

# Citizen Experiences in Cultural Heritage Archives: a Data Journey

Enrico Daga (The Open University)  
enrico.daga@open.ac.uk

7<sup>th</sup> Biennial ISKO UK Conference, Glasgow, 24-25 July 2023



## BACKGROUND

Enrico Daga has carried out R&D on Web Semantics first at the **Italian National Research Council** (CNR) and then at the **Knowledge Media Institute of The Open University** in the UK, where he obtained a **PhD**. Enrico is currently Senior Research Fellow in KMi.



## RESEARCH

His research is exploring novel methods for Knowledge Graph construction (**SPARQL Anything**) and infrastructure, with particular focus on metadata management, *policies* and *process* knowledge; applied to data-intensive scenarios in the smart cities, healthcare, humanities, and cultural heritage domains. He has so far supervised three PhD students to completion.

# Preliminaries

- Cultural heritage institutions (CHI) preserve artistic, historical, and cultural artefacts.
- Faro Convention (2005): Vision to *"Involve everyone in society in the ongoing process of defining and managing cultural heritage"*
- The Warwick Commission (2015): *UK residents from higher socioeconomic groups accounted for 87% of museum visitors, exclusion of BAME citizens*

# State of affairs

New forms of citizen participation in cultural heritage have emerged, producing a **wealth of material** from visitors' experiential feedback on exhibitions and cultural artefacts to **digitally mediated interactions**.



*Crowdsourcing* library metadata for users who are typically happy to volunteer (citizen as content curator)



*Decolonisation* of CH, introduces new perspectives to challenge dominant narratives



*Grassroot archives*: documenting and preserving the experiences of communities outside the mainstream (disconnected from "official" CHI)



*Citizen engagement*: new modalities of interaction (e.g. the gift project)



**Citizen curation** promotes the adoption of intelligent, extended technologies for cultural heritage engagement that mediate the production, collection, interpretation, and archiving of people's responses to cultural objects, favouring the emergence of multiple, sometimes conflicting viewpoints and motivating the users and memory institutions to reflect upon them.



Social cohesion, Participation, and Inclusion  
through Cultural Engagement

<https://spice-h2020.eu/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870811. The communication reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

# Case studies: DMH

Design Museum Helsinki / Aalto University

The Pop-Up VR Museum aims to bridge the **physical accessibility gap**, making it easier for people to experience art and culture



(Photo credit: Gautam Vishwanath)

# Case studies: GAM

Galleria d'Arte Moderna (GAM) /  
University of Turin / Fondazione  
Torino Musei

GAM-Game allows visitors to  
**create stories** to document their  
**moods** and **reactions** to the  
contents they encounter during the  
visit



<https://www.gamtorino.it/>  
<https://www.fondazionetorinomusei.it/>

Lieto, Antonio, Gian Luca Pozzato, Manuel Striani, Stefano Zoia, and Rossana Damiano.  
"Degari 2.0: A diversity-seeking, explainable, and affective art recommender for social  
inclusion." *Cognitive Systems Research* 77 (2023): 1-17.



# Case studies: HECHT

Hecht's Museum / University of Haifa

Historical artefact are linked to events; students elaborate short **essays** making emerge diverse **opinions** regarding historical and national issues



(Photo credit: Joel Lanir)

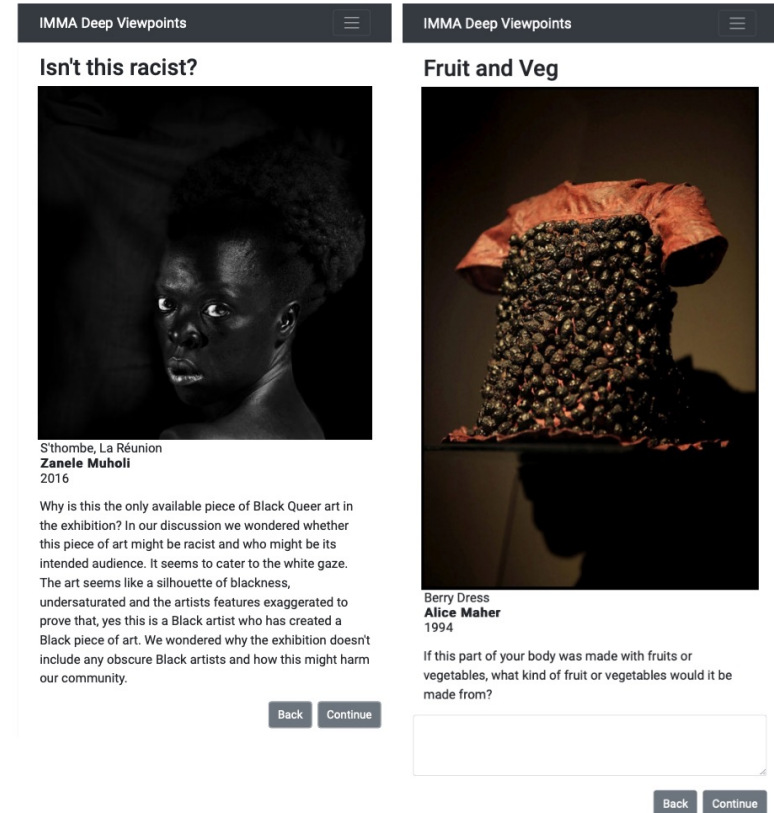
<https://mushecht.haifa.ac.il/>



# Case studies: IMMA

Irish Museum of Modern Art (IMMA)  
/ The Open University

The Deep Viewpoints system is based on the slow-looking methodology, where the experience of the **artworks** is mediated through **prompts and questions**

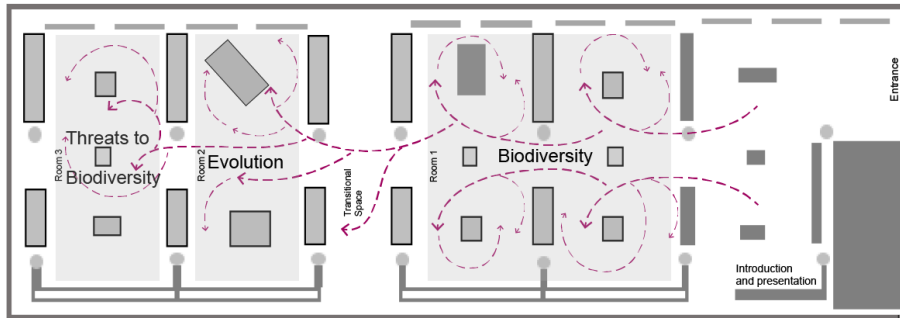


Mulholland, Paul, Adam Stoneman, Naomi Barker, Mark Maguire, Jason Carvalho, Enrico Daga, and Paul Warren. "The Sound of Paintings: Using Citizen Curation to Explore the Cross-Modal Personalization of Museum Experiences." In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, pp. 408-418. 2023.

# Case studies: MNCN

Museo Nacional Ciencias Naturales (MNCN) / Padaone Games

A treasure hunt game mixes **puzzles**, **quizzes** and **questions** linking objects in the natural history museum to themes of environmental sustainability



<https://www.padaonegames.com/>

(Photo credit: Lily Diaz-Kommonen)

# SPICE

## Linked Data Hub

Data infrastructure and tools to support Citizen Curation



### SOURCE DATA COLLECTIONS

IMMA  
ÁRÁS NUA-EALAINE  
NA HEIREANN  
IRISH MUSEUM OF  
MODERN ART

design museum  
helsinki

GAM  
GALLERIA  
CIVICA  
D'ARTE  
MODERNA  
CONTEMPORANEA  
TORINO



SPARQL ANYTHING

<https://sparql-anything.cc/>

### LINKED DATA HUB PORTAL



DATA CATALOGUE



METADATA



LICENSE NEGOTIATION



FILE TOOLS



API KEY MANAGEMENT



LOCATION TOOLS



JSON TOOLS



ACCESS NEGOTIATION



SPARQL TOOLS



NOTIFICATIONS



LINKED DATA HUB REST API



SPARQL ENDPOINT



JSON & FILE STORE



GRAPH DATABASE



KNOWLEDGE GRAPH

### ADD-ON TOOLS



CONTENT MONITORING

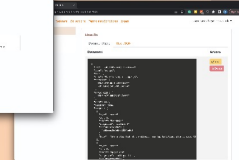
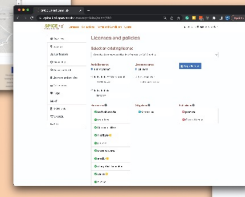
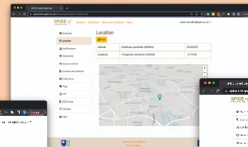
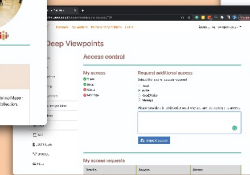
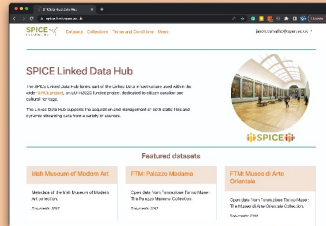
- Privacy and personal details
- Hate speech detection

External APIs and Services



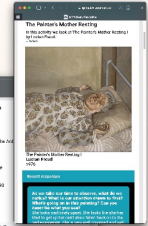
### LINKED DATA TOOLS

- JSON <=> RDF replication
- Graph management
- Custom graph construction



### PILOT APPLICATIONS

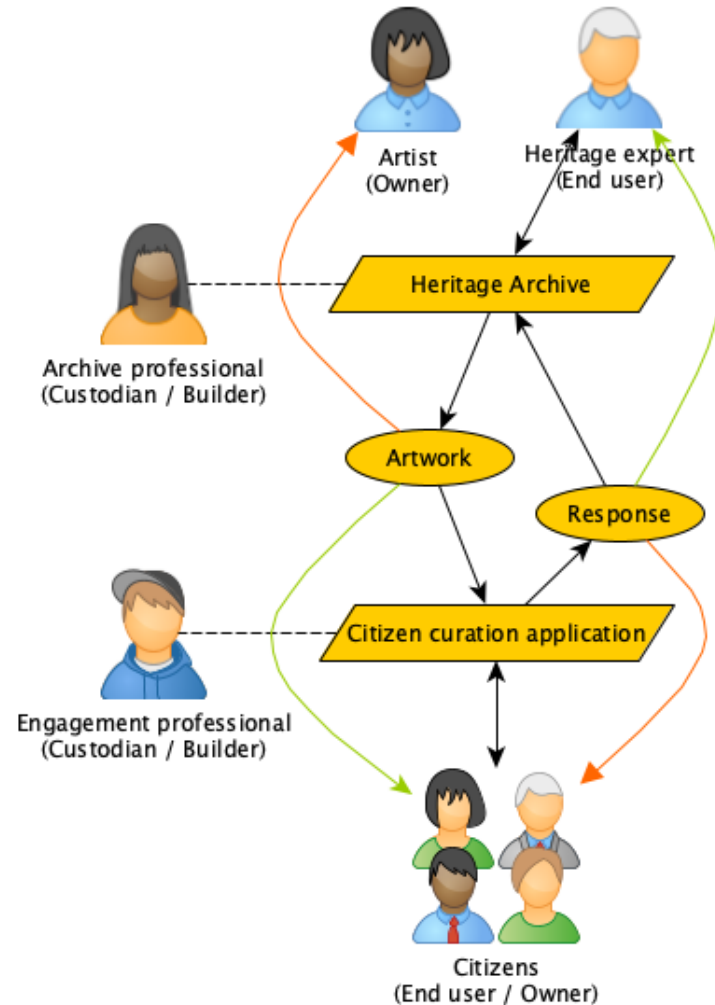
- IMMA Deep Viewpoints
- Helsinki Pop-UP VR museum
- Design Museum Deep Viewpoints
- GAM Game
- Madrid Treasure Hunt



Citizen curation generates a wealth of data that is of great interest to cultural heritage institutions (CHI)

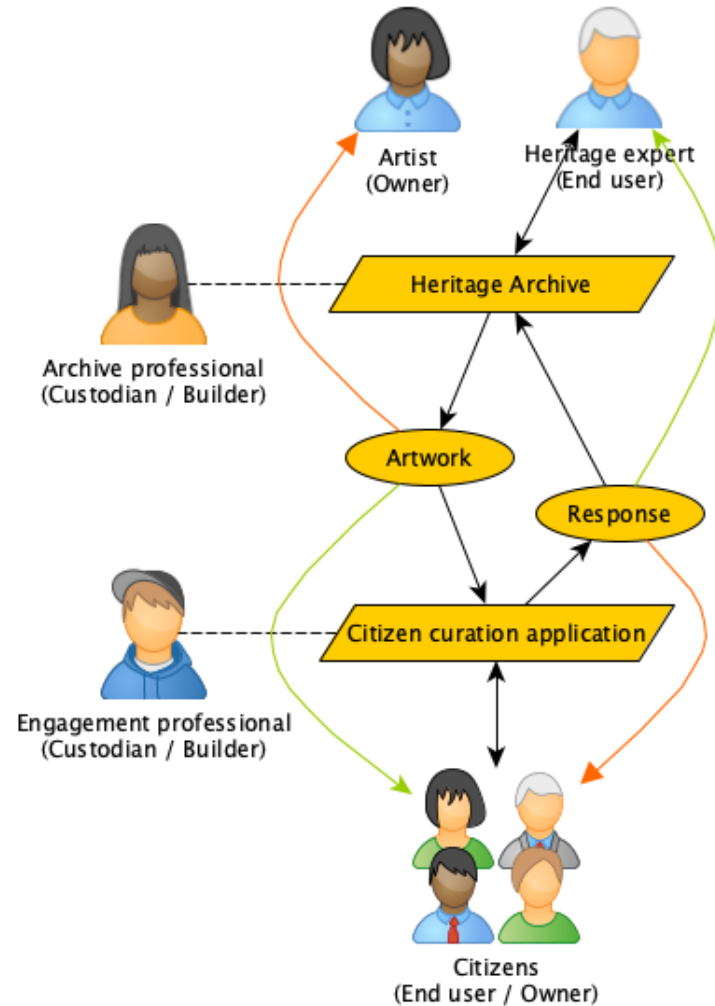
# Sense-making

To make sense of citizen responses, capturing the modality of their production is fundamental, we need to describe citizen curation applications alongside the generated content



# Ownership

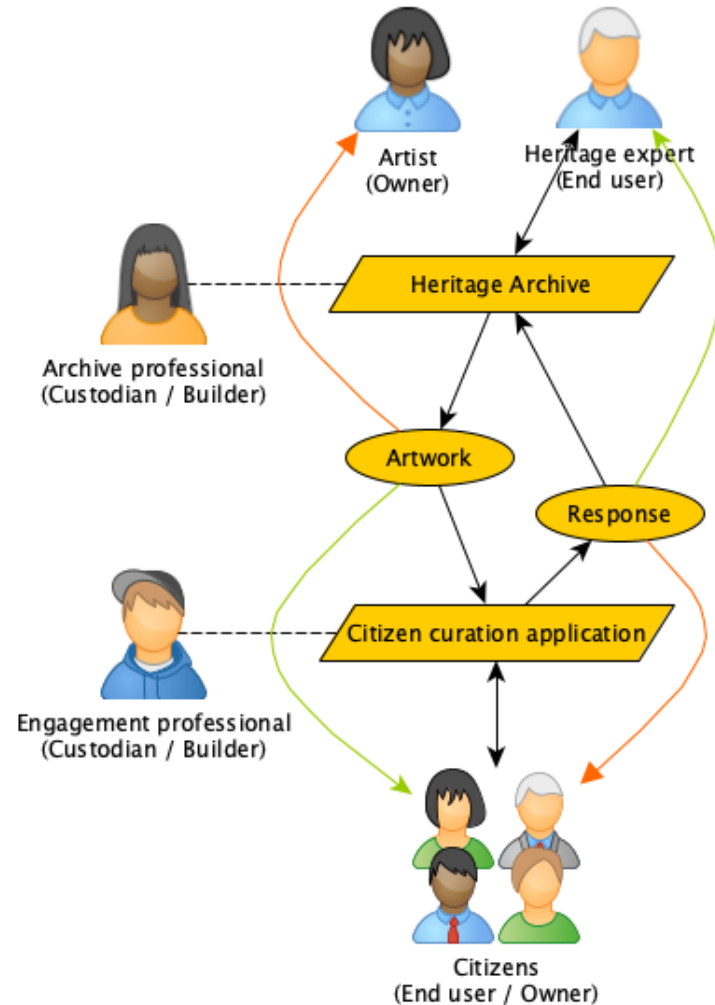
Users produce original content, and there is a question of whether they should be recognised as authors and whether such recognition should be promoted by the cultural heritage institution, for example, with attribution statements



# Sensitive content

The produced content can include inappropriate or sensitive information, personally identifiable information or may violate privacy regulations.

Museums could be seen as being responsible for, or endorsing, opinions found in user-generated content.

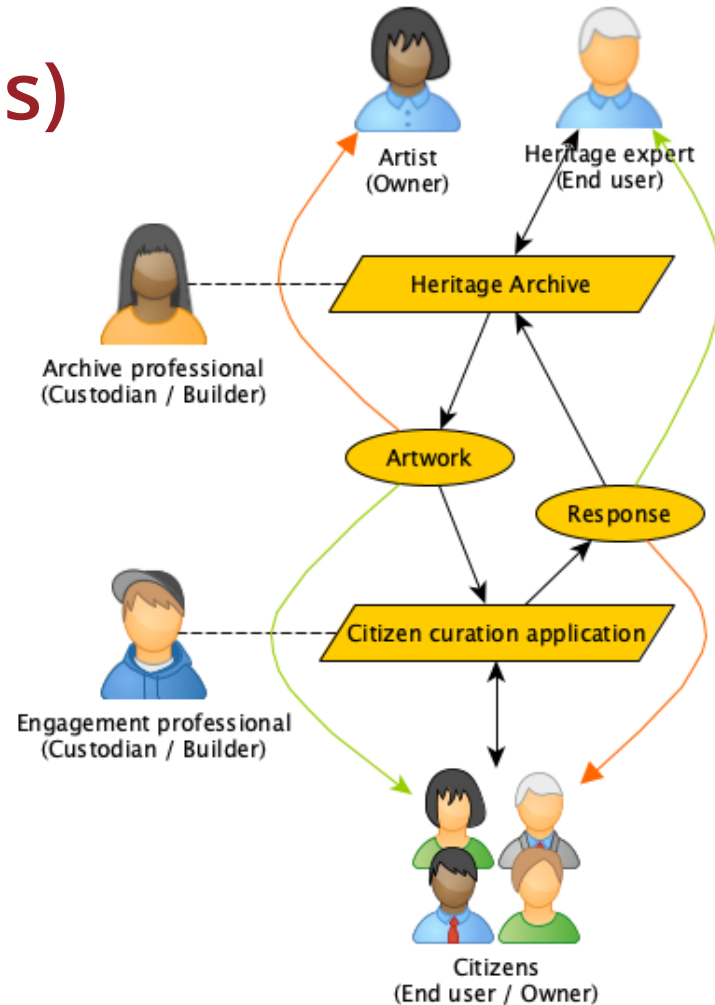




# Terms of use (policies)

Users should be **informed** of usage restrictions when designing the experiences.

These systems aim to support citizens in sharing their contributions with the museum and each other, opening essential issues regarding **rights** and terms of use of the *generated/combined* content.



What Knowledge Organisation System can support citizen curation?

# Data Journeys in the sciences

- In **data studies**, with the notion of data journeys, defined it as the *“movement of data from their production site to many other sites in which they are processed, mobilised and re-purposed.”*
- Strongly related to:
  - Provenance (Information Science)
  - Workflows (Computer Science)
- Semantic Web precursors:
  - Prov-O, Workflow Motifs, Datanode Ontology

# The Data Journeys Ontology (DJO)

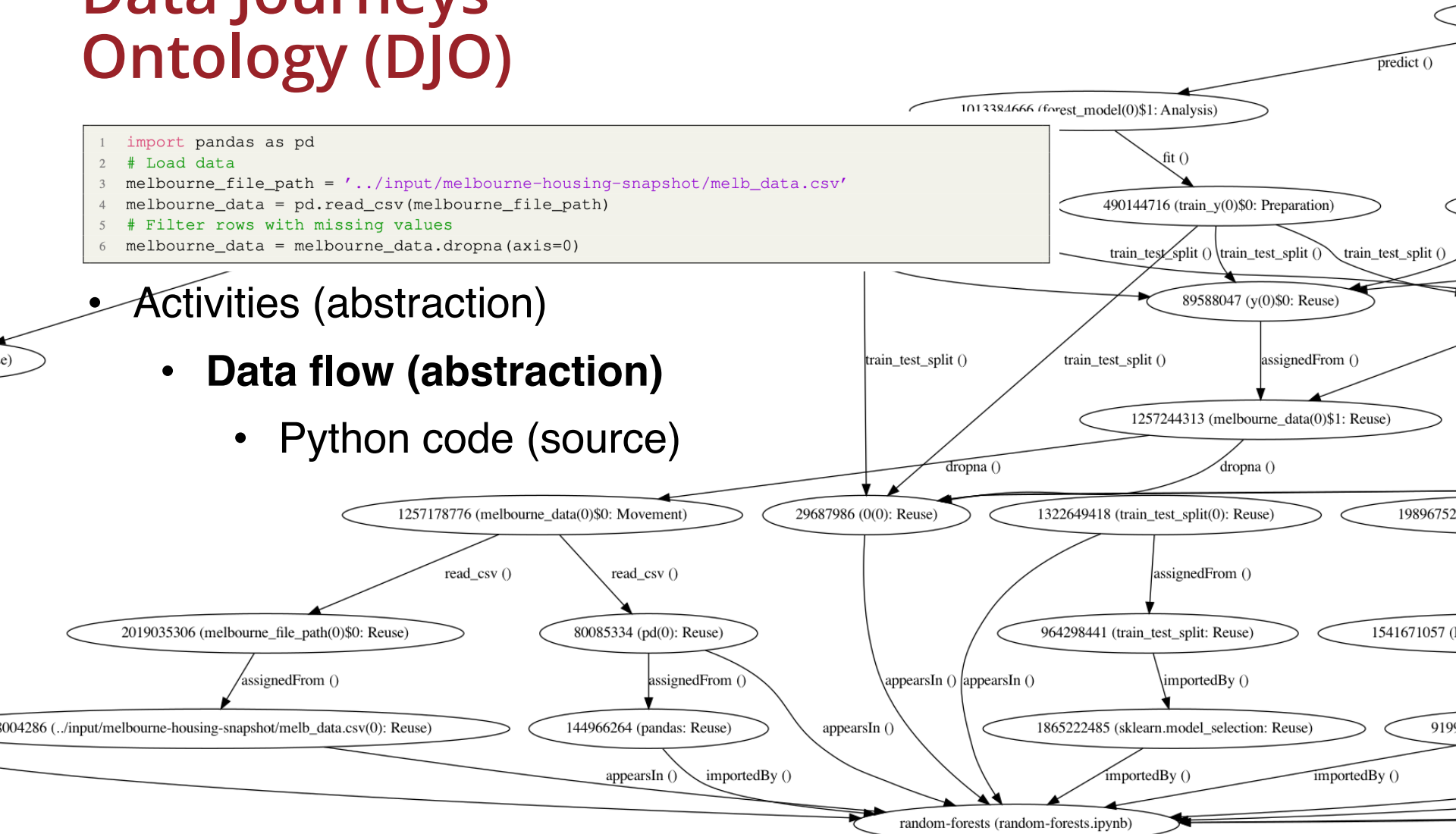
- Designed to **explain data science** pipelines
- A layered approach:
  - Activities (abstraction)
    - Data flow (abstraction)
      - Python code (source)

*Daga, Enrico, and Paul Groth. "Data journeys: explaining AI workflows through abstraction." Semantic Web Preprint (2023): 1-27.*

# Data Journeys Ontology (DJO)

```
1 import pandas as pd
2 # Load data
3 melbourne_file_path = '../input/melbourne-housing-snapshot/melb_data.csv'
4 melbourne_data = pd.read_csv(melbourne_file_path)
5 # Filter rows with missing values
6 melbourne_data = melbourne_data.dropna(axis=0)
```

- ~~Activities (abstraction)~~
  - **Data flow (abstraction)**
    - Python code (source)

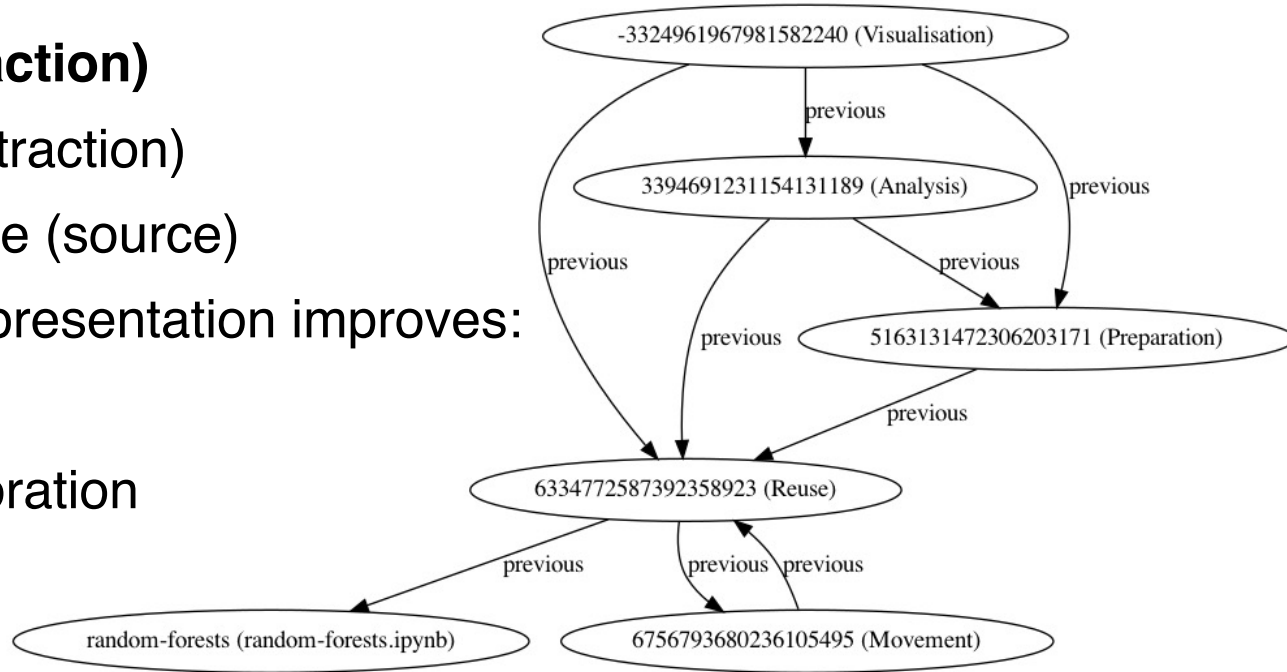


# Data Journeys Ontology (DJO)

- **Activities (abstraction)**
  - Data flow (abstraction)
  - Python code (source)

A more compact representation improves:

- Sense-making
- Navigation / Exploration
- Metadata curation

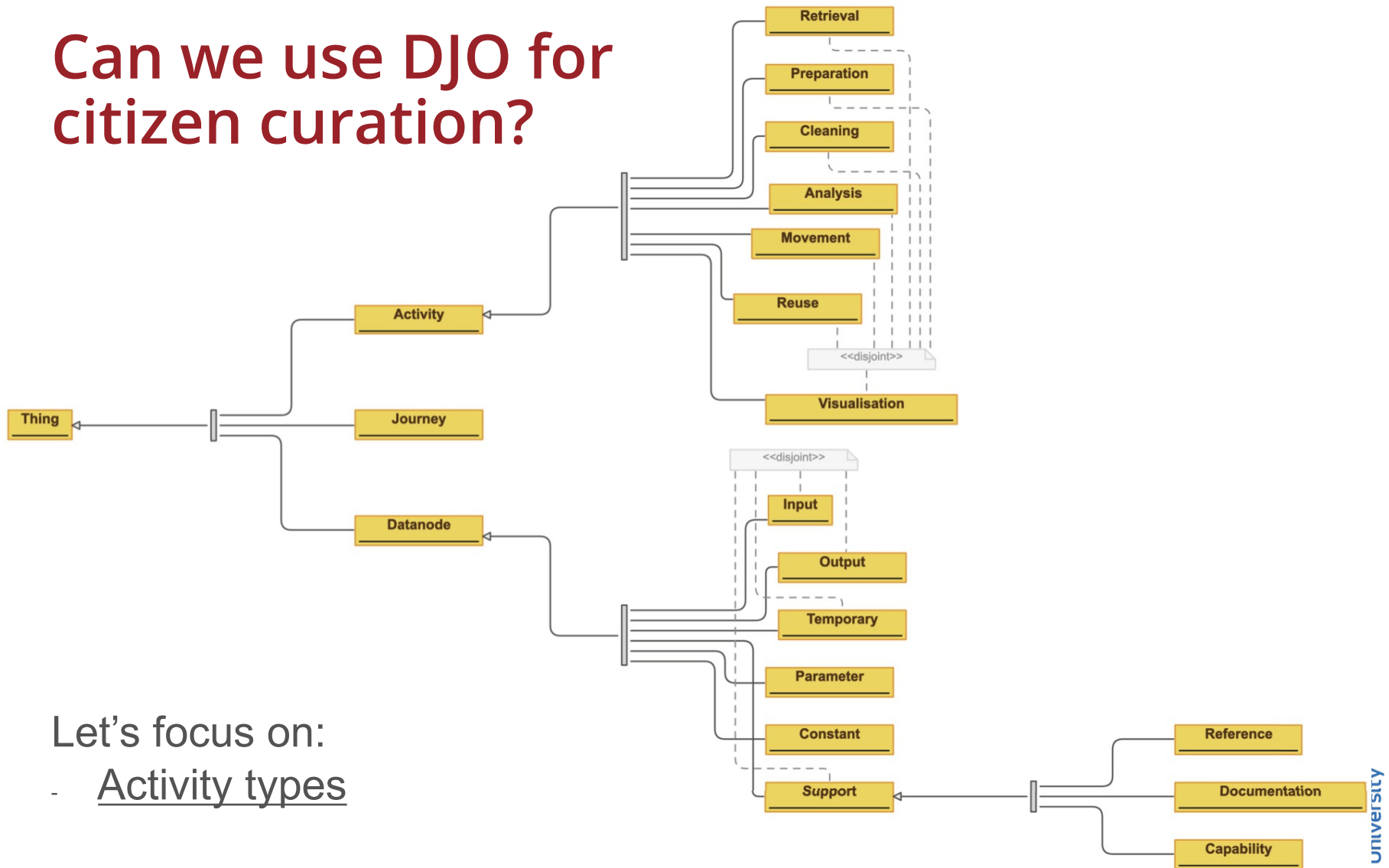


Daga, Enrico, and Paul Groth. "Data journeys: explaining AI workflows through abstraction." *Semantic Web Preprint* (2023): 1-27.

What are the similarities between DJO and CC?



# Can we use DJO for citizen curation?



Let's focus on:

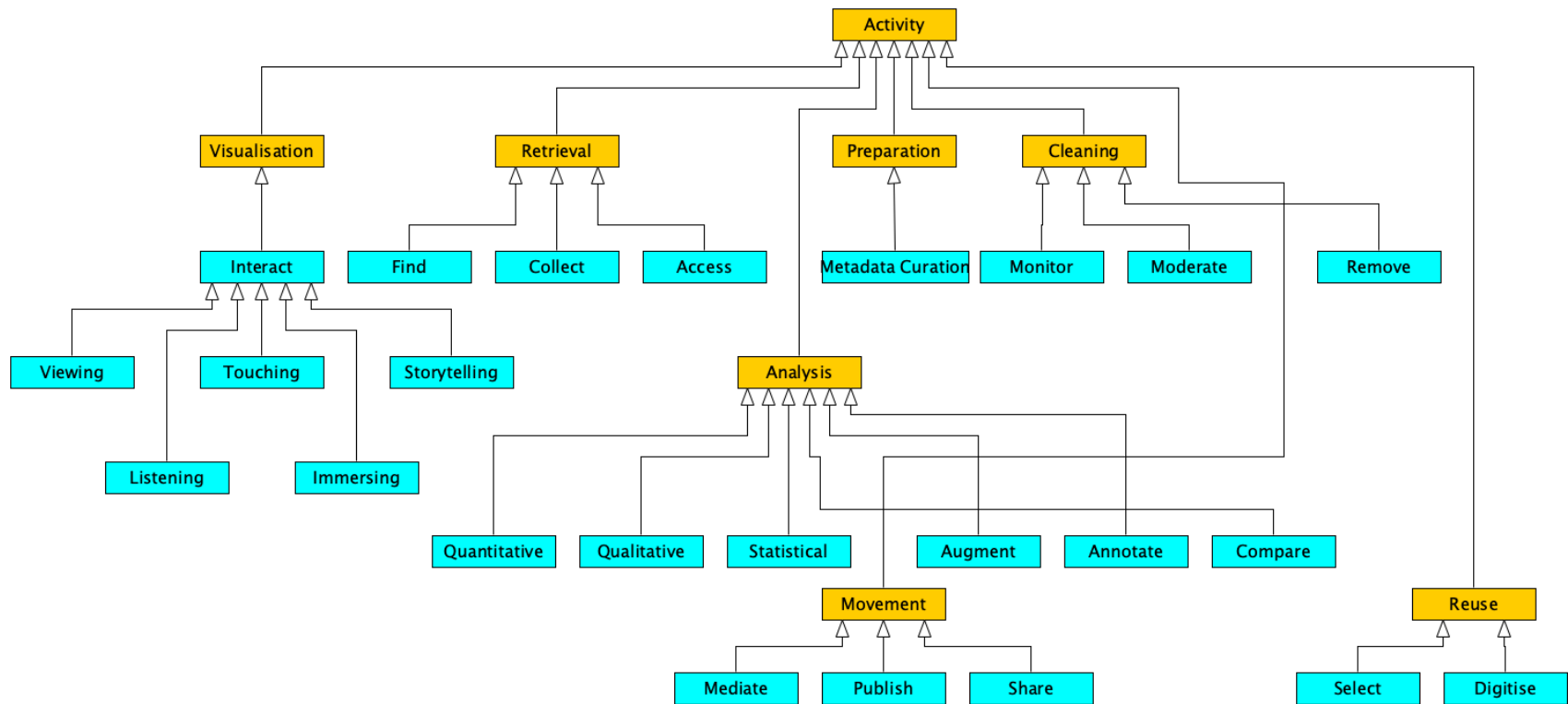
- Activity types

# Analysis of case studies

Case Study	Retrieval	Preparation	Cleaning	Analysis
<b>DMH</b>	Access Collect	VR/AR digital object setup (Curate) Metadata preparation (Curate)	-	Annotate (Human mediated analyses / Augment)
<b>GAM</b>	Access Collect	Metadata preparation (Curate)	-	Find similarities or differences (Compare) Annotate (Augment) Recommend
<b>HECHT</b>	-	-	-	Find similarities or differences (Compare) Statistical (Quantitative Analysis) Thematic analysis (Qualitative Analysis) Effect (Impact Assessment)
<b>IMMA</b>	Find Access Collect	Design script (Curate)	Monitor Moderate Delete responses (Remove)	Find similarities or differences (Compare) Thematic analysis (Qualitative Analysis) Statistical (Quantitative Analysis)
<b>MNCN</b>	Access Collect	Metadata preparation (Curate)	-	Find similarities or differences (Compare) Annotate (Augment)

# Analysis of case studies

Case Study	Movement	Reuse	Visualisation
DMH	Share Publish Data hub (Mediate)	Select Scan artifact (Acquire)	Engage Sense-making Multi-Modality Immersive VR-Interaction AR-Interaction
GAM	Share Publish Data hub (Mediate)	Select	Storytelling Explore Receive recommendations
HECHT	Share Publish Data hub (Mediate)	-	Explore
IMMA	Share Publish Data hub (Mediate)	Select	Explore Storytelling Wooden model (Cyber-physical)
MNCN	Share Publish Data hub (Mediate)	Select Scan artifact (Acquire)	Games, Puzzles, Treasure Hunt, Cyber-physical



CC activities nicely extend DJO activities

# Analysis of resource types

Case Study	Types
DMH	Artifact, Design Objects, Stories, Text, Audio, Video, 3D objects
GAM	Artwork metadata, Images, Comments, Emoticons, Emotions, Stories, Characters, Focus groups, Online survey, Ethnographic observation
HECHT	Dilemma (Prompt), Stories (Autoethnographies), Photos
IMMA	Artwork metadata, Artwork images, Interview, Survey, Question, Prompt, Stories (Autoethnographies), Scripts, Response, Text, Choice
MNCN	Images, Essay, Game, Puzzle

# Data journeys for CC ?

- CC differs from data science pipelines:
  - A considerable heterogeneity of activities
  - Operations are distributed on many systems
  - The process is not exposed
  - Not a linear activity (a lot of interaction)
  - A large variety of resource types (instead of just *data*)
- How data journeys can be applied to citizen curation?
- **How to capture data journeys for citizen curation?**

# Data journeys for CC: layers

## DJO:

- Activities (abstraction)
  - Data flow (abstraction)
    - Python code (source)

## CC:

- **Activity graph** (extended)
  - **Data flow graph**
    - **NEW! Event graphs:** *a mediator such as the LDH can preserve event traces. E.g. a user selects an artwork or a curator selects a response to be included in the archive.*
      - **Resources:** *artwork images, metadata records, data sources, licencing information, and terms of use*
      - **Policies graph:** *metadata about ownership, licences, and terms of use, ...*



# Why data journeys for CC

- Sense-making: DJs provide the context to interpret user responses;
  - Support analytics covering multiple dimensions of interest to museum professionals and researchers,
  - Preserve responses / emotions, preferences, and choices -- in context
  - Study similarities and differences across communities of users
- Monitoring sensitive content:
  - DJs allow to trace content and track changes
- Ownership:
  - DJs can trace who contributes what (event graphs)
- Terms of use:
  - DJs know what is used, where, how, and by whom

# Furthermore

- Sense-making, culture, and reception:
  - Heritage institutions are typically strongly characterised both geographically and culturally (eg. re. sensitive, offensive material)
  - DJs may support the study of how reception changes with changing cultural norms.
- Granularity of terms of use:
  - Linked data and related graph technologies can help specify terms with a high degree of granularity, pointing to collections, items, or their parts. There are ontologies for that as well (W3C ODRL)
- Composite objects.
  - DJs allow us to capture how assets are used and reason upon the compatibility of rights when joining content in composite objects.

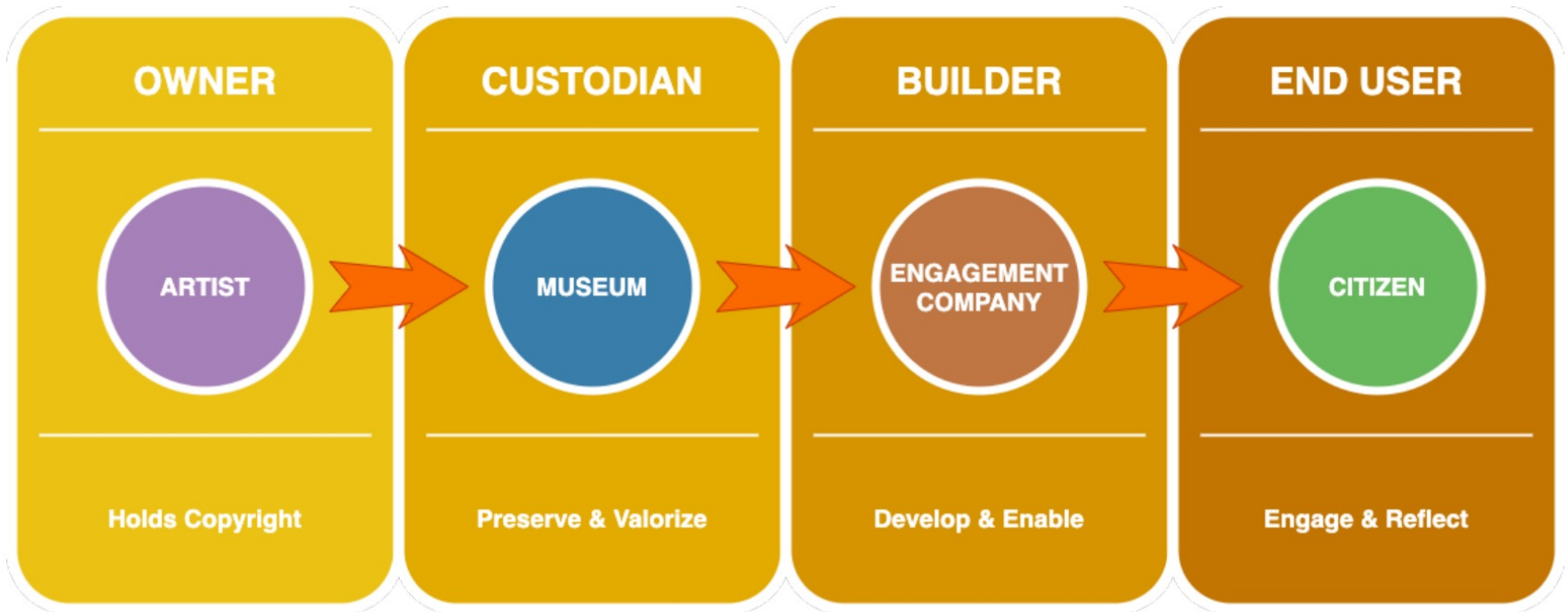
# Conclusions

- Citizen curation generates a wealth of user-generated content that is of great value to CHIs.
- A Knowledge Organisation System based on Data Journeys can help in capturing the complexity of Citizen curation to support sense-making, content moderation, and rights management.
- DJs can be captured by focusing on event traces
- To support this vision, the role of mediators is fundamental
- A distributed infrastructure supporting DJs would allow to capture user events and support CHI in reviewing and curating citizen-contributed content.
- It is an open question what type of infrastructure that could be (but there are some candidates, e.g. distributed social networks, solid, ...)

Thank you.

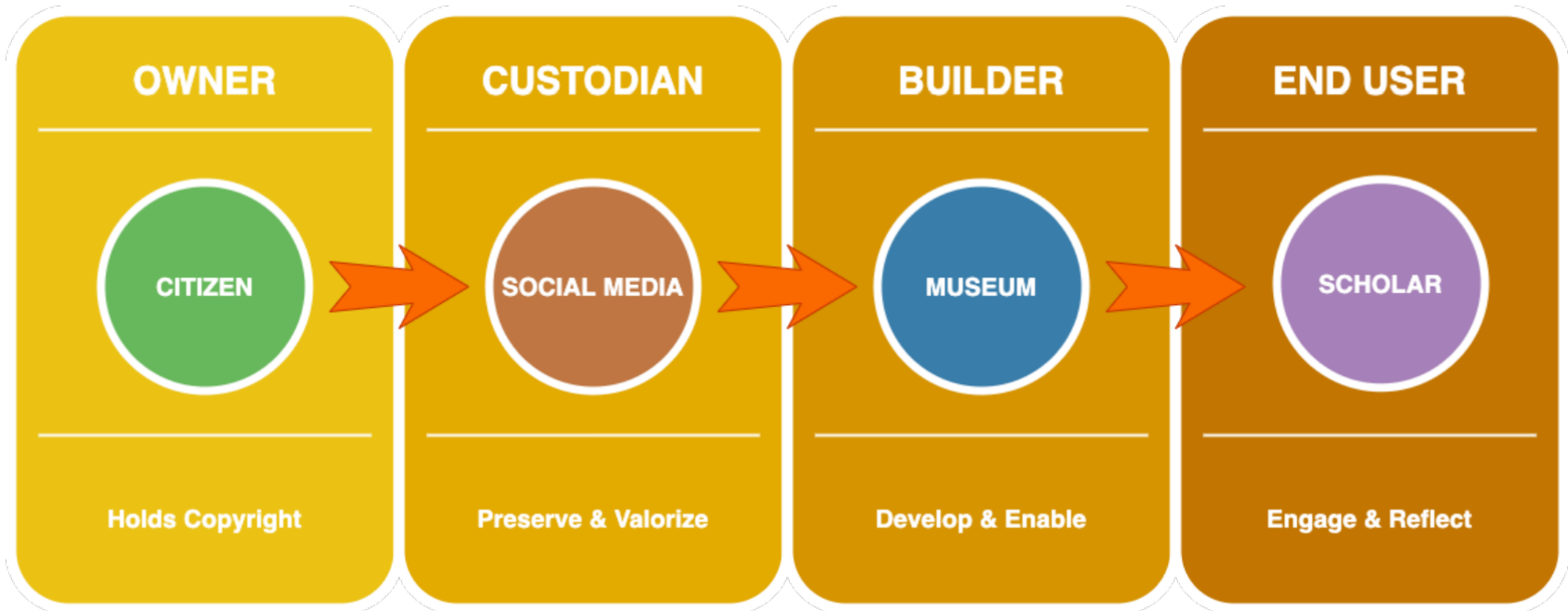
Questions?

# From artists to citizens



Daga, E et al. "Integrating citizen experiences in cultural heritage archives: requirements, state of the art, and challenges." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15, no. 1 (2022): 1-35.

# From citizens to scholars



Daga, E et al. "Integrating citizen experiences in cultural heritage archives: requirements, state of the art, and challenges." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15, no. 1 (2022): 1-35.