



PAGODE

FINAL CONFERENCE

23 SEPTEMBER

Automatic Enrichment

Mina Karatza
Vassilis Tzouvaras

**post
scriptum**



Automatic Enrichment

Activity 4: Content Quality Improvement of existing collections

ARTICLE I.3 – ACTIVITIES

I.3.1 Activities timetable

Activity number	Activity title	Indicative start date	Indicative end date	Milestone number
1	Project Management	01/04/2020	30/09/2021	4, 12
2	Content's digitisation, annotation and ingestion to Europeana	01/04/2020	30/09/2021	5, 6, 7, 9
3	Content's curation	01/06/2020	31/07/2021	2
4	Content Quality and Europeana	01/05/2020	31/08/2021	3, 10, 11

Metadata Curation and Enrichment Scenarios



scenario 1

metadata creation/update realised by content providers at the time of generation of the digitisation

scenario 2

metadata enriched via crowdsourcing annotation

scenario 3

a combination of AI and NLP technologies

- list of keywords/the semantic background
- enrich the metadata with wikidata and Getty URIs

Selected Content

Content Providers	No. of Records
Europeana Fashion	981
English metadata, excluding partners and Wellcome Collection, Rijksmuseum	6,104
Rijksmuseum, Wellcome Collection	7,722
Deutsche Digitale Bibliothek	3,771
Östasiatiska museet- Collection Photography	7,007
TOTAL No. or Records	25,590

Target and Results

31/08/2021 Report on improved thematic collections confirms the successful enrichment of metadata of 20,000 objects.



**Enriched
Records**

22.411

SAGE in PAGODE: automatic semantic enrichments of EDM records

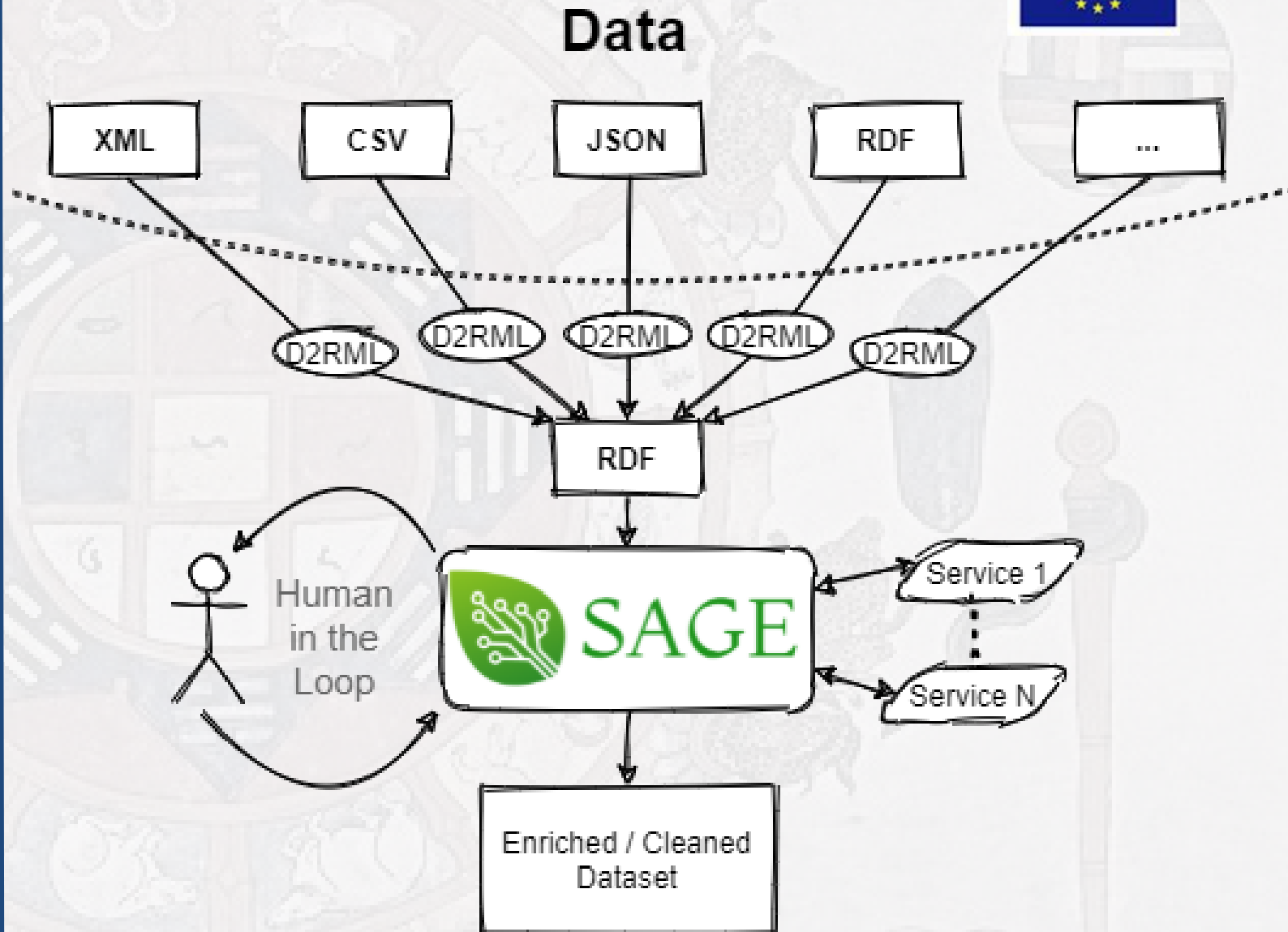
The Platform

SAGE is a Semantic Enrichment and Data management platform

It can import Heterogeneous Data (e.g. XML, CSV, JSON, RDF, etc.) from multiple sources.

Transform and Enrich the data through external services.

Overview and improvement through a Human in the Loop approach.



Main System Technologies



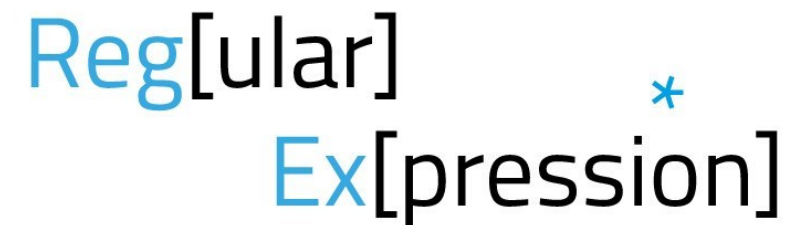
Services

SPARQL Annotators: Link to External Knowledge Bases with SPARQL Endpoint like Wikidata and Geonames.

NERD Annotators: Named Entity Recognition and Disambiguation.

Thesaurus Annotators: Smart string matching using lemmatizers and fuzziness to link text to vocabularies.

PreProcessing: RegEx

 Geonames The logo for Regular Expression, featuring the word "Regular" in blue, "ular" in black, and "Expression" in blue, with a small blue asterisk symbol above the "x" in "Expression".

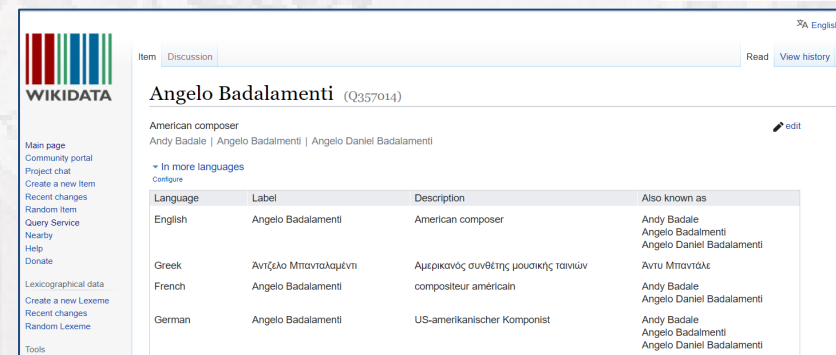
SPARQL Annotator

Creator: BADALAMENTI, ANGELO (ANDY)

RegEx

"Angelo Badalamenti"

SPARQL
Annotator

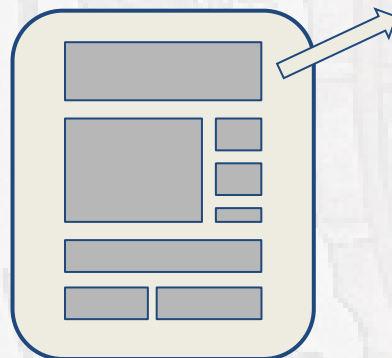


Language	Label	Description	Also known as
English	Angelo Badalamenti	American composer	Andy Badale Angelo Badalamenti Angelo Daniel Badalamenti
Greek	Άντζελο Μπαδάλαντεντι	Αμερικανός συνθέτης μουσικής ταινιών	Άντρυ Μπαδάλαντε
French	Angelo Badalamenti	compositeur américain	Andy Badale Angelo Daniel Badalamenti
German	Angelo Badalamenti	US-amerikanischer Komponist	Andy Badale Angelo Badalamenti Angelo Daniel Badalamenti

URI:
<http://www.wikidata.org/entity/Q357014>
Country: USA
Date of Birth:
22 March 1937
VIAF ID: [32187134](#)
...



NERD Annotator

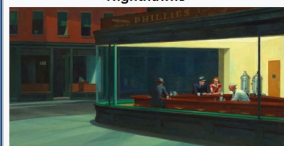
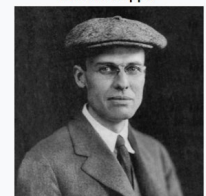
Description: Nighthawks is a 1942 oil on canvas painting by Edward Hopper that portrays people in a downtown diner late at night as viewed through the diner's large glass window. The artwork is displayed at the School of the Art Institute of Chicago.

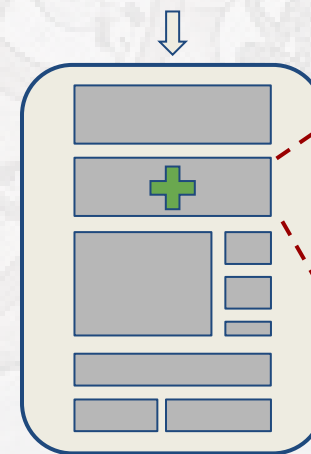


NERD
Annotator

Description: **Nighthawks** is a 1942 oil on canvas painting by **Edward Hopper** that portrays people in a downtown diner late at night as viewed through the diner's large glass window. The artwork is displayed at the **School of the Art Institute of Chicago**.

<p>Nighthawks</p>  <p>Artist: Edward Hopper Year: 1942 Medium: oil paint, canvas Movement: American realism Dimensions: 84.1 cm (33.1 in) × 152.4 cm (60.0 in) Location: Art Institute of Chicago Accession No.: 1942.51</p>	<p>Edward Hopper</p>  <p>Born: July 22, 1882 Nyack, New York, United States Died: May 15, 1967 (aged 84) Manhattan, New York, United States Nationality: American Known for: Painting Notable work: <i>Automat</i> (1927) <i>Chop Suey</i> (1929) <i>Nighthawks</i> (1942) <i>Office in a Small City</i> (1953)</p>	<p>SAIC</p> <p>School of the Art Institute of Chicago</p> <p>Type: Private art school Non-profit</p> <p>Established: 1866 President: Elissa Tenny Academic staff: 141 full-time 427 part-time Undergraduates: 2,894 (Fall 2018)^[1] Postgraduates: 745 (Fall 2018) Location: Chicago, Illinois, United States</p>
--	---	--

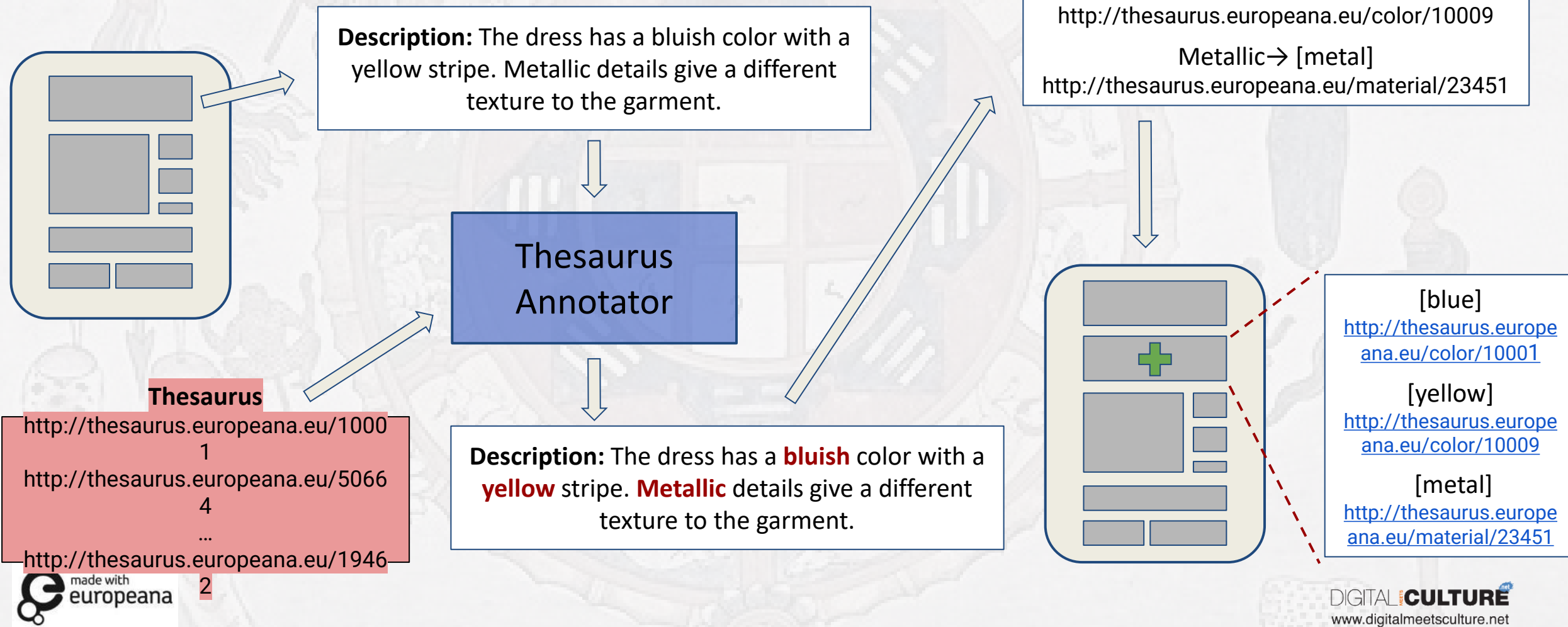


Nighthawks
Wikidata ID:
Q83872

Edward Hopper
Wikidata ID:
Q203401

School of the Art Institute of Chicago
Wikidata ID:
Q7432601

Thesaurus Annotator



Enrichments for PAGODE

NERD Annotators: Detection and Disambiguation of Named Entities such as persons, places, and organizations.

Thesaurus Annotators: Smart string matching using lemmatizers and fuzziness to link text to a custom vocabulary created from keywords provided by the University of Ljubljana.

All keywords were translated to various languages through the respective Wikidata and Getty links, in order to create a multilingual vocabulary to fit the needs of the project.

Sample Enrichments



Bildende Kunst, Porträt von Olga Körner, Arbeiten Dresdner Künstler zur Volkswahl, Arbeiten von Prof. Bernhard Kretschmar, entstanden während einer China-Reise, u.a., Juli 1954 - Januar 1955 Dresden & China

103

✓ ✗ <http://vocab.getty.edu/aat/300080094>
traveling

✓ ✗ <http://www.wikidata.org/entity/Q61509>
travel

✓ ✗ <http://www.wikidata.org/entity/Q148>
China

+

Berlin, 3. Weltfestspiele der Jugend und Studenten, Auftritte ausländischer Kulturensembles aus China und Albanien, August 1951 Berlin

70

✓ ✗ <http://www.wikidata.org/entity/Q64>
Berlin

✓ ✗ <http://www.wikidata.org/entity/Q148>
China

✓ ✗ <http://www.wikidata.org/entity/Q64>
Berlin

+

Dresden. "Tag der Opfer des Faschismus". Vereinigung der Verfolgten des Naziregimes (VVN), 9. September 1951 & Dresden. "Tag der Opfer des Faschismus". Chinesische Delegation anlässlich des Gedenktages in Dresden und Kundgebung vor dem Japanischen Palais - Kranzniederlegung am Ehrenhain für Opfer des Faschismus im Garten des Palais, 9. September 1951 DresdenVolksrepublik China

55

✓ ✗ <http://www.wikidata.org/entity/Q1731>
Dresden

✓ ✗ <http://www.wikidata.org/entity/Q5224691>
Darrell Opfer

✓ ✗ <http://www.wikidata.org/entity/Q315292>
Union of Persecutees of the Nazi Regime

✓ ✗ <http://www.wikidata.org/entity/Q1731>
Dresden

✓ ✗ <http://www.wikidata.org/entity/Q5224691>
Darrell Opfer

✓ ✗ <http://www.wikidata.org/entity/Q1731>
Dresden

✓ ✗ <http://www.wikidata.org/entity/Q5224691>
Darrell Opfer

✓ ✗ <http://vocab.getty.edu/aat/300008107>
botanic gardens

✓ ✗ <http://www.wikidata.org/entity/Q1107656>
garden

✓ ✗ <http://www.wikidata.org/entity/Q148>
China

+

Sample Enrichments



Coffee cup made from porcelain, transfer-printed in **underglaze blue** with the 'Temple' pattern of a Chinese pagoda and trees on a white ground with gilded decoration, manufactured by Caughley (or Salopian Works) c.1780.

11

✓ ✗ <http://www.wikidata.org/entity/Q98876293>
Underglaze blue

✓ ✗ <http://vocab.getty.edu/aat/300010662>
porcelain (material)

✓ ✗ <http://www.wikidata.org/entity/Q130693>
porcelain

Performer's note: Three instrumentalists and a singer from the Unit which performed the **Culture Show** at the Number One Hotel.
en

11

✓ ✗ <http://www.wikidata.org/entity/Q7728389>
The Culture Show

CRE unknown [19th century] **People's Republic of China**

10

✓ ✗ <http://www.wikidata.org/entity/Q148>
China

Tea bowl made from porcelain, transfer-printed in **underglaze blue** with the 'Temple' pattern of a Chinese pagoda and trees on a white ground with gilded decoration, manufactured by Caughley (or Salopian Works) c.1780.

10

✓ ✗ <http://www.wikidata.org/entity/Q98876293>
Underglaze blue

✓ ✗ <http://www.wikidata.org/entity/Q130693>
porcelain

✓ ✗ <http://vocab.getty.edu/aat/300010662>
porcelain (material)

Saucer made from porcelain, transfer-printed in **underglaze blue** with the 'Temple' pattern of a Chinese pagoda and trees on a white ground with gilded decoration, manufactured by Caughley (or Salopian Works) c.1780.

9

✓ ✗ <http://www.wikidata.org/entity/Q98876293>
Underglaze blue

✓ ✗ <http://vocab.getty.edu/aat/300010662>
porcelain (material)

✓ ✗ <http://www.wikidata.org/entity/Q130693>
porcelain



✓ ✗ <http://www.wikidata.org/entity/Q1422576>
saucer

✓ ✗ <http://vocab.getty.edu/aat/300195535>
scales

Validation Progress





Europeana Fashion - PAGODE

 100%

<http://purl.org/dc/elements/1.1/description>	(Term)		100%
<http://purl.org/dc/terms/spatial>	(Term)		100%
<http://purl.org/dc/elements/1.1/creator>	(Term)		100%
<http://purl.org/dc/elements/1.1/title>	(Term)		100%




Östasiatiska museet, Collection Photography - PAGODE

 100%

<http://purl.org/dc/elements/1.1/description>	(Term)		100%
<http://purl.org/dc/terms/spatial>	(Term)		100%
<http://purl.org/dc/elements/1.1/subject>	(Term)		100%
<http://purl.org/dc/elements/1.1/title>	(Term)		100%

Rijksmuseum, Wellcome Collection - PAGODE

 100%

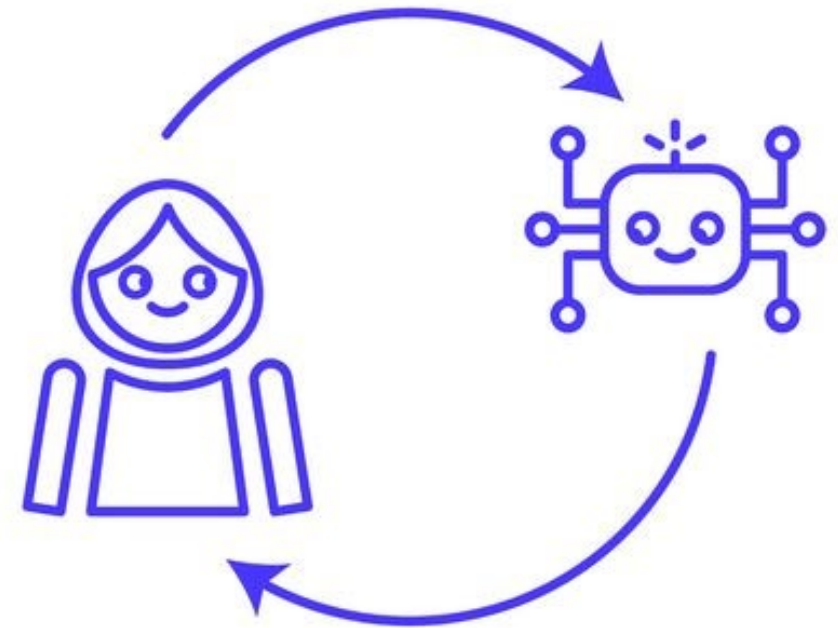
<http://purl.org/dc/elements/1.1/description>	(Term)		100%
<http://purl.org/dc/elements/1.1/creator>	(Term)		100%
<http://purl.org/dc/terms/spatial>	(Term)		100%

Human in the Loop

Useful feedback from validators and other experts.

Feedback was used to fine-tune the results.

Unfortunately, a second round of enrichment was not possible due to time limitations.



SAGE for PAGODE in Numbers



**Accepted
Enrichments**

94.000+



**Enriched
Records**

24.000+



**Active
Users**

7



**Validation
Progress**

100%

Ingestion to Europeana

A CSV file that contains all validated/accepted enrichments has been provided to Europeana mid-June.

