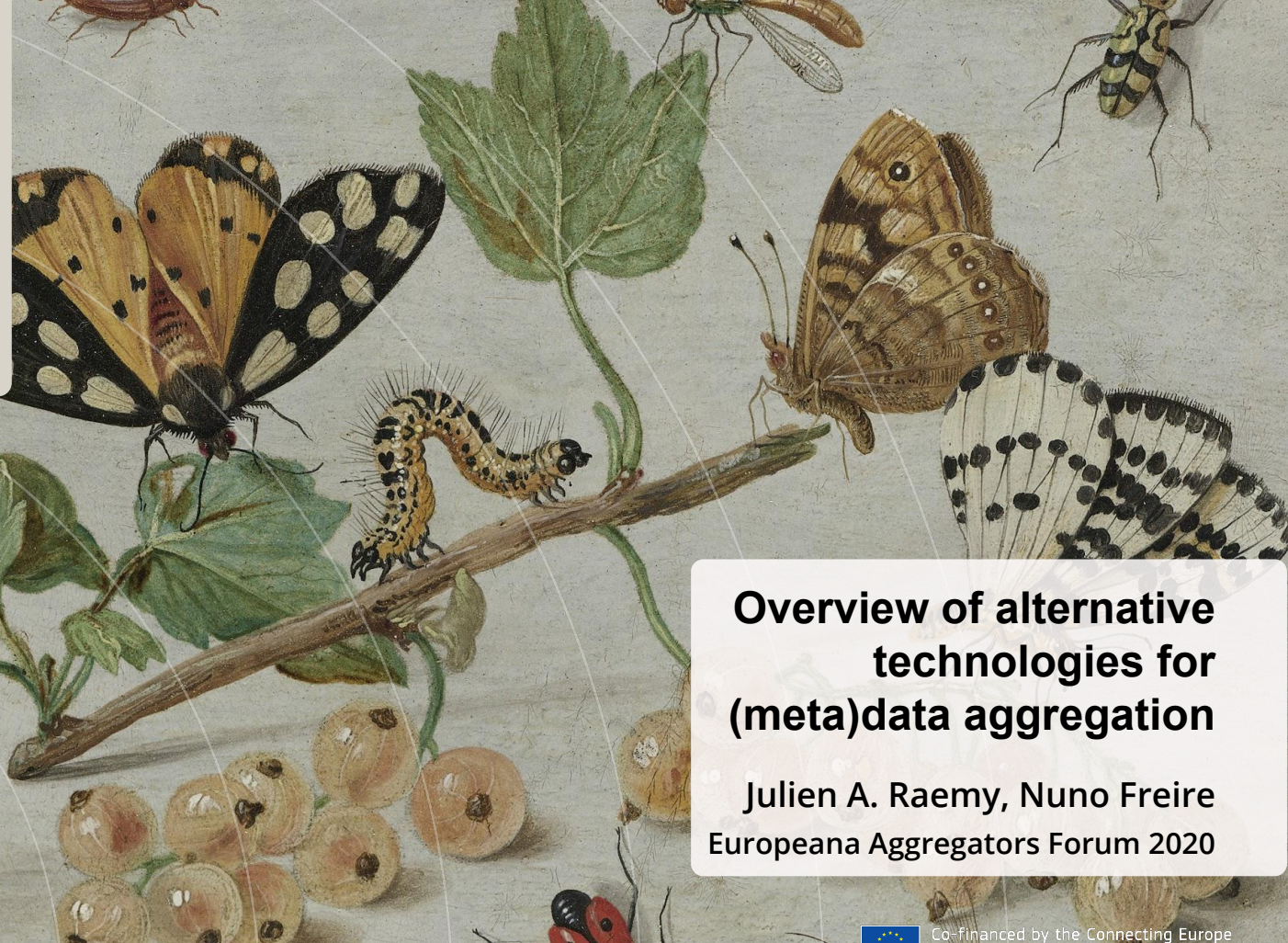




Europeana
Foundation



Overview of alternative technologies for (meta)data aggregation

Julien A. Raemy, Nuno Freire
Europeana Aggregators Forum 2020





BACKGROUND

Context and Objectives



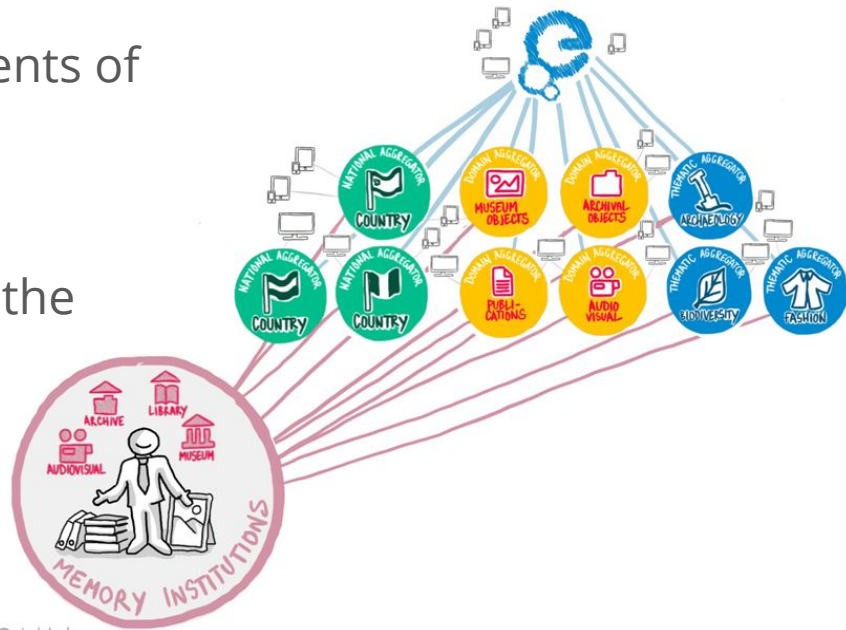
europaea
foundation

CONTEXT

- Master's thesis in Information Science with the partnership of Europeana's R&D team
 - *Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions*
 - Mid-February to Mid-August 2020
- Europeana Common Culture's Linked Data aggregation pilot (task 2.3) with Schema.org
 - <https://pro.europeana.eu/project/europeana-common-culture>
 - <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator>

OBJECTIVES

- Providing an overview of alternative technologies and methods to OAI-PMH
- Identifying the deployment requirements of the most promising technologies
- Conducting tests with interested data providers and aggregators to expand the work already undertaken in this area





OVERVIEW

Technologies and methods for
(meta)data aggregation

INVESTIGATION

How did we determine which technologies were relevant?

- **Mechanisms already used by data providers and aggregators for other purposes**
 - Search Engine Optimization (SEO)
 - Linked Open Data (LOD)
 - [International Image Interoperability Framework \(IIIF\)](#)
- **Methods with low-technical barriers for adoption**

schema.org



W3C®

THE HARVESTING MIX



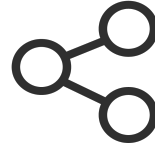
Dataset description or listing



Description of cultural heritage objects (CHOs)







Data transfer/synchronisation



Combination of technologies enabling aggregation

METHODS

	VoID, DCAT, Schema.org (<code>schema:Dataset</code>) Sitemaps, IIIF Collection
	Europeana Data Model (EDM) Schema.org
	ResourceSync Open Publication Distribution System (OPDS) IIIF Change Discovery API Activity Streams 2.0 (AS2)
	HTML and Structured Data Harvesting: <i>based on Sitemaps, RDFa, and Schema.org</i> Aggregation of IIIF resources: <i>based on Sitemaps, IIIF Collections, or on LDN</i> Aggregation via LOD datasets: <i>VoID, DCAT, or Schema.org</i>



SUCCESSES

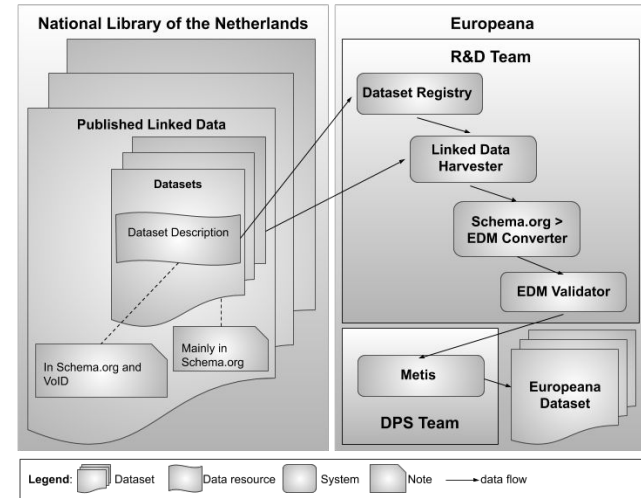
Findings from previous metadata
aggregation pilots
Data Aggregation Lab (DAL)



European
Foundation

THE RISE OF LITERACY PROJECT

Collection name	Dataset ID at Europeana	Number of ingested records	Ingestion technique	Source metadata vocabulary
Kinderboeken	50_KB_RiseOfLiteracy_Kinderboeken	381	Linked Data	Schema.org
Centsprenten	10_KB_RiseOfLiteracy_Centsprenten	1,255	Linked Data	Schema.org
Alba Amicorum	16_RoL_KB_AlbaAmicorum	2,155	Linked Data	Schema.org



IIIF AGGREGATION PILOTS

Collection name	Dataset ID at Europeana	Number of ingested records	Ingestion technique	Source metadata vocabulary
University College Dublin	2064501_Ag_IRE_UCD_IIIF	1,684	Sitemap	EDM
Wellcome Library	9200579_Ag_UK_WellcomeCollection_IIIF	84,405	IIIF Collection	Schema.org

DAL

- LOD Dataset
- IIIF Dataset
- WWW/HTML Crawling

Data Aggregation Lab

R&D Workbench for alternative aggregation mechanisms for Europeana

About the workbench

This system is a work in progress. It aims to gather information from data providers and aggregators, share experimental results, apply prototypes, and provide demom

Current R&D activities address:

- Aggregation mechanisms for linked data
- Aggregation mechanisms for IIIF based on [IIIF Change Discovery API v0.3](#)
- Aggregation mechanisms for IIIF based on IIIF Collections
- Aggregation mechanisms for IIIF based on Sitemaps
- Aggregation mechanisms specific for Schema.org metadata (e.g. HTML Crawlers)
- Aggregation mechanisms for the WWW (HTML meta, RDFa, RDFaLite, Microdata, etc)
- Conversion of Schema.org metadata to EDM
- Metadata profiling

Current operations available in the workbench (for data providers):

- [Register a LOD dataset for aggregation by Europeana:](#)
Allows data providers and aggregators to send us the description of a dataset where we will experiment the LOD aggregation mechanisms. The guidelines for describing a LOD dataset for Europeana are available in [this document](#).
- [Register a IIIF dataset for aggregation by Europeana:](#)
Allows data providers and aggregators to send us the description of a dataset where we will experiment IIIF harvesting mechanisms. IIIF datasets can be used for aggregation via several harvesting methods: [IIIF Change Discovery API v0.3](#), IIIF Collections, and Sitemaps.
- [Register a WWW dataset for aggregation by Europeana:](#)
Allows data providers and aggregators to send us the description of a dataset where we will experiment the WWW/HTML Crawling and aggregation of micro forma Schema.org data).
(a sitemaps.xml file is required)

Current operations available in the workbench (for Europeana aggregation):

- [Manage and process datasets:](#)
Allows the Europeana Aggregation team to manage the registered datasets, execute harvests, metadata analysis and conversion.





LOD aggregator

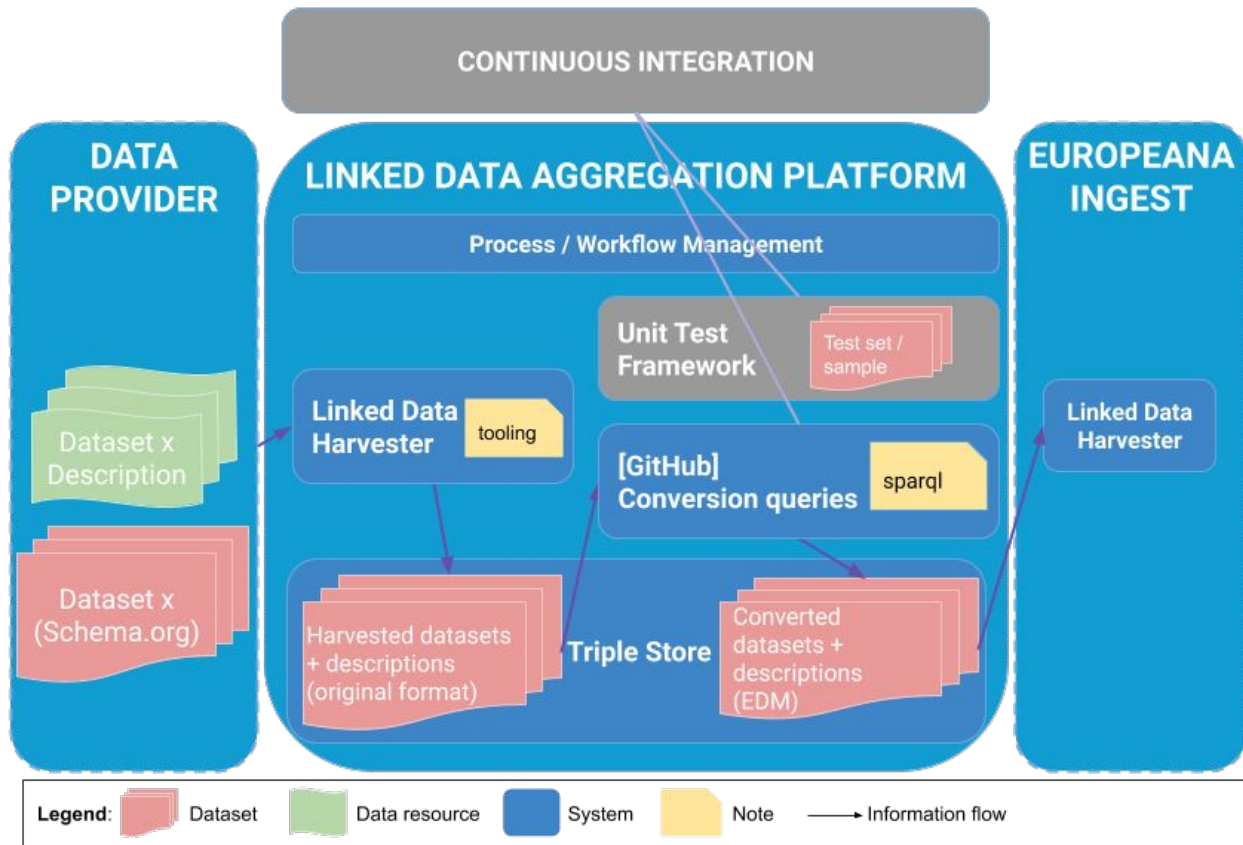
Aggregation of Schema.org for the
Europeana Common Culture (ECC) project



ECC - LOD pilot

- The LOD Aggregation pilot is aimed at bringing the LOD to EDM aggregation route into practice.
- The application setup uses linked data aggregation and Schema.org to EDM transformation





- <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator>
- [Guidelines for providing and handling Schema.org metadata in compliance with Europeana](#)
- [Specifying a linked data dataset for Europeana and aggregators](#)



SURVEY

Survey on alternative aggregation methods



europaea
foundation

SURVEY

- Survey to gauge the use and interest of alternative technologies other than OAI-PMH for (meta)data aggregation
- Target audience: any data providers and aggregators within the Europeana Network
- Further pilot experiments with interested organisations to take place
- <https://forms.gle/iq2fZ8wCgBMGTrDq6> - please fill it out by May 8

Alternative aggregation methods

This is a survey on the interest and possible implementations of alternative methods to OAI-PMH (such as methods based on Linked Open Data, the International Image Interoperability Framework - IIIF, ResourceSync, etc.) that organisations could adopt to have their data ingested on Europeana or other aggregator platforms.

It will also allow pilot experiments to be set up with interested organisations.

This survey, which should take you 10 to 15 minutes, is carried out in the context of a Master's thesis in Information Science with the partnership of Europeana's R&D team.

* Required

Name of the organisation *

Your answer _____

Institution type or institutional domain *

Please the most suitable option, several choices are allowed if your organisation belongs to or represents various domains.

- Gallery
- Library
- Archive
- Museum
- Research institute
- Industry
- Aggregator
- Other: _____

Next

Page 1 of 9



europæana
foundation




europæana.eu



@EuropeanaEU

Julien A. Raemy

 @julsraemy

julien.raemy@hesge.ch

Nuno Freire

nuno.freire@europæana.eu



Co-financed by the Connecting Europe
Facility of the European Union