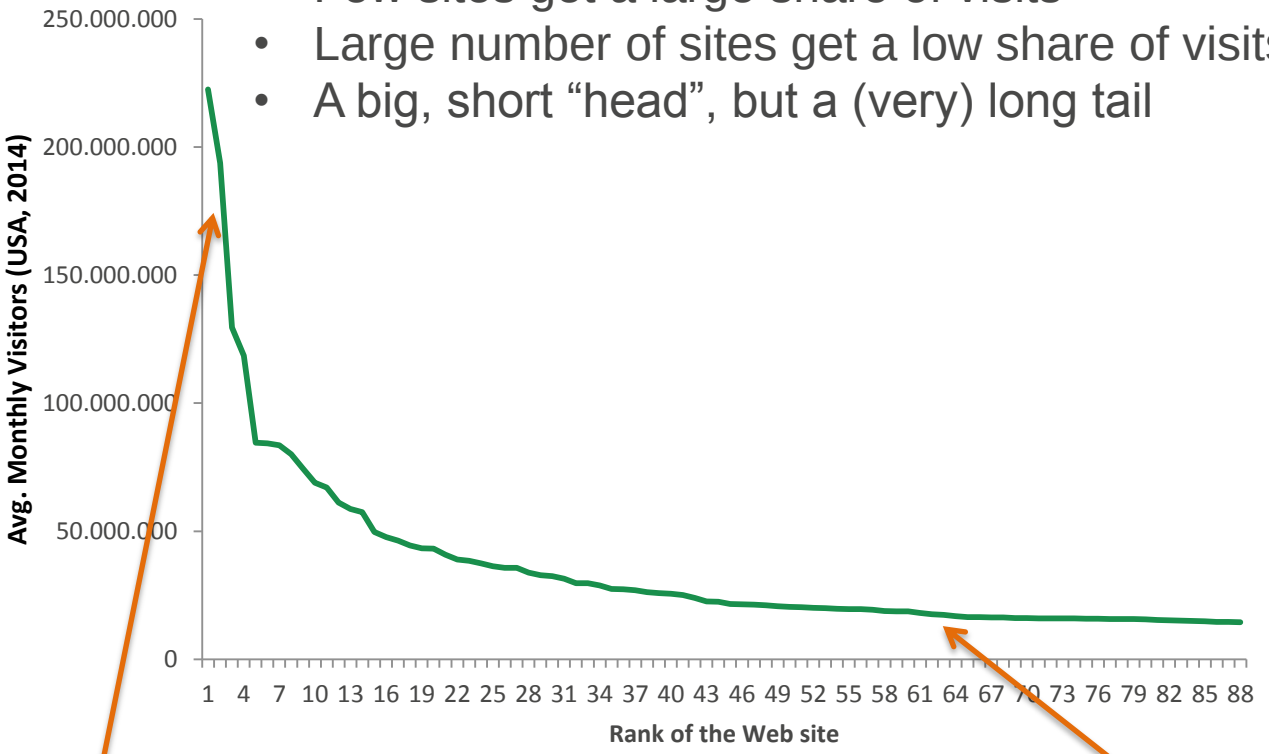
EEXCESS

Motivation

- Content provider strategies
  - Dedicated portals
  - Search engine optimisation
  - Social network marketing
- User strategies
  - Use major search engines
  - Use Wikipedia

The Long Tail Content

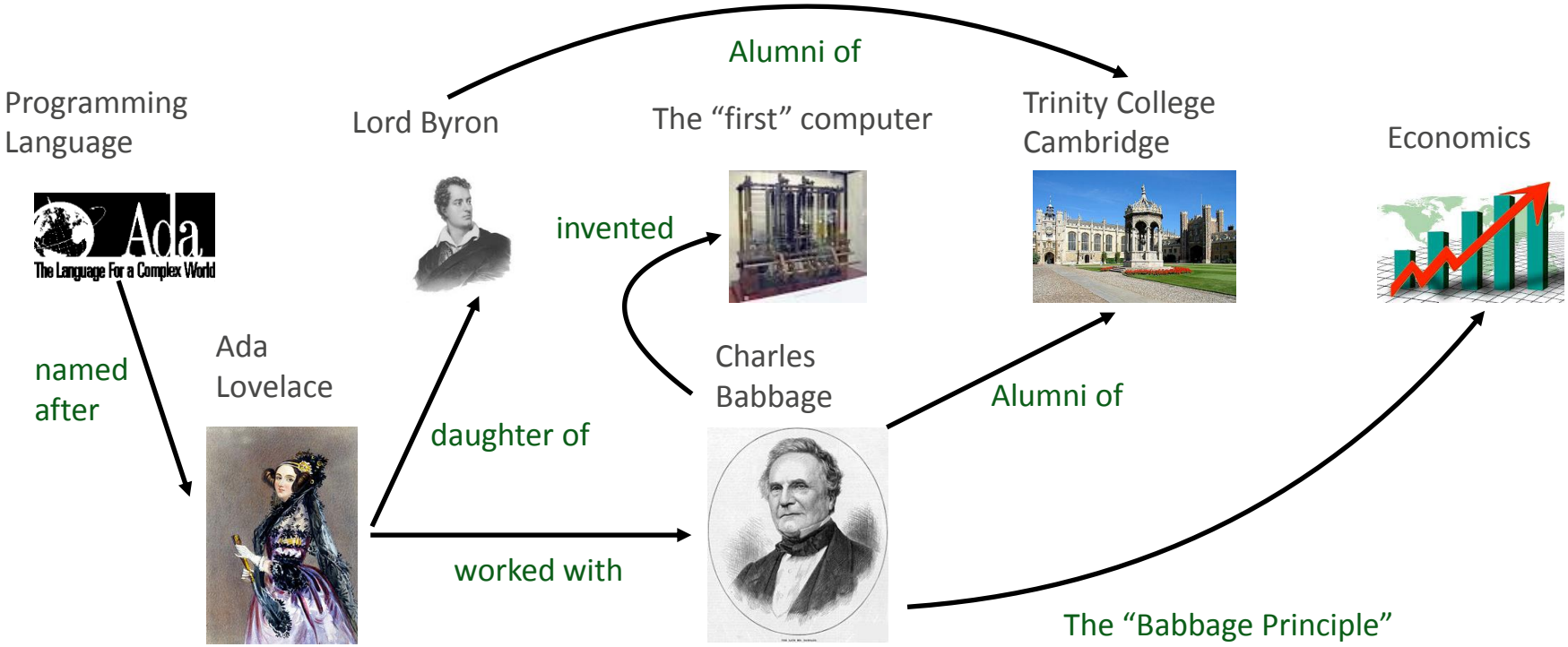
- Few sites get a large share of visits
- Large number of sites get a low share of visits
- A big, short “head”, but a (very) long tail



Challenges of the Long Tail

- High specialisation
- Low contextualisation
- Most items are unrelated
- Not easy to consume
- Low # of users per item

The value of long tail content



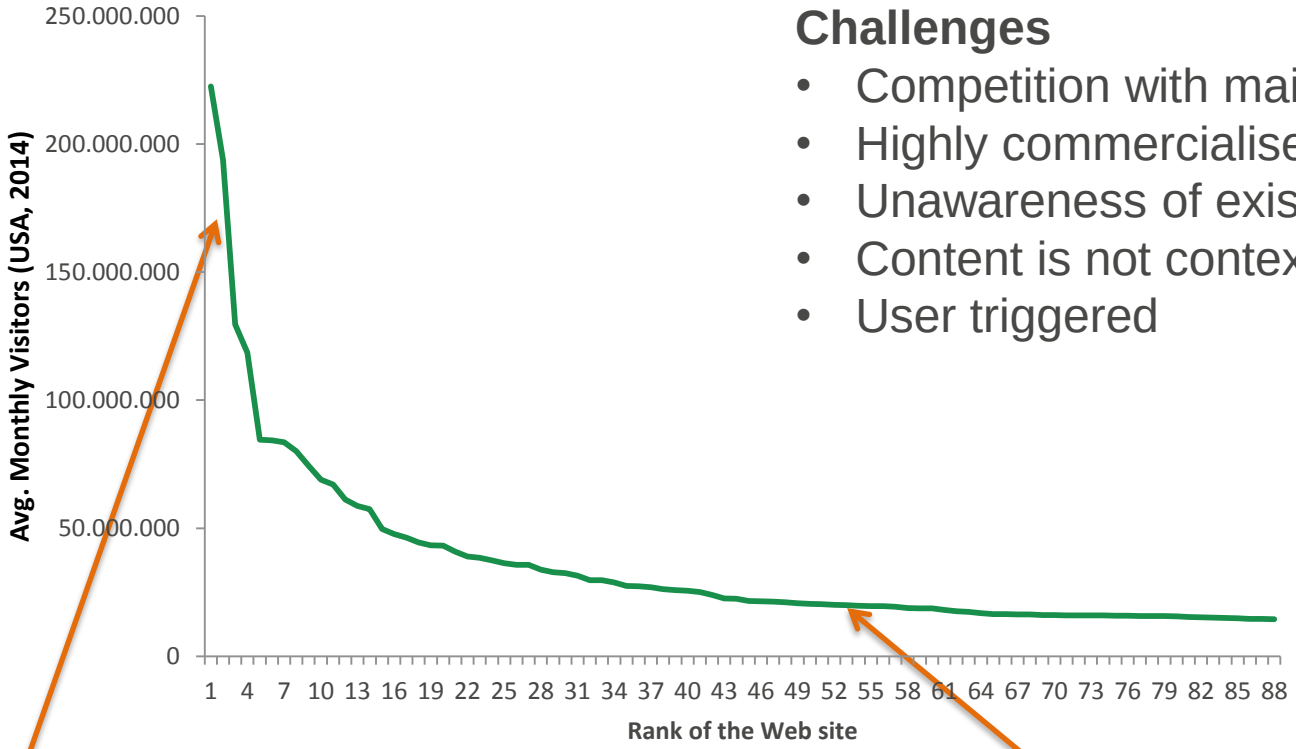
Cultural Heritage content
- Multimedia Artefacts
- Original Material
- Explanations

Value of Long Tail Content
- Discover new knowledge
- Verify information
- Enrich other content

Scholarly content
- Discourse
- Validated facts
- Additional explanations

EEXCESS

Long Tail content dissemination
Challenges of today‘s methods



Challenges

- Competition with mainstream content
- Highly commercialised
- Unawareness of existing portals
- Content is not contextualised
- User triggered

Search Engine Optimization
Social Media Marketing etc.

EEXCESS Vision

Unfold the treasure of cultural heritage and scholarly long-tail content for

- discovering new knowledge,
- triggering serendipitous effects,
- verifying consumed information,
- enriching new content

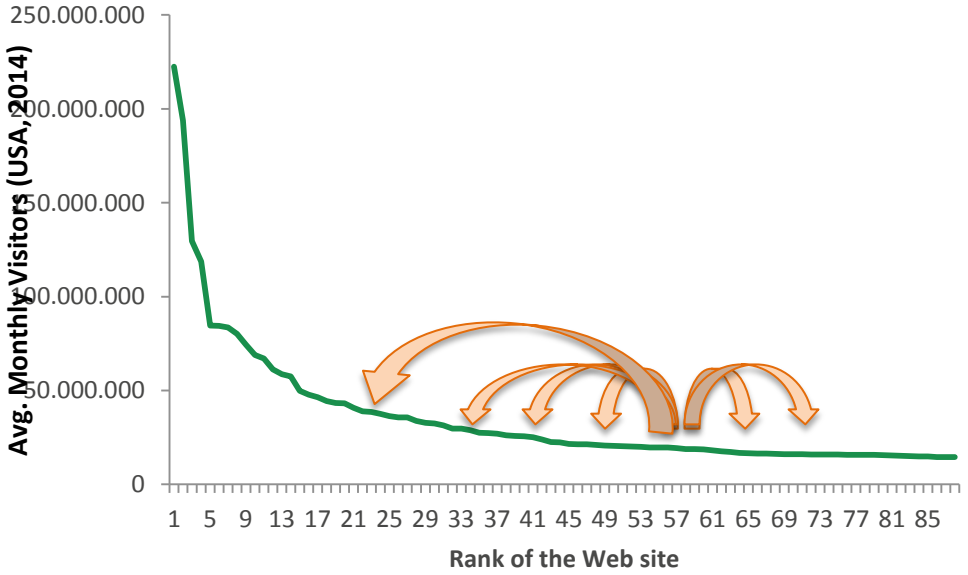
by “bringing the content to the user, not the user to the content”

EEXCESS

Approach
Idea

„Bring the content to the user, not the user to the content“

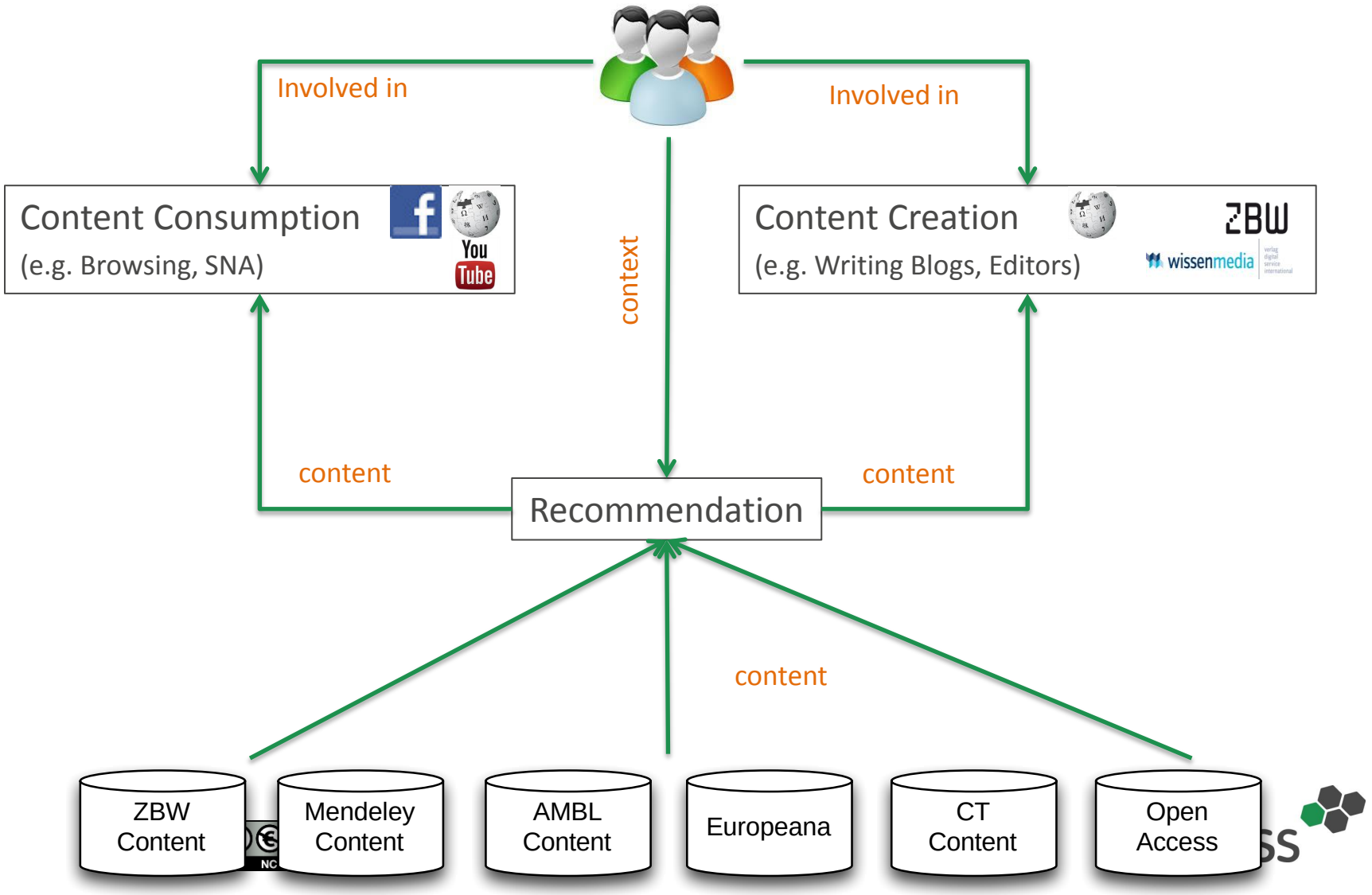
- Inject cultural and scientific content into existing web channels
  - Websites (Wikipedia, etc.)
  - CMS/LMS
  - Social media channels (Twitter, etc.)
  - Support “head-channels” as well as tail-channels
- Contextualise Long Tail content
  - Context of the web channel
  - User Context
  - User Task



- Gather user and usage feedback such that memory organisations can optimise their resource distribution

EEXCESS

Approach
Overview



Approach
Test Beds

3 User Groups as Test Beds

- Educational Support
  - Cultural/scientific resources injected to LMS
  - Pupils, teachers
- Scholarly Communication
  - Interconnecting cultural and scientific resources
  - Students, lecturers, researchers
- General Public Education
  - Disseminate cultural/scientific content to the general public
  - Regionally interested users, culturally interested users, media consumers

bit media
member of the bit group

MENDELEY

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

wissenmedia

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

EEXCESS

Objectives

- Adaptive Augmentation User Interfaces
- Personalized Recommendation
- Integration and Enrichment
- User and Usage Mining
- Privacy Preservation

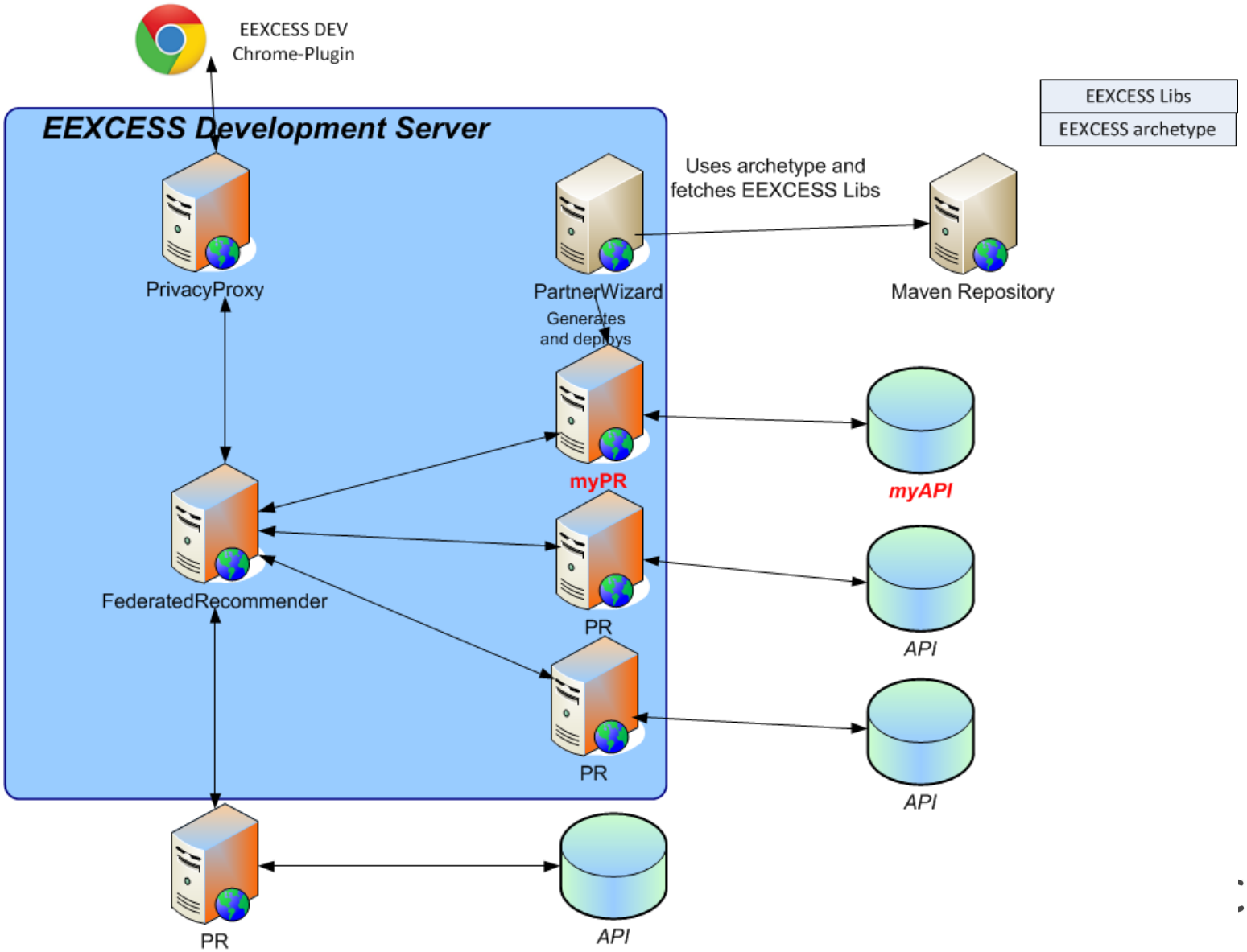
EEXCESS

Architecture

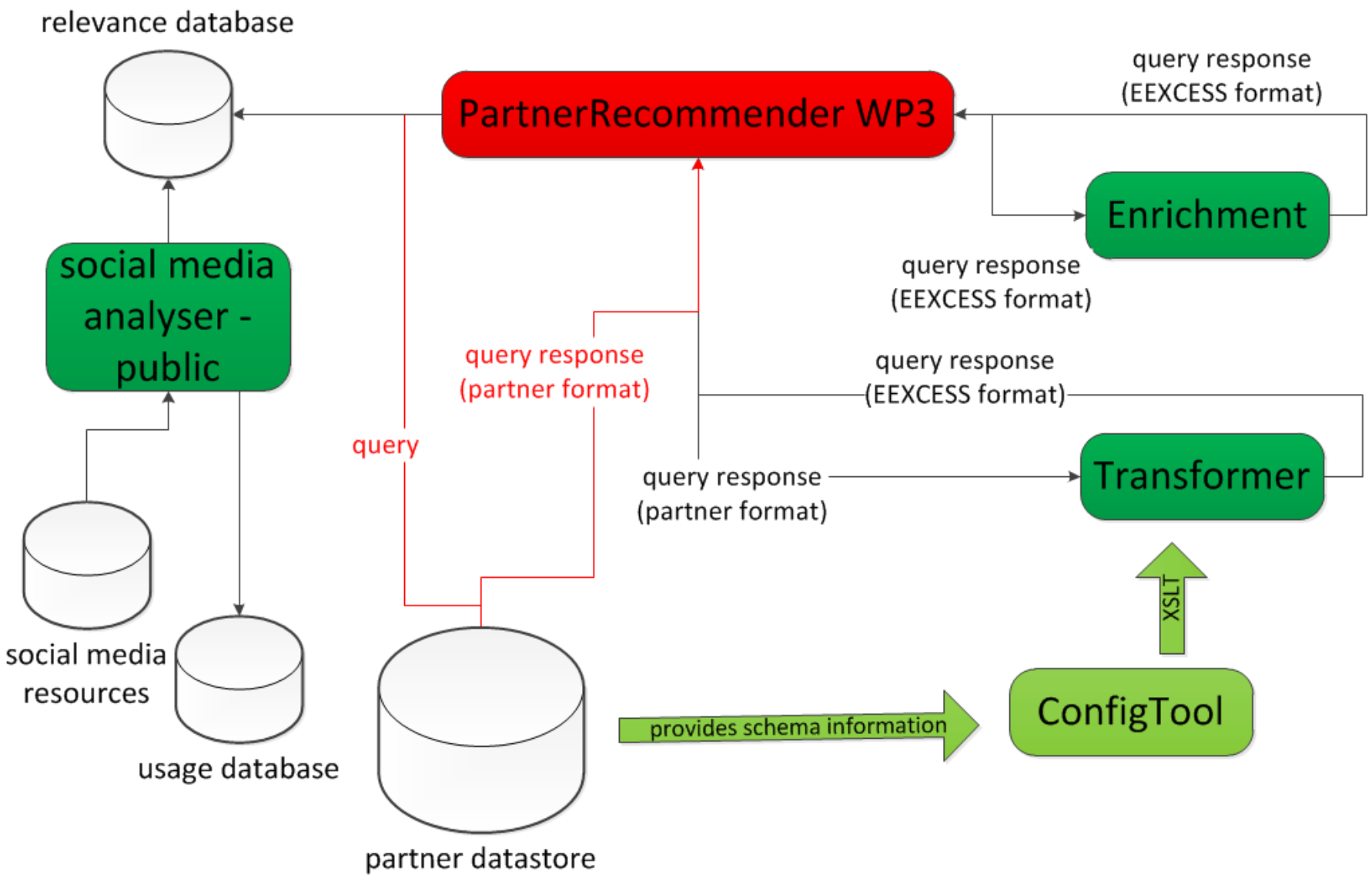
- Distributed data storage
  - Data remains with data providers
  - No central index
- Partner Recommender
  - Interface between data provider's API and EEXCESS system
- Federated Recommender
  - Aggregates and ranks results

EEXCESS

Architecture



Recommendation flow



Recommendation flow

- Implications from architecture
  - transformation and enrichment must work on the fly
  - configuration can be checked and revised manually, but transformation results cannot
  - no issues due to enrichment with resources that are no longer available

EEXCESS

Querying partner sites

- Two step process
  - Speed up retrieving initial results
  - Reduce load on partner sites
- Initial query
  - Get basic metadata of entries
- Detail query
  - Additional metadata
  - Images

EEXCESS

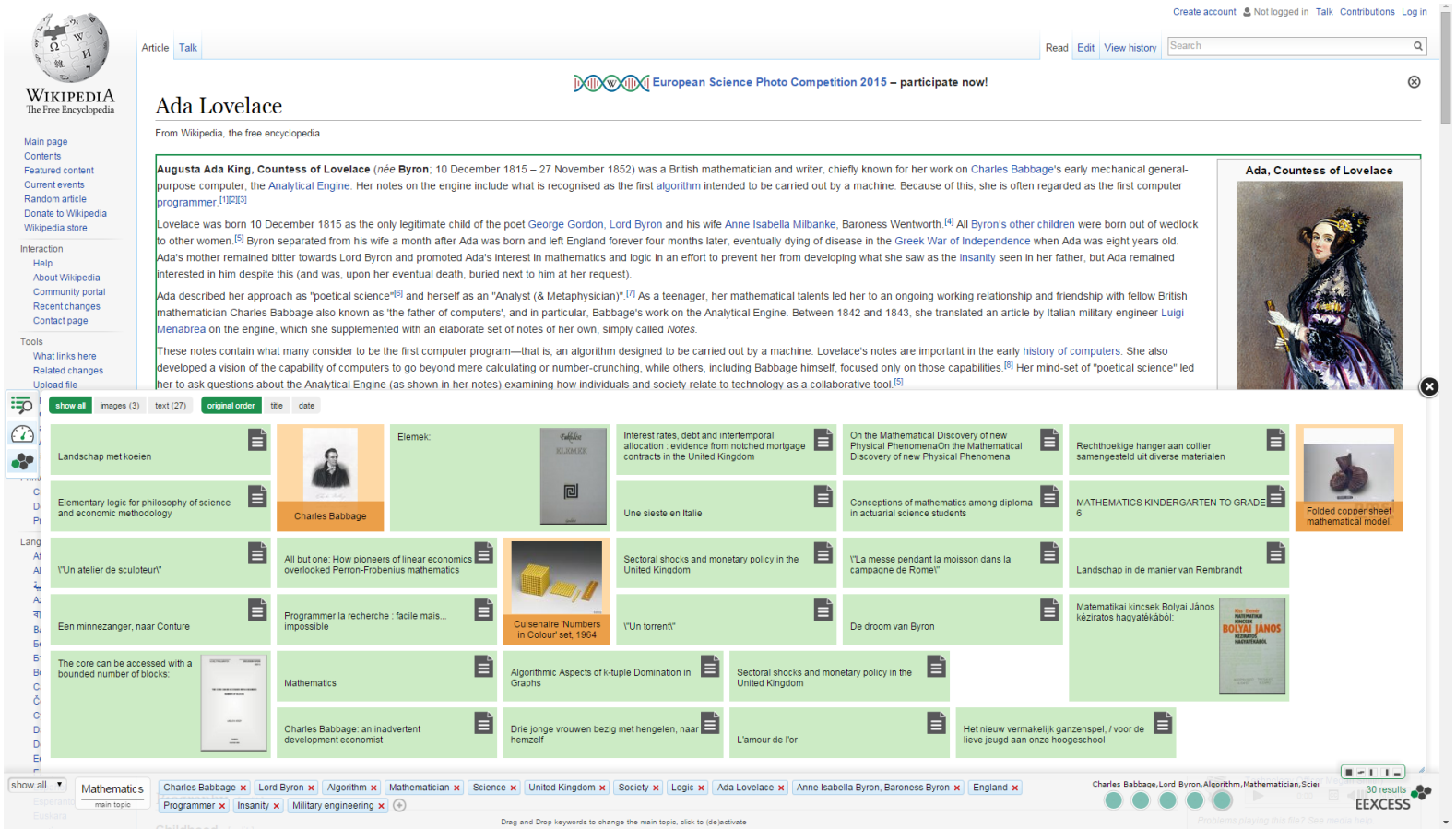
Metadata Enrichment

- Enriching textual information with named entities
- Type of metadata field is used to constrain entity type (e.g. persons) – search for entities with appropriate type
- Classify if words are entities in DBpedia
- Add synonyms using WordNet
- Add connected geographic terms using GeoNames

Content Injection – Chrome Browser Extension

Content Consumption

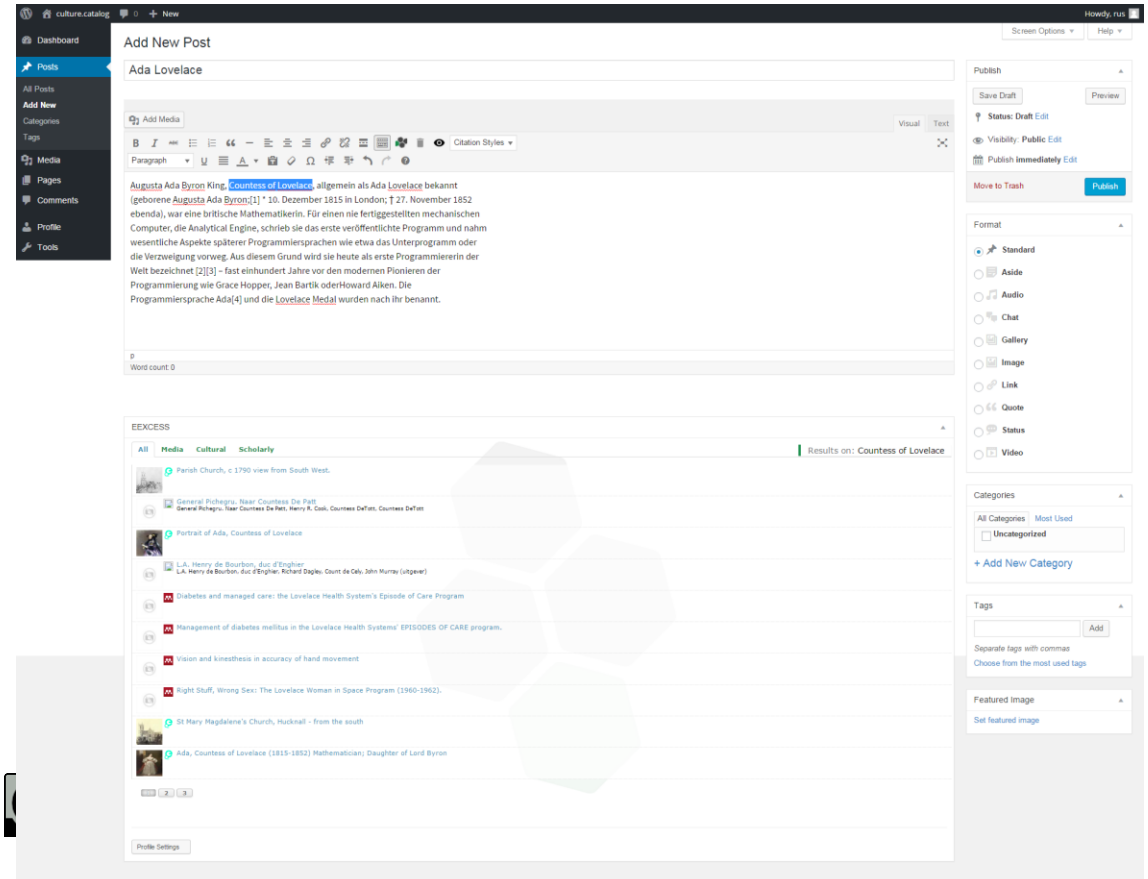
- A sidebar for recommending cultural/scientific content while browsing



Content Injection – Content Management Plugin (Wordpress)

Content Creation

- Inject cultural heritage and scholarly content into social media creation process
- Multiplier effect in the Blogging Community by providing a Wordpress Plugin

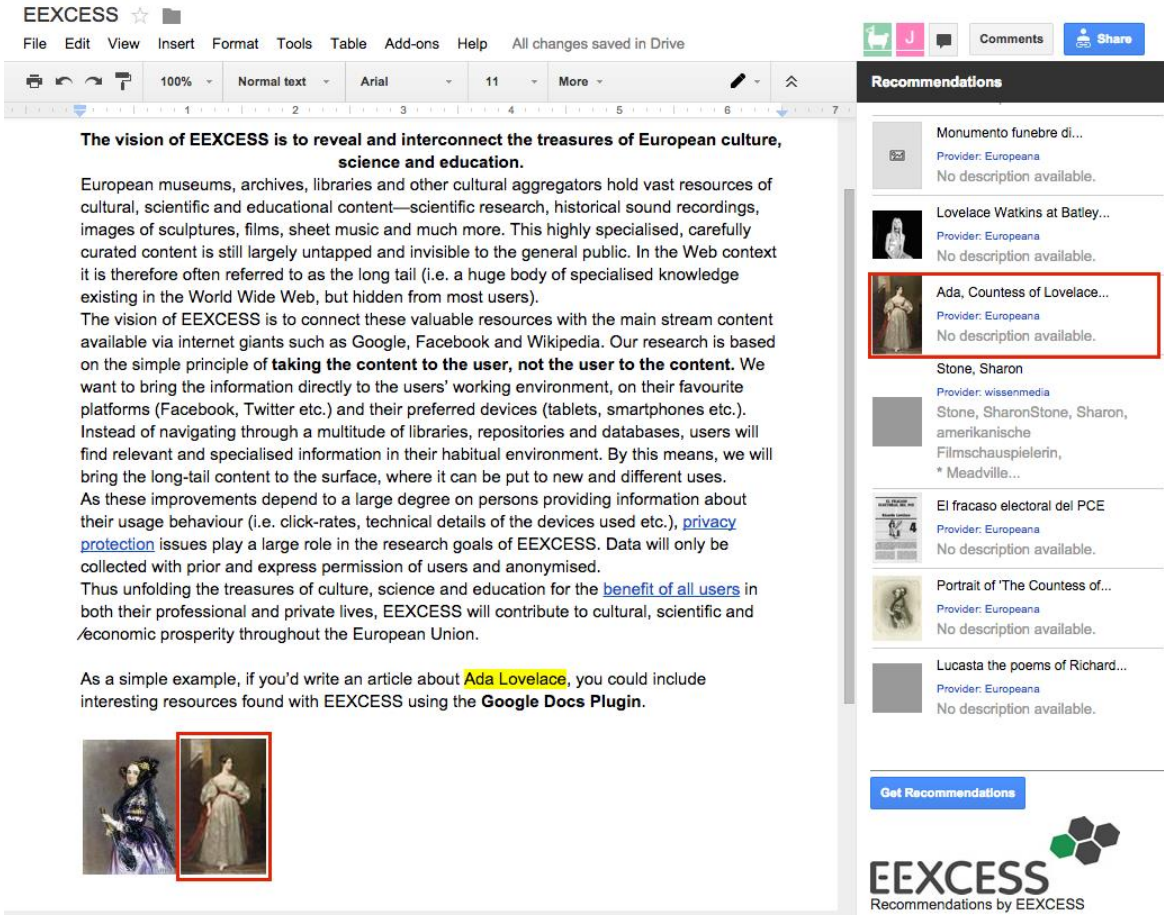


EEXCESS

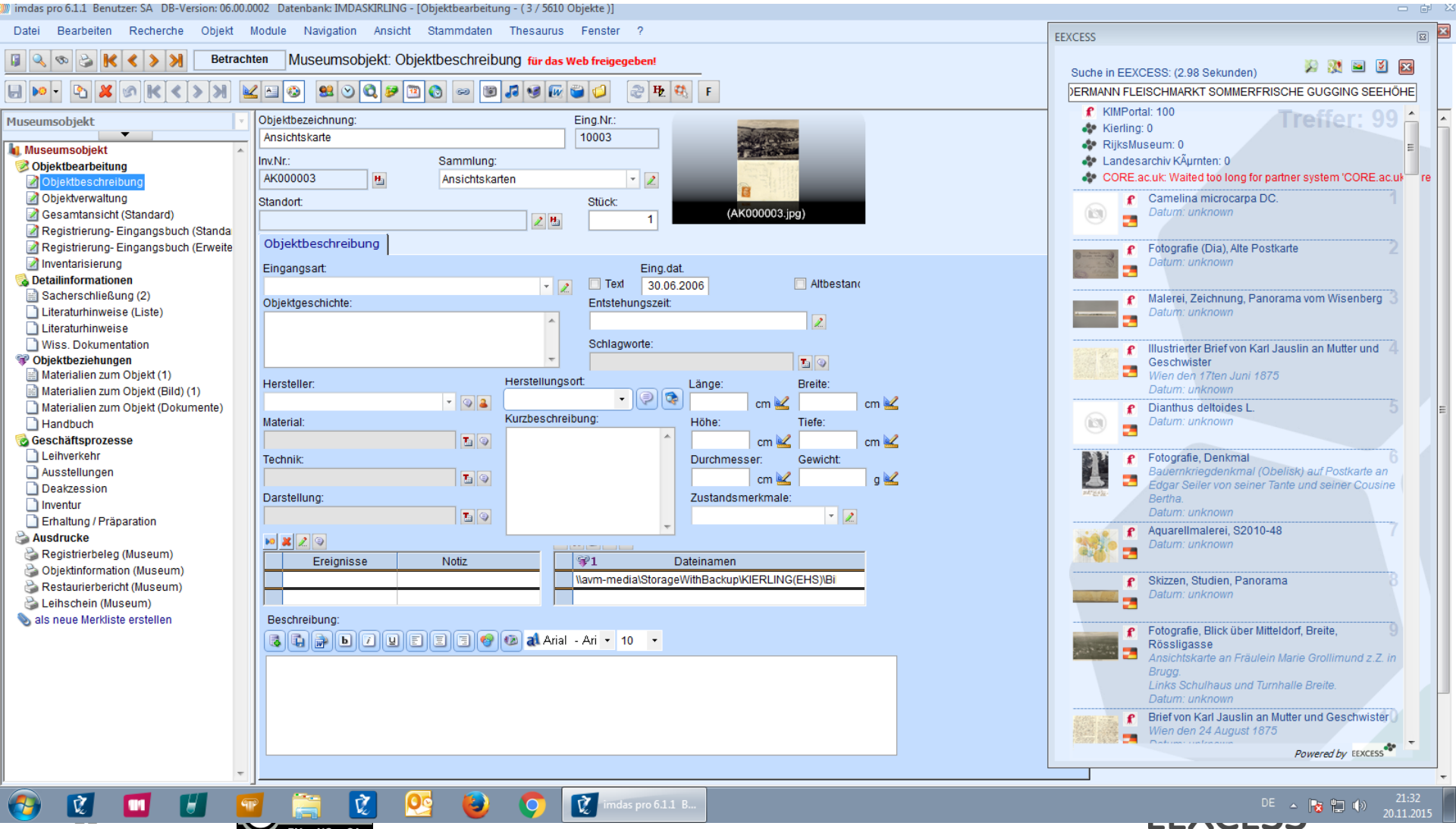
Content Injection – Google Docs App

Content Creation

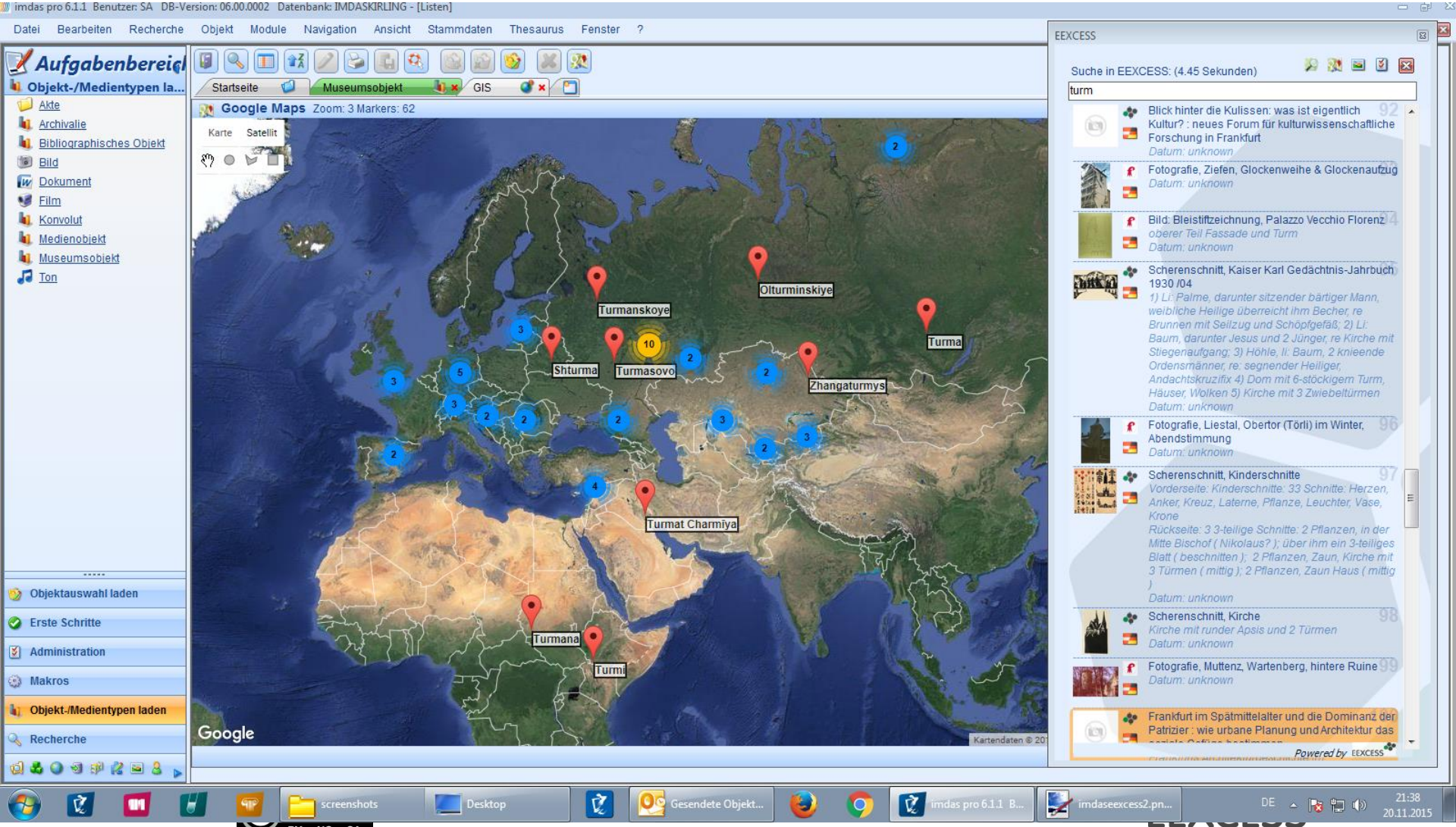
- Inject cultural heritage and scholarly content into collaborative word processing
- Support writing reports, grant requests, homeworks
- Google Apps Market for Google Documents as high-potential dissemination platform



Content Injection – Collection Management System



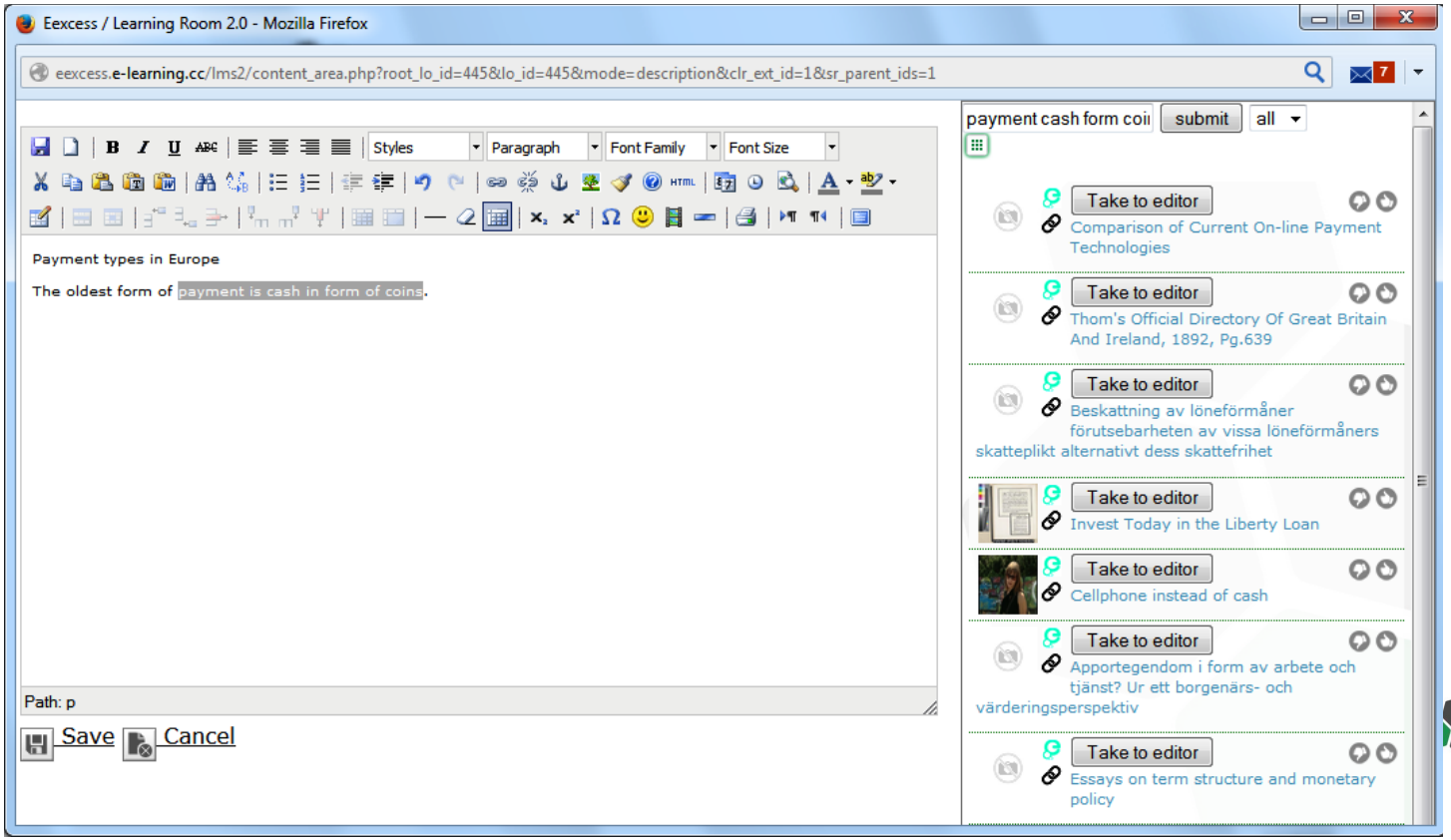
Content Injection – Collection Management System



Content Injection – Learn Management Systems

Content Creation for Educational Support

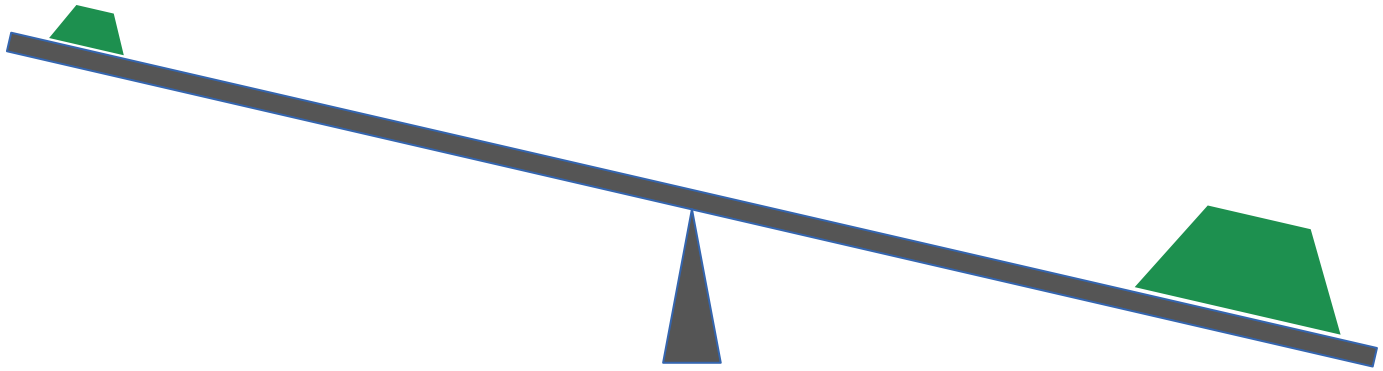
- Inject cultural heritage content into Learn Management Systems
- Moodle and BitMedia‘s SITOS LMS



Privacy vs. Personalisation trade-off?

Privacy

Personalisation/Quality

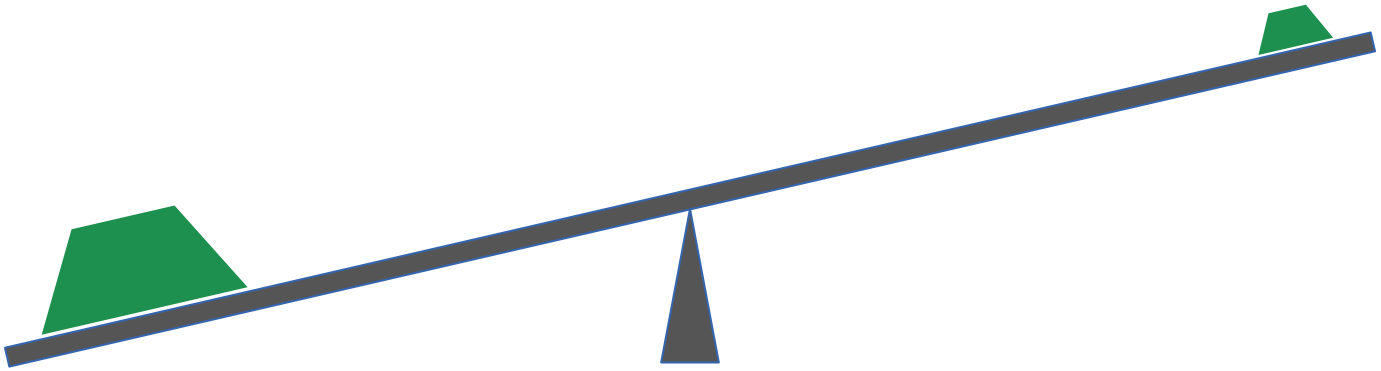


EEXCESS

Privacy vs. Personalisation trade-off?

Privacy

Personalisation/Quality



EEXCESS

Privacy vs. Personalisation trade-off?

User Awareness (and Transparency)

User Empowerment

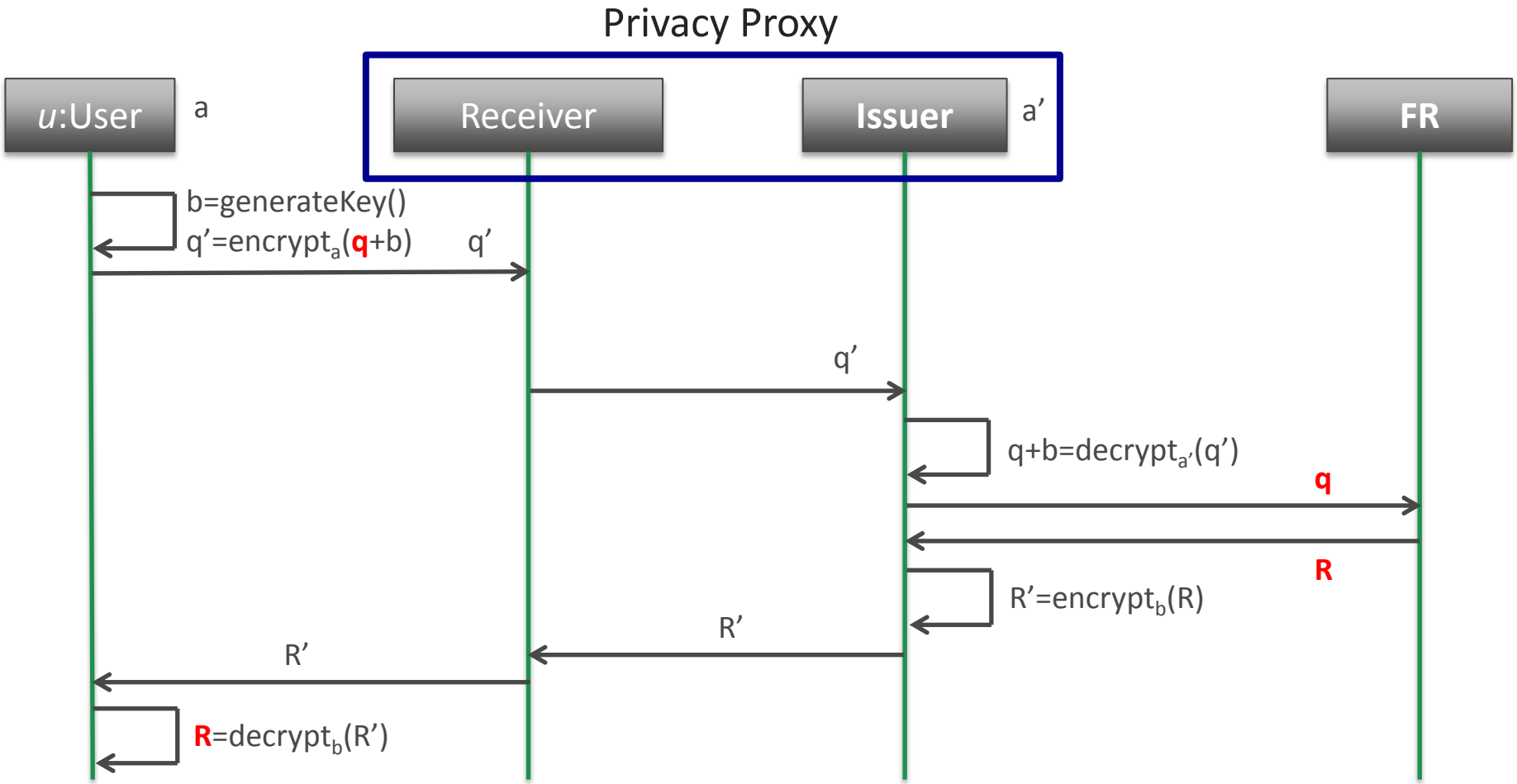
User Privacy Protection (Privacy Proxy)

EEXCESS

PEAS: Unlinkability Protocol

- PEAS: Private, Efficient, and Accurate web Search
- Hypothesis
  - only the user's device is trusted
- Split the Privacy Proxy into two pieces
  - Receiver: knows the user, but not the content of the query
  - Issuer: knows the content of the query, but not the user
  - Both are supposed "honest but curious" and do not collude

PEAS: Unlinkability Protocol (simplified)

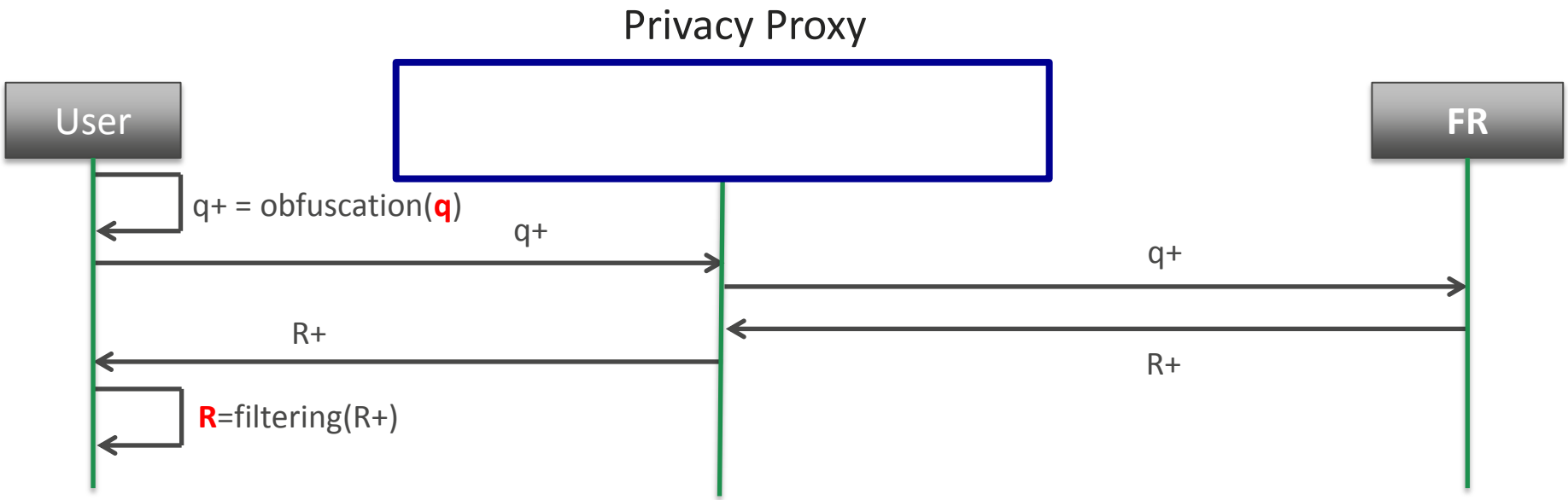


EEXCESS

PEAS: Indistinguishability Protocol (simplified)

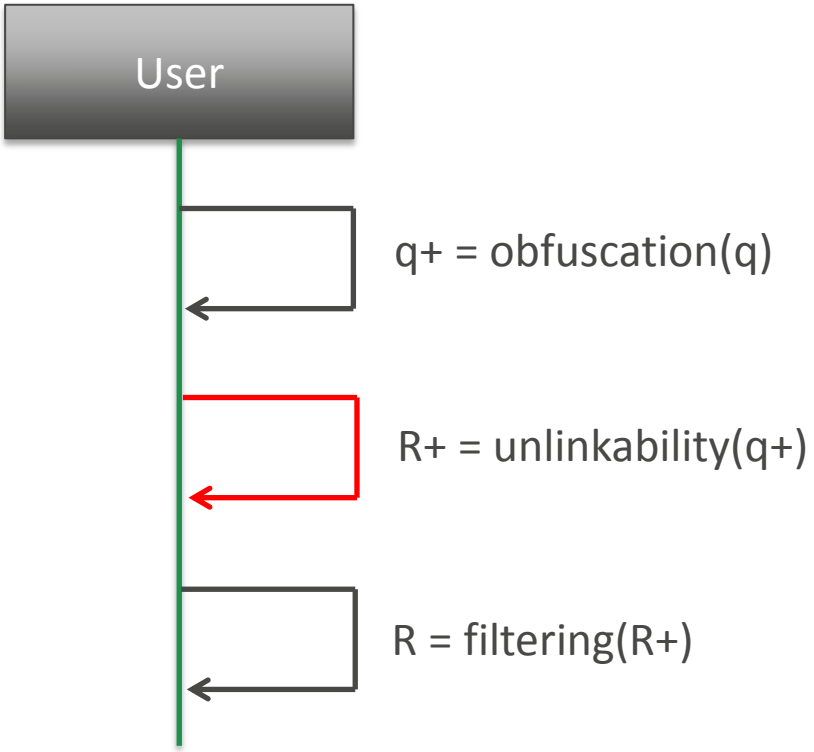
- Protocol divided into two parts
  - Obfuscation (done at the user's side): add fake queries
    - to mislead attackers, fake queries have the same structure as the original one, are built other users' queries, but are semantically different from the original query
  - Filtering: remove irrelevant results

PEAS: Indistinguishability Protocol (simplified)



EEXCESS

PEAS: Combination of Protocols



EEXCESS

Privacy Settings

- Transparent to user
- Choice which information to expose
- Choice to switch on/off different privacy features

EEXCESS

Information Society Technologies

Data Model

JOANNEUM RESEARCH

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

INSA LYON

wissenmedia

UNIVERSITÄT PASSAU

KNOW Center

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

bit media
member of the bit group

Data model

- Need to combine search results from different providers
- Perform duplicate removal, ranking
- Perform semantic enrichment
- Provide metadata in unified format to the client applications

EEXCESS Ontology

- Based on existing data models (EDM/PROV)
- Analysed data providers‘ formats
  - data providers investigated their data formats
  - identified overlaps and core metadata elements
- Defined EEXCESS Ontology
- Validated ontology by mapping data providers‘ formats

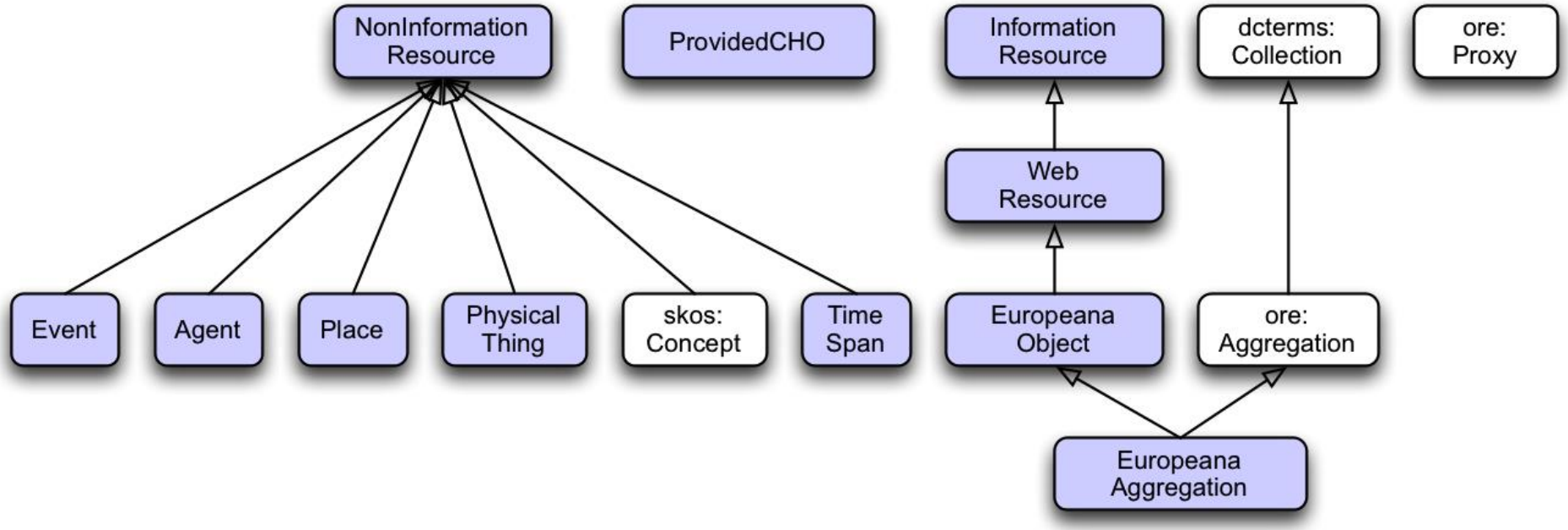
EEXCESS

EEXCESS Ontology

- Europeana Data Model - EDM
  - Represents metadata of cultural heritage objects (CHO)
  - **CHO**: real world resource
  - **Proxy**: representation CHO from one source
  - **Agent**: data provider
  - **Aggregation**: puts CHO, Agent and Proxy in relation
- EDM and EEXCESS
  - Objects are modeled as EDM CHOs
  - Annotations are modeled using EDM Proxies
  - Data providers are modeled as EDM Agents
  - Aggregation is used as in EDM

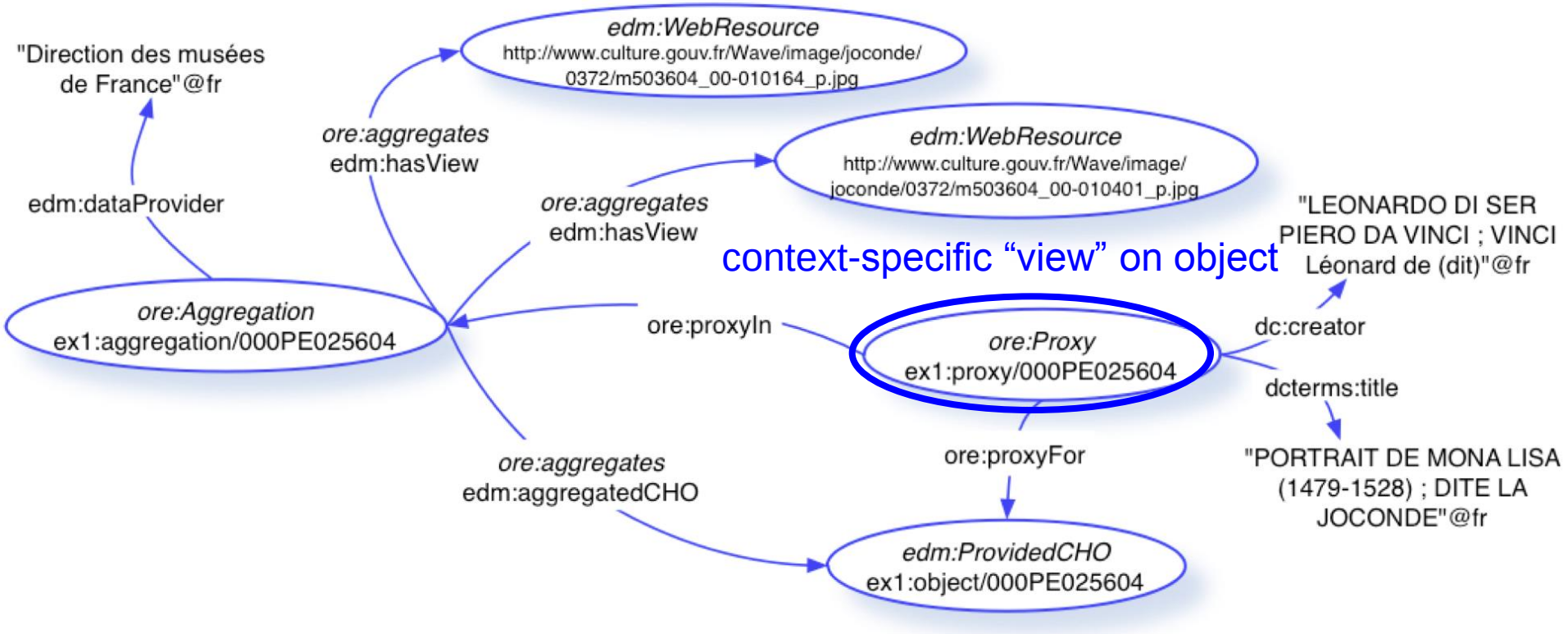
EEXCESS

EDM – Main entities



EEXCESS

EDM – Proxy example

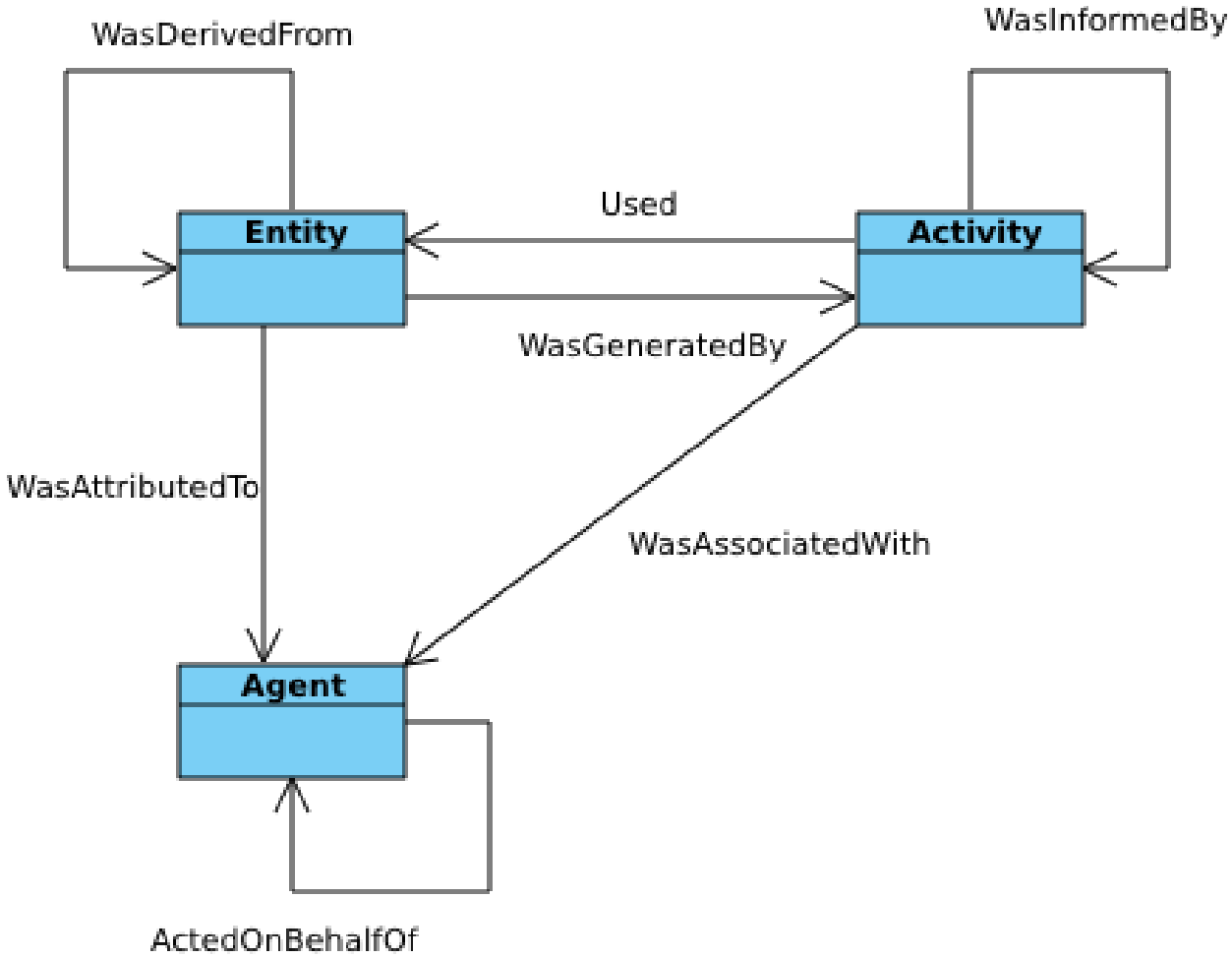


EEXCESS

EEXCESS Ontology

- W3C PROV
  - describes how things are created or delivered
  - **Entity**: physical, digital, conceptual, or other kinds of things
  - **Activity**: how entities are created or changed
  - **Agent**: takes a role in performing an activity
- PROV and EEXCESS
  - Objects and Proxies are modeled as PROV entities
  - Metadata creation is modeled as PROV activity
  - Creator of metadata is modeled as PROV agent

W3C PROV



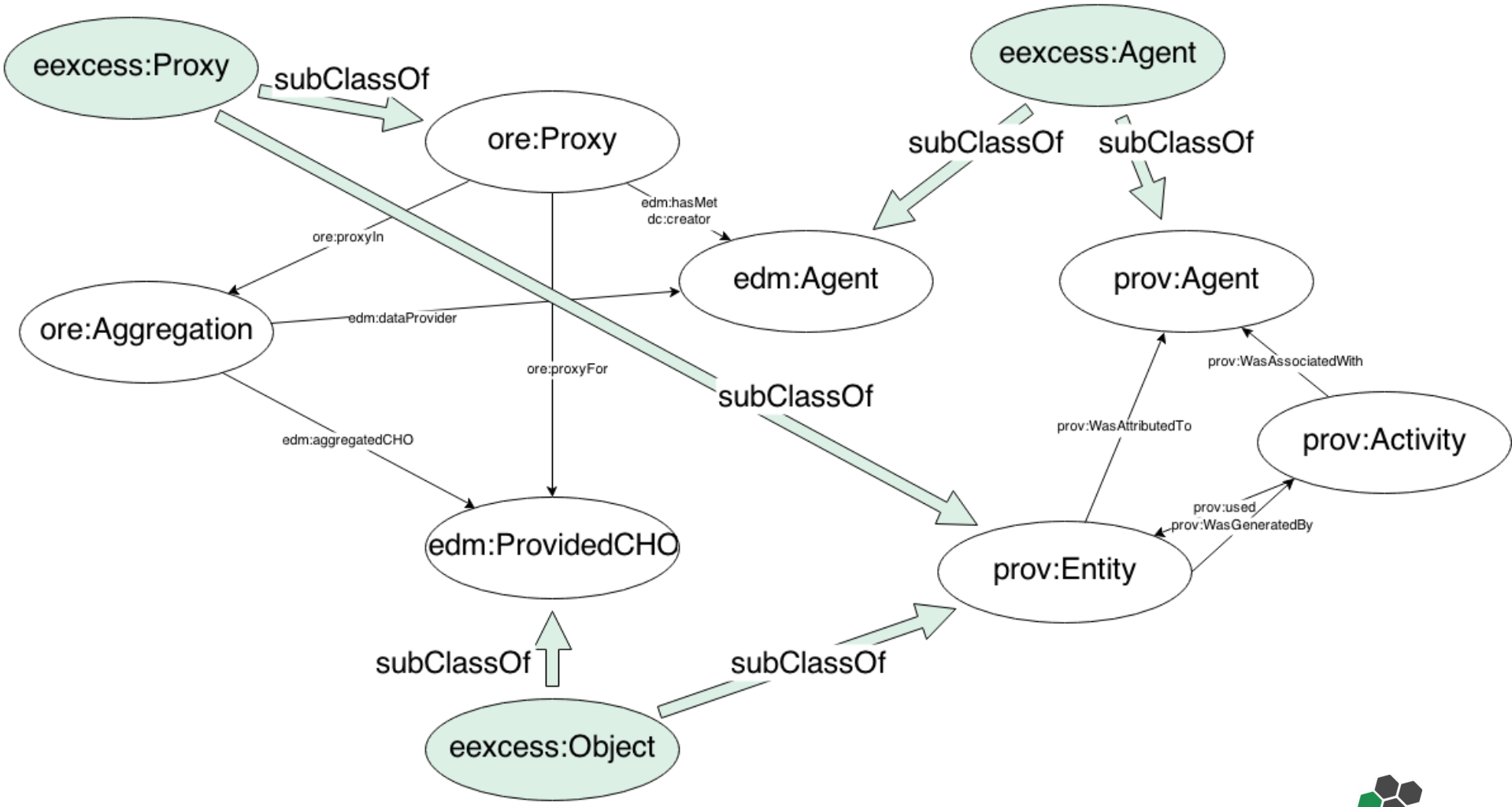
EEXCESS

EEXCESS Ontology

- eexcess:Object
  - Single item curated by a data provider
- eexcess:Agent
  - Data provider
  - Annotator of existing content
- eexcess:Proxy
  - Groups metadata from one source

EEXCESS

EEXCESS Ontology, EDM and W3C PROV



EEXCESS

Representation

- Serialisation
  - RDF/XML
  - JSON-LD
- Not stored, but exchanged between Partner Recommenders, Federated Recommender and clients

EEXCESS

EEXCESS

Information Society Technologies

PartnerWizard

JOANNEUM RESEARCH

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

INSA LYON

wissenmedia

UNIVERSITÄT PASSAU

KNOW Center

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

bit media
member of the bit group

Motivation

- Connect more data providers to the EEXCESS system
- Make it easy to achieve basic integration
- Allow setup without the need to write code
- Jump start software development by starting from a template

Overview

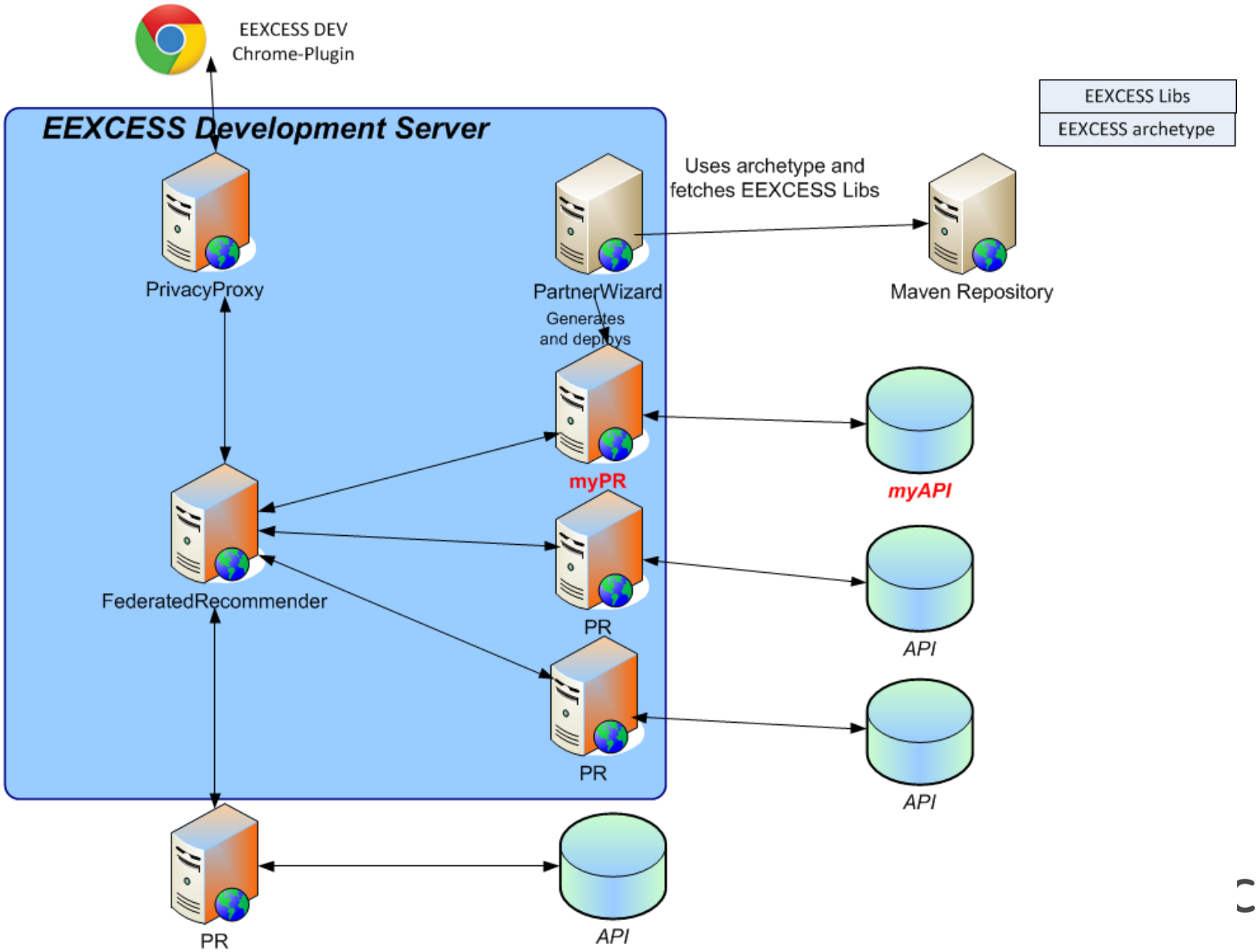
Build a new PartnerRecommender

- Create a new project
- Configure QueryGeneration, API-endpoints, …
- Implement special Classes e.g. QueryGeneration, Transformation,..
- Configure for EEXCESS-DEV-Server
- Deployment on local PC/Server
- New PartnerRecommender register on DEV-FederatedRecommender
- Download Chrome plugin from WebStore
- Configure Chrome plugin to EEXCESS-DEV-Server

User will see their data integrated in the Chrome plugin

EEXCESS

Architecture

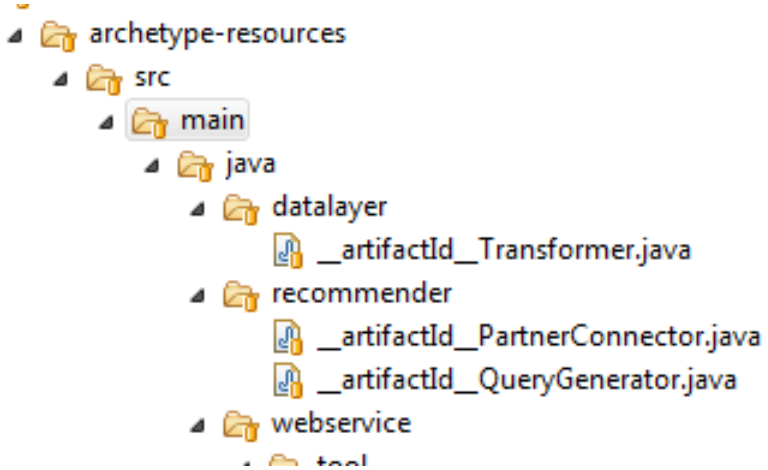


maven archetype

- Projects are built with maven
  - Defines dependencies incl. version of the lib
  - repositories
- maven archetype – project templating toolkit
- maven provides command to create an archetype from an existing project

maven archetype

- Existing PartnerRecommender as input
- Defining Parameters for the new archetype
- Replaced the specific code with placeholder




```
17  package ${package}.datalayer;
18
19⊕ import javax.xml.xpath.XPath;
36
37  public class ${artifactId}Transformer extends Transformer{
38
```


EEXCESS

maven archetype

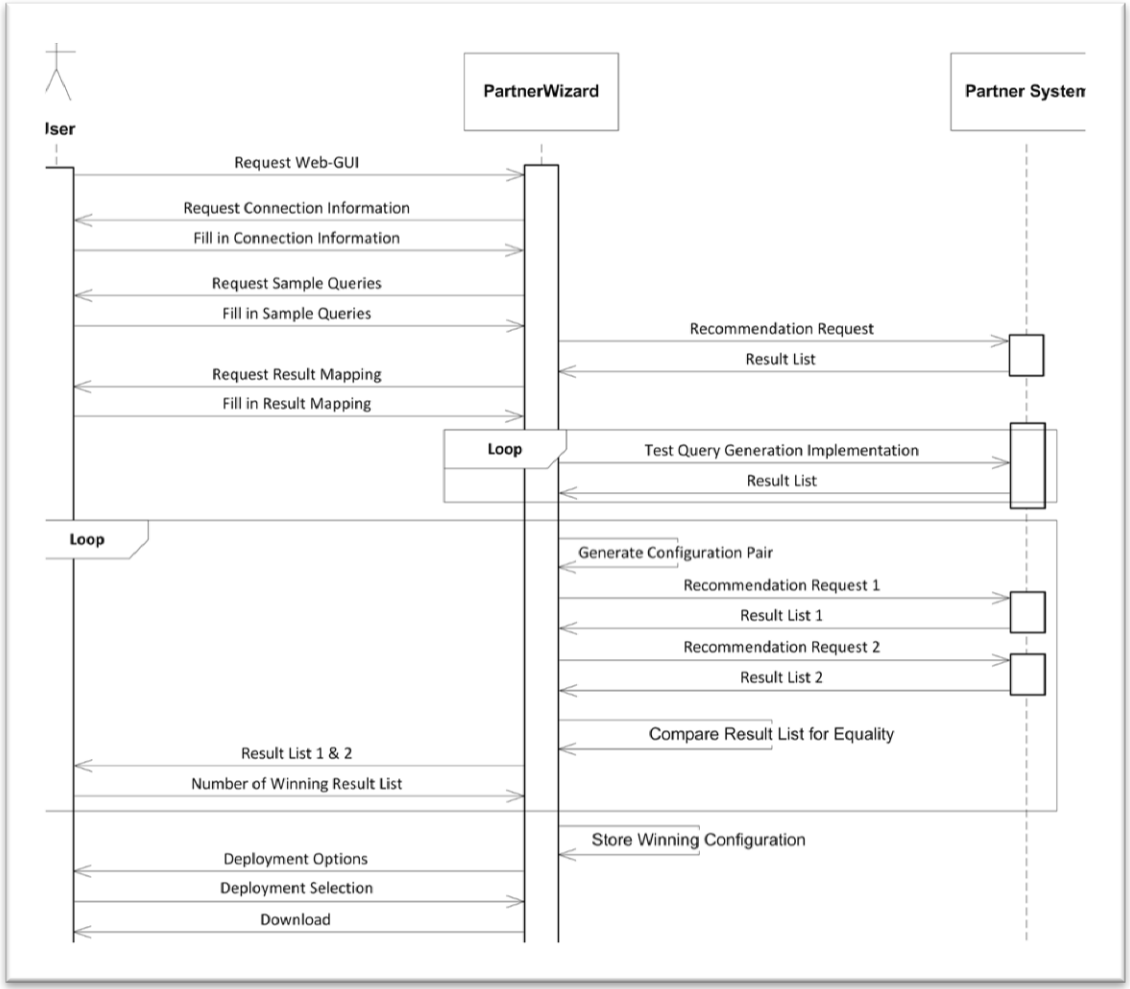
Parameters for maven archetype: EEXCESS archetype

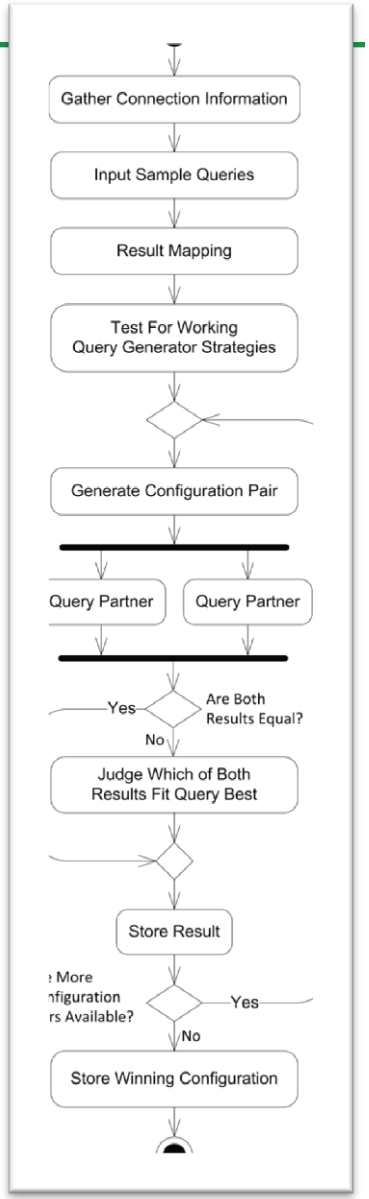
package=at.joanneum
version=0.1-SNAPSHOT
groupId=eu.eexcess
artifactId=myPRTest
partnerName=Partner Name
partnerURL=http://example.org/
dataLicense=unknown license
partnerAPIsearchEndpoint=https://kgapi.bl.ch/solr/kim-portal.objects/select/xml?q=_fulltext_:${query}&rows=${numResults}
partnerAPIsearchTerm=s
partnerAPIsearchMappingFieldsLoopXPath=/response/result/doc/
partnerAPIsearchMappingFieldsXPathID=str[@name='uuid']
partnerAPIsearchMappingFieldsXPathURI=str[@name='uuid']
partnerAPIsearchMappingFieldsXPathTitle=str[@name='_display_']
partnerAPIsearchMappingFieldsXPathDescription=str[@name='beschreibung']
partnerAPIdetailEndpoint=https://kgapi.bl.ch/solr/kim-portal.objects/select/xml?q=uuid:${detailQuery}
partnerAPIdetailTerm=s
partnerAPIdetailMappingFieldsLoopXPath=/response/result/doc/
partnerAPIdetailMappingFieldsXPathID=str[@name='uuid']
partnerAPIdetailMappingFieldsXPathURI=str[@name='uuid']
partnerAPIdetailMappingFieldsXPathTitle=str[@name='_display_']
partnerAPIdetailMappingFieldsXPathDescription=str[@name='beschreibung']

Query Optimiser

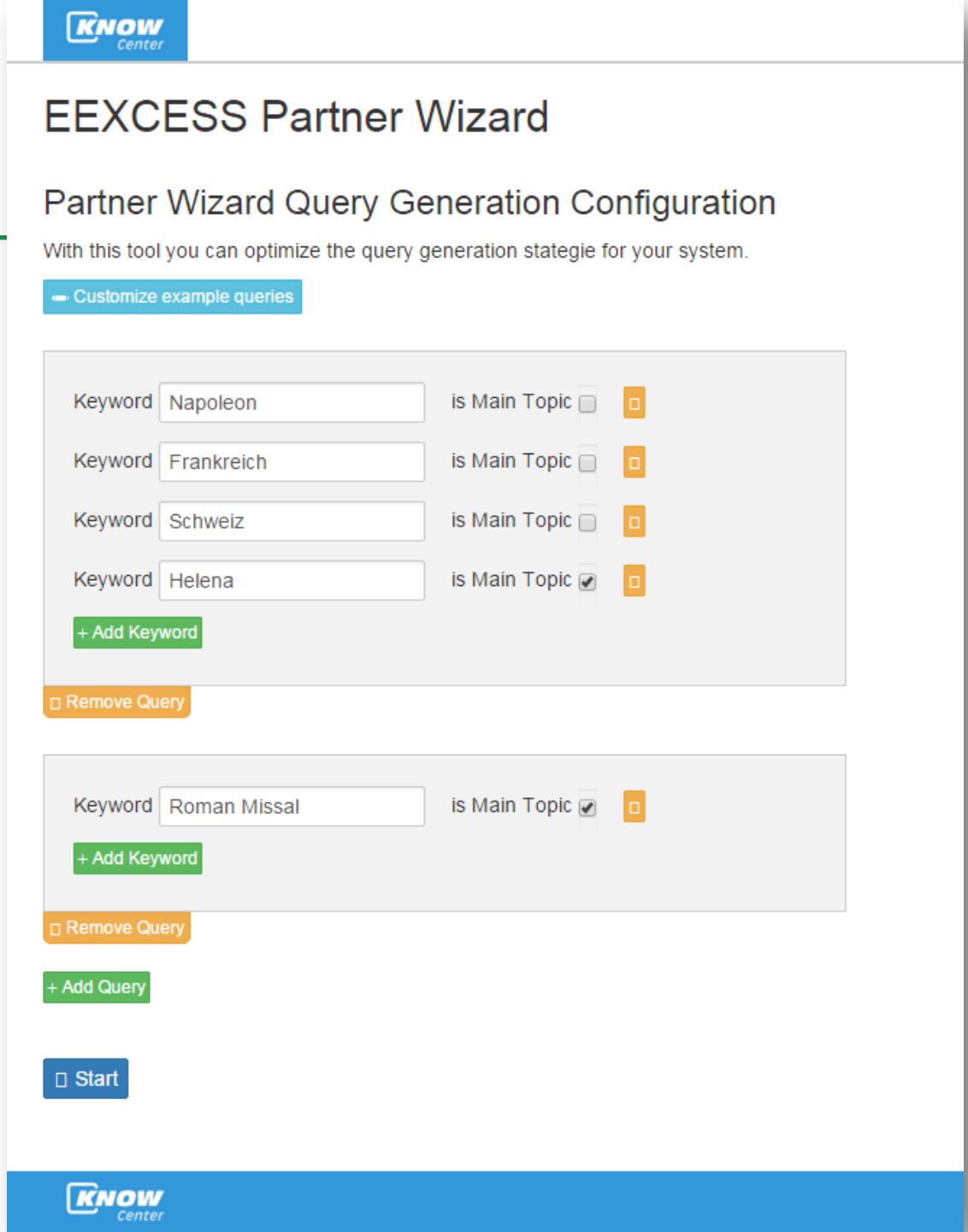
- Optimise query to partner sites
- Test different query options, e.g.
  - AND vs. OR of query terms
  - use of query expansion
- Expert selection from examples
- Automatically adjust query configuration of PartnerRecommender

Query Optimiser

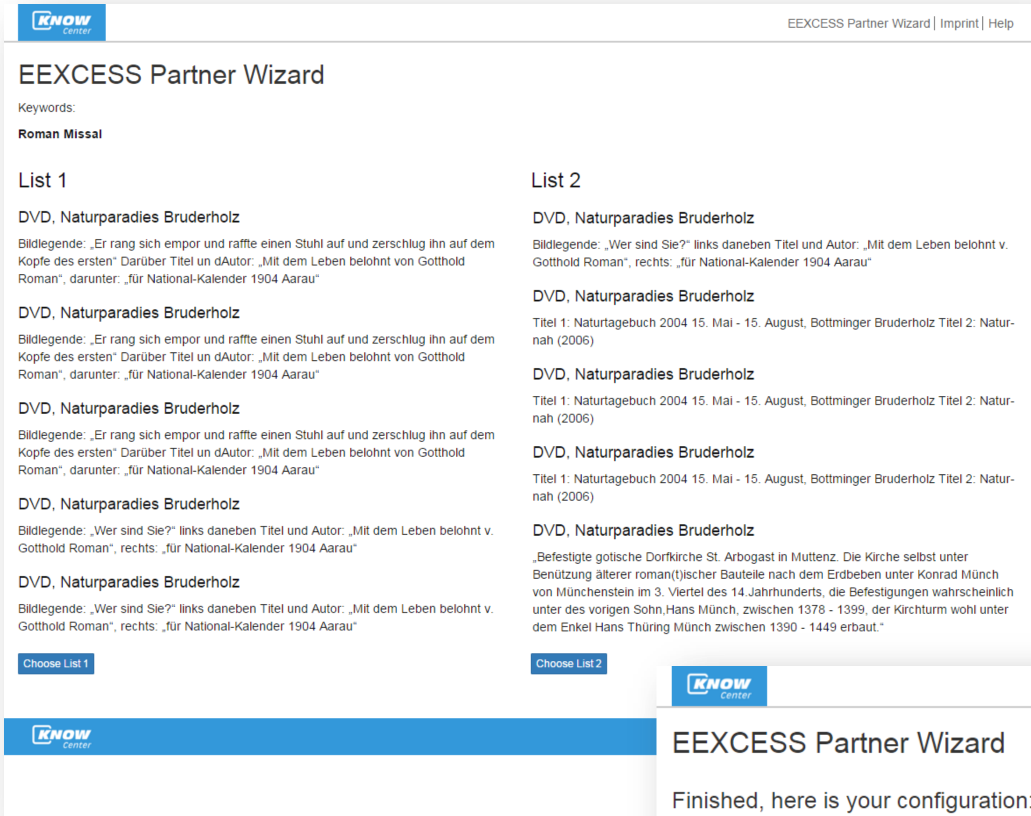


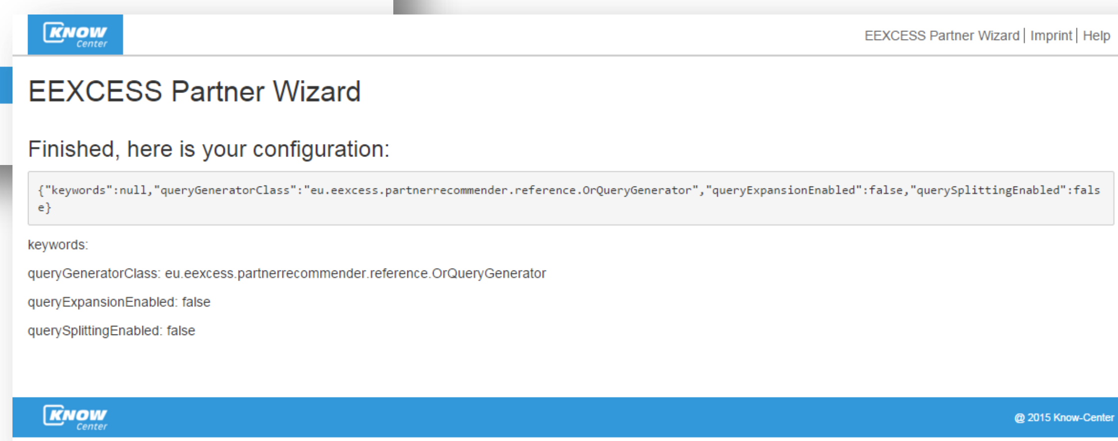


Query Optimiser



Query Optimiser





EEXCESS

Information Society Technologies

Metadata Mapping Configuration Tool

JOANNEUM RESEARCH

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

INSA LYON

wissenmedia

UNIVERSITÄT PASSAU

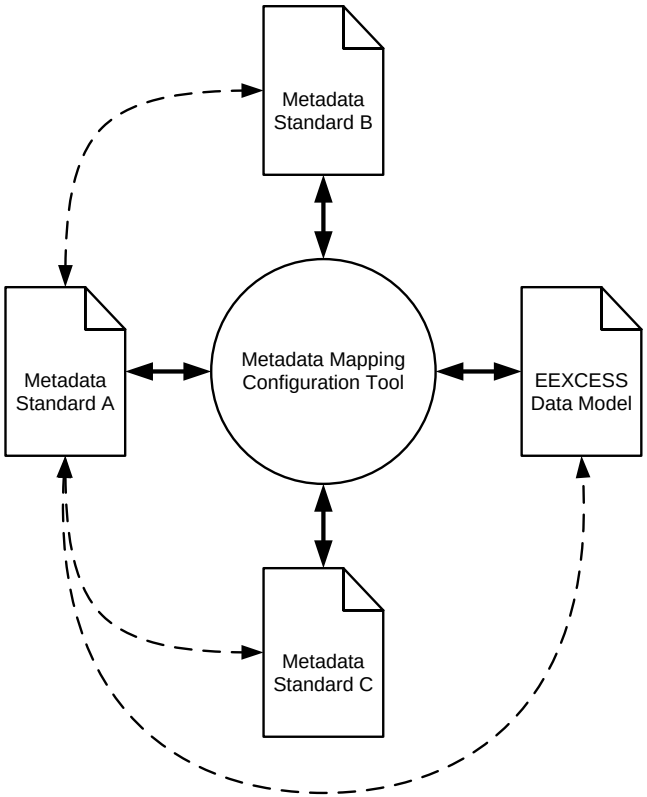
KNOW Center

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

bit media
member of the bit group

Motivation

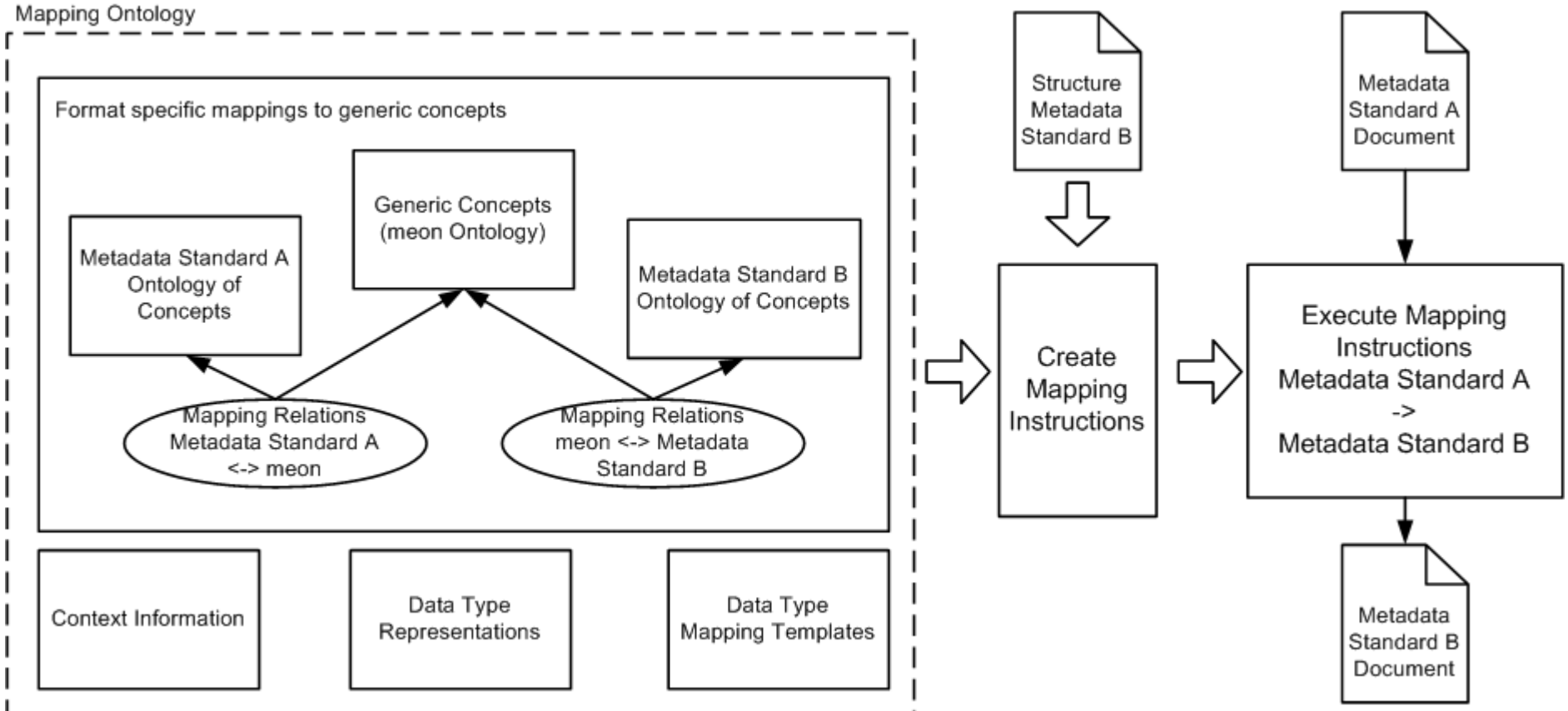
- Convert XML-based metadata documents between different metadata formats
  - Data providers' formats from and to the EEXCESS data model
- Define and configure mapping instructions
  - Avoid hand-crafted 1:1 mappings
  - Infer mapping instructions
  - Mappings are easier to maintain
  - Adding new metadata formats without side effects



EEXCESS

Metadata Mapping Configuration Approach

- Derive mapping instructions based on a mapping ontology



EEXCESS

Metadata Mapping Configuration Approach

- Mapping Ontology
  - Define mappings between metadata properties from different formats
  - Formalized with respect to on a conceptual representation of metadata properties serving as hub
  - Additional localization and context information
- Structural description of the target metadata format
- Result: XSL template

EEXCESS

Metadata Mapping Configuration Workflow

- Define format-specific metadata concepts
- Define mappings of the format-specific concepts to the conceptual representation
- Adding data type, localisation, structure information to format-specific concepts
- Create/edit structural representation of target format
- Create mapping instructions
  - Retrieve mapping parameters from mapping ontology
  - Merged into output structure

EEXCESS

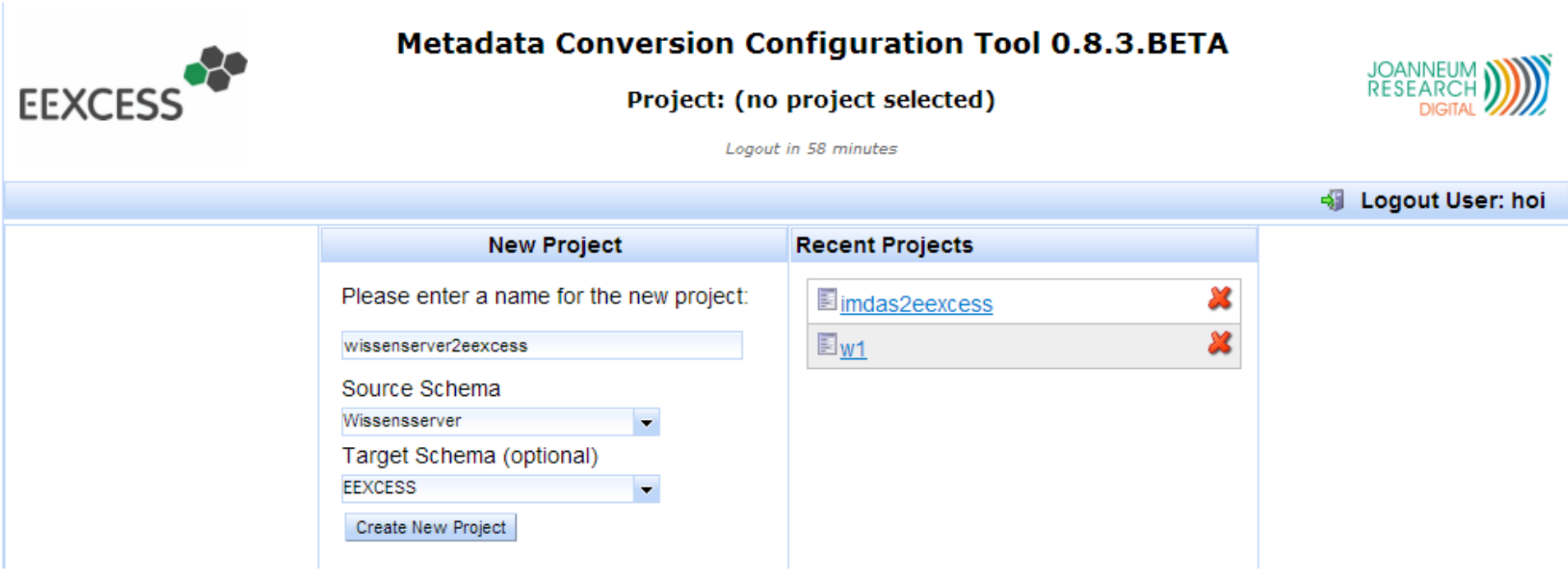
Metadata Mapping Configuration Tool

- Implemented as web application
- Configuration of metadata mapping
- Define relations between metadata fields by drag and drop
- Define data type mappings
- Define the output structure
- Preview of created mappings

EEXCESS

Metadata Mapping Configuration Tool

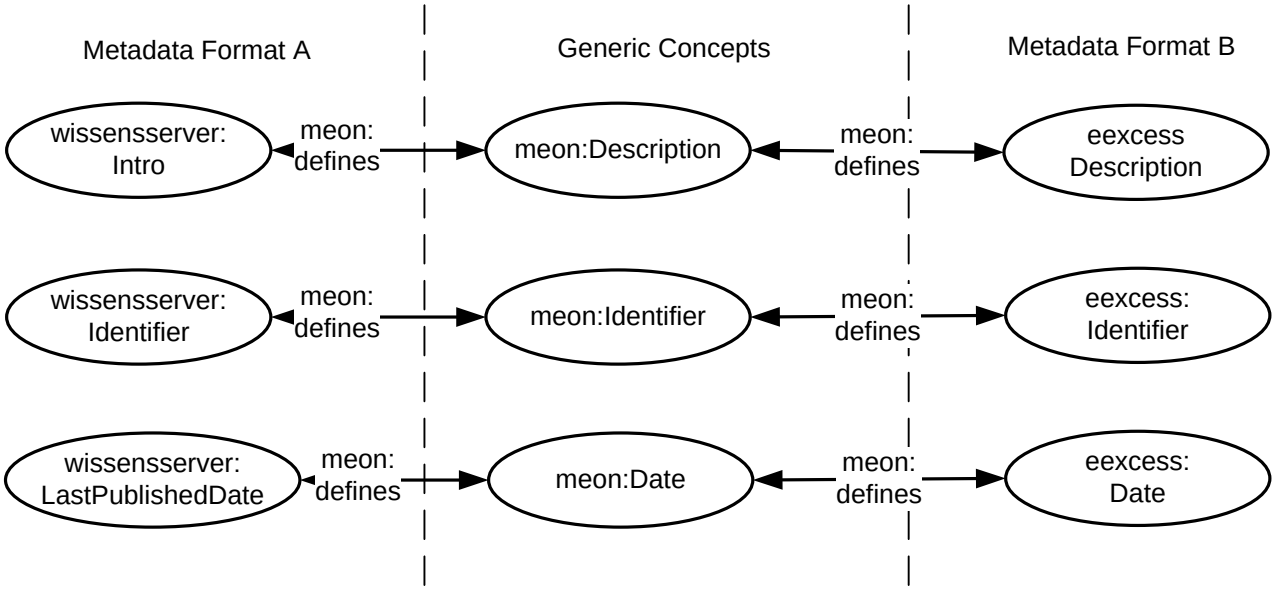
- Demo



EEXCESS

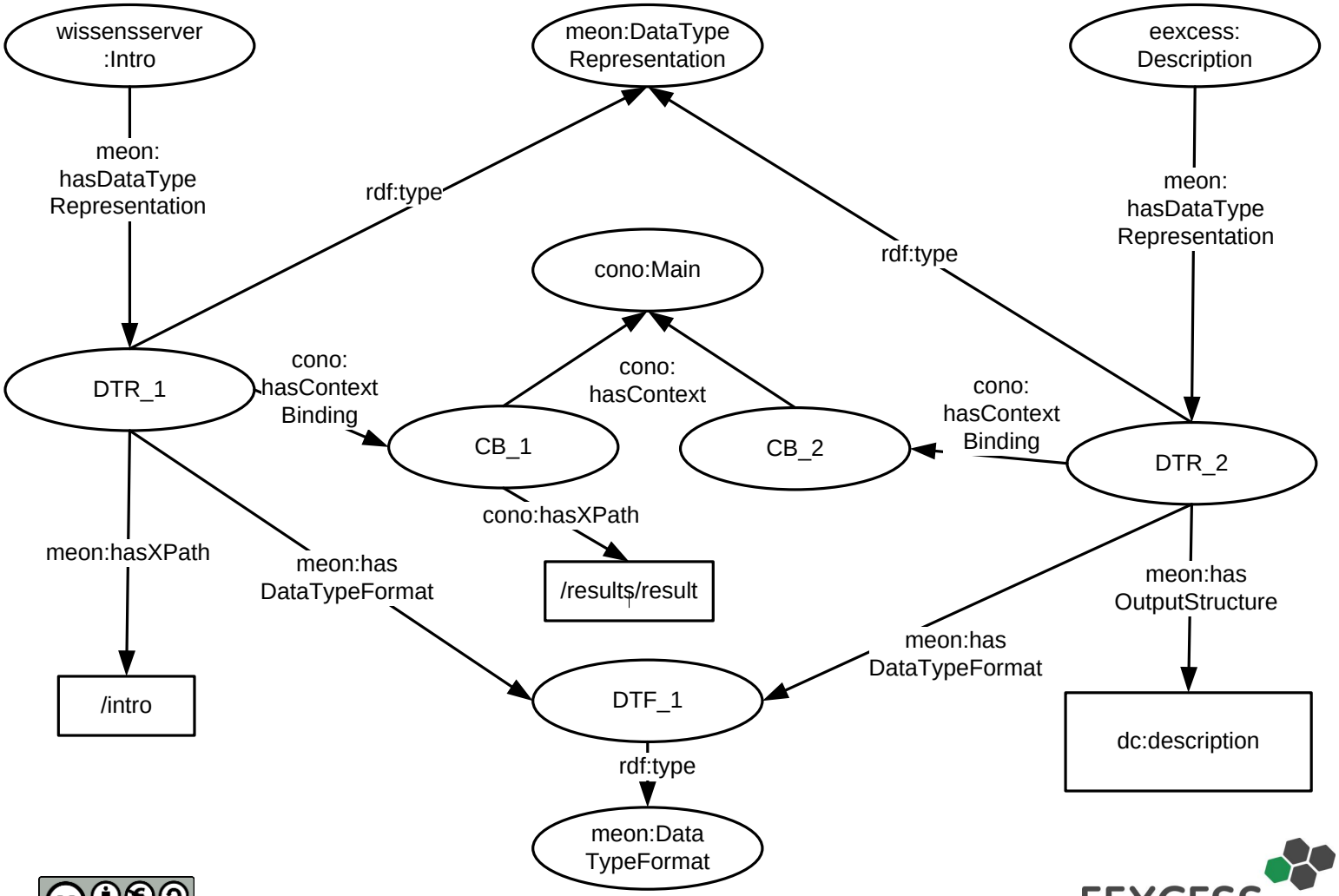
Metadata Mapping Configuration Workflow Concept Mappings

- based on meon ontology



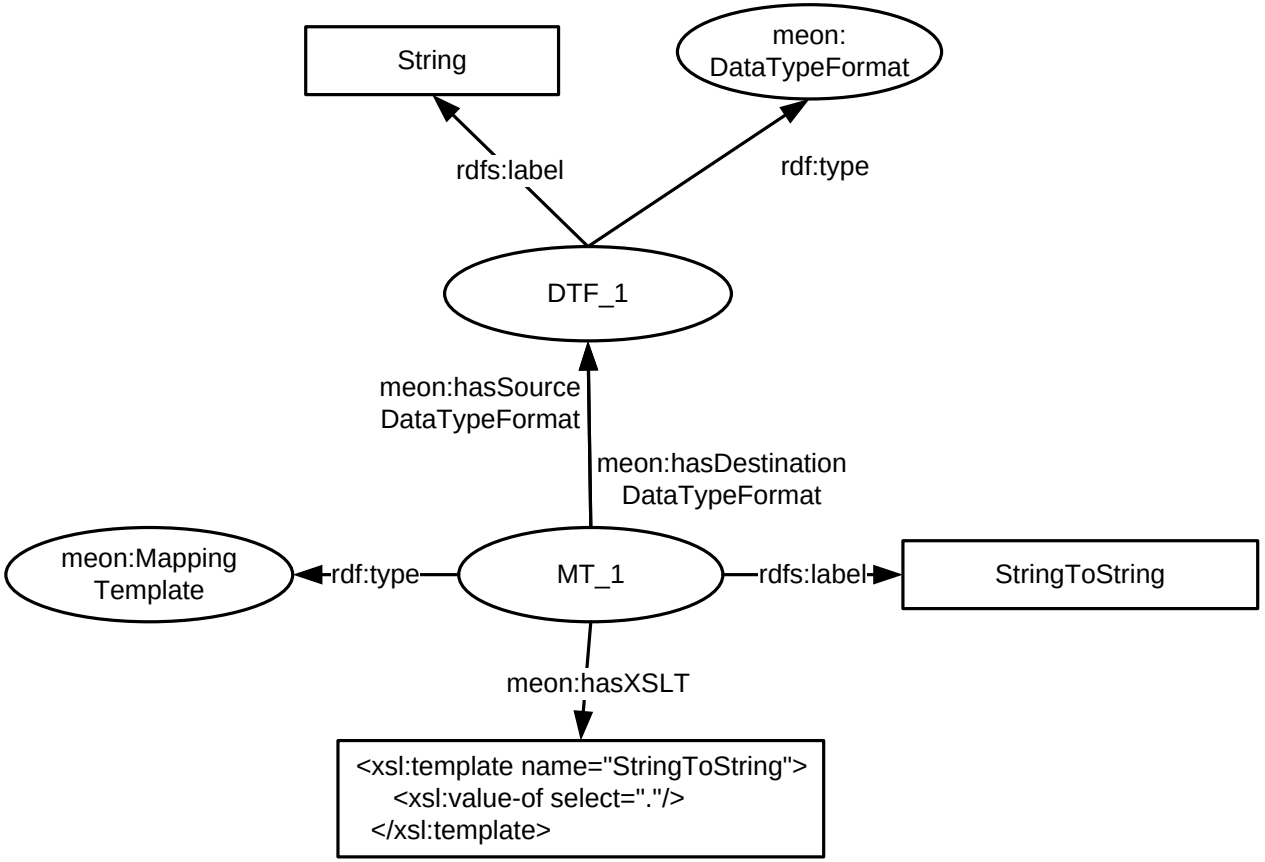
EEXCESS

Metadata Mapping Configuration Workflow Datatype Representations



EEXCESS

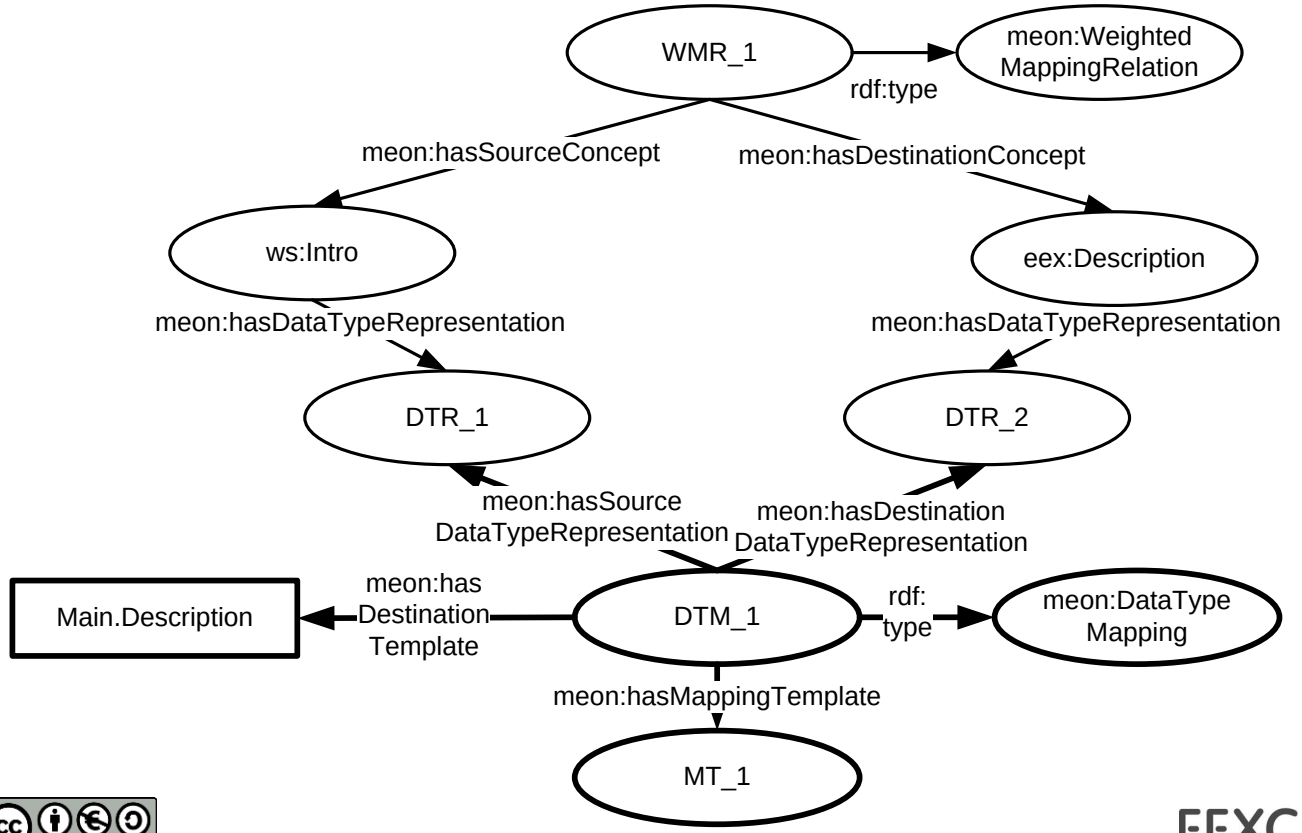
Metadata Mapping Configuration Workflow Mapping Template



EEXCESS

Metadata Mapping Configuration Workflow
Derive Mapping Parameters

- Mapping Parameters Inference



EEXCESS

Create Mapping Instructions Example

Output Structure:

```
<xsl:stylesheet>
  <xsl:element name="eexcess:Proxy">  …
    <xsl:call-template name="Main.Description"/>
…
</xsl:stylesheet>
```

Mapping Parameters:

Template Name: Main.Description

XPath: /intro

Output Structure: dc:Description

Mapping Template: StringToString

Mapping Instructions:

```
<xsl:template name="Main.Description">
  <apply-templates  select="intro"/>
</xsl:template>

<template  match="intro">
  <element name="dc:description">
    <call-template name="StringToString"/>
  </element>
</template>
```


EEXCESS

EEXCESS

SEVENTH FRAMEWORK PROGRAMME

Metadata Quality

JOANNEUM RESEARCH

Collections Trust

Archäologie und Museum
Bildungs-, Kultur- und Sportdirektion
Kanton Basel-Landschaft

MENDELEY

INSA LYON

wissenmedia

UNIVERSITÄT PASSAU

KNOW Center

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

bit media
member of the bit group

Motivation

- Metadata from many sources
- Heterogeneous formats
  (and thus conversions)
- Different workflows
- Context

Three subproblems

- Assessing Input Data Quality
- Assessing Enrichment Results
- Assessing Mapping Results

EEXCESS

Input data quality – metrics

- Statistics about input data
- Completeness of records
  - fields/record (min, max, average)
  - # empty fields/record
- Structuredness of data
  - for example the structuredness of date, name fields
  - Structured element or format specification (e.g. using XML Schema regular expressions)

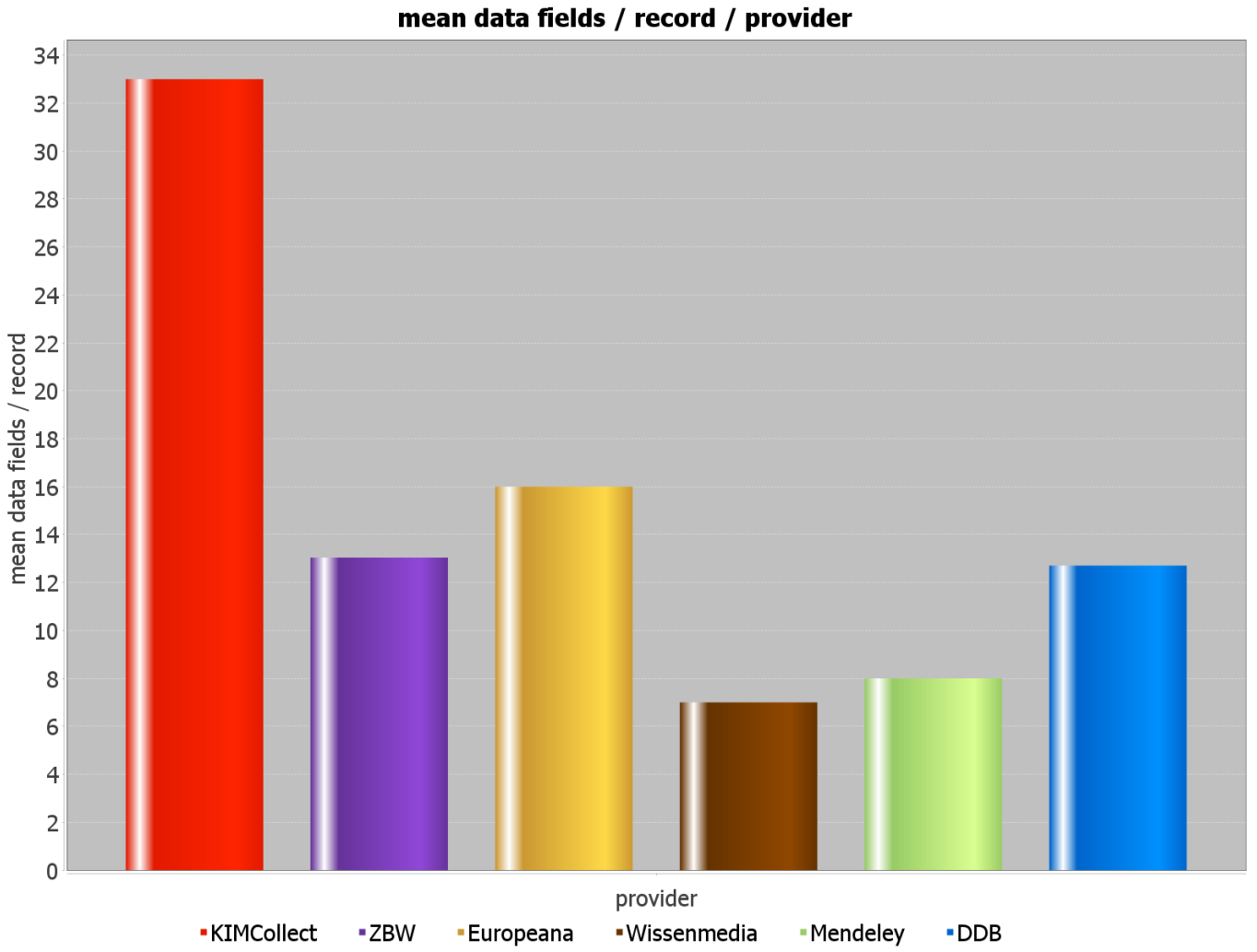
Input data quality – metrics

- Use of controlled vocabularies
- Availability of linked resources

- Evaluated on data collected during testbed on 6K records

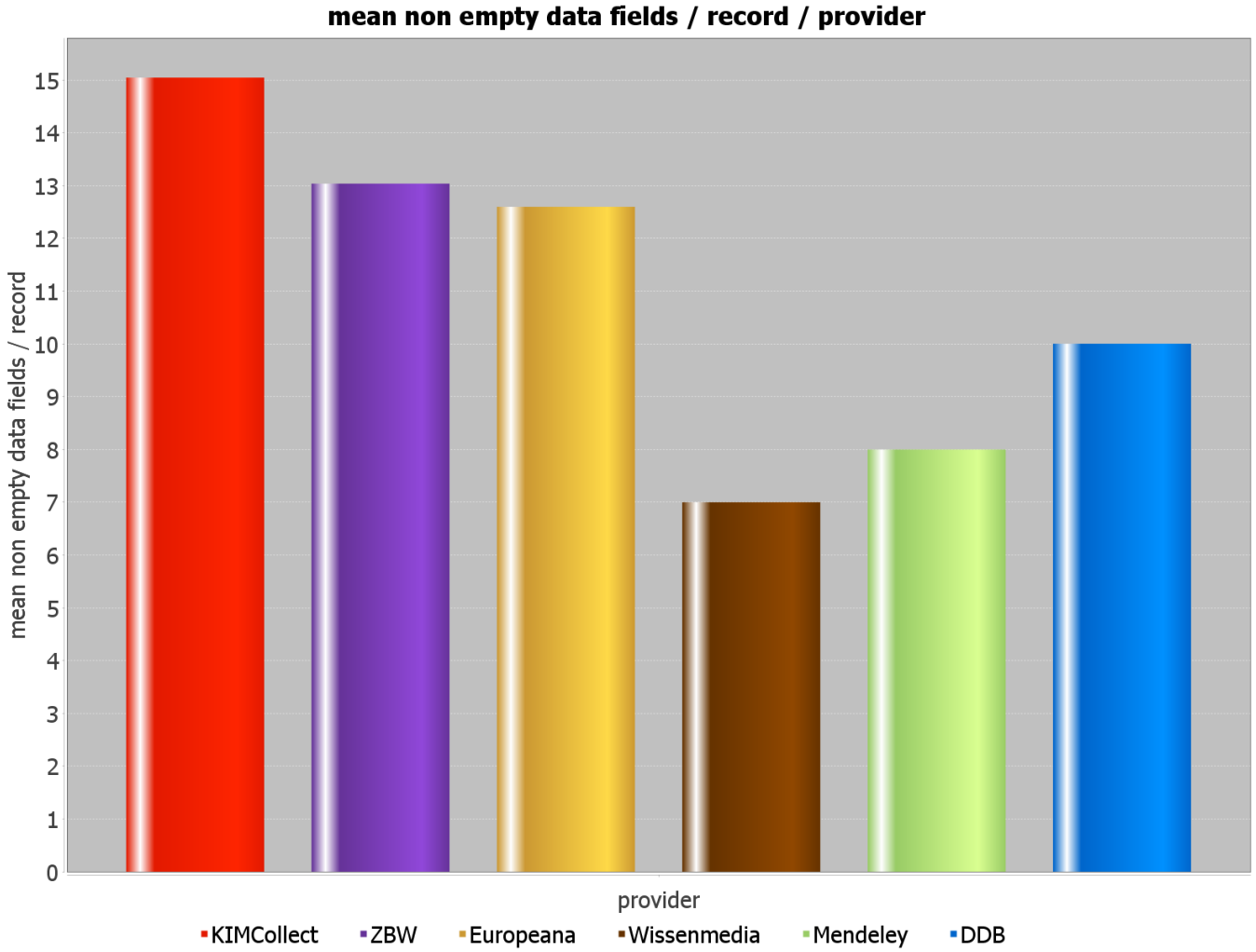
EEXCESS

Completeness



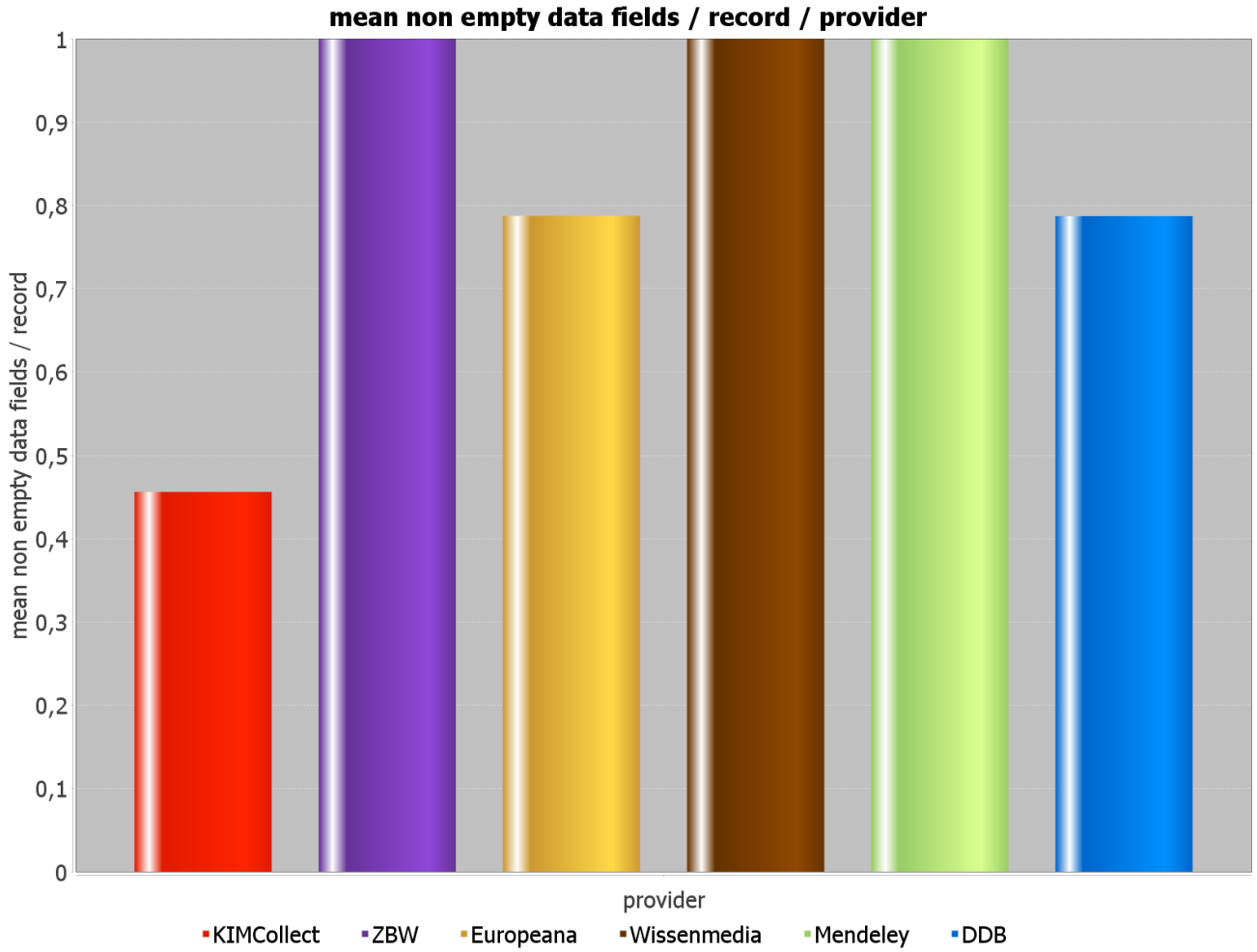
EEXCESS

Completeness



EEXCESS

Completeness



EEXCESS

Structuredness

- Length of value
  -> histogram
- Group characters and numbers
- Infer candidate patterns
  - e.g. Height: 00.0aa
    Width: 0.0aa
- Histogram of candidate patterns
- Detect known particles (e.g. SI unit abbreviations)

Time of origin	Start time of origin	End time of origin	Height	Width
1902	1902.0000	1902.0000	43.0cm	2.5cm
1868	1868.0000	1868.0000	35.0cm	1.7cm
2002			21.0cm	0.5cm
1904	1904.0000	1904.0000	47.0cm	2.7cm
1869	1869.0000	1869.0000	35.0cm	1.7cm
1870 - 1871	1870.0000	1871.0000	34.5cm	3.0cm
1872 - 1873	1872.0000	1873.0000	40.0cm	4.0cm
1874 - 1875	1874.0000	1875.0000	40.5cm	5.0cm
1876 - 1877	1876.0000	1877.0000	40.5cm	5.6cm
1878 - 1879	1878.0000	1879.0000	42.0cm	5.5cm
1880 - 1881	1880.0000	1881.0000	40.5cm	4.8cm
1882 - 1883	1882.0000	1883.0000	41.0cm	4.5cm
1884 - 1885	1884.0000	1885.0000	40.5cm	5.5cm
1886 - 1887	1886.0000	1887.0000	41.0cm	5.0cm
1888 - 1889	1888.0000	1889.0000	41.5cm	5.0cm
1890 - 1891	1890.0000	1891.0000	44.0cm	6.0cm
1892	1892.0000	1892.0000	44.3cm	2.5cm
1893	1893.0000	1893.0000	43.8cm	2.5cm

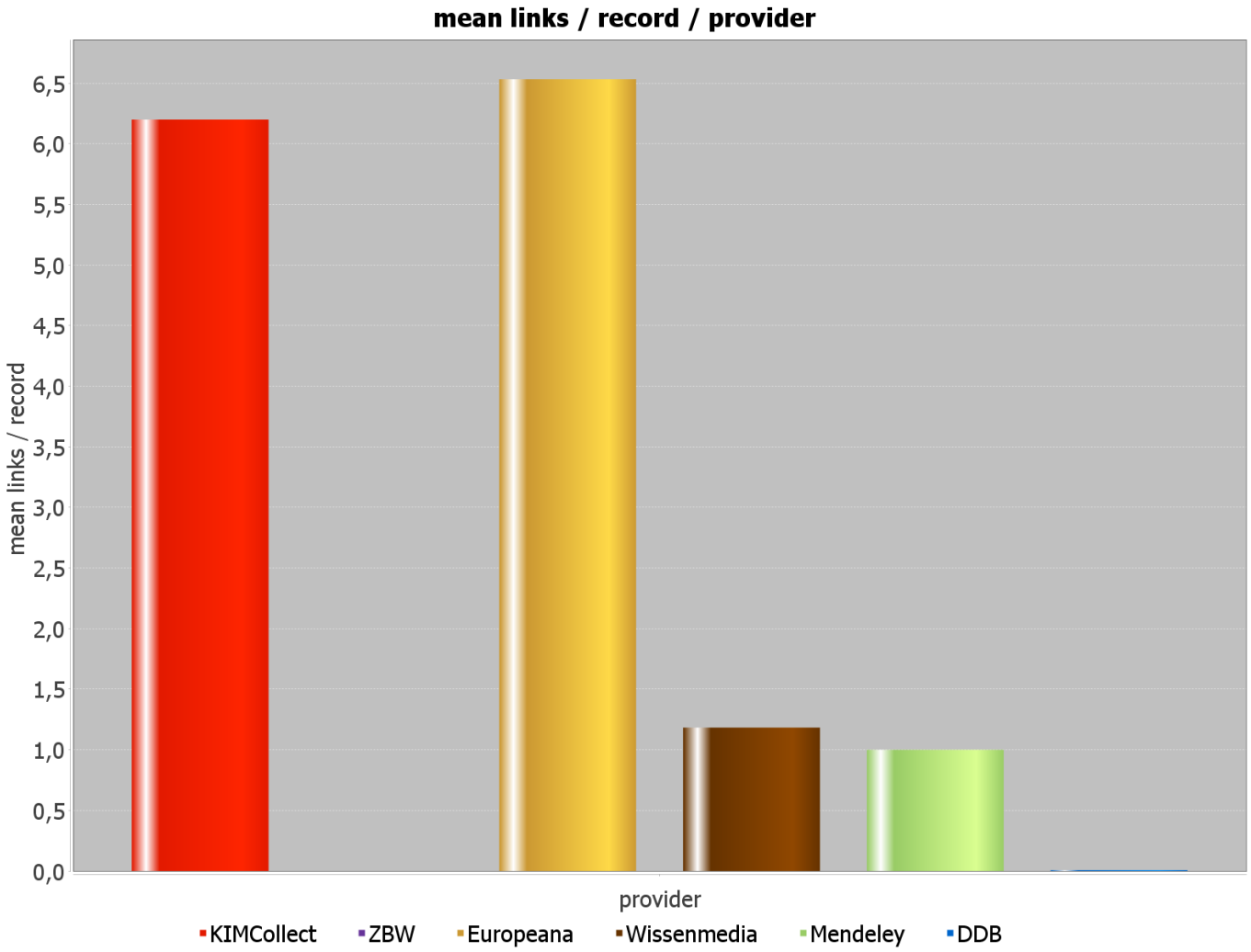
EEXCESS

URLs in record

- Counting URLs in responses
- Check if URL accessible
- Check type of response
  - XML/RDF, XML, HTML
  - determine if result is machine readable

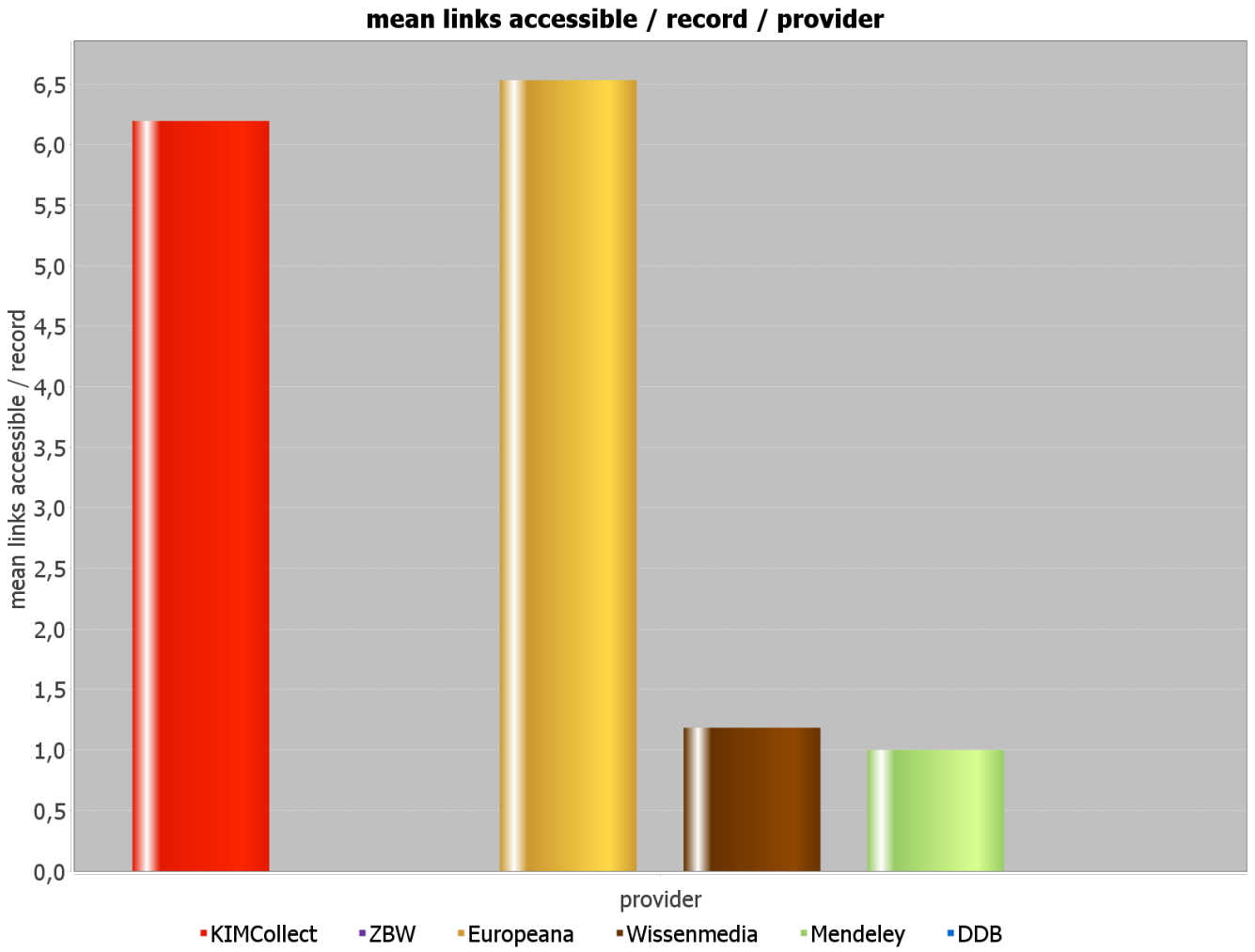
EEXCESS

URLs used in records



EEXCESS

URLs used in records (resolvable)



EEXCESS

Enriching and transforming data

- Apply the same metrics before and after transformation or enrichment
- Compare values, e.g.
  - decrease in number of empty fields
  - increase in use of controlled vocabularies
  - Increase in resolvable URLs in the data

EEXCESS

Use of input metadata quality results

- Statistics, completeness, etc.
  - Provide feedback to data provider
  - Improve result reprensentation returned by data providers
- Structuredness
  - More appropriate mapping
  - Detect outliers on the fly (avoid errors)

EEXCESS

Use of input metadata quality results

- Use of controlled vocabularies
  - Need for detecting/replacing named entities
  - Detect need to map vocabulary (to a standard and/or accessible one)

EEXCESS

Mapping Quality Assessment

- Assessment of mapping results
  - Comparison against an expert created reference
  - Round trip mapping via intermediate format
    - e.g., ZBW -> MEON -> ZBW
    - no expected loss
  - Round trip mapping via target format
    - e.g., ZBW -> EEXCESS -> ZBW
    - possibly expected loss

EEXCESS

Mapping Quality Assessment

Add Locator Association

Edit Locator Association

	Input XPath	Output XPath
Edit Delete	/objects/object/Hersteller	/objects/object/Hersteller
Edit Delete	/objects/object/Autor	/objects/object/Hersteller
Edit Delete	/objects/object/Fotograf	/objects/object/Hersteller

Recent XML Data Files

k1.xml (167B)

Mapping Quality Assessment

Input XPath	Output XPath	Result
/objects[1]/object[1]/Hersteller[1]	/objects[1]/object[1]/Hersteller[1]	Mapping ok
/objects[1]/object[1]/Autor[1]	/objects[1]/object[1]/Hersteller[2]	Mapping ok
/objects[1]/object[1]/Fotograf[1]	/objects[1]/object[1]/Hersteller	Mapping missing

EEXCESS

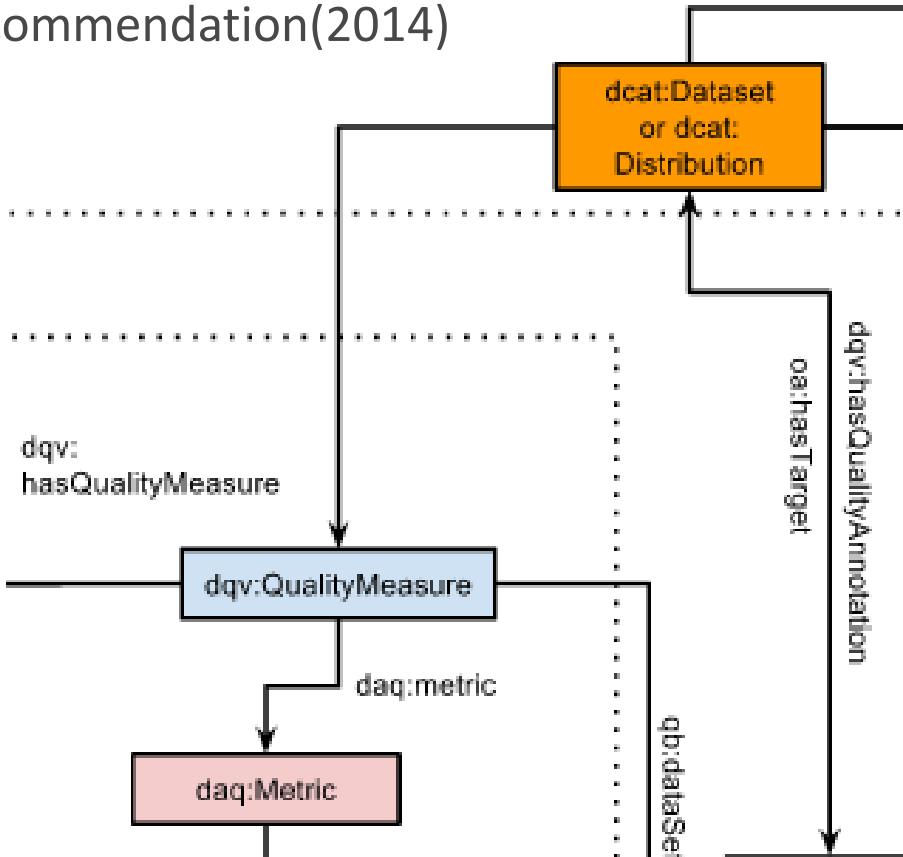
Data Quality Assessment – Result Representation

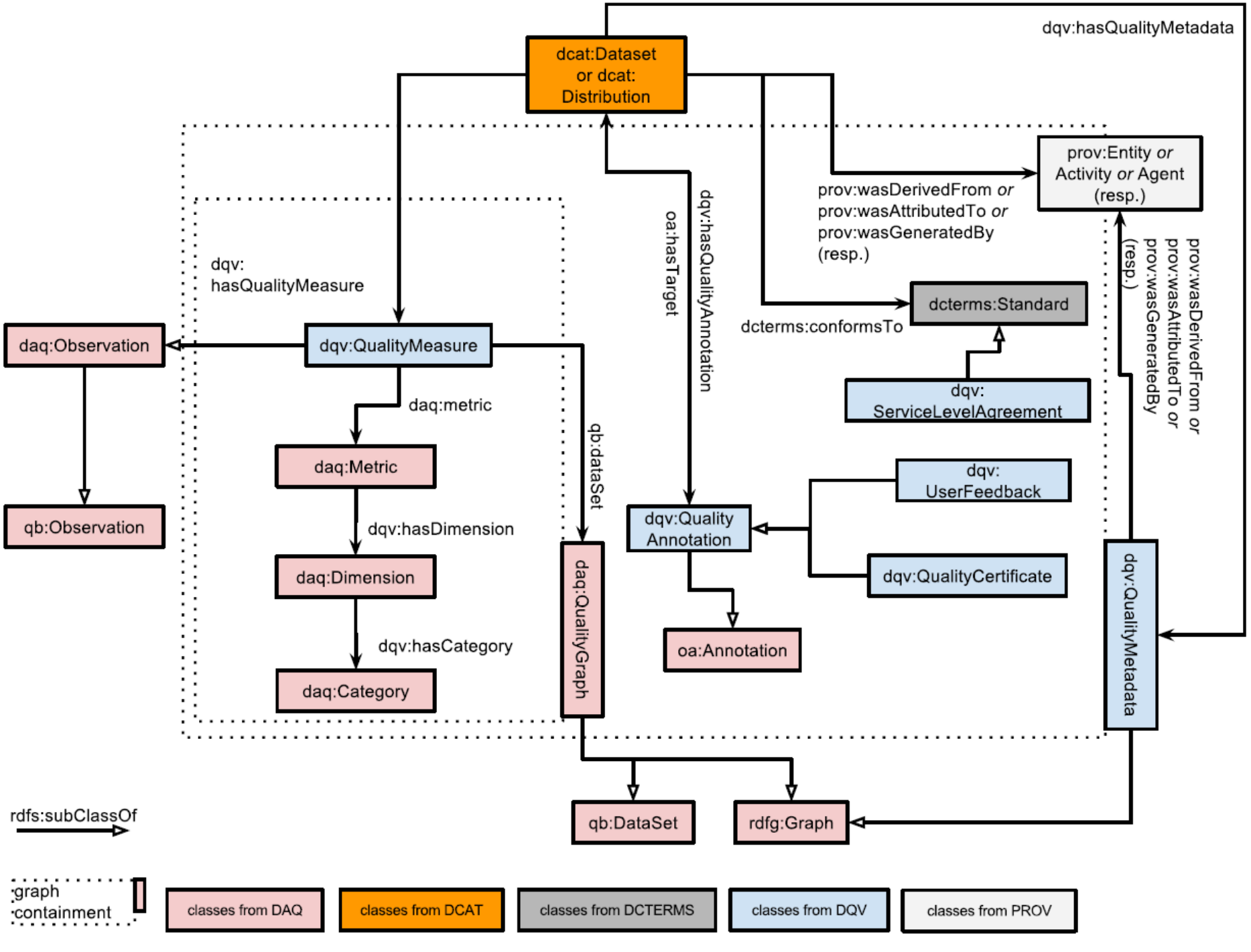
- Requirements
  - Well-defined
  - Structured
  - Machine-readable

EEXCESS

Data Quality Assessment – Result Representation

- W3C Data Quality Vocabulary (DQV) - First Public Working Draft 25 June 2015 http://www.w3.org/TR/2015/WD-vocab-dqv-20150625/
  - Data Catalog Vocabulary(DCAT) – Recommendation(2014)
- Dataset(DCAT)
- Distribution(DCAT)
- Metric(DQV)
- QualityMeasure(DQV)




dqv:hasQualityMetadata
dcat:Dataset or dcat: Distribution
prov:Entity or Activity or Agent (resp.)
prov:wasDerivedFrom or prov:wasAttributedTo or prov:wasGeneratedBy (resp.)
dqv: hasQualityMeasure
oa:hasTarget
dqv:hasQualityAnnotation
dcterms:Standard
dcterms:conformsTo
daq:Observation
dqv:QualityMeasure
daq:metric
dqv: ServiceLevelAgreement
prov:wasDerivedFrom or prov:wasAttributedTo or prov:wasGeneratedBy (resp.)
daq:Metric
qb:dataSet
dqv: UserFeedback
qb:Observation
dqv:hasDimension
dqv:Quality Annotation
daq:QualityGraph
dqv:QualityMetadata
daq:Dimension
dqv:QualityCertificate
dqv:hasCategory
oa:Annotation
daq:Category
qb:DataSet
rdfg:Graph
rdfs:subClassOf
graph containment
classes from DAQ
classes from DCAT
classes from DCTERMS
classes from DQV
classes from PROV

Data Quality Assessment – Result Representation

```
<dcat:Dataset rdf:about="#eexcessDataset">
    <dct:title>My EEXCESS dataset</dct:title>
    <dcat:distribution>
        <dcat:Distribution rdf:about="#eexcessDatasetZBWDistribution">
            <dct:title>My EEXCESS ZBW dataset</dct:title>
            <prov:wasGeneratedBy rdf:resource="#ZBW"/>
        </dcat:Distribution>
    </dcat:distribution>
    <dcat:distribution>
        <dcat:Distribution rdf:about="#eexcessDatasetZBWTransformationDistribution">
            <dct:title>My EEXCESS ZBW Transformation dataset</dct:title>
            <prov:wasGeneratedBy rdf:resource="#EEXCESSTransformation"/>
            <prov:wasDerivedFrom rdf:resource="#eexcessDatasetZBWDistribution"/>
        </dcat:Distribution>
    </dcat:distribution>
    <dcat:distribution>
        <dcat:Distribution rdf:about="#eexcessDatasetZBWEnrichmentDistribution">
            <dct:title>My EEXCESS ZBW Enrichment dataset</dct:title>
            <prov:wasGeneratedBy rdf:resource="#EEXCESSEnrichment"/>
            <prov:wasDerivedFrom rdf:resource="#eexcessDatasetZBWTransformationDistribution"/>
        </dcat:Distribution>
    </dcat:distribution>
</dcat:Dataset>
```


EEXCESS

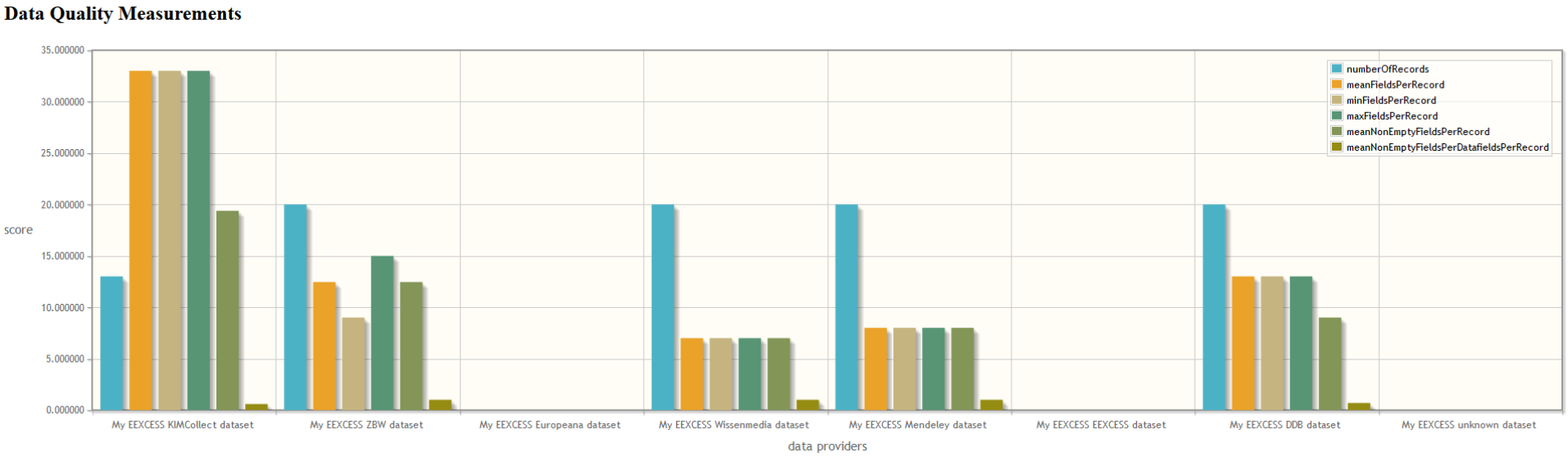
Data Quality Assessment – Result Representation

```
<daq:Metric rdf:about="#eexcessDataQMetricNumberOfRecords">
</daq:Metric>
<daq:Metric rdf:about="#eexcessDataQMetricNumberOfFields">
</daq:Metric>

<dqv:QualityMeasure rdf:about="#measureNumberOfRecordsZBW">
        <daq:value rdf:datatype="http://www.w3.org/2001/XMLSchemadouble">102</daq:value>
        <daq:computedOn rdf:resource="#eexcessDatasetZBWDistribution"/>
        <daq:metric rdf:resource="#eexcessDataQMetricNumberOfRecords"/>
</dqv:QualityMeasure>
<dqv:QualityMeasure rdf:about="#measureNumberOfFieldsZBW">
        <daq:value rdf:datatype="http://www.w3.org/2001/XMLSchemadouble">10</daq:value>
        <daq:computedOn rdf:resource="#eexcessDatasetZBWDistribution"/>
        <daq:metric rdf:resource="#eexcessDataQMetricNumberOfFields"/>
</dqv:QualityMeasure>

<dqv:QualityMeasure rdf:about="#measureNumberOfFieldsZBWAfterTransformation">
        <daq:value rdf:datatype="http://www.w3.org/2001/XMLSchemadouble">10</daq:value>
        <daq:computedOn rdf:resource="#eexcessDatasetZBWTransformation"/>
        <daq:metric rdf:resource="#eexcessDataQMetricNumberOfFields"/>
</dqv:QualityMeasure>
```


EEXCESS

Visualisation from DQV

- Generate diagrams using XSLT



EEXCESS