Discipline-specific ORDM: The case of Digital Humanities

ORDM – Train the trainers event

Georgios Artopoulos
The Cyprus Institute

Structure

- Which are the challenges in the Humanities
- How ORDM translates to the field of Digital Humanities
- Which are the best practices for each step of ORDM
- Examples

Who is involved:

- *Researchers, scholars and research performing organisations*
- *Content & service provides (GLAM)*

Structure

- Barriers to Open Access for Humanities & Cultural Heritage
- Demonstrated Benefits of Open Access for Humanities
- Clarifying “Open”
- Accessibility
- Open Licensing, Privacy, Sensitivity and Ethical Considerations
- GLAM good practices
- Technical Standards and Sustainable Infrastructures
  - Metadata
  - Digitization
  - Digital preservation
  - Interoperability
- New Areas of Focus through Examples and select cases

What is open access?

A widely used definition of “open” is provided by the Open Knowledge Foundation: “A piece of data or content is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike”.

Humanities

Humanities traditionally have not played a relevant role in the impact of research and valorization field at a time when this is increasingly necessary.

The publication in peer review journals is not a standardized practice, therefore there is no rigorous external evaluation of the results with the obligation to publish the data.

Open Science and the Humanities

Reasons:

- Values, conditions and methods of the humanities.
- Funding.
- The lack of licenses or patents.
- The hybrid ownership nature of humanities sources.

Incentives:

- Wide and fast dissemination.
- Increasing the impact and visibility of humanities research work.
- Comply with / prepare for directives.

Humanities

What is Data?

For the humanist, primary sources, secondary sources, theoretical texts, methodological tools, digital tools, notes, annotations, references … all of these comprise research data in the humanities.

Challenges

Unstructured data

Digitization initiatives

Store, share and analyse their data

Data ownership and rights management

Challenges: The unique case of the Humanities

Humanities data: analogue!

Multimodal knowledge creation systems

The semiotic systems of humanities data can be quite personal and individual

Open Access to Humanities content

"Cultural heritage belongs to everyone. It was created by – and for – all kinds of people. The digitization of physical heritage objects enables them to move out of storage rooms, library shelves, and file drawers, and land in the hands of the world's citizens. When cultural heritage is digital, there is nothing standing in the way of sharing and reusing it. […] When cultural heritage is digital, open and shareable, it becomes common property. It becomes a part of us."

Sanderhoff, M. (2014). Sharing is caring: Openness and sharing in the cultural heritage sector. Copenhagen: Statens Museum for Kunst.

Humanities content in the Public Domain

CH institutions limited access to additional funding.

Commercial use (licensing digital content) <-> Open Access <-> copyrights / rights clearance

Collections inaccessible online.

Creative Commons licensing

The basic set consists of six CC licenses, which all have as a basis the condition “Attribution” (BY)

The CC-BY license is the most liberal license which “lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation”
“ShareAlike”
“No Derivatives”
“NonCommercial”

CC0 license (i.e. “all rights granted”)
Public Domain Mark for works that are free of known copyright, typically very old works

The most adopted licenses

CC0 waiver
https://creativecommons.org/publicdomain/zero/1.0/

CC BY (Attribution only)
https://creativecommons.org/licenses/by/4.0/

CC BY-SA (Attribution ShareAlike)
https://creativecommons.org/licenses/by-sa/4.0/

Licenses for data and databases: Open Data Commons.

IPR and licensing issues: UK Strategic Content Alliance (JISC 2009).

Open educational resources

A lot of success has been achieved by the Open Access movement in the field of education, training and lifelong learning (Geser 2007; OECD 2007; Kanwar et al. 2011; SURF 2013).

UNESCO,
OER Commons,
Open Education Consortium,
Commonwealth of Learning.

Opportunities

Open Access collections have many advantages for cultural institutions: Cultural institutions should be mediators. Open access and “social web” tools can play a major role in meeting this goal.

Opportunities

Cultural heritage & other humanities collections in the Public Domain have a great potential to enable creativity and economic growth: The re-use of the cultural content that is available in the Public Domain allows cultural and creative businesses to create new products.

Amended European Union Directive on the Re-use of Public Sector Information

Cultural heritage institutions that are public sector bodies according to the definition of the Directive 2013/37/EU are obliged to allow re-use of the content they hold.

This includes:
(a) any content, whatever its medium (written on paper or stored in electronic form or as a sound, visual or audio-visual recording); or
(b) any part of such content.

A report by the ePSI Platform (2014) provides more information for cultural heritage institutions LAPSI 2.0 project.

Challenges

The amended Directive on Public Sector Information does not solve the practical issues faced by cultural institutions concerning IPR.

Opportunities

Europeana standards and guidelines: Europeana has played an important role in creating standards and guidelines for legitimate re-use and redistribution of cultural content.

- Europeana Licensing Framework
- Digital Public Library of America

Examples & good practices: OpenGLAM

An OpenGLAM institution

1. Releases digital information about the artefacts (metadata) into the public domain using an appropriate legal tool such as the Creative Commons Zero Waiver.

Examples & good practices: OpenGLAM

An OpenGLAM institution

2. Keeps digital representations of works for which copyright has expired (public domain) in the public domain by not adding new rights to them.
Creative Commons Public Domain Mark.

see:
The Rijksmuseum
The British Library
The Walters Art Museum

For more detailed documents and charters on the importance of the digital public domain see:
Europeana Public Domain Charter
Communia Public Domain Manifesto

Examples & good practices: OpenGLAM

An OpenGLAM institution

3. When publishing data makes an explicit and robust statement of expectations with respect to reuse and repurposing of the descriptions, the whole data collection, and subsets of the collection.

see:
The Rijksmuseum
The British Library
The Walters Art Museum
National Library of New Zealand

Examples & good practices: OpenGLAM

An OpenGLAM institution

4. When publishing data uses open file formats, which are machine-readable.

Examples & good practices: OpenGLAM

An OpenGLAM institution

5. Opportunities to engage audiences in novel ways on the web should be pursued.

FURTHER READING

Christine Borgman “Big Data, Little Data, No Data: Scholarship in the Networked World” 2015, MIT Press, London https://mitpress.mit.edu/big-data-little-data-no-data

Science Europe Guidance Document “Presenting a Framework for Discipline-specific Research Data Management” January 2018 http://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

“Linked Open Data – What is it?” from Europeana

DARIAH-ERIC (https://www.dariah.eu/)
PARTHENOS (https://training.parthenos-project.eu/)

What are Digital Cultural Heritage Assets?

Digital Cultural Heritage Assets such as digital photographs, high-resolution scans of manuscripts, 3D objects and 'born digital' items can be hosted by Cultural Heritage Institutions (CHIs) or by dedicated Data Centres.

The Cultural Heritage Data Charter

The Cultural Heritage Data Reuse Charter is a particular initiative that was developed by DARIAH-EU and now involves a wide community of interest that includes infrastructures like CLARIN-EU, E-RIHS, Europeana and affiliated projects such as HaS, IPERION-CH, EHRI, PARTHENOS.

FAIR Data in Trustworthy Data Repositories Webinar
Webinar proceedings from December 2016, from an event organised by DANS, EUDAT and OpenAIRE
https://www.eudat.eu/events/webinar/fair-data-in-trustworthy-data-repositories-webinar

Reviewing practices, tools and services within a Cultural Heritage Institution

Europeana is Europe's platform for digital cultural heritage with a mission to 'transform the world with culture'.

Europeana Collections

In 2014, Europeana launched its five-year strategy "We transform the world with culture": Europeana Strategy 2015-2020. In it, three key priorities were declared for the Foundation to focus on:

- Improve data quality
- Open the data
- Create value for partners

Reviewing practices, tools and services within a Cultural Heritage Institution

"A Call to Culture: Europeana 2020 Strategic Update" identifies three priorities:

- Make it easy and rewarding for Cultural Heritage Institutions to share high-quality content
- Scale with partners to reach target markets and audiences
- Engage people on Europeana websites and via participatory campaigns

The Europeana Data Model
The Europeana Data Model (EDM) (2000s).

Brings together multiple metadata standards using Linked Open Data

EDM is *not* built on any one community standard.
This allows users to access the information, and for Europeana and CHIs to forge more meaningful links to data in other European CHIs.

FURTHER READING

EDM Mapping Guidelines

EDM case studies: http://pro.europeana.eu/case-studies-edm

EDM object templates & XML schema: http://europeanalabs.eu/wiki/EDMXMLSchema

EDM Factsheet: EDM factsheet

FURTHER READING

Charles, V. (2016) “Building a framework for semantic cultural heritage data”, presentation given at VALA2016 – CC BY-SA https://www.vala.org.au/direct-download/vala2016-proceedings/vala2016-slides/734-vala2016-plenary-3-charles-slides/file (accessed 27th Nov 2017)

Europeana Data Model Primer (published 14 July 2013) https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf (accessed 29th Nov 2017)

Europeana Data Model Documentation (Published 18th Nov 2014) https://pro.europeana.eu/page/edm-documentation (accessed 29th Nov 2017)

Research Data Management and Data Management Plan

Communication:
Structure:
Vocabularies (ontologies; taxonomies; thesauri)

Digital Curation Centre: DCC wizard

Data Collection

What data will you collect or create?

Questions to consider:

What type, format and volume of data?
Do your chosen formats and software enable sharing and long-term access to the data?
Are there any existing data that you can reuse?

You can select type and format from the documents suggested by:

Archaeology Data Service (ADS) (http://tiny.cc/wo9lcz), or
Data Archiving and Networked Services (KNAW-DANS) (http://tiny.cc/xq9lcz)

How will the data be collected or created?

Questions to consider:
What standards or methodologies will you use?
How will you structure and name your folders and files? How will you handle versioning?
What quality assurance processes will you adopt?

Documentation and Metadata

What documentation and metadata will accompany the data?

Questions to consider:
What information is needed for the data to be read and interpreted in the future?
How will you capture / create this documentation and metadata?
What metadata standards will you use and why?
Do you follow naming conventions? File naming best practices http://tiny.cc/op8lcz

Data Sharing

How will you share the data?

Questions to consider:
How will potential users find out about your data? **Which methodologies are used to make data findable?**

- Common metadata-based discovery
- Ontology-based discovery
- Content-based discovery (e.g. text, images)
- Other

With whom will you share the data, and under what conditions?
Will you share data via a repository, handle requests directly or use another mechanism?
When will you make the data available?
Will you pursue getting a persistent identifier for your data?

Data Documentation

Use appropriate file formats

If you want your data to be re-used and sustainable in the long-term, you will use open, non-proprietary formats.

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

Depending on the expected usage of each type of data, the data will be divided in levels such as:

- Quantitative data: user input data, e.g., 3D models, photographs, text, etc.;
- Quantitative data: generated data, e.g., GIS data, point clouds, user profile, etc.;
- Qualitative data: social media posts, publications of other people's uploads, social group suggestions and information.

Specify the granularity of the data to be archived

- Single items (i.e. one page of a manuscript, one fieldwork report)
- Datasets (a set of homogeneously structured data records, consisting of fields carrying data values)
- Collections (an aggregation of resources, a collection may include e.g. a textual document, a set of images, one or more datasets and other collections)
- Corpora
- Raw data
- Other

Specify software tools are needed to access the data

- Linked Open Data
- SPARQL access point
- Standard Visualizer (e.g. browser, Acrobat Reader, Image Viewer)
- Local Visualizer (3DHOP, Potree)
- Download
- GIS software
- Other

Interoperability of the data

- RDF
- OWL
- DAML+OIL
- JSON LD
- XML
- KML
- GML
- GeoJSON
- CSV
- Other

Data vocabularies

- Art and Architecture Thesaurus (AAT): http://www.getty.edu/research/tools/vocabularies/aat/
- PeriodO: http://perio.do/en/
- Pleiades: https://pleiades.stoa.org/

Data Sharing

- Research Data Journal for the Humanities and Social Sciences: https://brill.com/view/journals/rdj/rdj-overview.xml?rskey=2G8kx3&result=1
- Journal of Open Archaeology Data: http://openarchaeologydata.metajnl.com/
- Make use of discipline specific, institutional or European **repositories** to deposit data/publications (e.g. Zenodo: https://zenodo.org/)
- Use tools to **register research data** (e.g. re3data: https://www.re3data.org/)

Find a repository, for humanities data:

- DARIAH EU (https://hal.archives-ouvertes.fr/, https://de.dariah.eu/en/repository)
- CLARIN (https://www.clarin.eu/content/repositories)
- GESIS (www.gesis.org)
- IANUS (www.ianus-fdz.de)

Are any restrictions on data sharing required?

Questions to consider:
What action will you take to overcome or minimise restrictions?
For how long do you need exclusive use of the data and why?
Will a data sharing agreement (or equivalent) be required?

Describe which information you gather on the rights holder

- The rights owner is recorded in the metadata form
- Documentation is gathered from their holders
- Permissions are gathered from their holders
- Agreement with each content provider
- The data creator is responsible for recording any rights
- If rights are held by third parties, the creator is responsible for ensuring permissions are given, or content removed
- Support standards for data citation
- Provide proper attribution and credit information in an external metadata record where a dataset is implemented by different individual contributors
- Not available

Storage and Backup

How will the data be stored and backed up during the research?

Questions to consider:
Do you have sufficient storage or will you need to include charges for additional services?
How will the data be backed up?
Who will be responsible for backup and recovery?
How will the data be recovered in the event of an incident?

How will you manage access and security?

Questions to consider:
What are the risks to data security and how will these be managed?
How will you control access to keep the data secure?
How will you ensure that collaborators can access your data securely?

What tools does your system use to provide access to users?

- FEDORA container-based OS
- Dspace
- Locally developed system
- Other

Responsibilities and Resources

Who will be responsible for data management?

Questions to consider:
Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?
Who will be responsible for each data management activity?
How will responsibilities be split across partner sites in collaborative research projects?
Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?

Responsibilities and Resources

What resources will you require to deliver your plan?

Questions to consider:
Is additional specialist expertise (or training for existing staff) required?
Do you require hardware or software which is additional or exceptional to existing institutional provision?
Will charges be applied by data repositories?

Ethics

When should you consider ethics in research?

Two main types of research in Arts, Humanities and Social Sciences:

Participatory research (interviews, anonymous surveys, testing, crowd-sourcing).
Non-participatory research (manuscripts, archives, etc.).

Ethics in Participatory Research

Research data that would usually be subject to ethical approval:

- Any recorded interviews (either video or audio).
- Surveys or questionnaires.
- Research where the participant is asked to reveal or reflect on instances from their past.
- Anything that involves the participation of minors.
- Personal information/narratives.
- Tests, or test-like scenarios.

Ethics in Non-Participatory Research

Still responsible for the ethical management of your data.

Ethics and Data Management Plans: the case for Data Sovereignty

The concept of indigenous data sovereignty and its application to DH projects that deal with artefacts, data, images, and concepts that may be considered sacred, secret, or culturally sensitive by certain groups of people.

Open Access principles <-> Intellectual Property Rights/Management.

FURTHER READING

Bernard, H. R. (2011). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira.

Chu, H. (2015). Research methods in library and information science: A content analysis. *Library and Information Science Research*, *37*, 36–41. https://doi.org/10.1016/j.lisr.2014.09.003

Josselson, R., & Lieblich, A. (2001). Narrative research and humanism. *The Handbook of Humanistic Psychology*, 275–289
Sieber, J. (2012). *The Ethics of Social Research: Surveys and Experiments*. Springer Science & Business Media

Archaeological research DMP: http://vast-lab.org/dmp/ariadneplus/form/

Examples & Practical Guidance: https://book.fosteropenscience.eu/en/05ExamplesAndPracticalGuidance/

Sample DMP Arts & Humanities, UK: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/AHRC-Scots-syntactic-atlas-DMP.pdf

NEH (National Endowment for the Humanities) makes example grants available, including DMPs, e.g. https://www.neh.gov/grants/odh/digital-humanities-advancement-grants

Formal Ontologies

PUTTING A STRUCTURE ON MESSY DATA

Challenging characteristics of humanities data:

- Variety in formats.
- Not structured.

Data Heterogeneity: the mixing of all these difference data types and standards, and how we can make it 'Interoperable'.

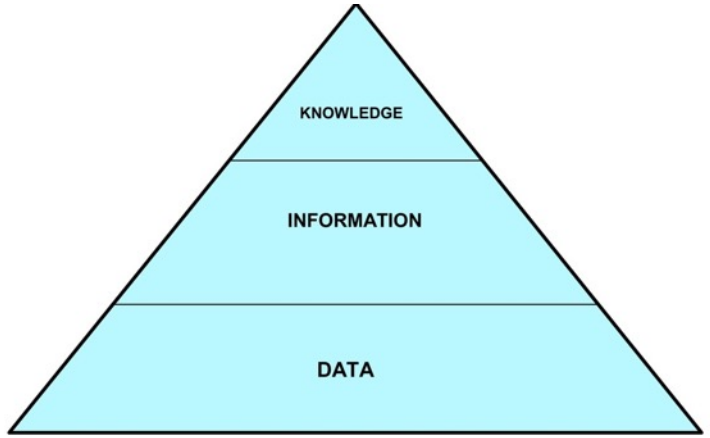
Ontologies: categorise data in a way that is understandable to a human, as well as to a computer.

Semantic data: applying human-motivated categorisation and meaning to data in a way that a computer can understand it, structure it, and re-represent it back to another human while retaining that meaning.

Formal Ontologies - putting a structure on messy data

Why would someone want to develop an ontology?

- Share common understanding
- Enable reuse of knowledge
- Make domain assumptions explicit
- Analyze domain knowledge



Rowley (2007), “The wisdom hierarchy: representations of DIKW hierarchy”.

Formal Ontologies - putting a structure on messy data

What is Data Heterogeneity?

- Incomprehensible data silos.
- Different standards**.**
- Data curation in incompatible ways.

Different institutions, by adopting different standards for different ends, end up potentially blocking the researcher in his/her investigation.

The differences between data and how it is managed and represented in a knowledge representation system leading to incompatible data that cannot be compared is what we call **data heterogeneity.**

Formal Ontologies - putting a structure on messy data

Causes

- flexibility of the medium
- rapid pace of change
- difference of tools
- difference of actors
- difference of means
- differences of questions and traditions
- different schemas for data representation
- different formats for data representation
- lack of serious policies regarding data

Data schemas

Implementation of standardized vocabularies or thesauri and reference resources (e.g., AAT, TGN, Geonames, ULAN, VIAF, DBPedia)

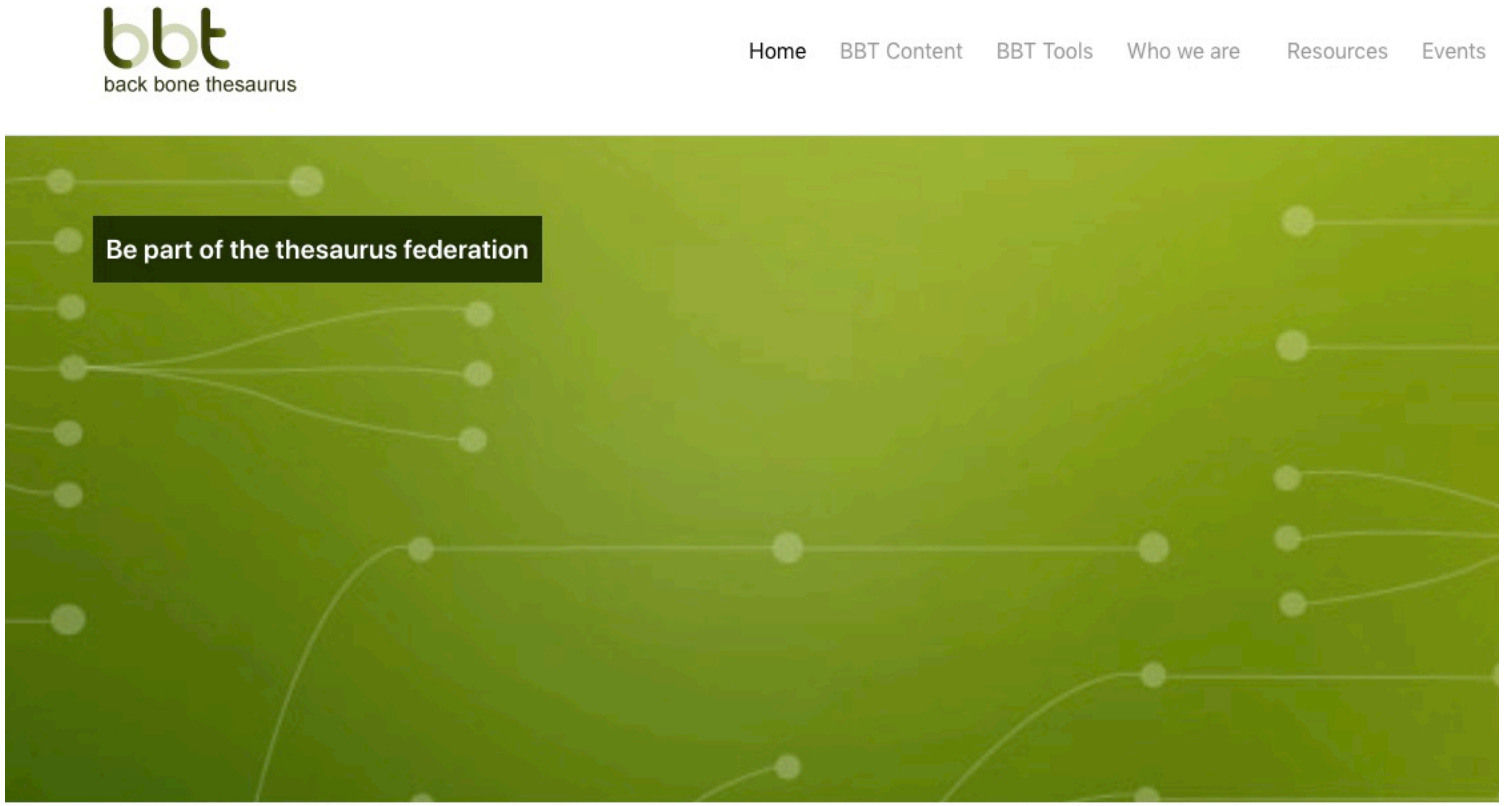
Formal Ontologies - putting a structure on messy data

Thesauri files: standardization of data

Controlled thesauri are generally curated by a specific community and provide a list of terms and their (un)official spellings for those concepts that are recognized and used for describing some aspect of reality.

Getty Art and Architecture Thesaurus

Examples



https://vocabs.dariah.eu/backbone_thesaurus/en/

- An overarching thesaurus for the humanities
- Common basis
- Enables interdisciplinarity

Formal Ontologies

Formal ontologies are designed to represent the basic elements of information within a broad domain and formalize these into a logical language that consists of classes and relations.

> "An ontology is a description (like a formal specification of a program) of the concepts and relationships that can formally exist for an agent or a community of agents". (Gruber, 2001)

Subject-Verb-Object

No one "correct" way or methodology for developing ontologies.

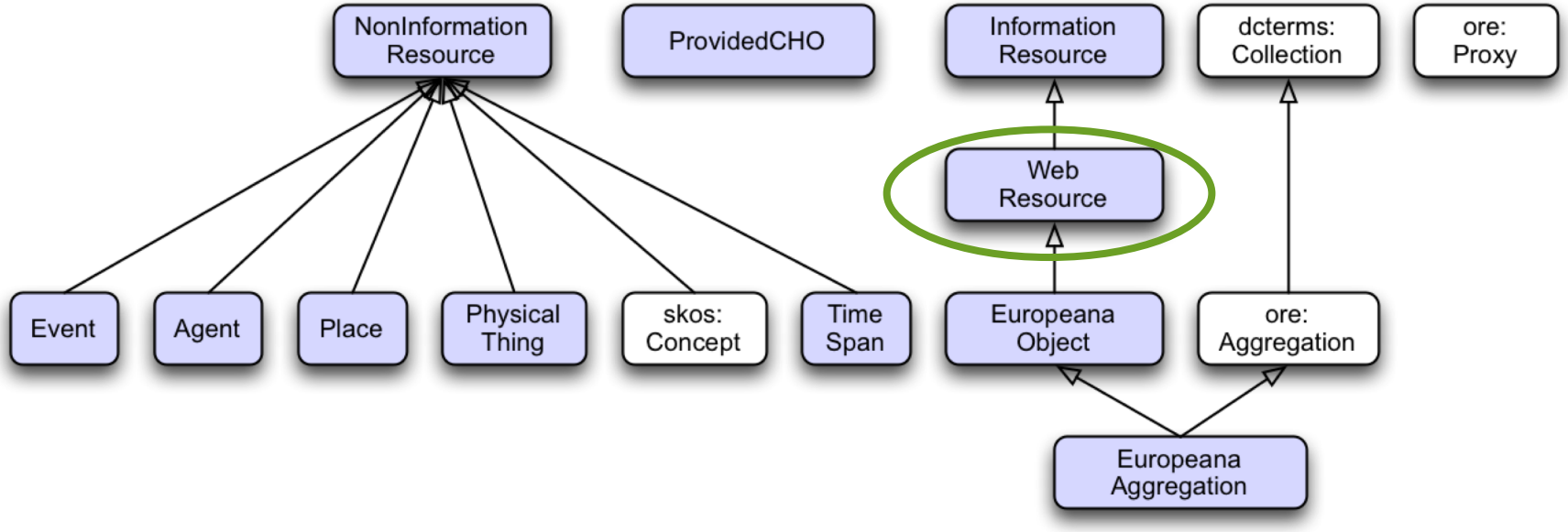
Formal Ontologies – the case of the EDM

A formal ontology is composed of the following elements:

•a scope: The **scope declaration** of an ontology indicates the subject matter that the ontology aims to represent.

•class declarations: **Class declarations** provide the 'noun' classes available for use in the proposed formal language.

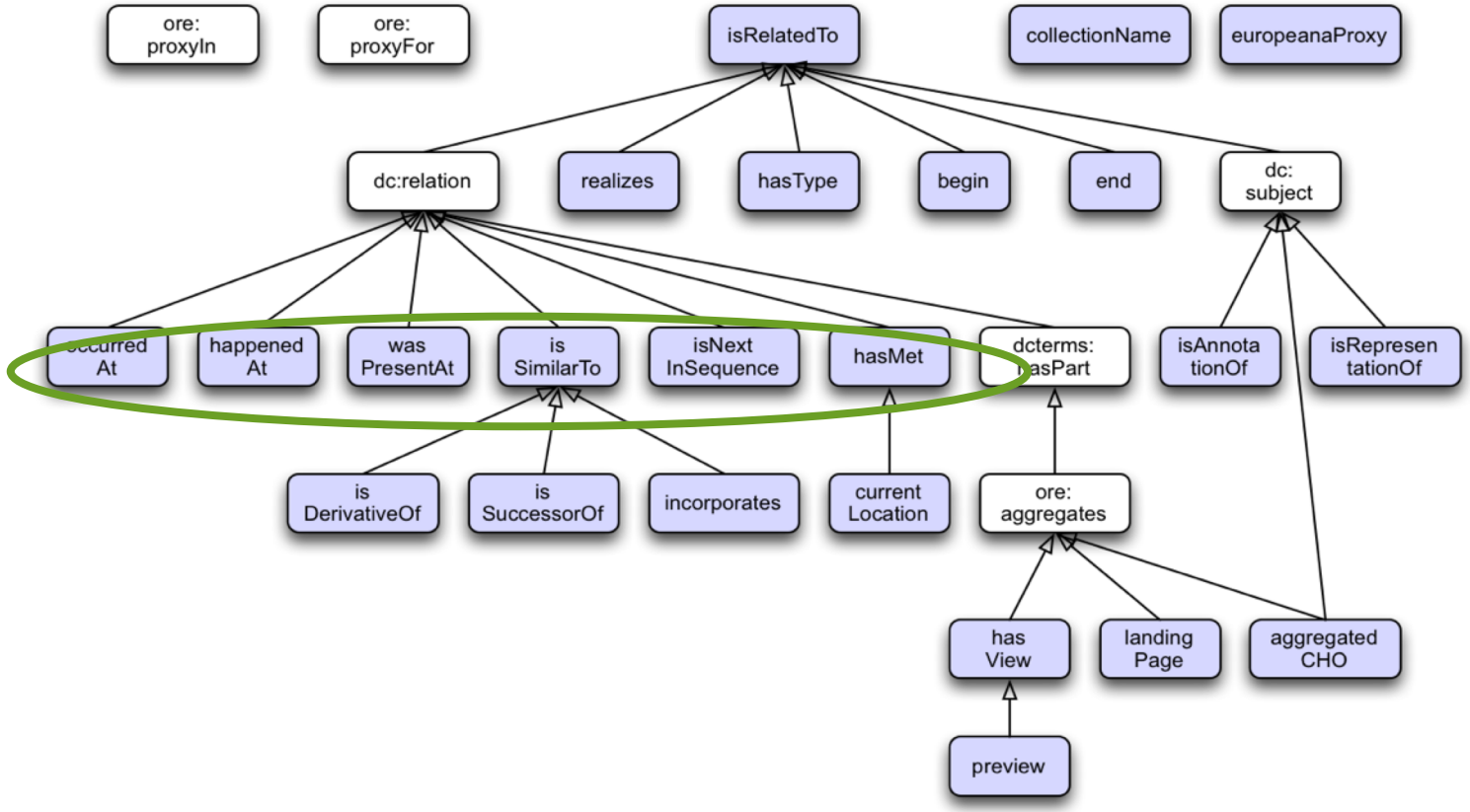
EDM main classes:
Groups of things that have common properties,
e.g., web resources.



Formal Ontologies – the case of the EDM

•relation declarations: **Relation declarations** provide the syntax of the language defined by the ontology, that is the 'sentence structure', to continue the analogy with natural language.
•a set of logical rules.

EDM main properties:



FURTHER READING

Doerr, Martin. 2003. “The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata.” *AI Magazine* 24 (3):75.

Gruber, Thomas R. 1995. “Toward Principles for the Design of Ontologies Used for Knowledge Sharing?” *International Journal of Human-Computer Studies* 43 (5):907–928.

Guarino, Nicola. 1997. “Understanding, Building and Using Ontologies.” *International Journal of Human-Computer Studies* 46 (2):293–310.

Gruber, Thomas R. 1995. “Toward Principles for the Design of Ontologies Used for Knowledge Sharing?” *International Journal of Human-Computer Studies* 43 (5):907–928.

Guarino, Nicola. 1997. “Understanding, Building and Using Ontologies.” *International Journal of Human-Computer Studies* 46 (2):293–310.

Uschold, Michael, and Martin King. 1995. *Towards a Methodology for Building Ontologies*. Citeseer. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.480.1214&rep=rep1&type=pdf.

Ciula and Tupman (2016), Session 7: Ontologies and Data Modelling.

Eide and Ore (2018), Ontologies and Data Modelling.

Standards to be used for metadata creation

FAIR Humanities!

- **TEI** (Text Encoding Initiative): www.tei-c.org
- **CEI** (Charter Encoding Initiative): http://www.cei.lmu.de/index.php
- **MEI** (Music Encoding Initiative): https://music-encoding.org/
- **CMDI** (Language Resources, CLARIN):
- **IIIF** (International Image Interoperability Framework): https://iiif.io/
- **EAD** (Encoded Archival Description, for finding aids): https://www.loc.gov/ead/
- **Dublin Core** (description of digital documents): http://dublincore.org/

- PARTHENOS Entities
- ARIADNE model
- CARARE
- CIDOC CRM
- EDM - Europeana Data Model
- LIDO
- DCAT

Standards to be used for metadata creation

Standardization Survival Kit (SSK)

• Overlay platform developed by PARTHENOS dedicated to promoting a wider use of **standards** (TEI, Dublin Core, etc.) within the Arts and Humanities.

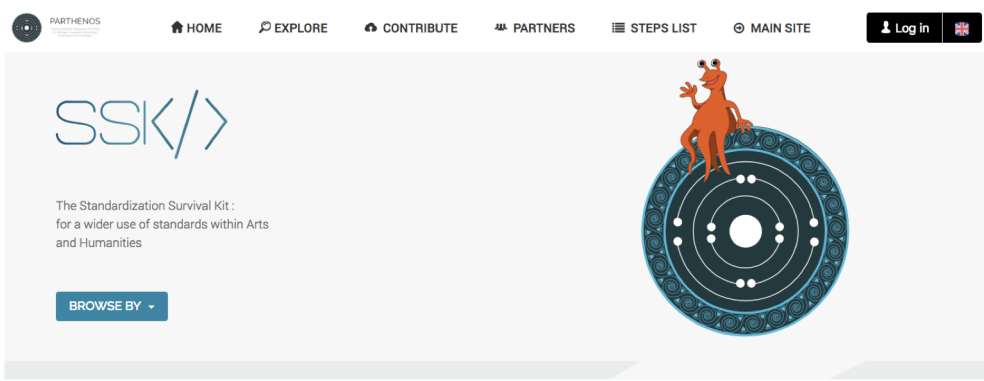
• Aims:

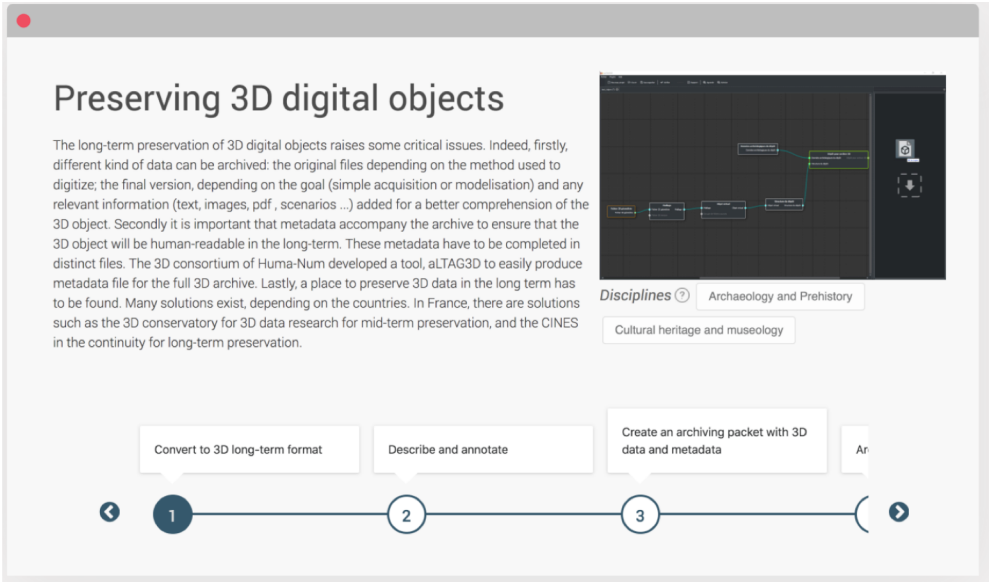
– Selecting and using the appropriate standards for their particular disciplines and work flows

– Documentation of existing standards by providing reference materials

– Foster the adoption of standards

– Communication with research communities

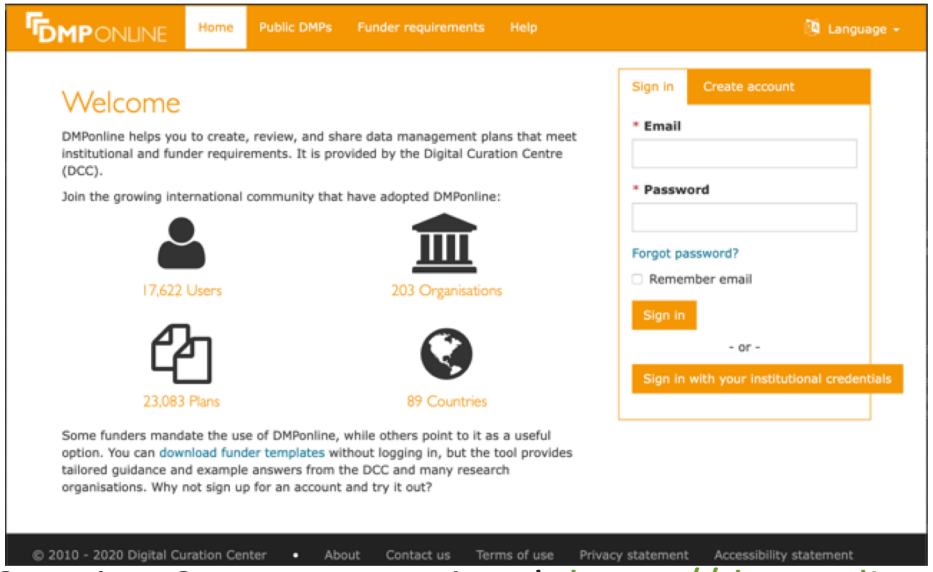




PARTHENOS Standardization Survival Kit!

https://ssk-application.parthenos.d4science.org/ssk/#/

Standards to be used for metadata creation



Digital Curation Centre DMP wizard: https://dmponline.dcc.ac.uk/

Selection and Preservation

- What metadata will you maintain?
- What standards will you use for this purpose?
- Will any standard thesauri, vocabularies or methods be applied?
- If you are creating your own metadata schema, vocabulary or other convention, will a mapping to commonly available alternatives be made available?
- Will particular software tools be required to access it / query?
- If so, can the source code for this software be made available as well?
- How long can you commit to the data being accessible for?

Formal Ontologies

Agnostic to data formats

- Resource Description Framework (RDF)
- Web Ontology Language (OWL)
- Ontology-Based Data Access (OBDA)
- Mappings connect a query platform to the data sources
- Mapping process

Formal Ontologies

Uniform Access to Heterogeneous Sources

Ontology + mappings = access to heterogeneous sources.

Interface to access data: the user-oriented ontology.

User > Ontology > Data > System returns queries.

Protégé:

- Import, edit and save existing ontologies.
- Create new ontologies.
- Save ontologies in formats.
- Visualize ontologies.
- Populate ontologies.

In practice: metadata standards

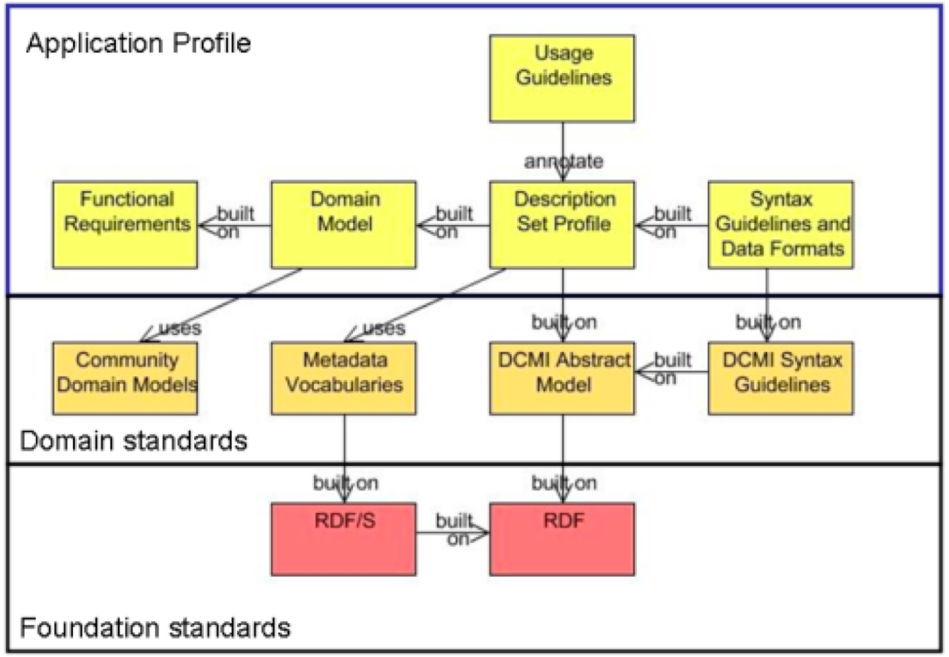
Dublin Core is the most prominent metadata used to describe electronic resources
Dublin Core™ Application Profiles

Functional Requirements

Domain Model

Defining Metadata Terms

RDF vocabularies

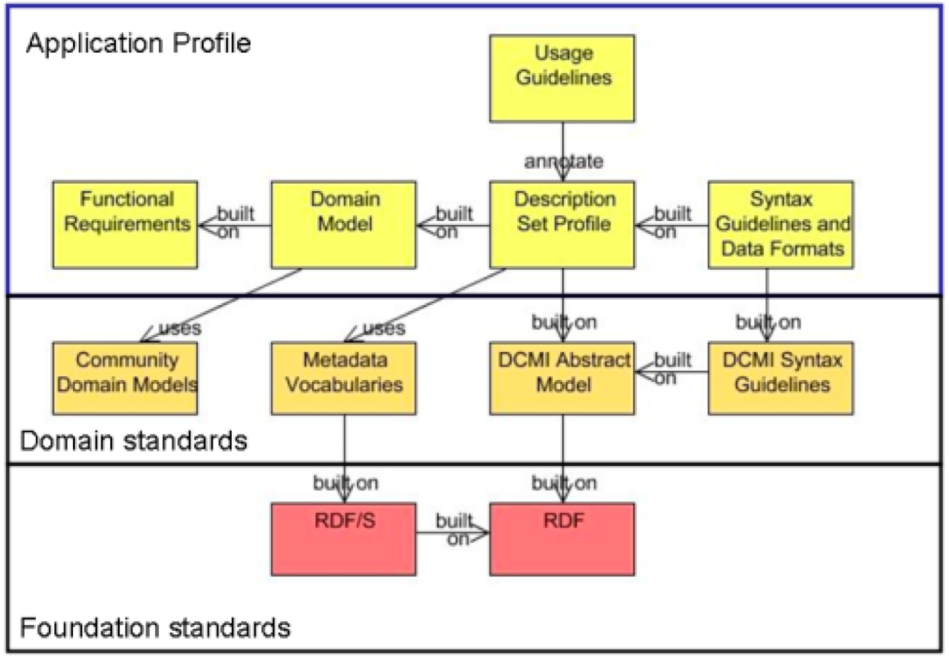


(https://www.dublincore.org/specifications/dublin-core/profile-guidelines/)

In practice – step A

Functional Requirements

- Purpose of the application
- Limits of the application
- Need to perform specific actions
- Key characteristics of your resources
- Target users
- Existing community standards

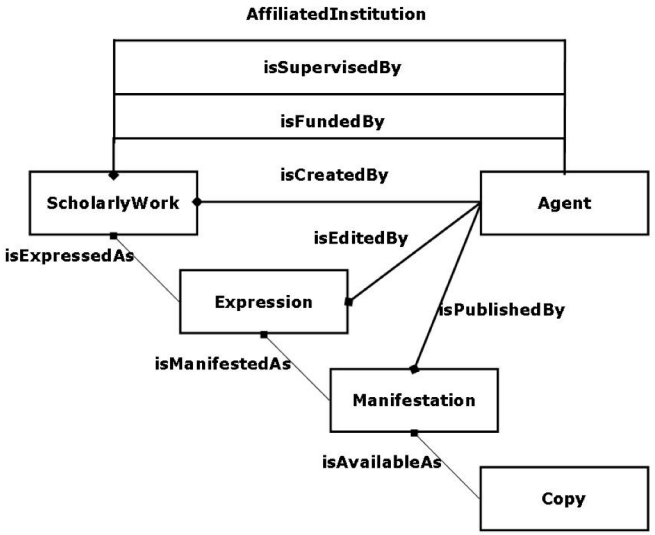


(https://www.dublincore.org/specifications/dublin-core/profile-guidelines/)

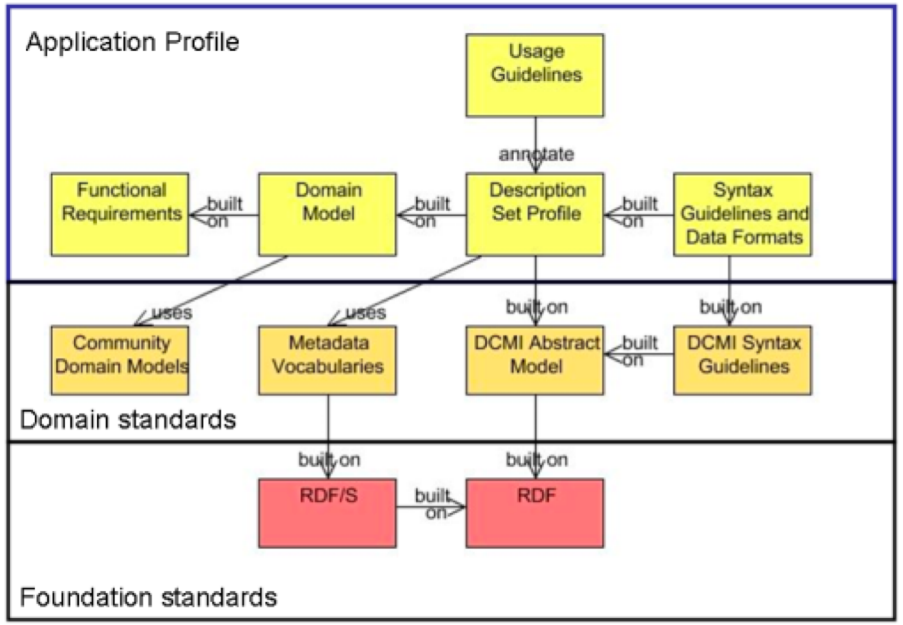
Domain Model

A domain model is a description of what things the metadata will describe, and the relationships between those things.

A simple model.



A more complex model.



In practice – step C

Defining Metadata Terms

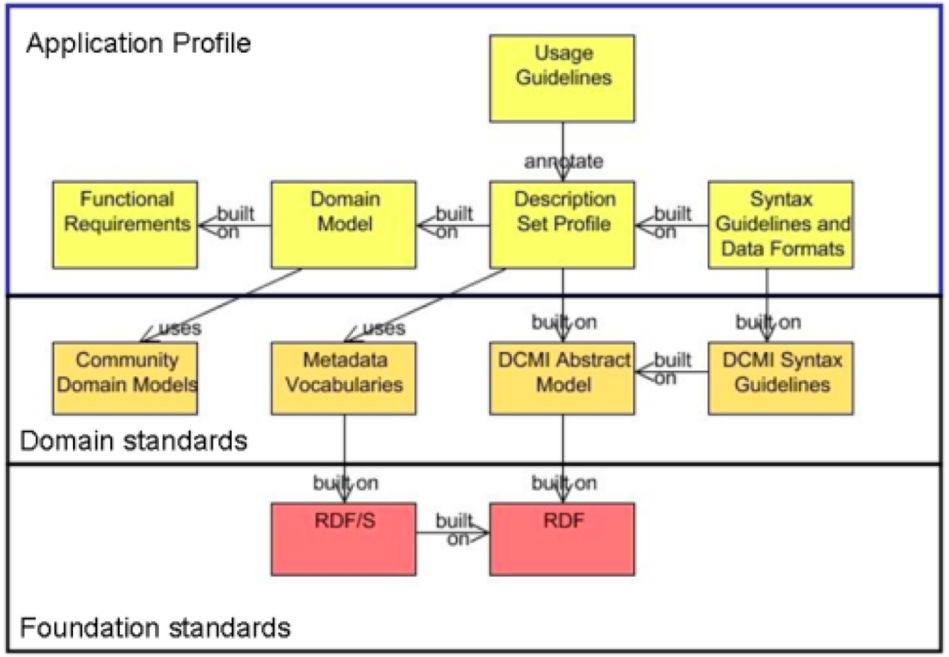
- Choose properties

RDF vocabularies

- Use existing properties

Dublin Core Element set
(http://dublincore.org/documents/dces/)

Contributor, Coverage, Creator,
Date, Description, format
Identifier. Language, Publisher
Relation, Rights, Source
Subject, title, type



(https://www.dublincore.org/specifications/dublin-core/profile-guidelines/)

FURTHER READING

Resource: OWL2 (W3C Recommendation)
The World Wide Web Consortium maintains the technical specification of OWL2. Go to resource

Resource: SPARQL (Wikipedia)
The Wikipedia page for SPARQL contains a list of advantages of the language, as well as descriptions of it's most basic features. Go to resource

"RDF – Semantic Web Standards." n.d. Accessed January 23, 2018. https://www.w3.org/RDF/.

"RDF Tutorial." n.d. Accessed January 23, 2018. http://w3schools.sinsixx.com/rdf/default.asp.htm.

"ICS – X3ML Toolkit." n.d. Accessed January 23, 2018. http://www.ics.forth.gr/isl/index_main.php?l=e&c=721.

Resource: Linking Data to Ontologies (Scientific Article)
"In this paper [the authors] present a solution to the problem of designing effective systems for ontology-based data access." Go to resource

https://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Formal Ontologies – Example

Formal ontologies relevant to digital humanities

Ontologies allow integration of data on an extremely high level, sometimes including logical rules within the structure that allow for automated reasoning over datasets.

- CIDOC CRM: originally designed in the museological community, has been extended to account for cultural heritage and e-sciences data.

> The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. [...] In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (CIDOC CRM)

- CIDOC CRM international standard (ISO 21127)

Formal Ontologies – Example

CIDOC Conceptual Reference Model (CRM)

Shared understanding of cultural heritage information by:

- a common and extensible semantic framework;
- a common language for domain experts.

CIDOC-CRM examples

ICCD (Central Institute for Catalogue and Documentation) schema:

- CRMba
- CRMarchaeo
- CRMsci

Setting up a CIDOC CRM: https://www.youtube.com/watch?v=ou8Y43XD2g4&feature=emb_title

The Europeana Data Model

OAI ORE (Open Archives Initiative Object Reuse & Exchange) for organizing an object's metadata and digital representation(s)
Dublin Core for descriptive metadata
SKOS (Simple Knowledge Organization System) for conceptual vocabulary representation
CIDOC-CRM for event and relationships between objects
Adopts Semantic Web representation principles (RDF)
- Re-use and mix different vocabularies together
- Preserve original data and still allow for interoperability

FURTHER READING

http://ceur-ws.org/Vol-1117/paper2_slides.pdf
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34424
http://www.ics.forth.gr/proj/isst/Agtivities/CIS/GIDOG/DOGS/GRm_definition_15_6_00.rtf
http://www.geneva-city.ch:80/musinfo/GIDOG/OOModel/CRMdefinition_040999.rtf
http://www.geneva-city.ch:80/musinfo/Gidog/oomodel

Formal Ontologies: DATA Standards – Example of good practice

MIDAS Heritage - the UK Historic Environment Data Standard is a British cultural heritage standard for recording information on:

- buildings,
- archaeological sites,
- shipwrecks,
- parks and gardens,
- battlefields,
- areas of interest, and
- artefacts.

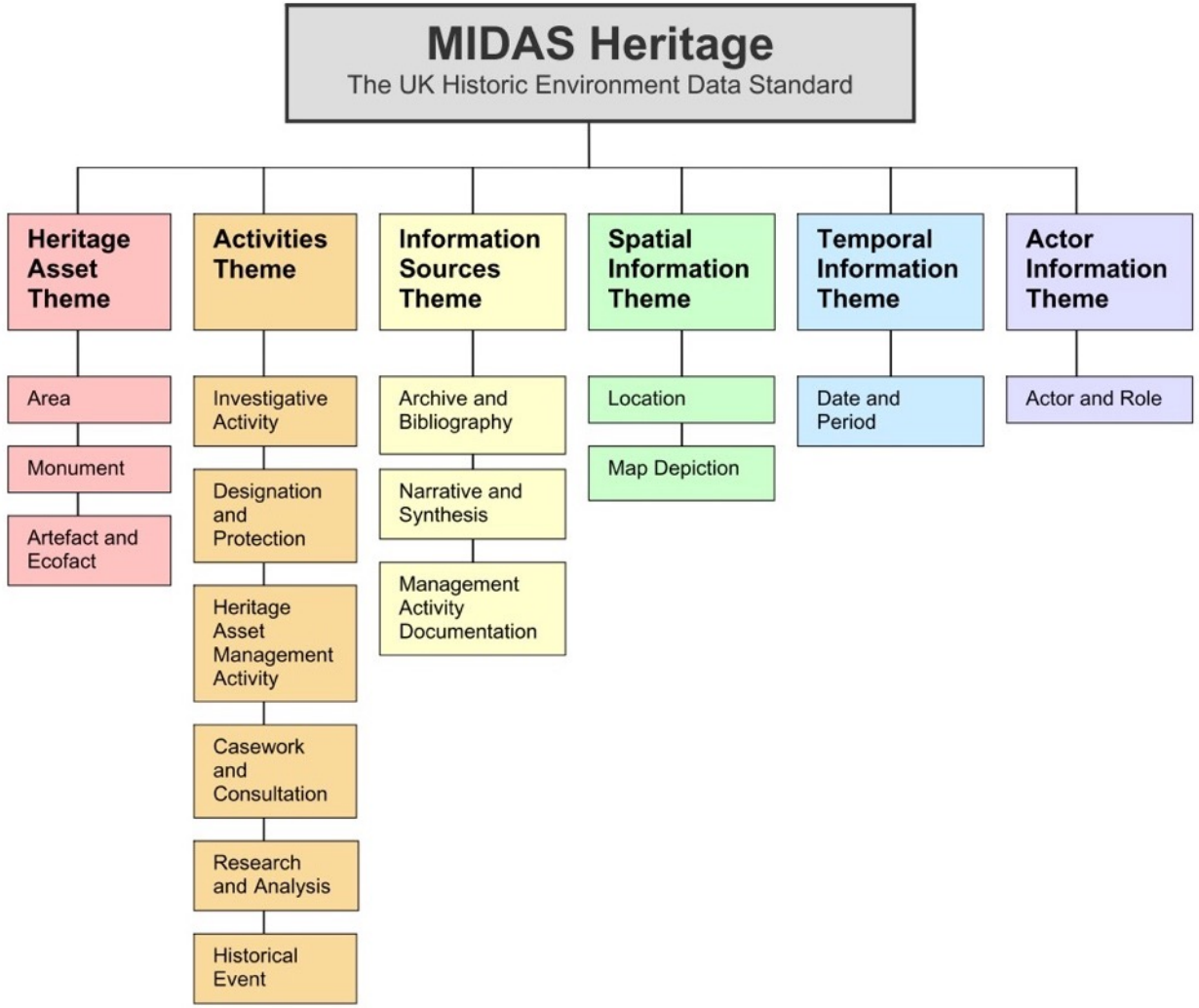
Data standard
Guidelines

https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/

http://www.english-heritage.org.uk/professional/archives-and-collections/nmr/heritage-data/midas-heritage/links-to-other-resources/

Formal Ontologies

DATA Standards – Example of good practice



Formal Ontologies

Formal ontologies relevant to digital humanities

Focused ontologies:

- FOAF: for tracking social relations.
- SPAR: for organizing citation data, article structure and context.
- NeMO: used by humanities researchers to track the workflow of their scholarly practices.
- GOLD for linguists to use to codify language according to linguistic elements.

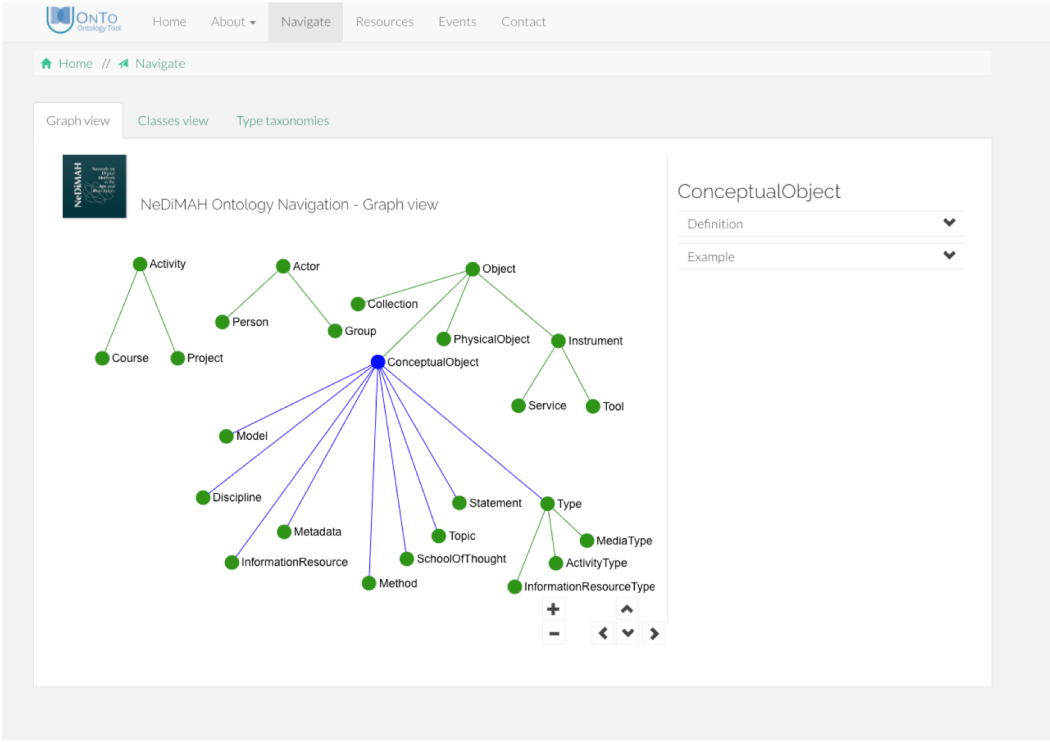
schema.org
bartoc.org

Examples

NeMO NeDiMAH Methods Ontology

- TaDIRAH;
- the arts-humanities.net and Oxford taxonomies of ICT methods;
- DHCommons;
- CCC-IULA-UPF;
- DiRT.

The scope of abstract entities encompassed by the method's ontology may include classifications of scholarly disciplines and fields, methodologies, theoretical approaches, research techniques, procedures, research data and resources, epistemic objects, research actors, as well as research environments, mechanisms, tools, systems, services, and infrastructures.



FURTHER READING

Home | CIDOC CRM.” n.d. Accessed January 23, 2018. http://www.cidoc-crm.org/.

“Home – Schema.org.” n.d. Accessed January 23, 2018. http://schema.org/.

“Laboratory for Applied Ontology – DOLCE.” n.d. Accessed January 23, 2018. http://www.loa.istc.cnr.it/old/DOLCE.html.

“Linked Open Vocabularies (LOV).” n.d. Accessed January 23, 2018. http://lov.okfn.org/dataset/lov/.

“NeDiMAH Methods Ontology: NeMO.” n.d. Accessed January 23, 2018. /content/nedimah-methods-ontology-nemo.

Formal Ontologies

Key questions:

- Is it a good fit with my research questions?
- Has it been used by other users in my research community?
- How easy is it to find learning resources for this ontology?
- What is the long term support of the ontology itself (sustainability)?

It is important that the community that develops and maintains the ontology is active and responsive.

Humanities

Concluding, Open Science applied to Humanities could mean:

- Sustainable projects and initiatives
- Valorization of Humanities knowledge and practices
- Bigger research impact
- Promotion of new discoveries and paradigms

Open Data Intensive Applications in the Humanities

Online repositories and data management systems, e.g., digital libraries with support for metadata semantic structuring.

3D documentation workflows and online visualization tools, e.g., interactive visualization and online viewers of 3D models and the products of computational imaging processes (e.g., photogrammetry, point-cloud models, RTi files).

Data processing, e.g., unsupervised feature learning on results of photogrammetric techniques, remote sensing. Data processing services for:

- semantic referencing;
- image matching; and,
- geo-referencing.

Data Intensive Applications in the Humanities

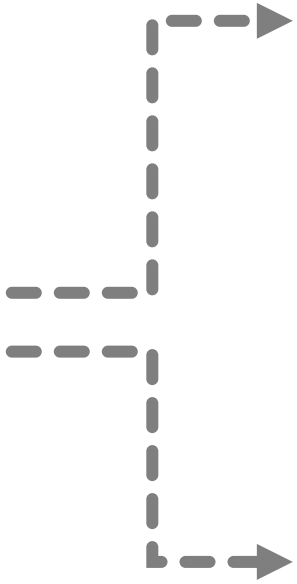
The primary uses of these applications include:

- Archaeological studies;
- Museological studies;
- Historical studies;
- Cataloguing and documentation;
- Other scholarly research activities; and
- Public outreach and education.

Use of data in the Humanities

Purpose

- Research activity
- Data curation
- Safeguarding
- Engagement / management
- Education



Parameters differ according to the application:

- Documentation
- Visualisation
- Representation
- Modelling
- Simulation
- Interpretation

Technology used:

- Interactive data visualization
- Cloud services, & repositories
- Digital libraries & GLAM industry (galleries, libraries, archives, and museums)
- Virtual Research Environments & web based content management systems

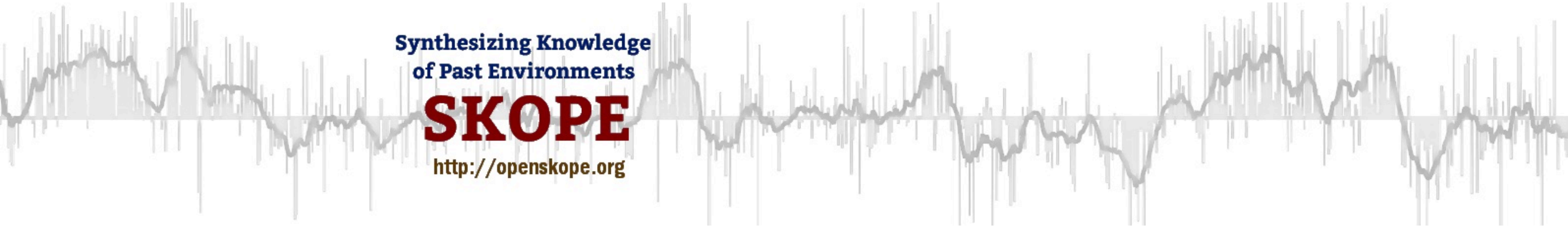
Use of data in the Humanities

1 - Online Services and Access to Repositories

Digital Libraries/Repository to provide access to services via portals in order to:

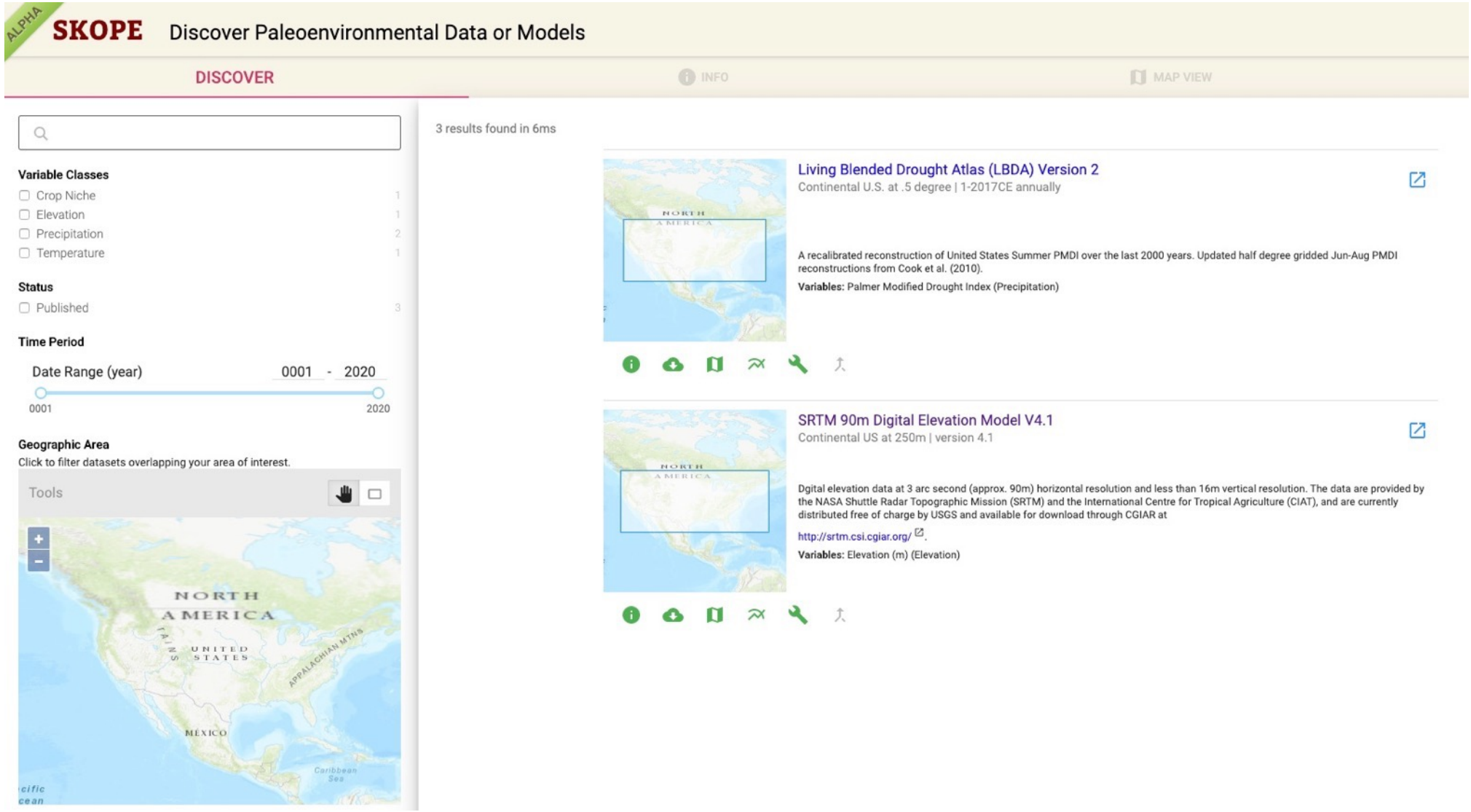
- Enable topical research inquiries and data related studies;
- Share research results & datasets;
- Visualise datasets.

Examples: Online Services

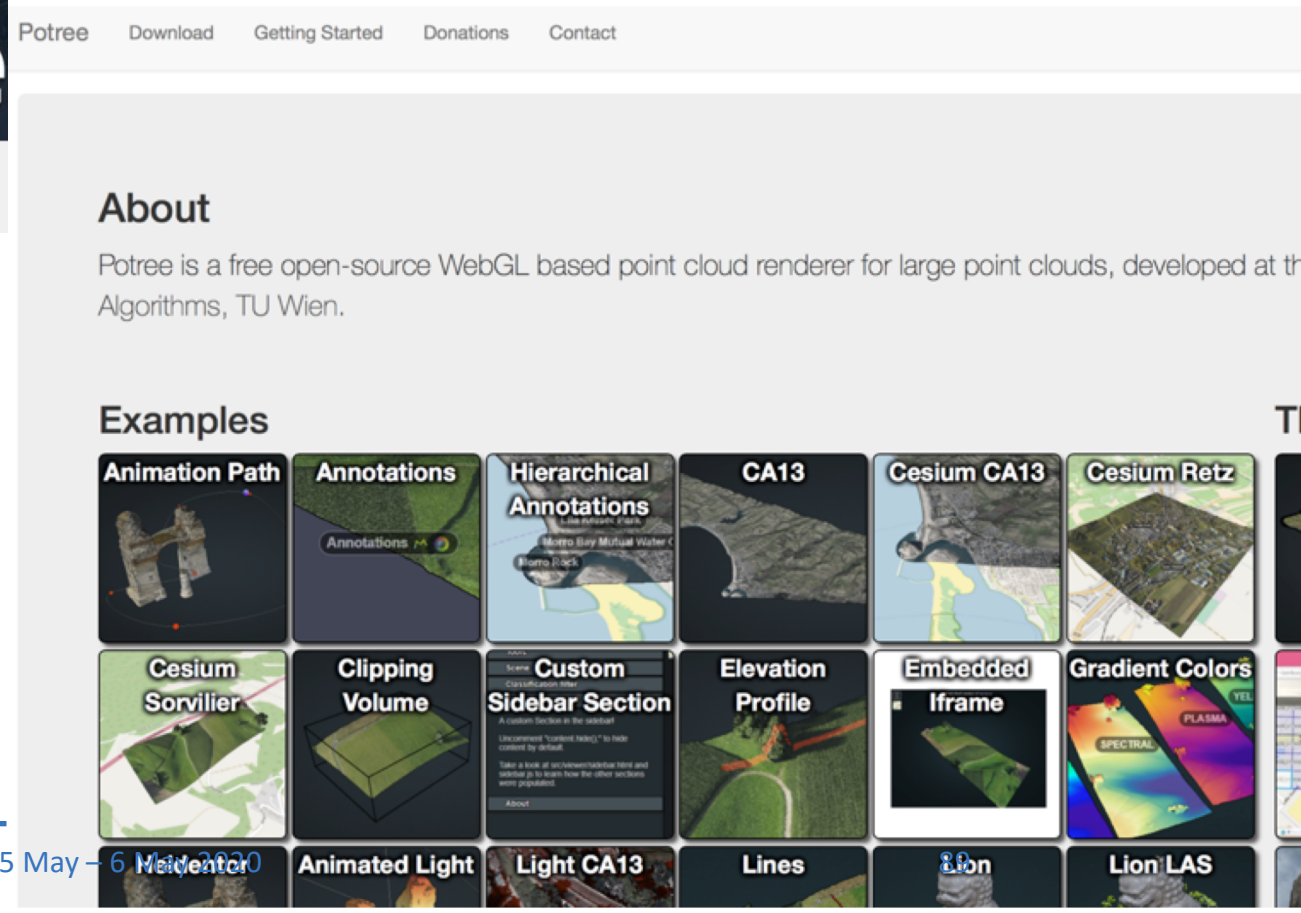


SKOPE (Synthesizing Knowledge of Past Environments) is an online resource for paleoenvironmental data and models. It enables scholars to easily discover, explore, visualize, and synthesize knowledge of environments in the recent or remote past.

Examples: Online Services

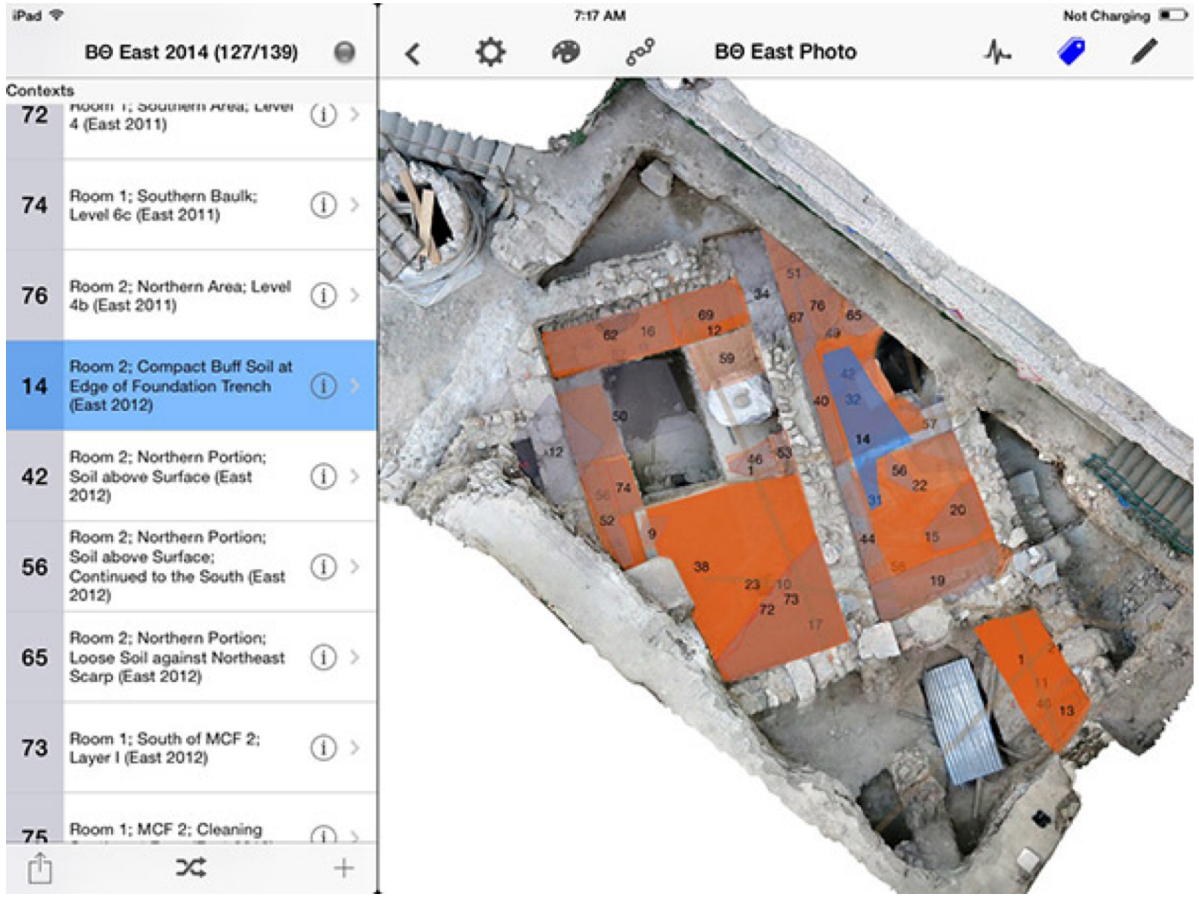


Examples: Open 3D data vis & repositories



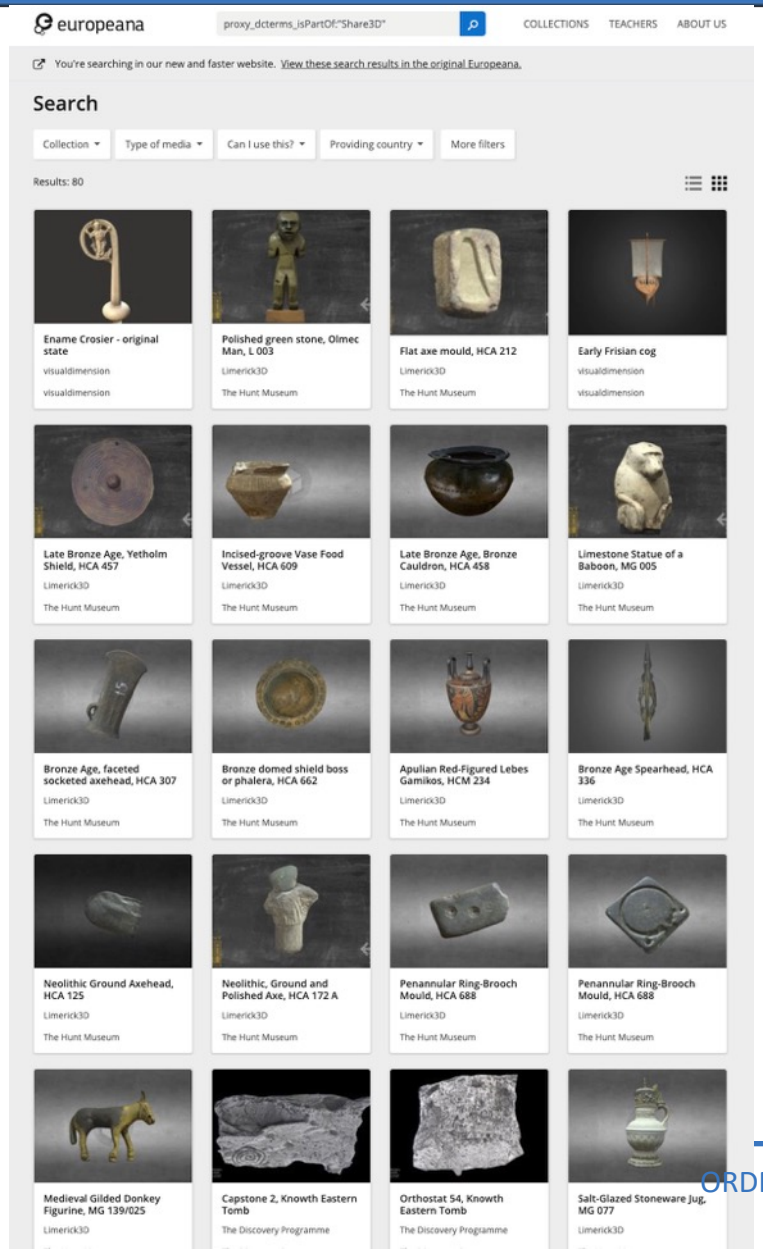
Examples: Digitization of assets / linked data

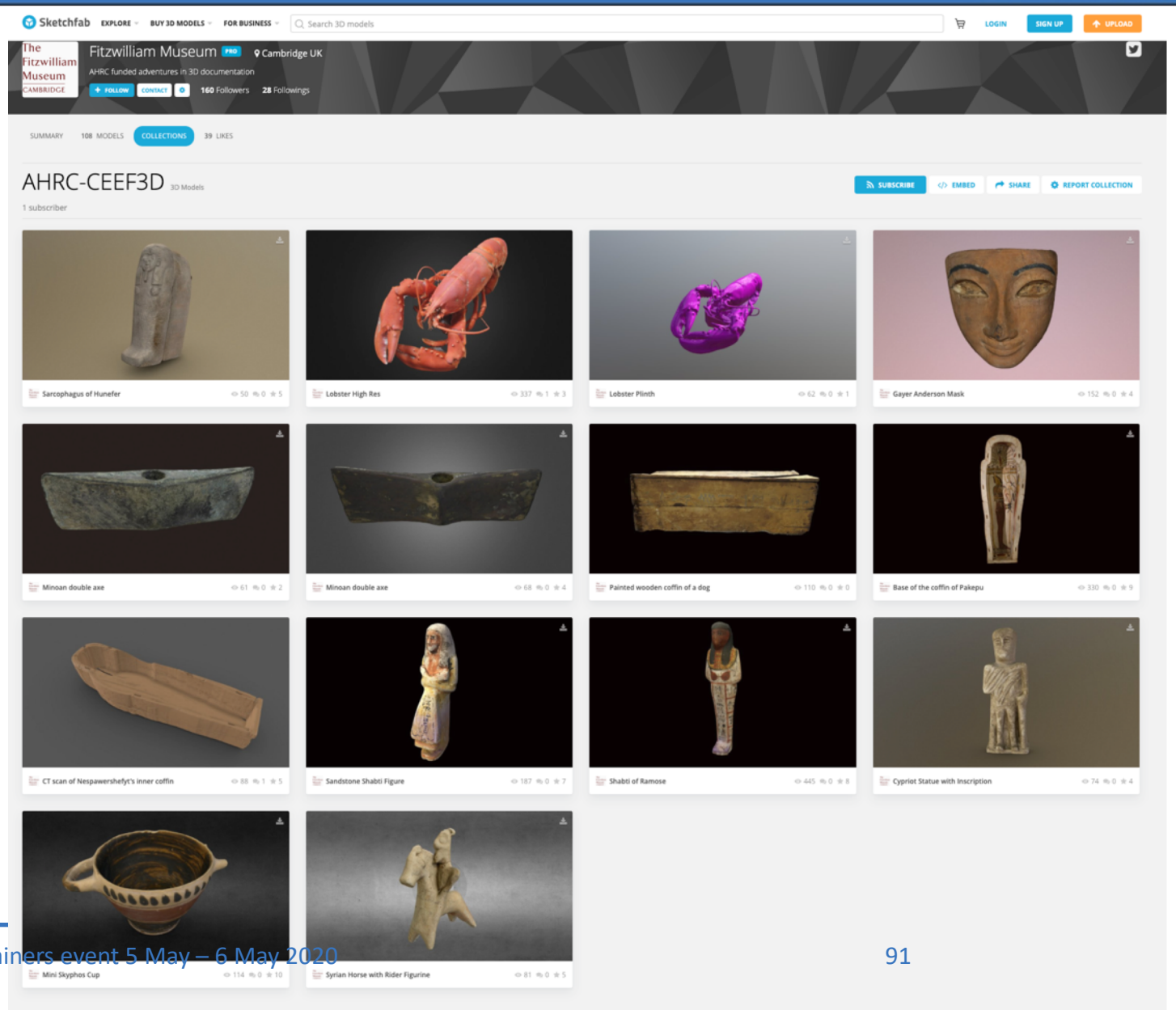
2 – Digital Documentation Methods and Tools



Bruce Hartzler, iDig is a digital app that aspires to transform archaeological recording and analysis; (http://www.ascsa.edu.gr/index.php/news/newsDetails/bruce-on-idig).

Examples - Structure for Open 3D data





Examples – Workflows & Collections

Digitization workflows

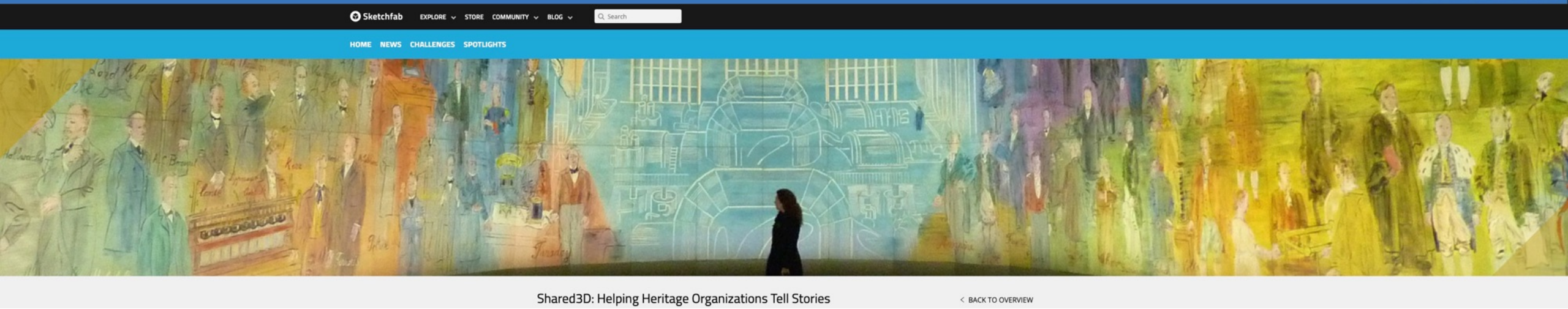
https://unity3d.com/files/solutions/photogrammetry/Unity-Photogrammetry-Workflow_2017-07_v2.pdf
https://eulacmuseums.net/images/eulac_docs/eulac_3dmanual_final.pdf

Sketchfab-Powered Online Collections

Here are 10 examples for you to explore:

- Médiathèque by archeologie.culture.fr (scroll down on linked page)
- Storymaker by share3d.eu
- Małopolska's Virtual Museums (WMM) (click the '3D plus' icon on an object page)
- Morbase | Museu Virtual
- Minneapolis Institute of Art (collection search for 3D)
- *Explore the IS Tunnels* by BBC News
- The British Museum (scroll down on linked page)
- Natural History Museum London Data Portal
- 3D Scanning Project by Harvard Peabody Museum
- Idaho Virtual Museum

Examples: Structure for Open 3D data



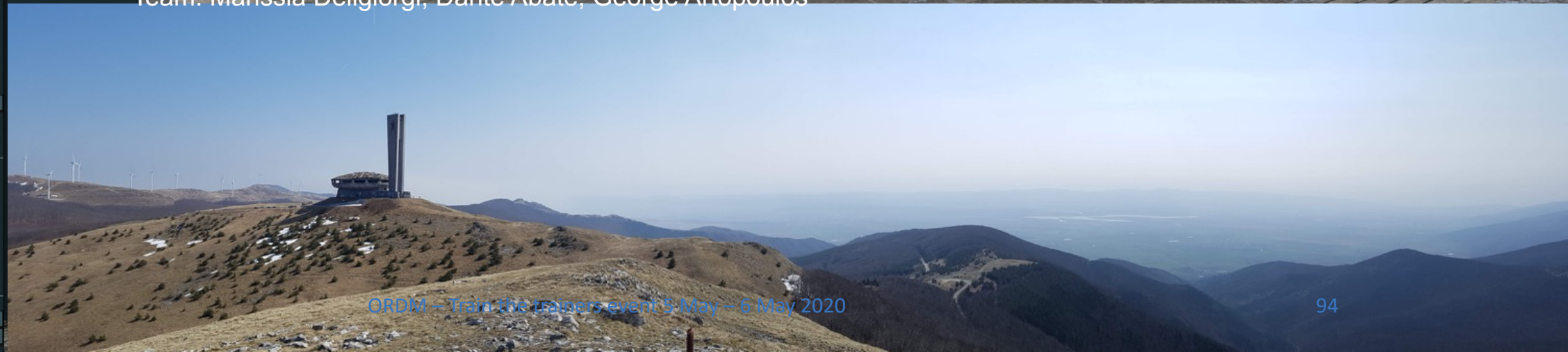
The Share3D Dashboard is a cloud-based service that uses Sketchfab's API. The edited metadata records can then be submitted to Europeana from the Share3D Dashboard via CARARE's aggregation service.

The dashboard is ideal for:

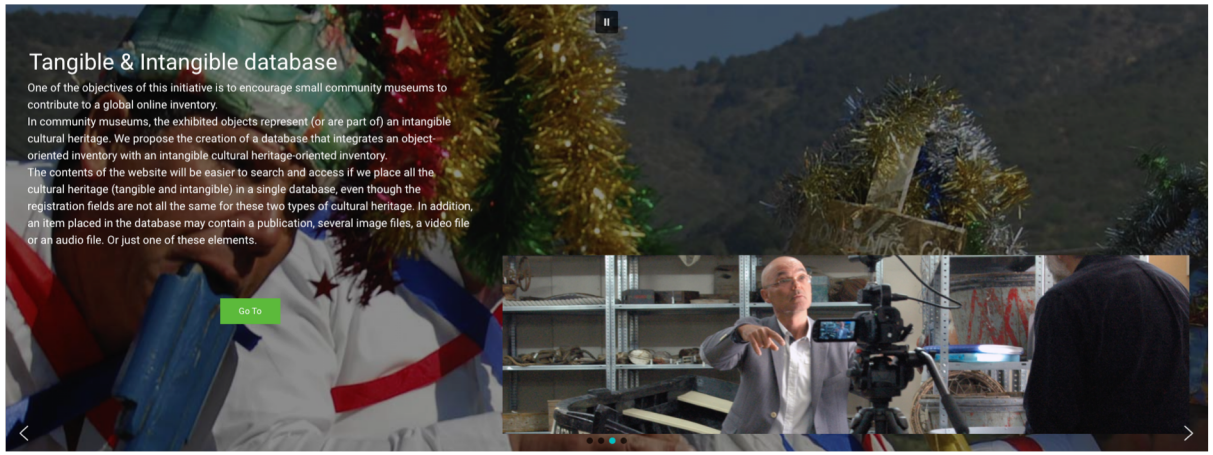
- Cultural institutions (galleries, libraries, archives and museums).
- Researchers (archaeologists, historic buildings experts, curators) who are capturing monuments, landscapes, excavation findings, historic buildings and urban areas in 3D.
- Creators of 3D models that depict cultural heritage objects.
- All those who want to create an XML formatted metadata record that conforms to Europeana Data Model (EDM) schema.

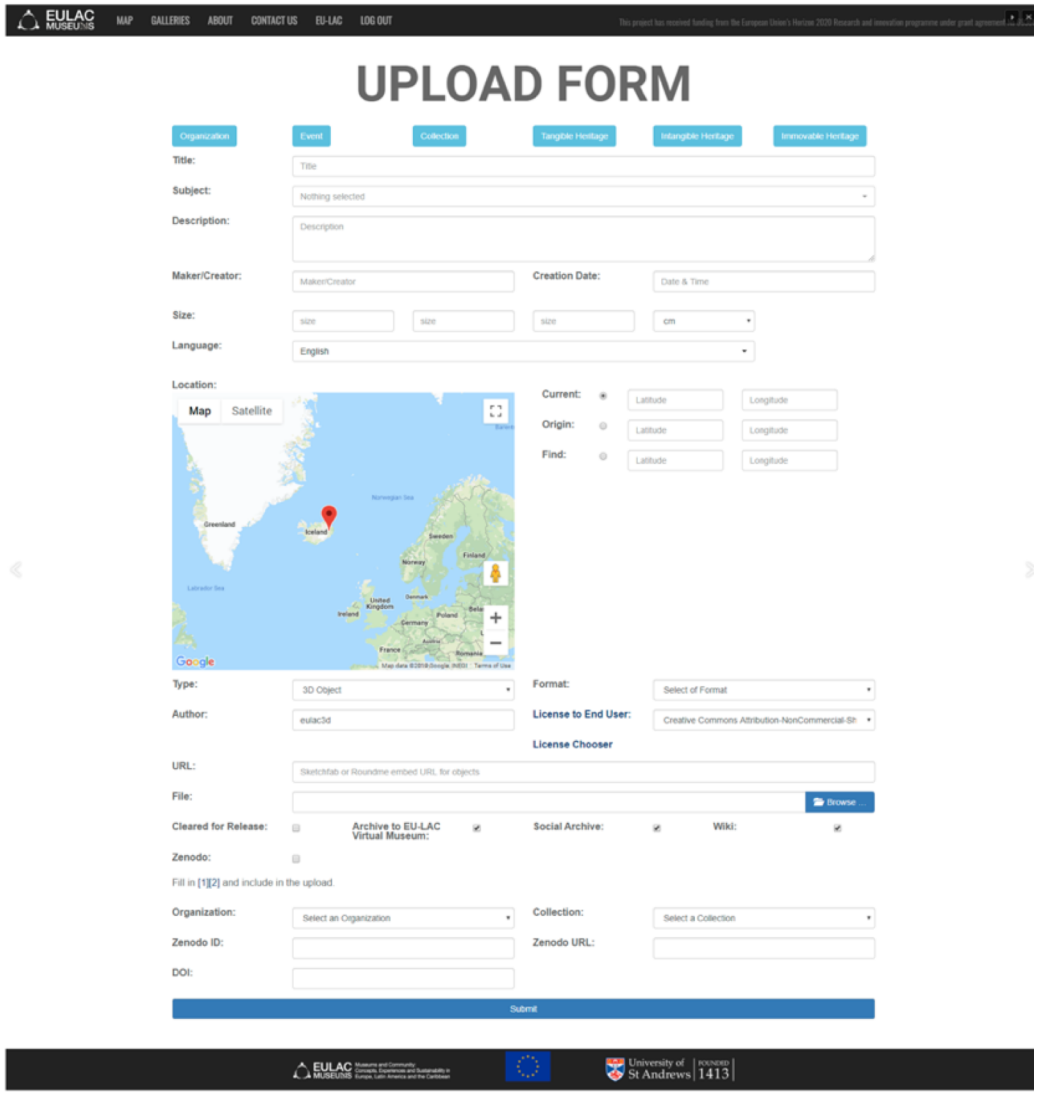
Examples: Analysis for Open 3D data

Team: Marissia Deligiorgi, Dante Abate, George Artopoulos



Examples: Input metadata





https://eulacmuseums.net/index.php

Examples: Data Processing

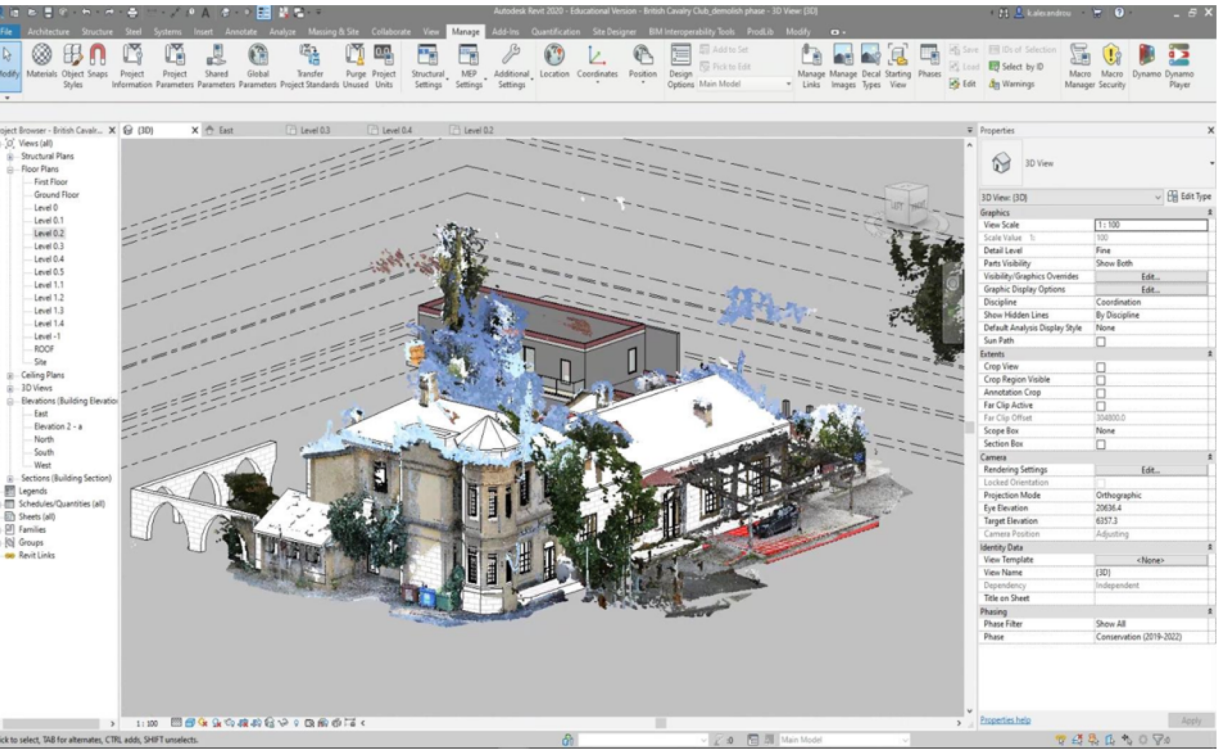
3 - Data Processing*:* image matching software integration

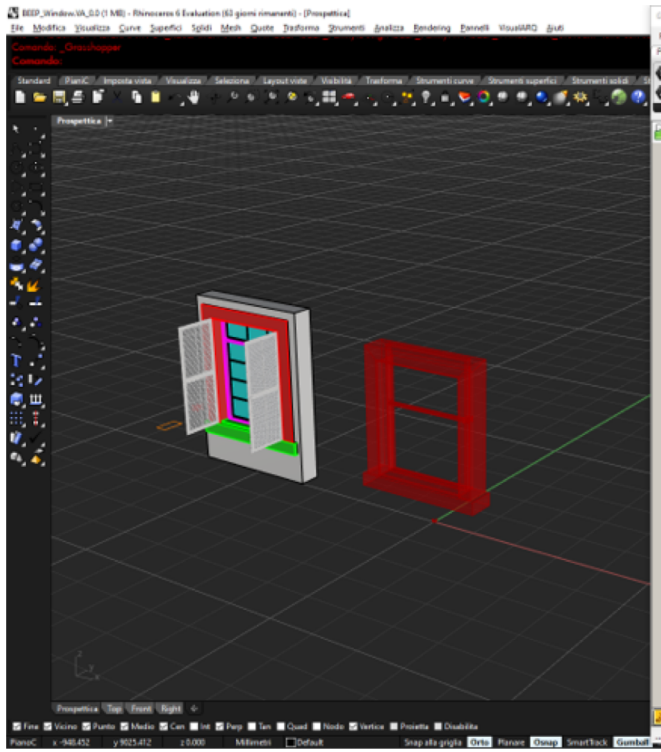
Digital cultural heritage repositories require image classification and retrieval techniques.

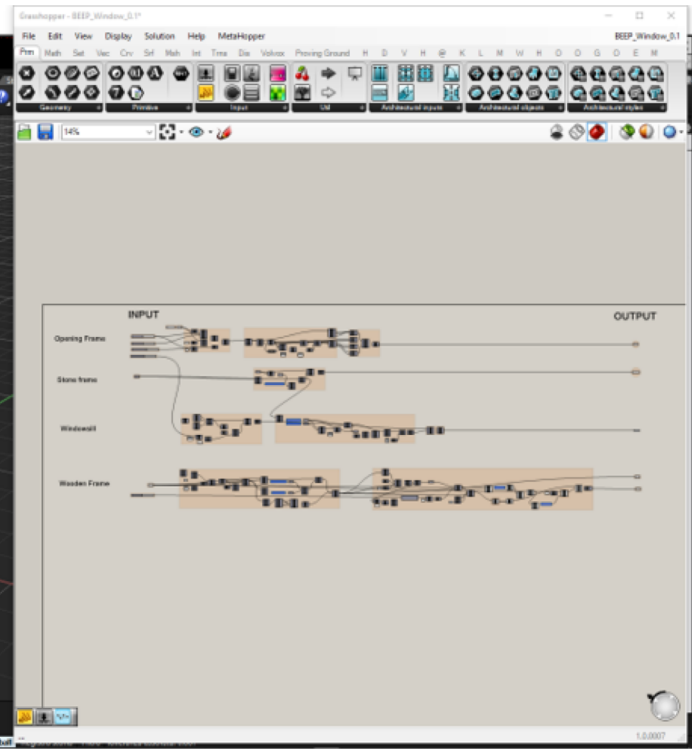
E.g., photogrammetry

Examples

ENI-CBC-MED project BIM for Energy Efficiency in the Public sector (BEEP): http://www.enicbcmed.eu/projects/beep



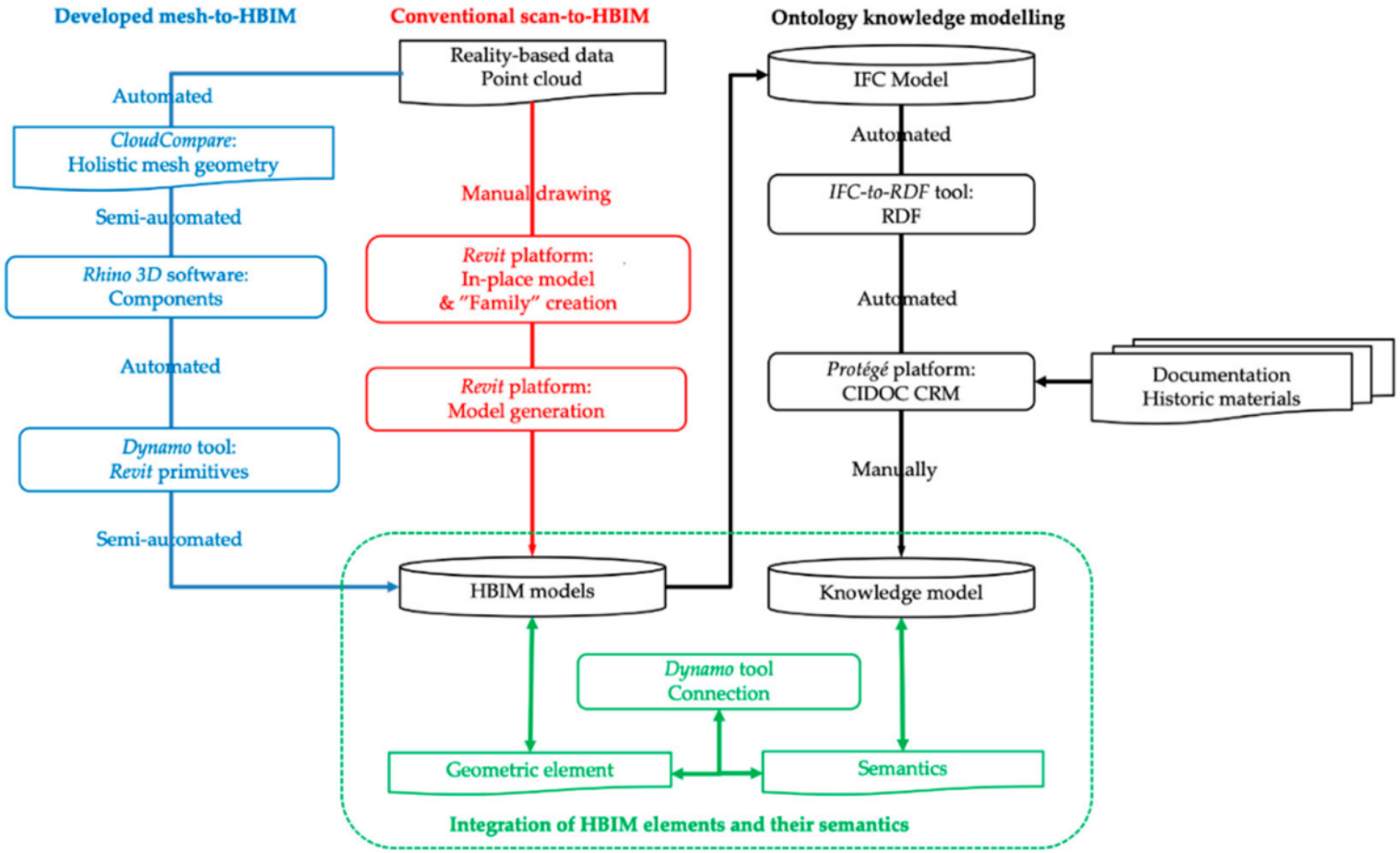




Team: Kristis Alexandrou, Stavroula Thravalou, George Artopoulos

ENI-CBC-MED project BIM for Energy Efficiency in the Public sector (BEEP): http://www.enicbcmed.eu/projects/beep

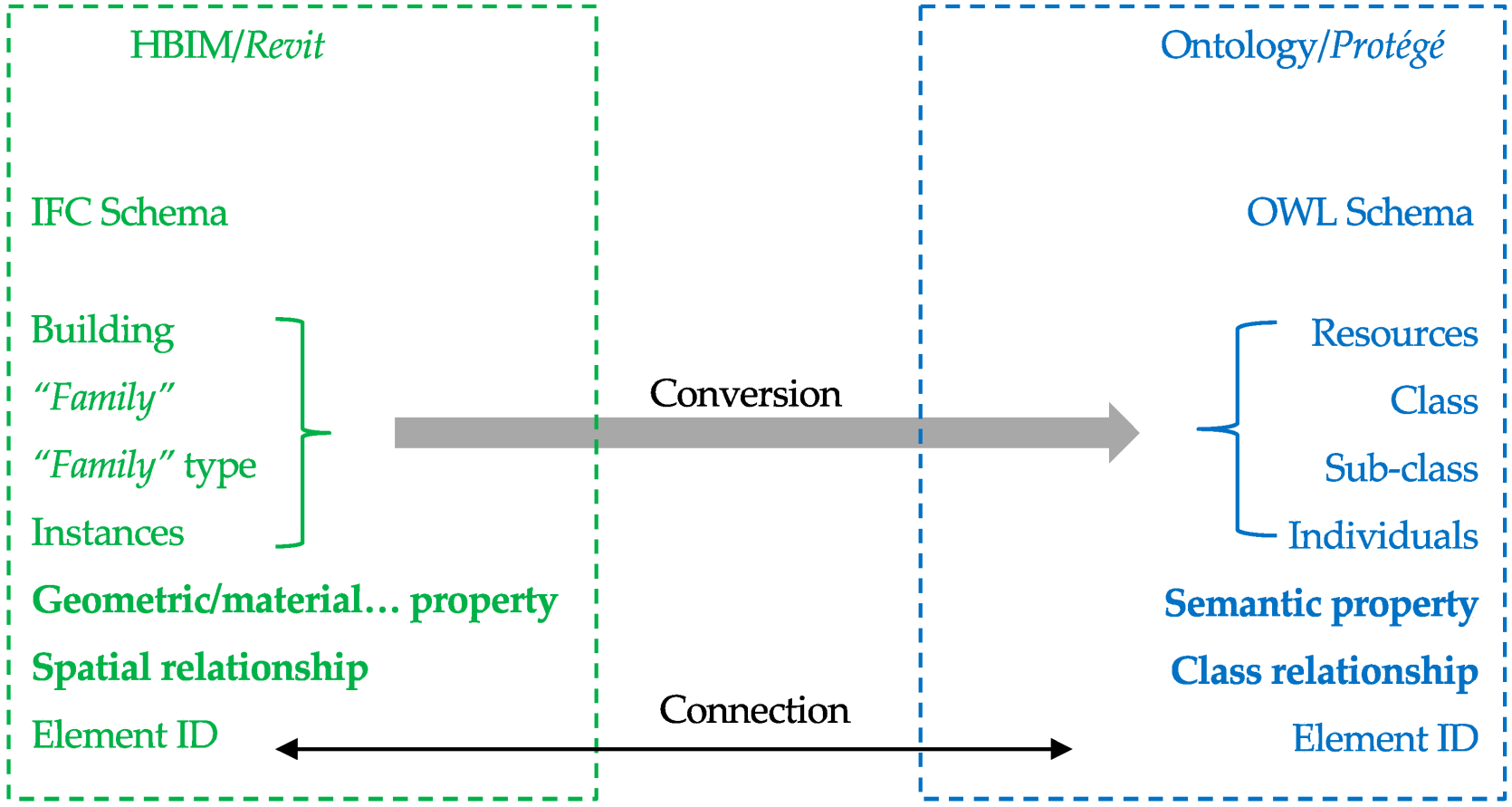
Challenges: Building Information Modeling for Heritage



Examples

ENI-CBC-MED project BIM for Energy Efficiency in the Public sector (BEEP): http://www.enicbcmed.eu/projects/beep

Data mapping (architecture <> DH)



Examples

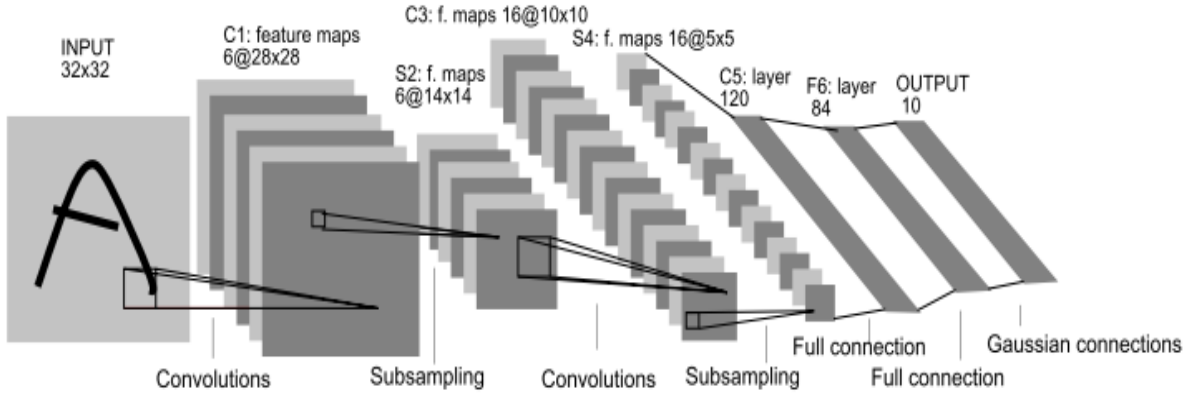
An Artificial Neural Network Framework for understanding historical monuments Architectural Structure and Style

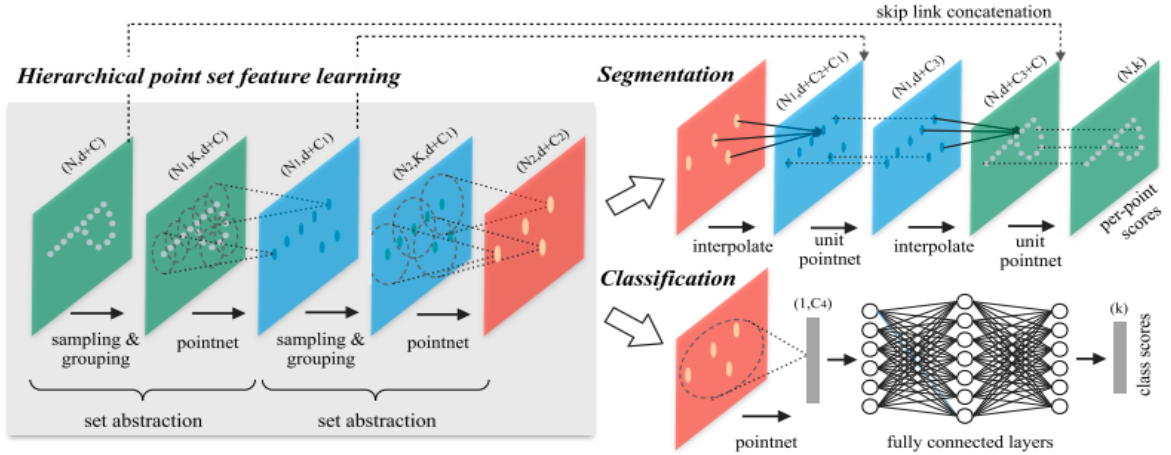
Team: Maria Igarievna Maslioukova, Marissia Deligiorgi, Melinos Averkiou, George Artopoulos, Evangelos Kalogerakis, Gustavo Patow, Yiorgos Chrysanthou.

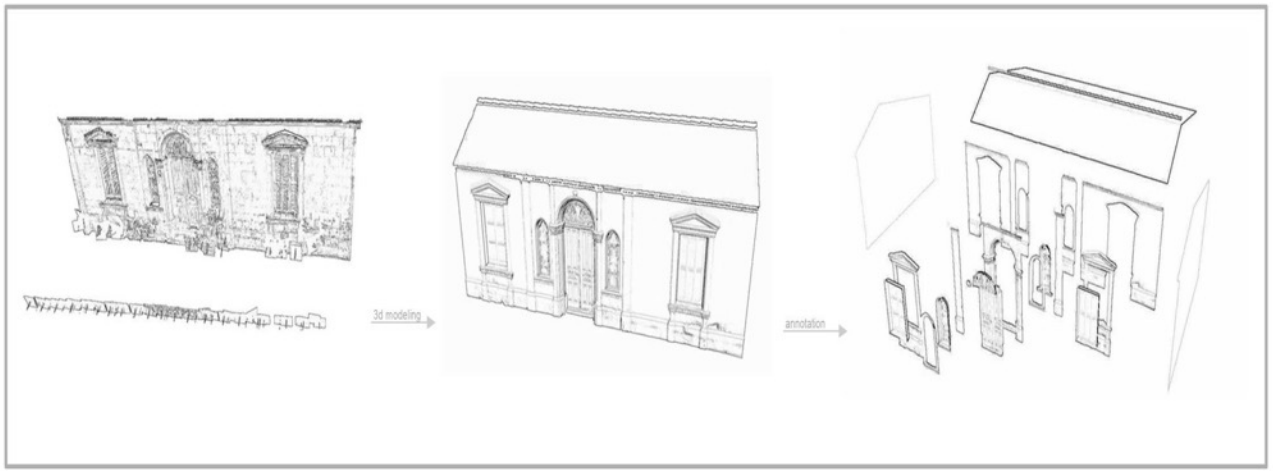
http://annfass.cs.ucy.ac.cy/annotation/admin_labelling.php

Examples

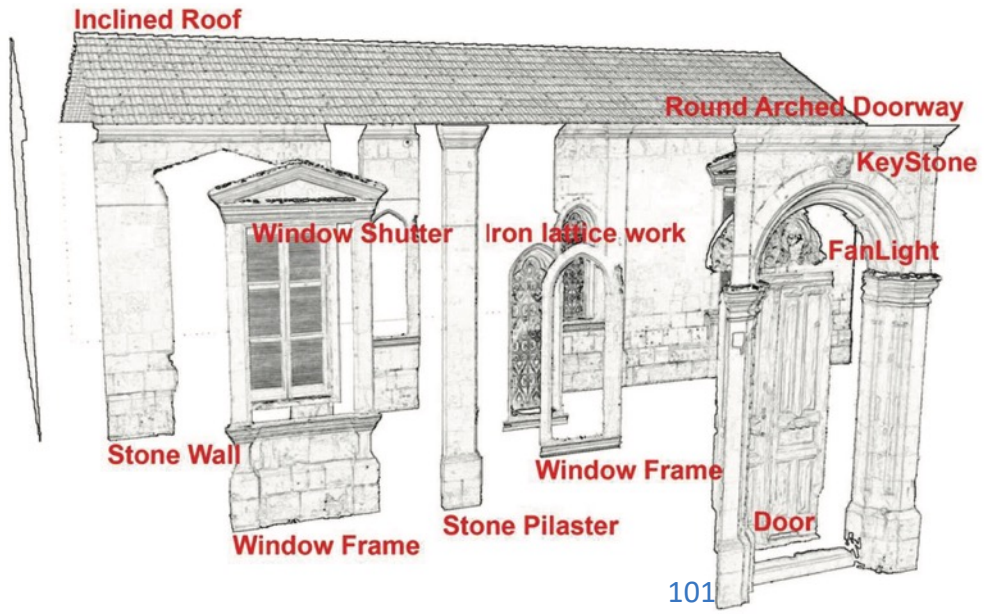
An Artificial Neural Network Framework for understanding historical monuments Architectural Structure and Style







CityHouse in Ayioi Omoloyites



Working Group in DARIAH on Research Data Management

1. Data management best practices in the humanities.
2. Concise, discipline- or data type-focused case studies.
3. Facilitating access to Cultural Heritage data (Heritage Data Reuse Charter).

https://www.dariah.eu/2020/04/16/new-dariah-working-group-on-research-data-management/