

h e g

Haute école de gestion
Genève

28 August 2020

ENABLING BETTER AGGREGATION AND DISCOVERY OF CULTURAL HERITAGE CONTENT FOR EUROPEANA AND ITS PARTNER INSTITUTIONS

Julien A. Raemy

Master's thesis oral defence
Information Science Department, Haute école de gestion de Genève

Presentation overview



1. Rationale



2. Key results



3. Recommendations

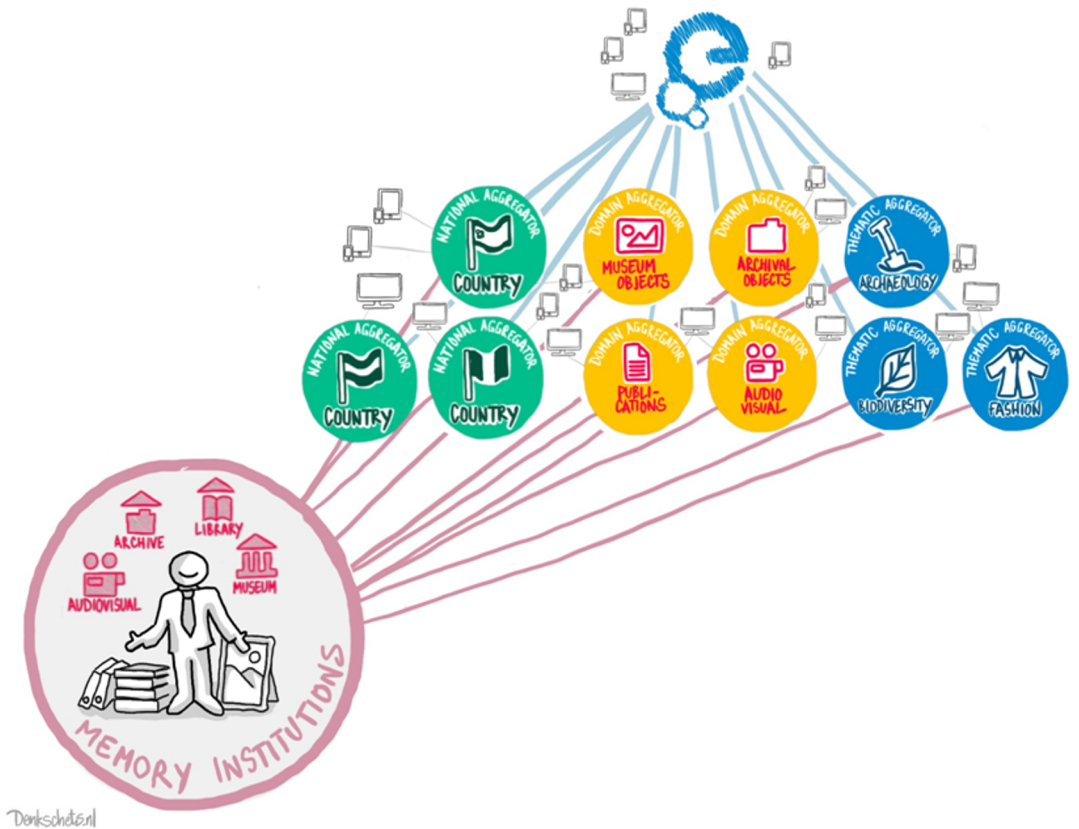


4. Main challenges and difficulties



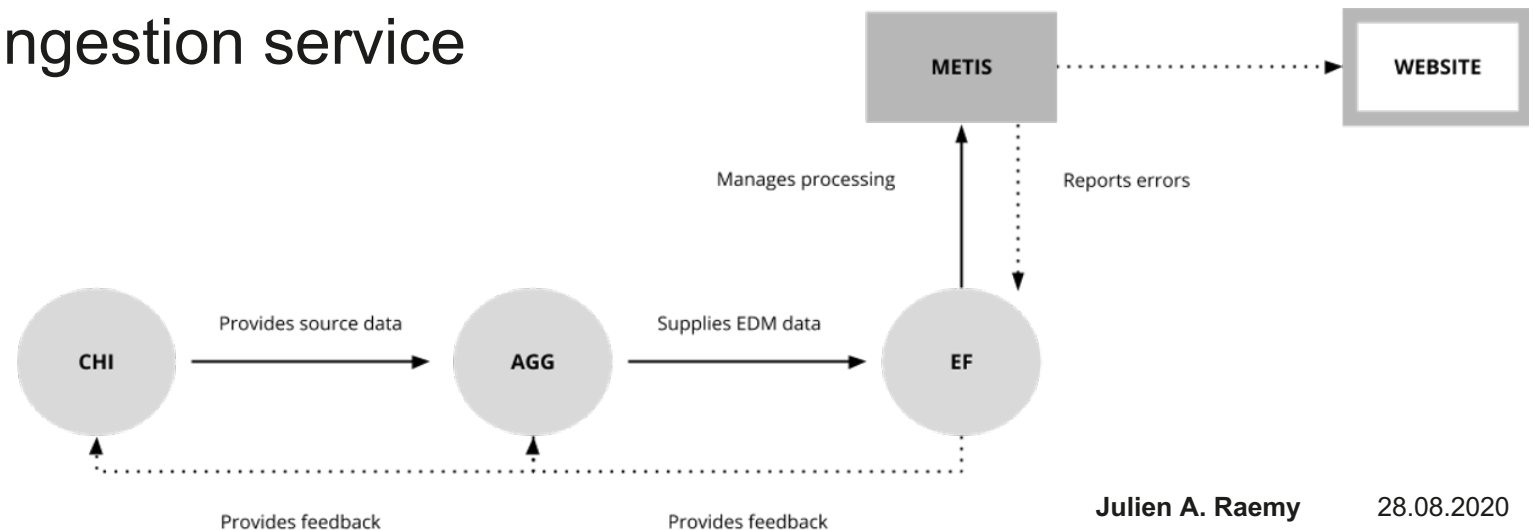
5. Perspectives

1. RATIONALE



Aggregation at Europeana

- Aggregation of metadata and links to digital cultural heritage objects (CHOs) held by over 3700 providers
 - Cultural heritage institutions (CHIs): libraries, archives, museums
 - Intermediary aggregators
- Current aggregation mechanism
 - Data transfer: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
 - Data modelling: Europeana Data Model (EDM)
- Metis – Europeana’s ingestion service



Research scope

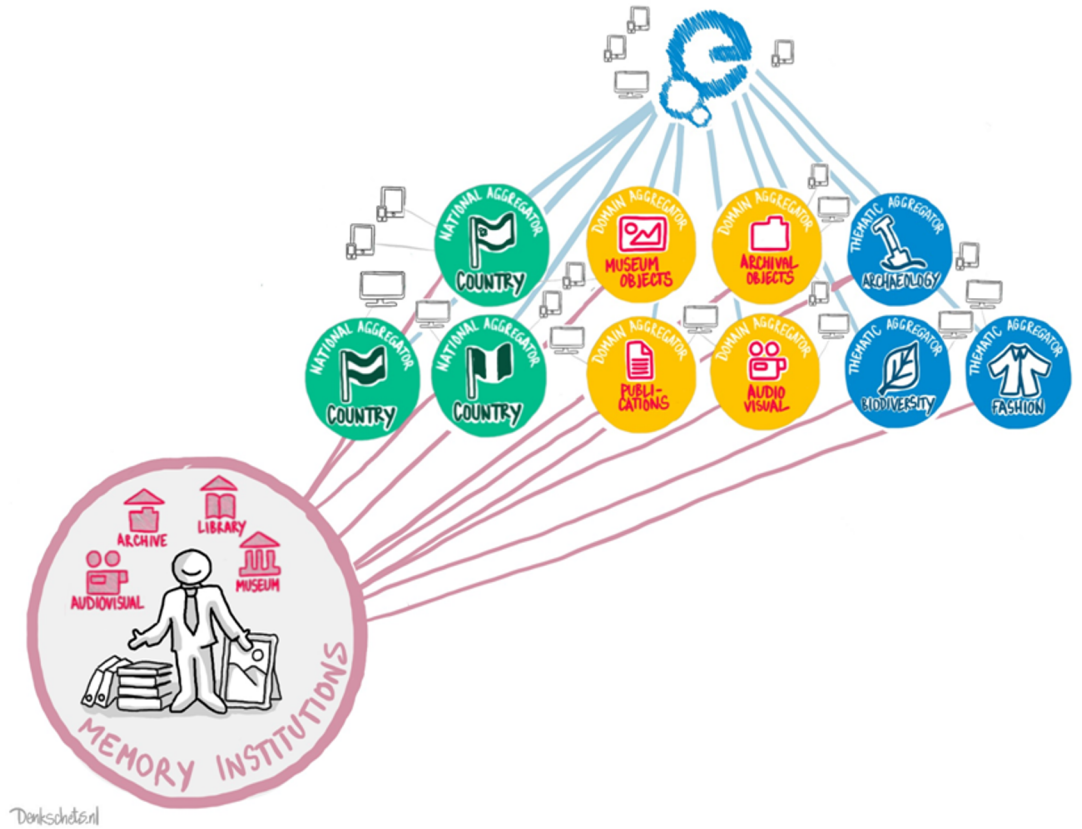
Expectations

- Extension of experiments that Europeana R&D has already been carried out
- Support Europeana's decisions on the directions for improved data aggregation

Research questions

- Suitable alternative mechanisms to OAI-PMH
- Deployment conditions of these technologies

2. KEY RESULTS



Involvement in the Europeana Common Culture (ECC) project



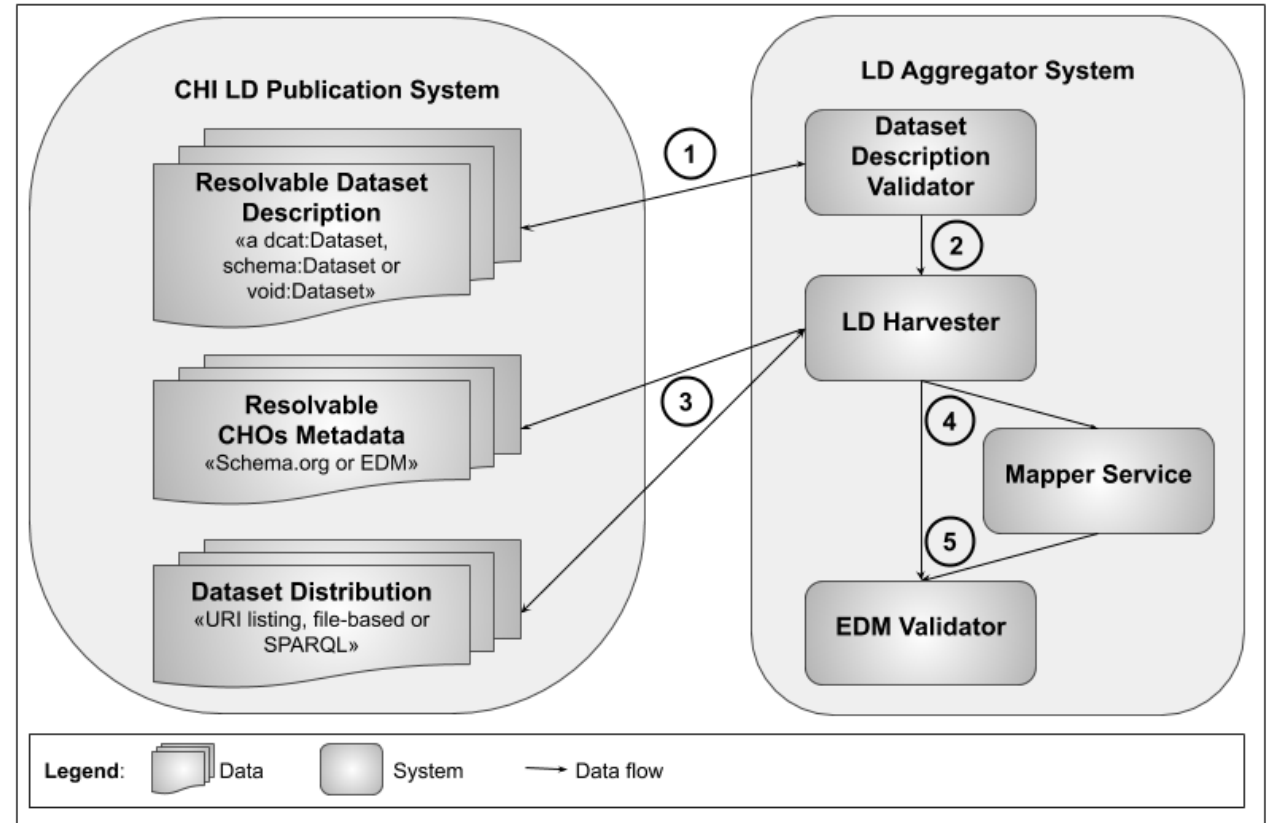
Documentation revision of the LOD-aggregator with the NDE*



Sustainability discussions



Submission of a conference paper



Freire et al. 2020, *Metadata aggregation via linked data: results of the Europeana Common Culture project*. Submitted paper for the Metadata and Semantics Research (MTSR) Conference 2020

Online survey

Survey to gauge the awareness, interest and use of technologies other than OAI-PMH for (meta)data aggregation



20 April – 8 May
Google Forms



9 Sections
15 questions
(10 mandatory + 5 optional)

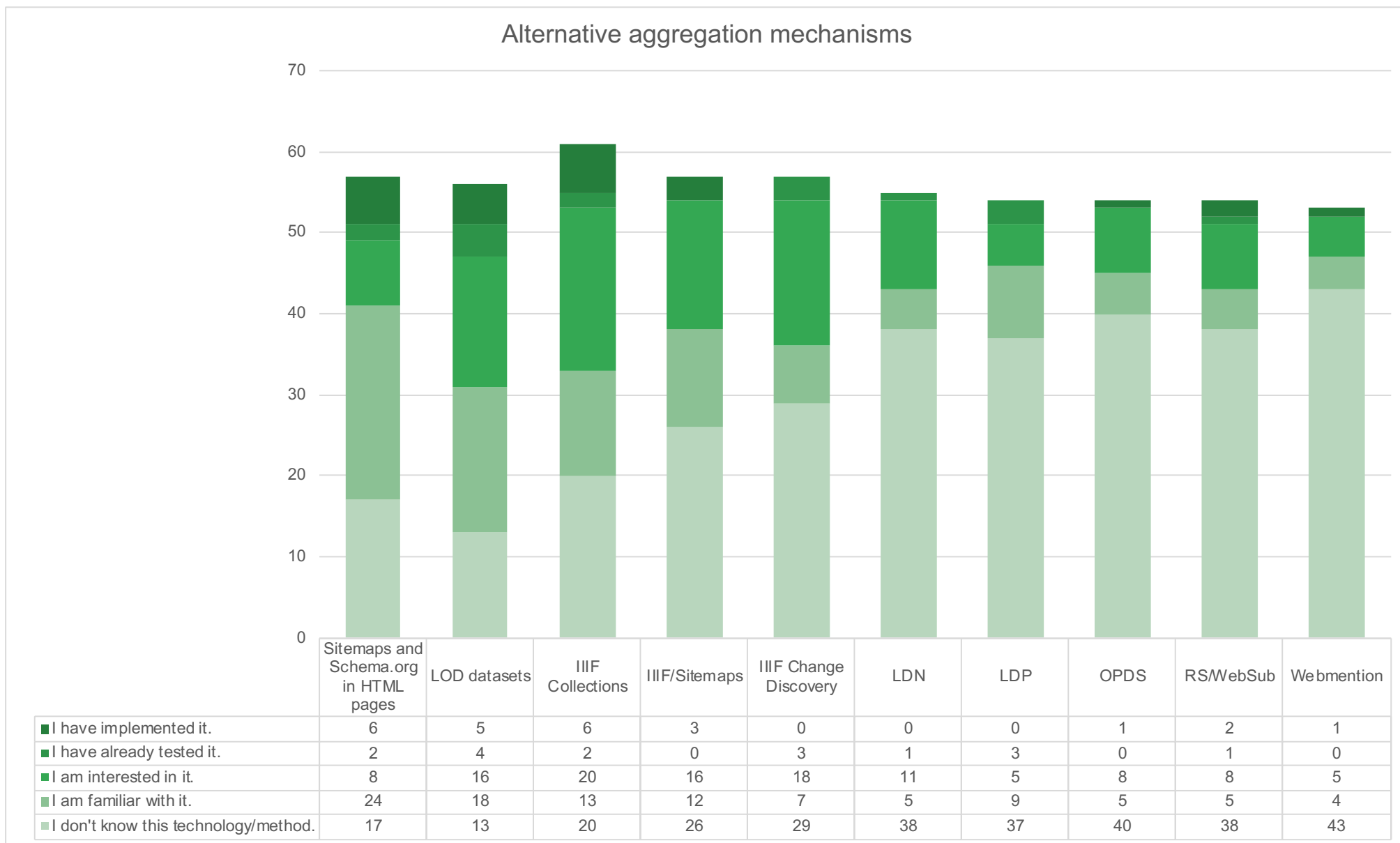
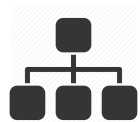


52 participants
20+ Countries
38.5% aggregators



35.1% who use OAI-PMH → only for Europeana
23 interested in an aggregation pilot

Awareness, use, and interest in alternative aggregation mechanisms



Aggregation pilots



Parameters for
defining and
assessing
potential pilots

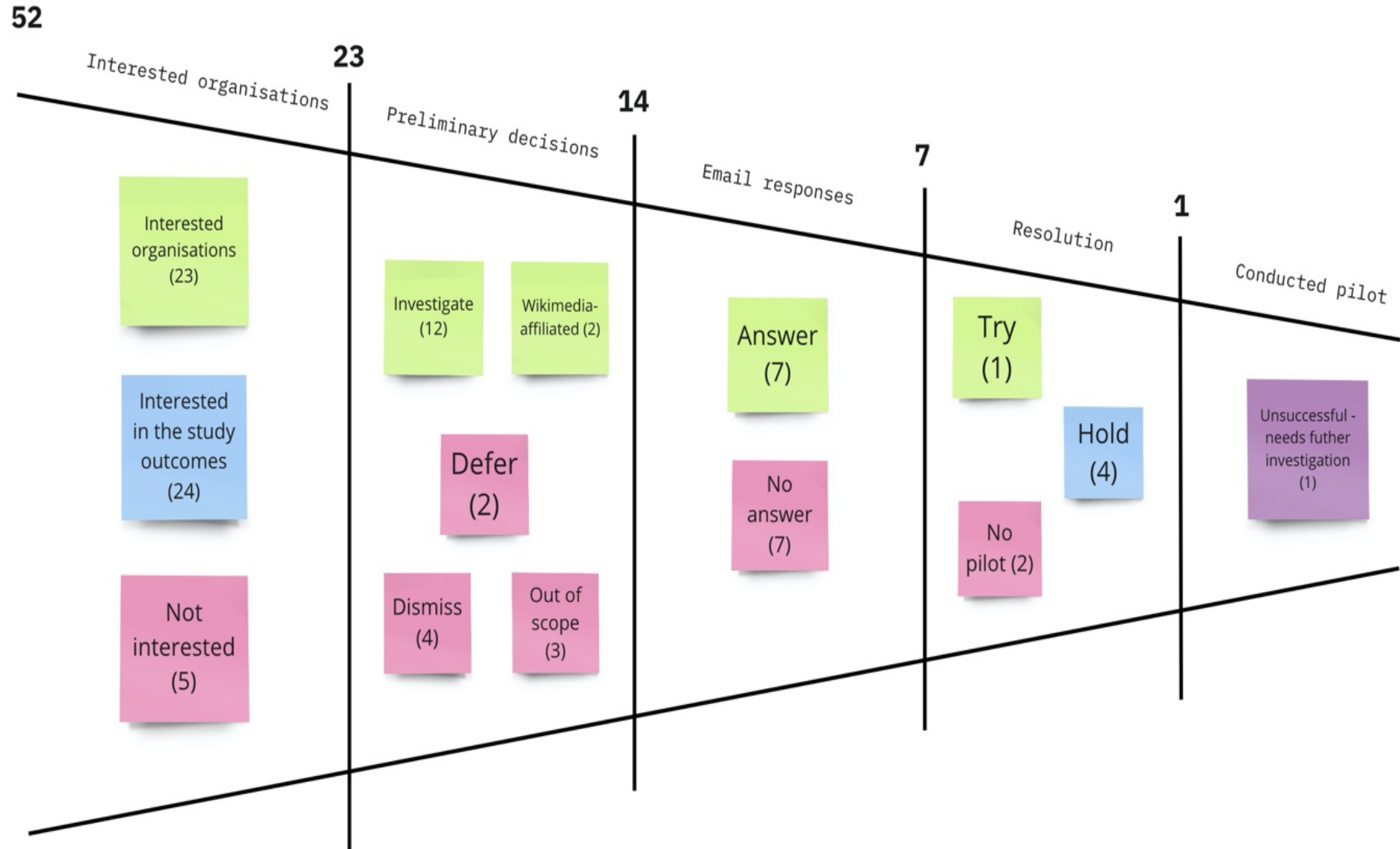


Identifying
aggregation
routes

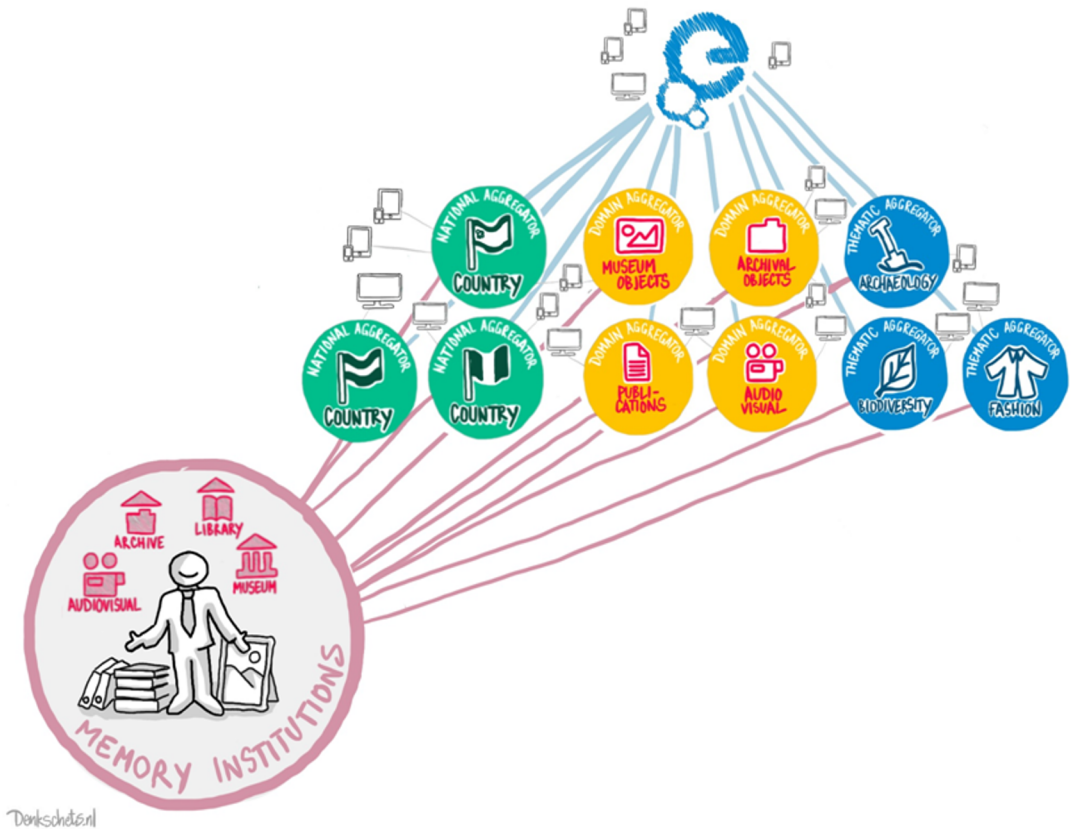


Assessment of
potential
aggregation pilots

Assessment of potential aggregation pilots



3. RECOMMENDATIONS



Recommendations (steps)



1) Target levels

digital object
associated metadata
providing institution



2) Opportunity Solution Tree

1 desired outcome
3 opportunities
5 solutions
13 experiments



3) Alignment with the European Strategy 2020-2025

strengthen the infrastructure
improve data quality
build capacity



4) Suggestions for implementing the identified solutions

20+ detailed suggestions



Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions

Showcasing CHOs by making them more engaging to end users

- IIIF
 - Aggregation of IIIF based on Sitemaps / IIIF Collections / Change Discovery API / LDN
 - IIIF and Europeana Working Group
 - Involvement of Europeana collaborators in the relevant committees and groups

Enriching the overall quality of CHOs' associated metadata

- LOD
 - Aggregation via LOD based on listing of URIs / dataset distribution / SPARQL
 - Aggregation via LDN and storage over LDP

Increasing the online visibility of providing institutions

- Semantic SEO
 - Aggregation via Sitemaps and Schema.org in HTML pages
 - Aggregation via ResourceSync in conjunction with WebSub
- Use of CMS or static sites for small and medium-sized organisations
 - Aggregation via Webmention
 - Aggregation via ActivityPub Delivery
 - Aggregation via an agreed REST API
- Integration of Europeana data in third-party platforms
 - Wikimedia Commons
 - Flickr
 - Social networks (Instagram, Twitter, Facebook)

Opportunity 1

Solution

Experiments

Showcasing CHOs by making them more engaging to end users

IIIF

Aggregation of IIIF based on Sitemaps / IIIF Collections / Change Discovery API / LDN

IIIF and Europeana Working Group

Involvement of Europeana collaborators in the relevant committees and groups

Opportunity 2

Solution

Experiments

Enriching the overall quality of
CHOs' associated metadata

LOD

Aggregation via LOD based on listing of URIs / dataset distribution / SPARQL

Aggregation via LDN and storage over LDP

Increasing the online visibility of providing institutions

Semantic SEO

Aggregation via Sitemaps and Schema.org in HTML pages

Aggregation via ResourceSync in conjunction with WebSub

Use of CMS or static sites for small and medium-sized organisations

Aggregation via Webmention

Aggregation via ActivityPub Delivery

Aggregation via an agreed REST API

Integration of Europeana data in third-party platforms

Wikimedia Commons

Flickr

Social networks (Instagram, Twitter, Facebook)

Europeana Strategy priorities and suggestions

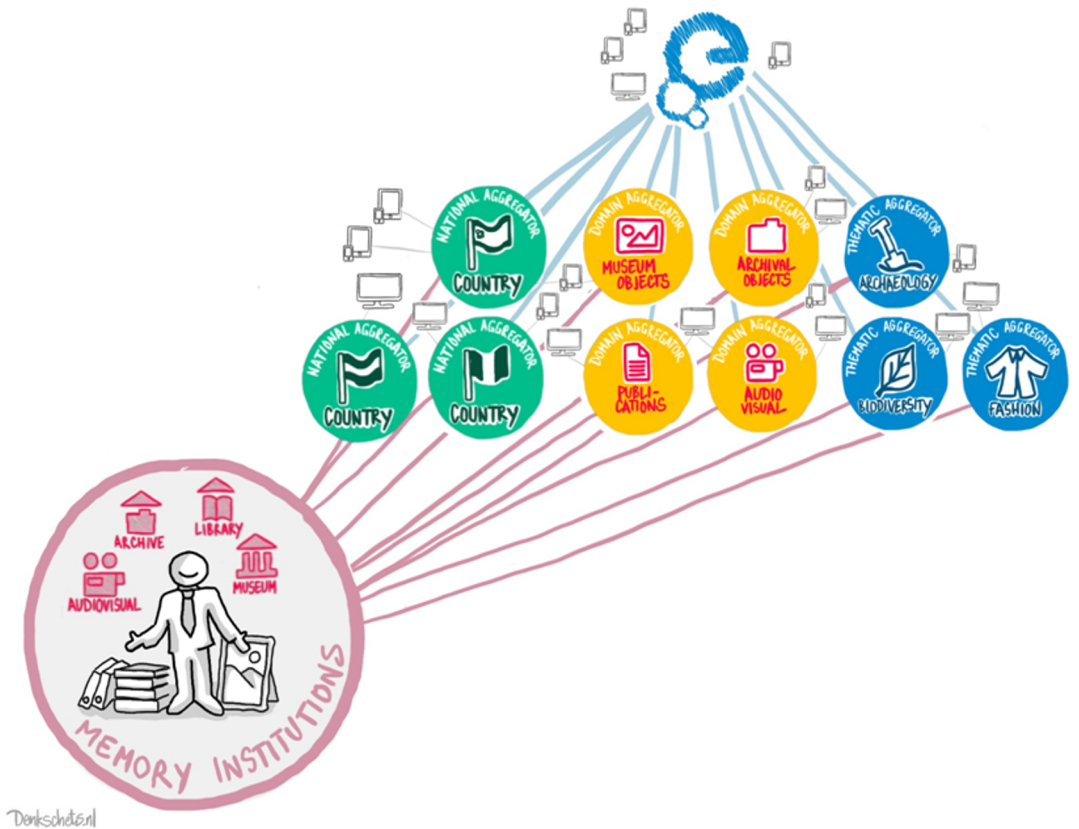
Alignment with the strategy

	Strengthen the infrastructure	Improve data quality	Build capacity
IIIF	✓		✓
LOD	✓	✓	✓
Semantic SEO	✓	✓	
CMS / static sites	✓		
Third-party platforms		✓	✓

Some suggestions for implementing the solutions

- Produce a guideline on best practices on how and which structured metadata should be linked within IIIF Manifests
- Provide a tool for data providers and aggregators that is able to quickly build Sitemaps with relevant pages (for CHOs)
- Develop easy-to-deploy frameworks for data providers
- (...)

4. MAIN CHALLENGES AND DIFFICULTIES



Challenges and difficulties



Diverse expectations and needs - often hard to capture



Lack of knowledge and expertise in Social Web Protocols



Aggregation pilots were more complex to assess, organise and conduct than expected



Inconclusive error reporting

Survey feedback

Difficult assessment of what they have already implemented

Institutions with limited technical expertise

OAI-PMH should be maintained

Non-exhaustiveness in the identification of mechanisms

Very knowledgeable professionals

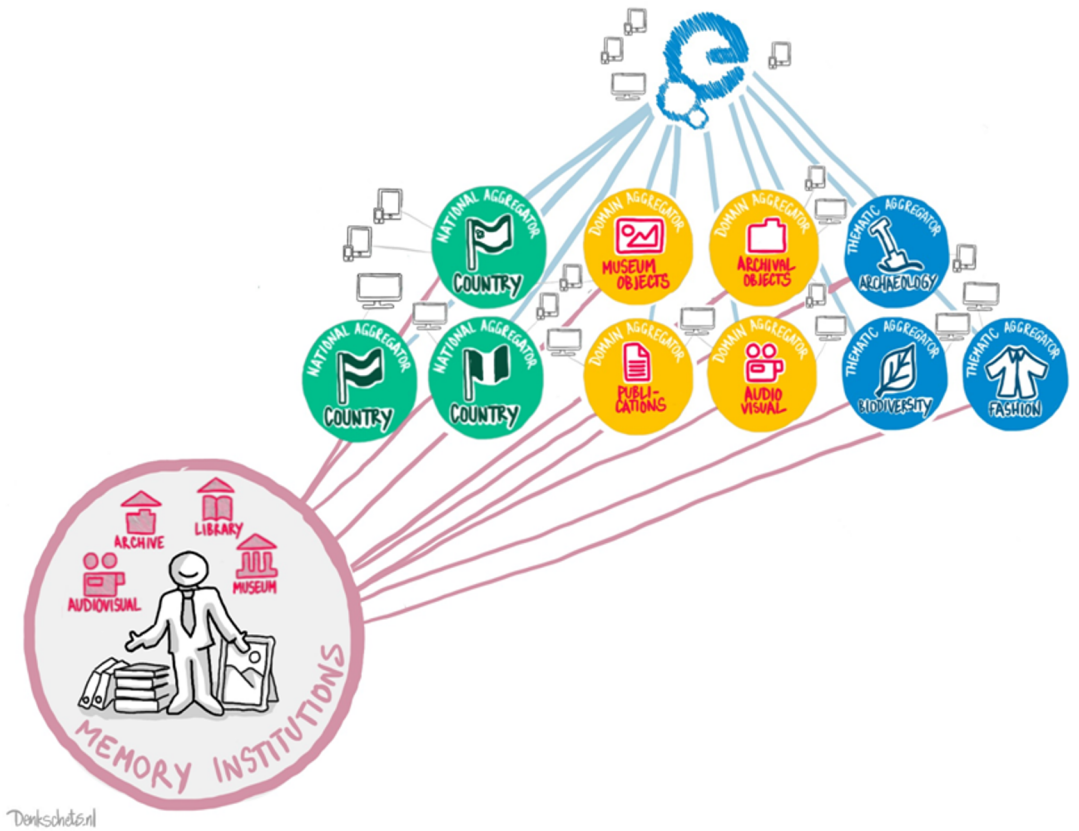
Full Linked Data

Extracting Schema.org from web pages

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "CreativeWork",
  "creator": {
    "@type": "Person",
    "name": ""
  },
  "image": "http://frontend.mnm.monguz.hu/JaDoX_Portlets/displayContent?outerId=oai-aggregated-bib6799262&size=tv&cast=jpg",
  "name": "1. honv&acute;dz&acute;szl&acute;alj z&acute;szl&acute;ja, 1848",
  "material": "slink",
  "dateCreated": "1848, 19. sz&acute;zad",
  "locationCreated": "",
  "contentLocation": "",
  "about": "1. honv&acute;d z&acute;szl&acute;alj",
  "identifier": {
    "@type": "PropertyValue",
    "propertyID": "Inventory Number",
    "value": "93/Z1"
  }
}
</script>
```

Page source of <https://www.museumap.hu/record/-/record/oai-aggregated-bib6799262>

5. PERSPECTIVES



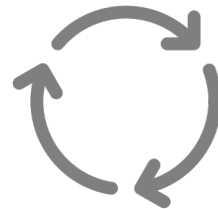
Future work



Advisory effort



Metis Sandbox



Implementation of the
thesis' recommendations



Compliance with alternative
aggregation mechanisms

Recent collaboration between the NDE and the Swedish National Heritage Board (SOCH)

SOCH datasets added to the LOD-aggregator

README.md

Test run with one of the SOCH datasets

SOCH provided several dumps from different organizations. The dataset from the Livrustkammaren Museum (LSH) was used for this test. The dumps contained a directory structure with a XML/RDF file for each resource and agent description. The dumps were repackaged to remove the directory structure and to use the default zip format (*.zip).

The lod-aggregator expects the data to be online and described by a dataset description. For this test a simple dataset description was created for [the collection](#) and for [the agent descriptions](#). The dataset descriptions and the dump files were placed online on a testserver to perform the test.

Temporary URL of the datasets:

- http://cclod.netwerkdigitaalergoed.nl/soch_lsh.ttl
- http://cclod.netwerkdigitaalergoed.nl/soch_lsh_agents.ttl

Initialize the environment

In the config file .env the variable for the provider VAR_PROVIDER was set to 'SOCH'.

For running the scripts described below we must first run the following command in the root of the lod-aggregator:

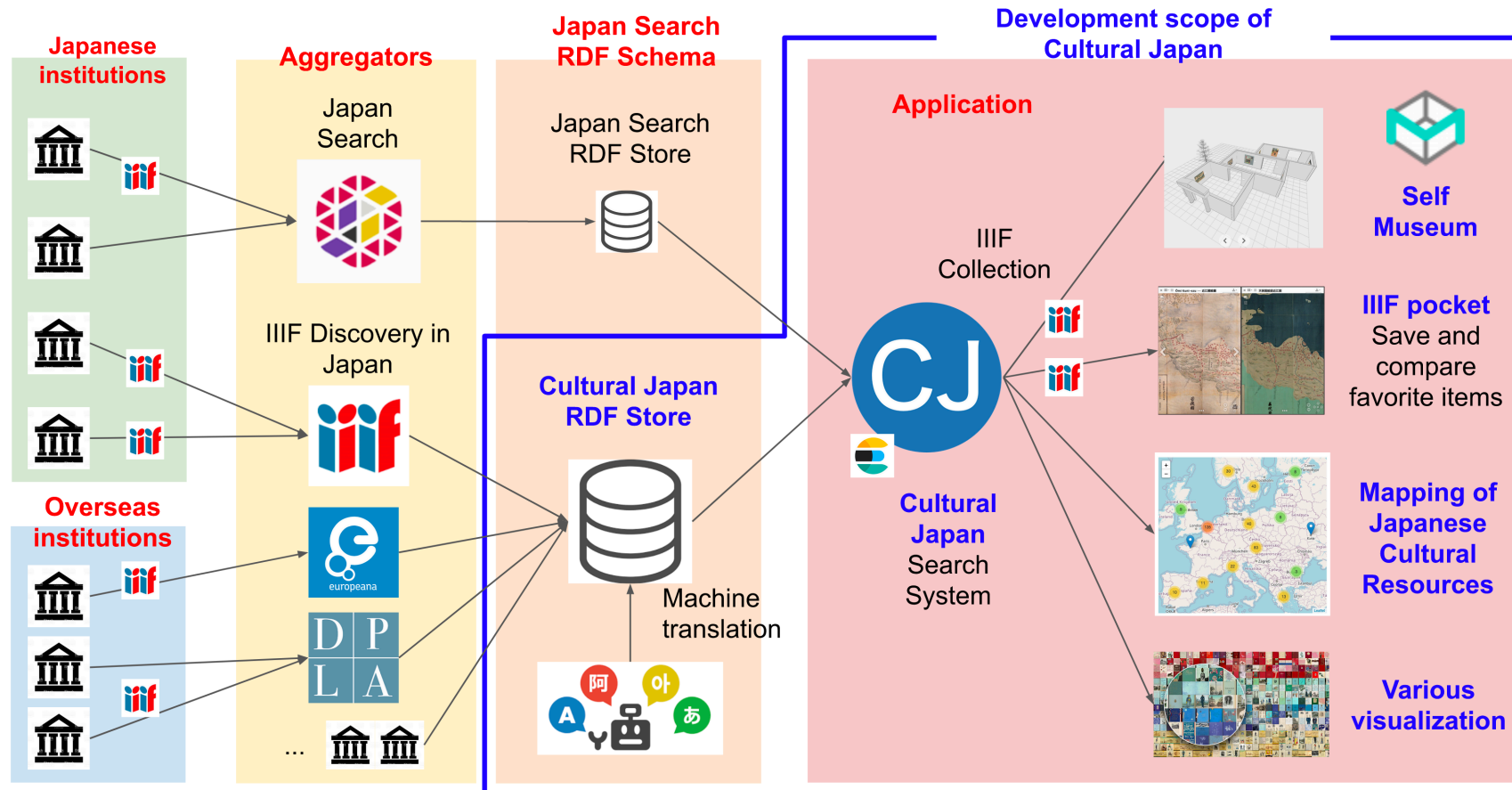
```
source bin/setpaths
```

Download the datasets

```
crawl.sh --dataset-uri http://cclod.netwerkdigitaalergoed.nl/soch_lsh.ttl --output soch-lsh.nt
```

```
crawl.sh --dataset-uri http://cclod.netwerkdigitaalergoed.nl/soch_lsh_agents.ttl --output soch-lsh-agents.nt
```

Cultural Japan's system overview



<https://cultural.jp/en/about>

h e g

Haute école de gestion
Genève

Julien A. Raemy

Master's student in Information Science

julien.raemy@hesge.ch

<https://julsraemy.github.io/>



[@julsraemy](https://twitter.com/julsraemy)