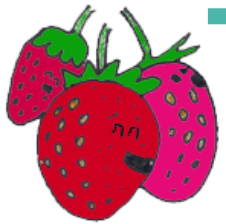
Europeana Data Model Workflows in Archipelago

Open Repositories 2021

Workshop Playground: https://studio.archipelago.nyc/
github.com/esmero/archipelago-deployment



OpenRepositories

Metropolitan New York Library Council
Diego Pino - dpino@metro.org
Allison Lund - alund@metro.org

Links and facts:

Current stable branch is **1.0.0-RC2**
Archipelago is GNU V3
Archipelago has an Advisory Board
Archipelago has a lovely caring Community
Archipelago extends and empowers Drupal

archipelago.nyc
Main Playground (Open year-round!)
Google Group
Technical roadmap
Active development repos
Archipelago 1.0.0-RC2 Specs and features

Link for this workshop:

Today's Playground https://studio.archipelago.nyc

You should have gotten an Account Invitation via email?
Did not? Ask in the Chat and we will resend it.

Workshop Agenda

- Introduction and overview of Archipelago: Greatest Hits Tour & Showcase
- Brief overview of Europeana Data Model (EDM)
- Demonstration, explanation of, and playing with EDM implementation in Archipelago
- Hands-on interactions in Archipelago
- Questions, open discussions, feedback

Housekeeping

- Gentle reminder to **please keep your microphone muted** throughout the presentation portion of the workshop session. We will ask participants to to unmute themselves, if comfortable, during the latter Questions and Open Discussion portion of the workshop (last 15 minutes). **Thank you!**

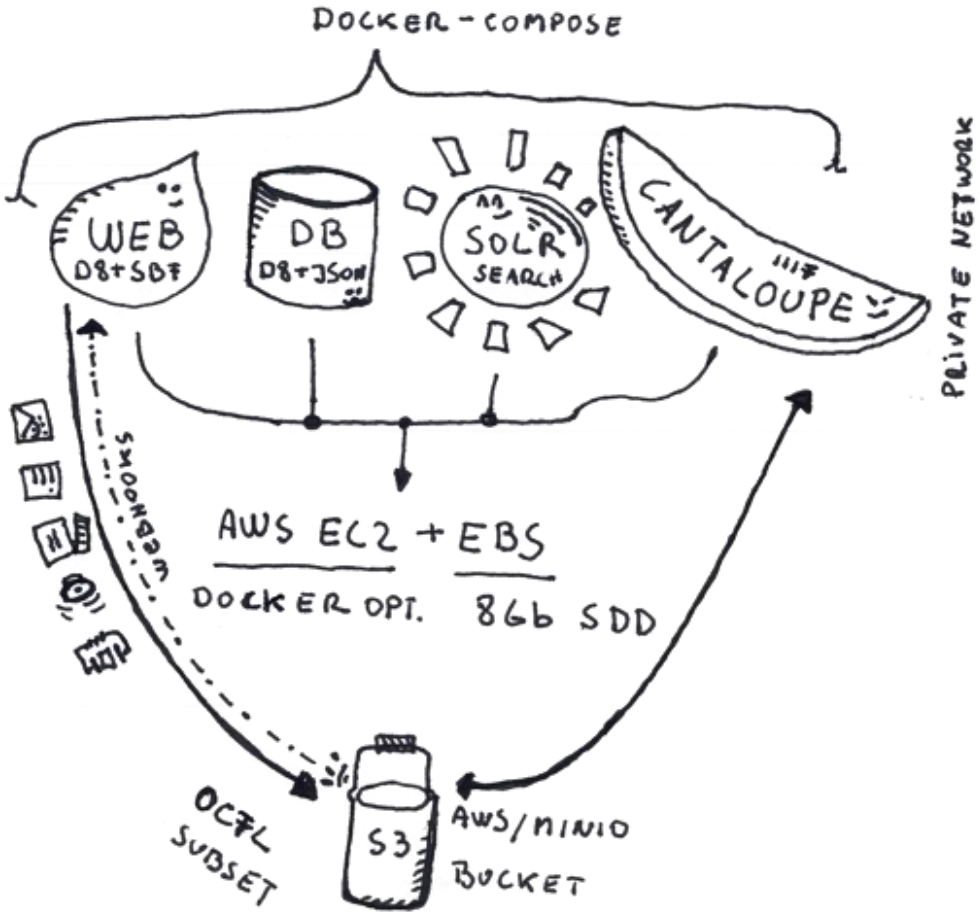
Archipelago is a Digital Objects Repository Architecture

- It has code across 3 Core Drupal 8/9 Modules that take care of the complete ADO Lifecycle
  - **Strawberry Field** (Everything data logic/persistence, low level **CRUD**)
  - **Webform Strawberry Field** (Everything UI/UX Data **Editing**)
  - **Format Strawberry Field** (Everything UI/UX Data **Display**)
- And additional Powerful Drupal 8/9 Modules that extend Archipelago even further
  - **AMI** (Tabular UI/UX Batch driven Importer/updater; Search and Replace using VBO/Actions)
  - **StrawberryRunners**
    - Plugin Driven **Post Processor** for ADOs; Hierarchical plugable Processing
    - Can generate files (WARC to WACZ), new JSON, Search API Index data

Archipelago is a Digital Objects Repository Architecture

- It has a way of **Deploying** and **Running** (we do not like the word devops)
  - A well defined choice of Services (Web/DB/File Storage/Media Access/Discovery)
  - A *Docker Compose* ensemble
  - A set of pure Drupal Configurations to get started
  - Guides from Zero to Running
- It has a public roadmap that evolves with needs, mistakes, lessons learned.

https://github.com/esmero/archipelago-deployment



Archipelago and Metadata

- Metadata for Libraries, Archives, Museums, Cultural Heritage Organizations:
  - **Many** shapes and possibly **hierarchical** / deeply nested
  - **Many Standards and Schemas**
  - **Many** Serializations
  - May **not be totally clear** during “Design”
  - Quite **intimate** and close to the creator’s **identity**
  - **Content** creators know always more than **Developers**

Archipelago and Metadata

- Archipelago's Approach to Metadata:
  - Allow the **shape** of the **metadata** be defined by the creator's needs.
  - Metadata is *plastic, evolving*. Let's honor that **reality.**
  - Remove **Hardcoded Data** models as much as possible from the **Software**
  - Give users **Tools** to decide based on their **own expertise**

Archipelago:

Many Islands, many Realities.

Software as a means of navigating.

In between water, the currents and tides, cohesion.

Canned Food Metadata paradigm

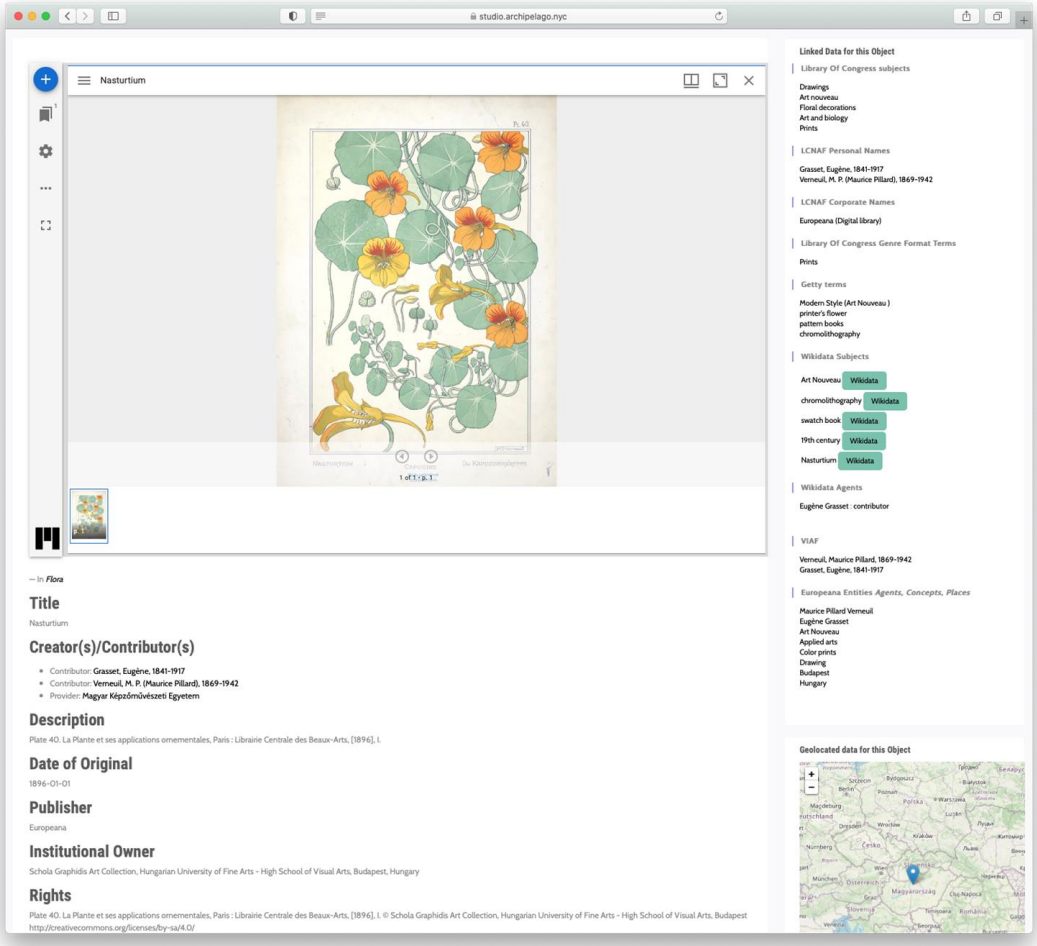
- You can heat it
- You can eat it
- Still MODS
- (or Still Drupal!)

Ingredients + Recipe(s)

- What you already have in your fridge
- Recipes can get better
- Worst case scenario: it tastes like canned food the first time

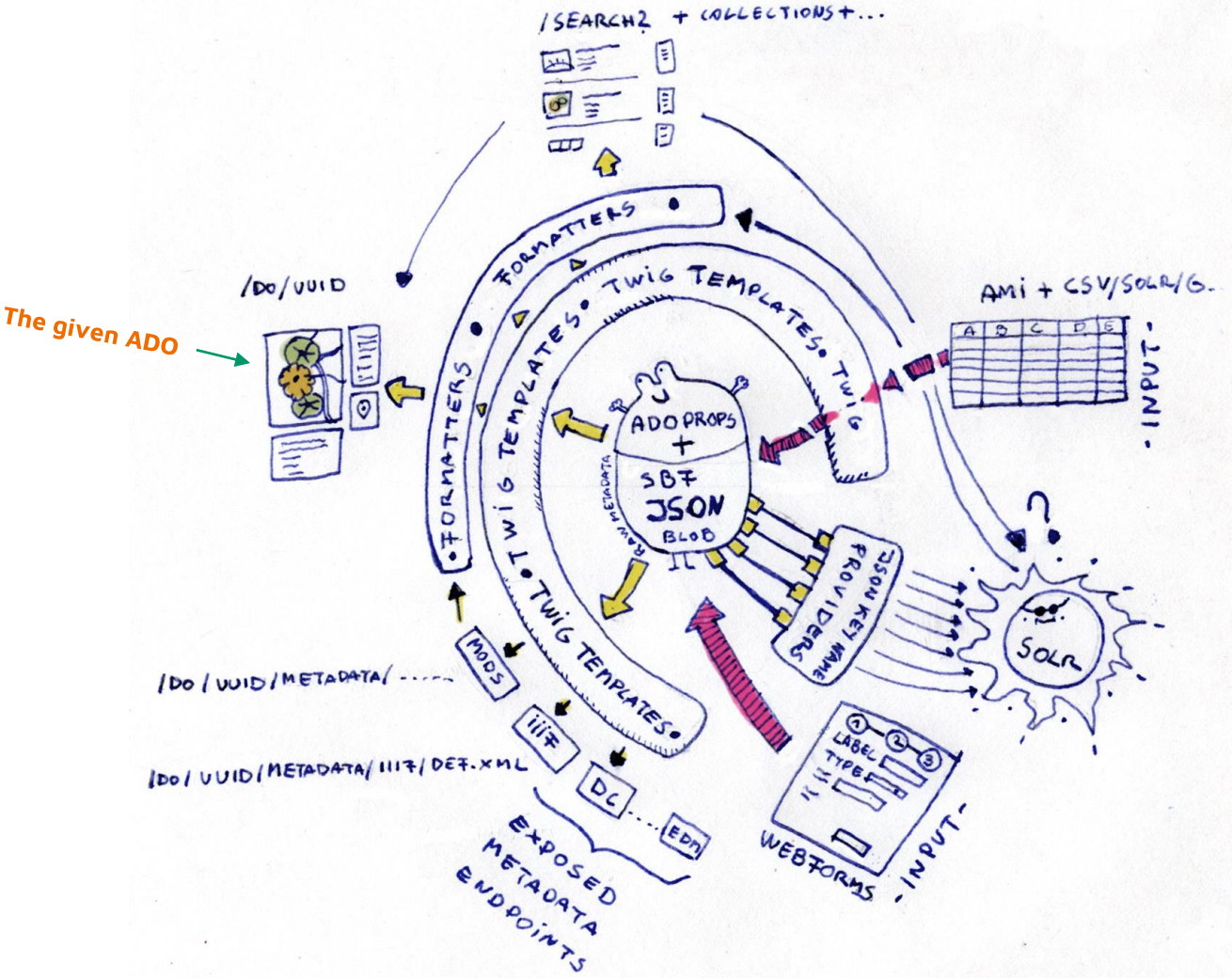
https://studio.archipelago.nyc/do/7012eab5-e754-4519-b8b1-c6be9ceaa24d

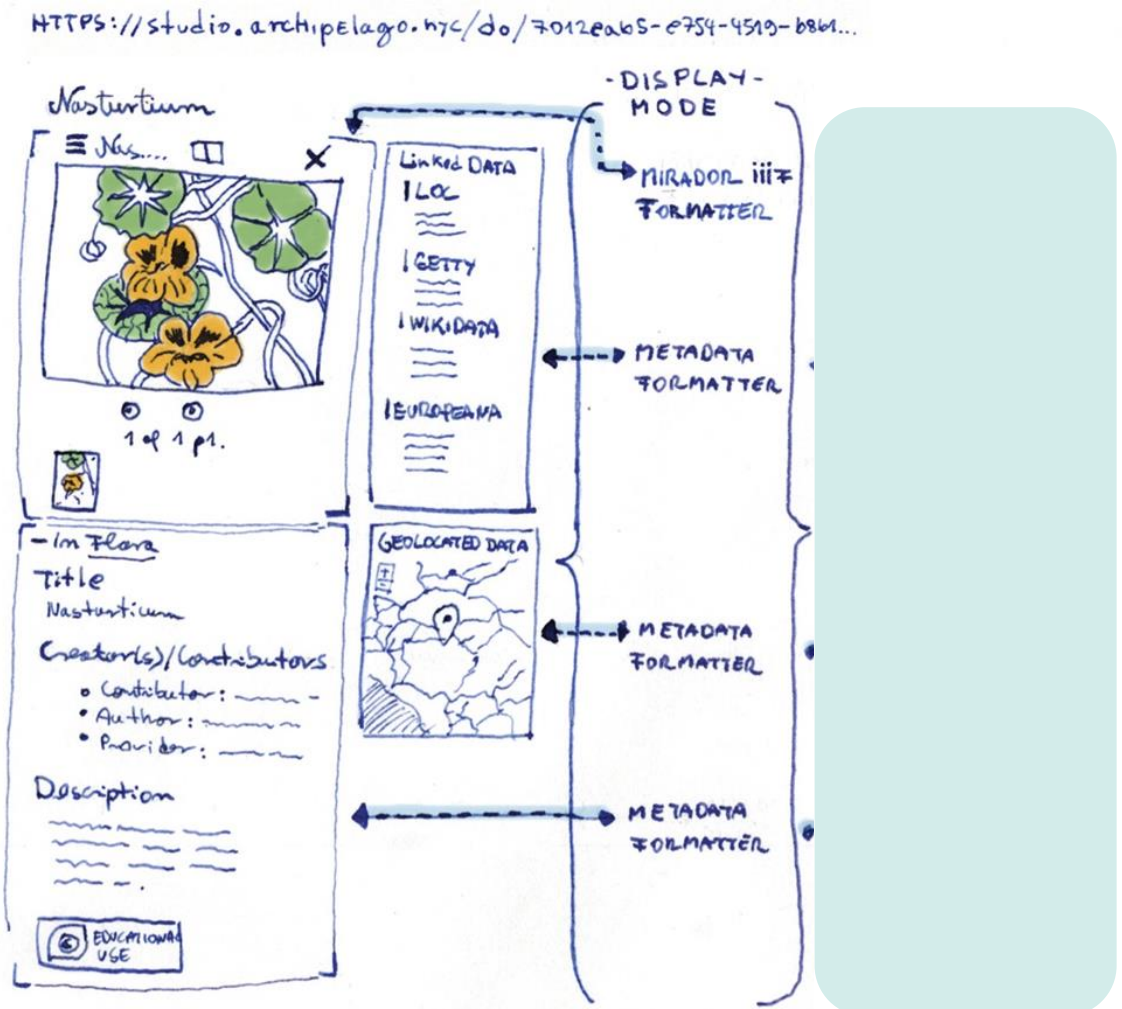
Given this ADO...



Star Diagram

The given ADO





One leg of the Star

This Workshop

What a metadata template is

Digital Object JSON stored in a SBF

```
{
  "type": "Book",
  "label": "Canto General",
  "owner": "Metropolitan New York Library Council",
  "edm_agent": [
    {
      "name_uri": "http://www.wikidata.org/entity/Q34189",
      "role_uri": "http://www.wikidata.org/entity/Q36180",
      "name_label": "Pablo Neruda ",
      "role_label": "writer "
    }
  ],
  "ismemberof": null,
  "description": "Canto General is Pablo Neruda's tenth book of poems. It was first published in Mexico in 1950, by Talleres Gráficos de la Nación. Neruda began to compose it in 1938.\r\n\r\n\"Canto General\" (\"General Song\") consists of 15 sections, 231 poems, and more than 15,000 lines. This work attempts to be a history or encyclopedia of the entire American Western Hemisphere, or New World, from a Hispanic American perspective.",
  "subject_loc": [
    {
      "uri": "http://id.loc.gov/authorities/subjects/sh85103704",
      "label": "Poetry"
    }
  ],
  "local_identifier": "BOOK_001",
  "subject_wikidata": [
    {
      "uri": "http://www.wikidata.org/entity/Q482",
      "label": "poetry "
    },
    {
      "uri": "http://www.wikidata.org/entity/Q1033840",
      "label": "Canto General "
    },
    {
      "uri": "http://www.wikidata.org/entity/Q37922",
      "label": "Nobel Prize in Literature "
    },
    {
      "uri": "http://www.wikidata.org/entity/Q34189",
      "label": "Pablo Neruda "
    }
  ]
}
```

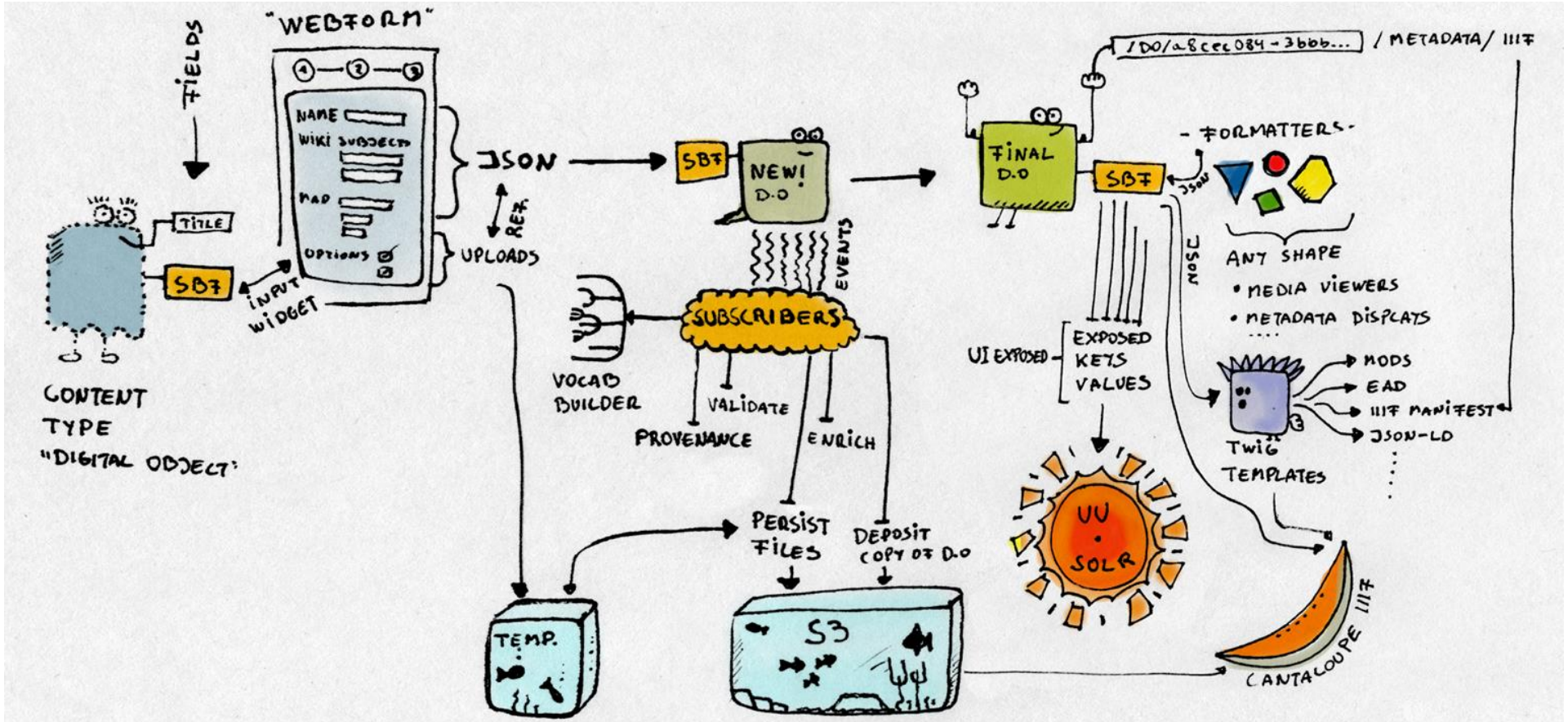
Metadata Display template for EDM

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements
{# -- edm:ProvidedCHO -- #}
<!-- edm:ProvidedCHO -->
{%~ if node.id %}
    {%~ set nodeurl = url('entity.node.canonical', {'node': node.id}, {'absolute': true}) %}
<edm:ProvidedCHO rdf:about="{{ nodeurl }}">
{%~ else %}
    {%~ set nodeurl =  url('<current>') %}
<edm:ProvidedCHO rdf:about="{{ current }}">
{%~ endif %}
{# Required Properties #}
{# dc:Title #}
  <dc:title>{{ data.label|escape('html_attr') }}</dc:title>
{# dc:Description #}
{%~ if data.description|length > 0 %}
    {%~ if data.description is iterable %}
        {%~ for description in data.description %}
  <dc:description>{{ description }}</dc:description>
        {%~ endfor %}
    {%~ else %}
  <dc:description>{{ data.description }}</dc:description>
    {%~ endif %}
{%~ endif %}
{# dc:Language #}
{%~  if data.language|length > 0 %}
    {%~  if data.language is iterable %}
        {%~  for language in data.language %}
  <dc:language>{{ language }}</dc:language>
        {%~  endfor %}
    {%~  else %}
  <dc:language>{{ data.language }}</dc:language>
    {%~  endif %}
{%~  endif %}
{# dc:Subject #}
{# Wikidata Subjects #}
{%~  if data.subject_wikidata|length > 0 %}
    {%~  for subject in attribute(data, 'subject_wikidata') %}
  <dc:subject>{{ subject.label }}</dc:subject>
    {%~  endfor %}
{%~  endif %}
```

About TWIG

- https://twig.symfony.com/doc/3.x/
- https://github.com/esmero/archipelago-documentation/blob/1.0.0-RC2/docs/metadatatwigs.md
- tinyurl.com/archipelagoandtwig

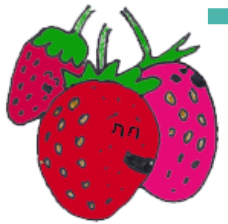
The long and meaningful life of an ADO



Quick Break (3-5 minutes)

Questions? Coffee? User Account Login Issues?

There will be more time for Questions & Comments towards the end of the session. Please also feel free to write directly in the Workshop Agenda & Notes document as well/instead by Friday, June 11th. We will do our best to respond to all questions written there.
Thank you very much!



Europeana - https://www.europeana.eu/

- **Mission**: ***"Europeana empowers the cultural heritage sector in its digital transformation. We develop expertise, tools and policies to embrace digital change and encourage partnerships that foster innovation."***
- About: "Europeana works with thousands of European archives, libraries and museums to share cultural heritage for enjoyment, education and research." … "Access to millions of books, music, artworks and more – with sophisticated search and filter tools to help you find what you're looking for."
- Europeana Pro: "find out about Europeana's technical infrastructure, and how Europeana empowers the cultural heritage sector in its digital transformation."

Basic Overview of the Europeana Data Model (EDM)

- **Europeana Data Model**:
  - https://pro.europeana.eu/page/edm-documentation
    - Factsheet: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Factsheet.pdf
    - Mapping Guidelines: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017
    - Presentation: *https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_slides_130714.ppt

**Sources for main content of slides*

Basic Overview of the Europeana Data Model (EDM)

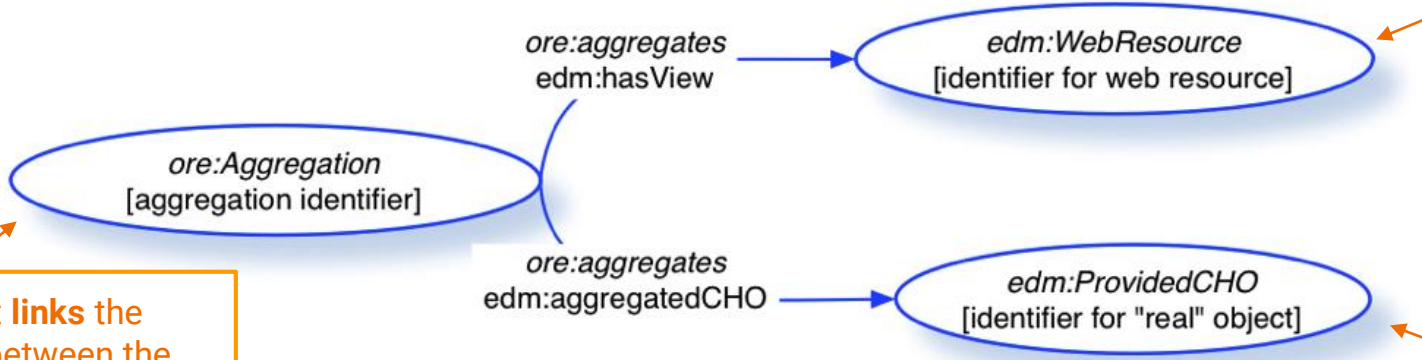
Three Core EDM classes

From the EDM Mapping Guidelines

- "Three core classes: an **Aggregation** with a **Provided CHO** and a **Web Resource**
- Each submission of an item to Europeana should give rise to the creation of three types of "resource" (section of metadata):
  a. one for the source cultural heritage object - **edm:ProvidedCHO** [Cultural Heritage Object]
  b. one for the digital object being submitted - **edm:WebResource** [can be one or more]
  c. one for the overall package - **ore:Aggregation**
- Each will have its own associated metadata, giving three core "metadata sections" per item submitted."

EDM basic pattern

- A data provider submits to Europeana a “bundle” of an object and its digital representation(s)



Class that **links** the properties between the associated classes ; Has additional properties that apply to the whole group (such as edm:dataProvider)

*From: EDM presentation

Descriptive metadata for a cultural heritage object

The ProvidedCHO is the cultural heritage object which is the subject of the package of data that has been submitted to Europeana.

Properties:

dc:contributor, dc:creator, dc:date, dc:format
dc:publisher, dc:relation, dc:source, dcterms
dcterms:temporal, dcterms:medium, dcterms
dcterms:issued, dcterms:conformsTo, dcterm
dcterms:isFormatOf, dcterms:hasVersion, dc
dcterms:hasPart, dcterms:isPartOf, dcterms
dcterms:references, dcterms:isReplacedBy,
dcterms:isRequiredBy, dcterms:requires, dc
edm:isNextInSequence, edm:isDerivativeOf

Mandatory Properties for the ProvidedCHO:
**Included in Archipelago EDM template*

dc:title **or** dc:description
dc:language for text objects
dc:subject **or** dc:type **or** dcterms:spatial **or** dcterms:temporal
edm:type

**Recommended Properties included*
dc:creator, dc:date, dc:identifier;
dc:publisher, dc:source; dc:relation; dc:rights

**Fuller list of properties available for edm:ProvidedCHO on page 8 of the EDM Mapping Guidelines*

**From: EDM presentation*

Digital representations of the object

One or more WebResources are provided for the cultural heritage object.

Properties:

dc:rights
edm:rights
dc:format
dc:description
dcterms:isPartOf
edm:isNextInSequence…

edm:WebResource - *there are no mandatory properties*

XML Snippet from Archipelago EDM template:

```
<edm:WebResource rdf:about="https://studio.archipelago.nyc/do/20/iiif/7691d53e-71d5-45
  <ebucore:hasMimeType>image/jpeg</ebucore:hasMimeType>
  <ebucore:width rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3059</ebucore
  <ebucore:height rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">2470</ebucor
</edm:WebResource>
```

*Fuller list of properties available for edm:WebResource on page 9 of the EDM Mapping Guidelines

*From: EDM presentation

CC BY SA

Aggregations organize data of a provider

The Aggregation represents the set of related resources about one real object contributed by one provider.
It carries the metadata that is about the whole set

Properties for: **ore:Aggregation**

Mandatory:

- edm:aggregatedCHO
- edm:dataProvider
- edm:isShownBy or
- edm:isShownAt
- edm:provider
- edm:rights

Optional:

- edm:hasView
- edm:object
- dc:rights
- edm:ugc

● = *Included in Archipelago EDM template*

edm:dataProvider = The name or identifier of the data provider of the object (i.e. the organisation providing data to an aggregator)

edm:Provider = The name or identifier of the provider of the object (i.e. the organisation providing data directly to Europeana).
**edm:Provider set to "Archipelago Commons Digital Repository" in Archipelago EDM template*

**Fuller list of properties available for ore:Aggregation on page 9 of the EDM Mapping Guidelines*

**From: EDM presentation*

Contextual entities

= *Included in Archipelago EDM template*

Representing (real-world) entities related to a provided object as fully fledged resources, not just strings

edm:Agent
- foaf:name
- skos:altLabel
- rdaGr2:biographicalInformation
- rdaGr2:dateOfBirth

skos:Concept
- skos:prefLabel
- skos:altLabel
- skos:broader
- skos:related
- skos:definition….

cc:License
ordl:inheritFrom *(required)*
cc:deprecatedOn

edm:TimeSpan
- skos:prefLabel
- dcterms:isPartOf
- edm:begin
- edm:end
- ….

edm:Place
- wgs84_pos:lat
- wgs84_pos:long
- skos:prefLabel
- skos:note
- dcterms:isPartOf….

*See page 9 of the EDM Mapping Guidelines for updated list of Contextual Classes and properties

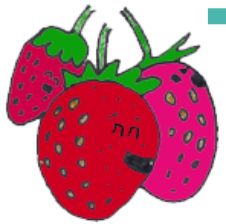
*From: EDM presentation

Europeana Data Model (EDM) in Archipelago

- @https://studio.archipelago.nyc
- Hands-on interactions in Archipelago
  - Using your provided User Accounts
  - Add your own objects, collections, dive deeper
  - Append "**/metadata/edm/default.xml**" to the end of any Digital Object or Collection URL to view the corresponding EDM XML document
    - example: https://studio.archipelago.nyc/do/0fd746cf-e3c4-4f44-a278-b9047a18a5da/metadata/edm/default.xml)
  - View the EDM Metadata Display Entity (TWIG template) here: **must be logged in* https://studio.archipelago.nyc/metadatadisplay/17

Questions? Comments?

Please feel free to write directly in the Workshop Agenda & Notes document as well/instead by Friday, June 11th. We will do our best to respond to all questions written there. Thank you very much!



Thank you!

Cast:

Diego Pino Navarro, Software Developer and A.D. for Digital Strategy at METRO
Allison Lund, Digital Projects & Metadata Librarian at METRO
Giancarlo Birello, Software Developer at CNR Italy
Nate Hill, Director at METRO

Links:

pre-release site
Google Group
technical roadmap
active development repos