Towards a Comprehensive Assessment of the Quality and Richness of the Europeana Metadata of food-related Images

Yalemisew Abgaz[1], Amelie Dorn[2], José Luis Preza Díaz[2] & Gerda Koch[3]

[1]Adapt Centre DCU, [2]ACDH-CH OeAW, [3]Europeana Local - AT

WWW.OEAW.AC.AT

ACDH-CH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

Background

ChIA

- Interdisciplinary Digital Humanities project (2019-2021)

- Involved expertise:
  Digital Humanities, AI & NLP (ACDH-CH OeAW, AT)
  Semantic technologies (Adapt Centre, IE)
  Cultural Image aggregation (Europeana Local - Österreich, AT)

- Projet aim & results: **the ChIA system -** enabled increased access and analysis possibilities of cultural (food) images for content providers and educational purposes
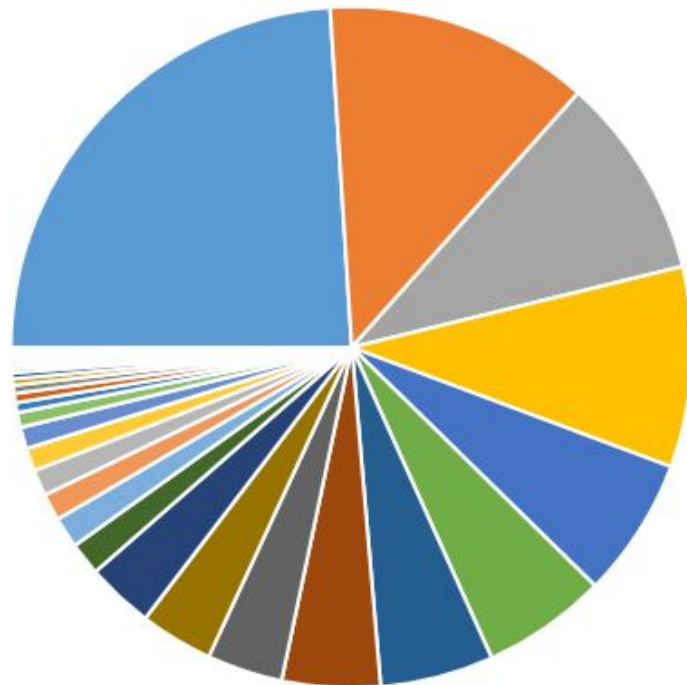
Background

Europeana data set

Total: 58.6 Mio digital objects

Includes: 34.2 Mio digital images

from: 3.500 institutions in 42 countries

- Netherlands
- United Kingdom
- Sweden
- Germany
- France
- Italy
- Norway
- Belgium
- Spain
- Denmark
- Austria
- Czech Republic
- Hungary
- Finland
- Poland
- Europe
- Greece
- Estonia
- Lithuania
- Slovenia
- Portugal
- Switzerland
- Ireland
- Croatia
- Latvia
- Romania
- Malta
- Bulgaria
- Slovakia
- Cyprus
- Israel
- Serbia
- Iceland
- Ukraine
- Turkey
- Russia
- Macedonia
- Moldova
- Montenegro
- Georgia
- Bosnia and Herzegovia
- Luxembourg
- Albania

ÖAW AUSTRIAN ACADEMY OF SCIENCES

The Problem

Dataset:

Selection based on food context of images

42.969 images (available with Free Access licenses) were selected in form of various sets (baskets) for later download & analysis of metadata and images

Prepare all | Download Metadata | Edit basket | ← Return

Title:	Painting-Fruit
Description:	Search for images with the description "painting" and "fruit"
Author:	Gerda Koch

484 objects in basket

View objects 20 to 30

Prev | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | … | 49 | Next

Remove | View Object
Remove | View Object
Remove | View Object

The Problem

Metadata

Although descriptions are available they seldom tell all details on what is depicted in the images. In most cases content descriptions use iconographic phrases like "fruits", "flowers", "still life".

about:	/proxy/provider/2059517 /data_foodanddrink_WAGMU_op725	
dcCreator:	def:	STEELE; E
dcDate:	def:	1889
dcDescription:	def:	Oil painting showing a vase of flowers and a bowl of fruit on a table, more fruit lies on the table.
dcFormat:	def:	image/jpeg

The Problem

Vocabularies

- Some institutions deliver metadata to Europeana that already includes vocabulary URIs.
- Europeana enrich semantic connections with vocabularies such as AAT, ULAN, IconClass, VIAF, LCSH

  http://data.europeana.eu/concept/base/222

  http://iconclass.org/41A671

- No specific food or drink related vocabularies are used.
- Irregularity in the use of semantics across the dataset.

ÖAW AUSTRIAN ACADEMY OF SCIENCES

The problem

The current image collection

- mostly have metadata that is focused on bibliographic and format related but lacks domain-specific metadata

- most records are not interlinked on the basis of content (only when joint vocabularies are used)
- Thus,the current metadata needs analysis on the basis of
  - The quality of the current metadata
  - The use of multiple domain-specific vocabularies
  - The gap between what the image depicts and the metadata express

Proposed Solution

Analyse the richness of the metadata using

- Quantitative approach
  - using objective quality assessment metrics

- Qualitative approach
  - using expert judgement on the expressiveness of the metadata
- Semantic enrichment to fill the gap
  - Computer vision
  - Semantic annotation

Quality Analysis Metrics

Metrics in four categories are selected

- Contextual
  - *indication of used vocabularies*
- Intrinsic
  - *extensional conciseness*
- Accessibility
  - *links to external LD providers*
- Representational

ÖAW | AUSTRIAN ACADEMY OF SCIENCES

Semantic Richness Analysis

Semantic Richness: The availability of multiple descriptors of a resource particularly representing the main concepts represented by the target.

about:	/proxy/provider/2059517 /data_foodanddrink_WAGMU_op725
dcCreator:	def: STEELE; E
dcDate:	def: 1889
dcDescription:	def: Oil painting showing a vase of flowers and a bowl of fruit on a table, more fruit lies on the table.
dcFormat:	def: image/jpeg

More semantics for this image
- Fruits
    - apple
    - grapes
    - Rose flower? etc
- Objects
    - vase
    - bowl?
- Culture
    - what culture does it represent?
        - rich/poor?

Results so far

Our initial analysis shows that the metadata

- is rich in **bibliographic information**
- provides labels with **multiple languages**
- however, lacks **semantic richness**

Our Current work

- use of selected vocabularies to quantify semantic richness
- Analysis of the images with computer vision has a potential to address the richness problem
- Preparation of a training set is underway

Conclusion

- It is important to provide quality metadata to improve search and retrieval of historical images

- Semantic richness is a key for the search and exploration of historical images

- Understanding the gap is crucial to do the semantic annotation

- The use of computer vision combined with expert annotation and evaluation has a potential to improve both semantic richness and quality.

Thank you for listening

Any Questions?

Or ideas for collaboration?

#chia4dh

@yalemisew
@adooorn
@Europeanaeu
@jlprezadiaz

Check out our website → https://chia.acdh.oeaw.ac.at

Yalemisew.Abgaz@adaptcentre.ie
Amelie.Dorn@oeaw.ac.at
kochg@europeana-local.at