

Making Archaeological Data FAIR

Julian Richards



European Archaeological Association, Bern
5 Sept 2019

Outline

- Research data and digital corpora
 - How should we be developing them?
 - The Bigger Picture – infrastructure frameworks
- Importance of FAIR data
- Research e-infrastructures, and ADS role:
 - E-RIHS (European Research Infrastructure for Heritage Science)
 - SSHOC (Social Sciences and Humanities Open Cloud)
 - ARIADNEplus
 - Archaeologists need to engage with these initiatives
 - Introducing the ARIADNEplus taskforces

FAIR DATA

- Findable
- Accessible
- Interoperable
- Re-usable

Introduction to FAIR principles

Findable

Easy to find by both humans and computer systems;

Accessible

Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions;

Interoperable

Ready to be combined with other datasets by humans as well as computer systems;

Reusable

Ready to be used for future research.

In practice....

- Being FAIR when archaeological information is MEAN:
 - Miscellaneous
 - Exceptional
 - Arbitrary
 - Nonconformist

(Isto Huvila, Centre for Digital Heritage conference, University of Leiden 2017)

Guidelines to *fairify* your data management



PARTHENOS FAIR guidelines

20 GUIDELINES to FAIRify data management and make data reusable



1 Invest in people and infrastructure

An important prerequisite to be able to implement the rest of the nineteen guidelines in this guide, is to invest in data infrastructures and in hiring and educating data experts.



Get acquainted with best practices in research data management. Check out the PARTHENOS training modules on data management or have a look at the CESSDA Data Management Expert Guide.



Invest in hiring and educating data experts and define a budget for making investments in technical infrastructure and staff.

FINDABLE

Research data should be easy to find by both humans and computer systems and based on mandatory descriptions of the metadata that allows the discovery of interesting datasets.

2 Use persistent identifiers

Locating data is a necessary condition for any other step from access to reuse. To be findable, any data object and dataset should be uniquely and persistently identifiable over time with a persistent identifier (PID). A PID continues to work even if the web address of a resource changes. PIDs can take different forms, such as a Handle, DOI, PURL, or URN.



Reference the PID which was assigned to your dataset in your research output.



Select the appropriate form of persistent identification schema and assign a PID to every resource. Use the PID Guide from NCDD to decide on the right PID for your research infrastructure.

3 Cite research data

If research data have a persistent identifier and are cited in accordance with community standards, the corresponding data objects or datasets are more easily found.



Get acquainted with data citation guidelines that are specific to your field or discipline and cite research data accordingly.



Provide information about best practices in data citation to research communities and make it easy for data users to cite data, e.g. by using a standardised button which says 'How to cite this dataset'.

ACCESSIBLE

Research data should be easily accessible and retrievable with well-defined access conditions using standardised communication protocols.

4 Choose a trustworthy repository

A certified repository offers a trustworthy home for datasets. Certification is a guarantee that data are stored safely, and will be available, findable and accessible over the long-term. Examples of certification standards are CoreTrustSeal, nestor seal and ISO 16363 certification.



Make your data accessible through a trustworthy repository. In addition, if you follow the repositories' standards (on preferred file formats, metadata schemas etc.) you can make sure that all requirements for making data FAIR are met.



Clearly state the level of certification on your website. If you are not (yet) certified, state how you plan to ensure availability, findability, accessibility and reusability in the long-term.

7 Clearly state accessibility

Access information specifies how a data user may access a dataset. When depositing data in a data repository, it should be clear which access options a data depositor can choose.



When choosing an access option, consider legal requirements, discipline-specific policies and ethics protocols when applicable. Choose Open Access when possible. When you collect personal data, ask yourself whether it contains any information which might lead to participants' identities being disclosed, what participants consented to and which measures you have taken to protect your data. If your data cannot be published in Open Access, the metadata should be, allowing data discovery.



Encourage (metadata to be published in Open Access. Clearly state restricted access options for sensitive (metadata that should not be part of the publicly accessible (metadata). In this case, strive to make the (metadata available through a controlled and documented access procedure.

4 Use persistent author identifiers

A persistent author identifier (e.g. VIAF, ISNI or ORCID) helps to create linkages between datasets, research activities, publications and researchers and allows recognition and discoverability.



Distinguish yourself from any other researcher or research group. Apply for an author identifier if you do not already have one and reference it in your dataset.



Reference author identifiers in the metadata.

5 Choose an appropriate metadata schema

Metadata is essential in making data findable, especially the metadata which is used for citing and describing data. A metadata schema is a list of standardised elements to capture information about a resource, e.g. a title, an identifier, a creator name, or a date. Using existing metadata schemas will ensure that international standards for data exchange are met.



To enable the discovery of content, describe research data as consistently and completely as possible. Include enough information for the data to be accessed and understood later on. If possible, use an existing metadata schema which fits the type of data object or dataset you are describing.



Clearly state which metadata schema you apply and recommend to the research community. To enrich datasets at data deposit, consider having a data submission form which collects additional metadata, e.g. about the provenance of the data.

6 Use a data embargo when needed

During a data embargo period, only the description of the dataset is published. The data themselves are not accessible. The full (meta)data will become available after a certain period of time.



Clearly state why and for what period a data embargo is needed. Make the (meta)data openly available as soon as possible.



Specify whether a data embargo is allowed and what conditions apply.

9 Use standardised exchange protocols

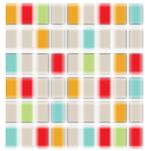
By using standardised exchange protocols, research infrastructures can make (meta)data publicly accessible and harvestable by e.g. search engines, vastly improving accessibility.



Use standardised protocols such as SWORD, OA-PMH, ResourceSync and SPARQL. Convert metadata schemas into XML/RDF. Maintain a registry for protocol endpoints, the path at which research data can be accessed, and publish them.

Research e-infrastructures





E-RIHS

EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE



synchrotron, laser, ion beam,
neutron, microscopy
dating, biogeochemistry
genomics, proteomics



portable instruments
material sampling

FIXLAB
Large-scale facilities

MOLAB
Fleet of advanced mobile instruments

DIGILAB
Data Infrastructure

ARCHLAB
Scientific archives
access to scientific archives
access to reference collections

data documentation and sharing
high performance computing



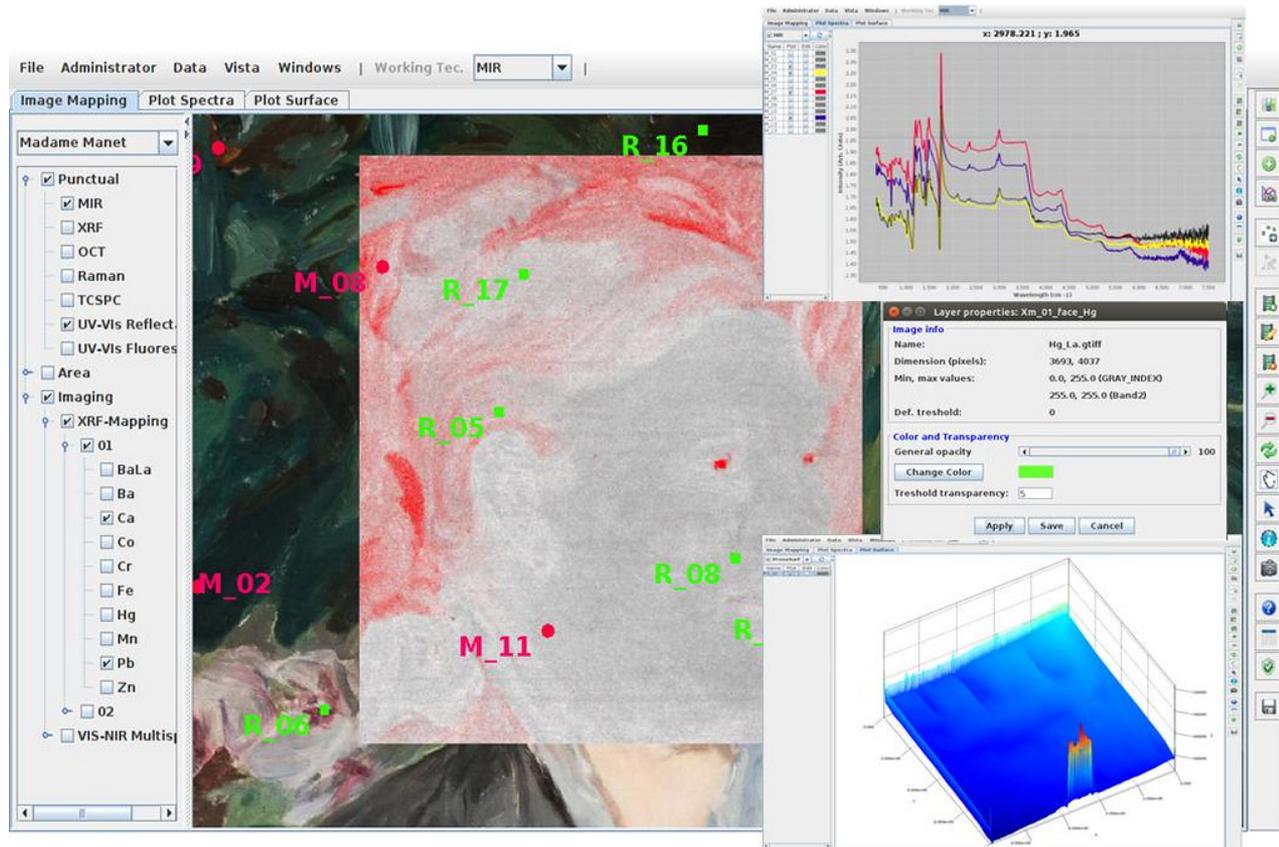


E-RIHS

EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

DIGILAB

Digital data and tools: Virtual access to scientific data concerning tangible heritage



- **(Task 3.3) Financial aspects of data policy and management**

Heritage data sets are potentially very large and are collected in a variety of formats, including born-digital. However, they have a long term re-use value and European funding bodies are increasingly adopting Open Data policies. This can lead to significant long term costs – an aspect that is rarely considered. This task explores the cost models for the long term preservation of heritage data

- **(Task 5.4) Data curation**

Issues concerning data curation for heritage science:

- data quality assurance
- data life-cycle
- data management and preservation

The task defines policies to be adopted and will provide guidelines for researchers, e.g. for the creation of data plans within research projects.



SSHOC

social sciences & humanities open cloud

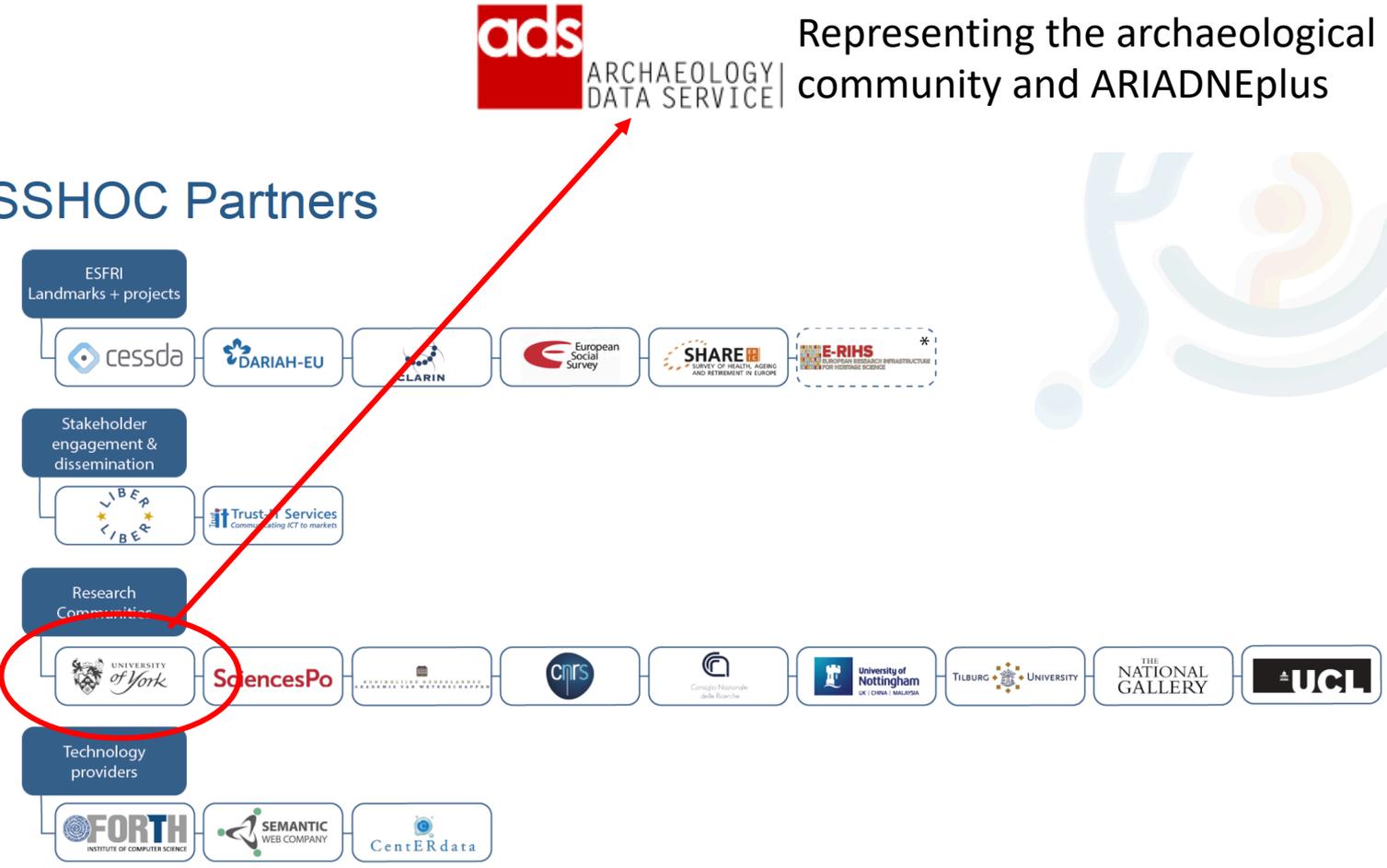
Social Sciences and Humanities Open Cloud

Realising the Social Sciences and Humanities part of
the European Open Science Cloud

SSHOC Partners

Representing the archaeological community and ARIADNEplus

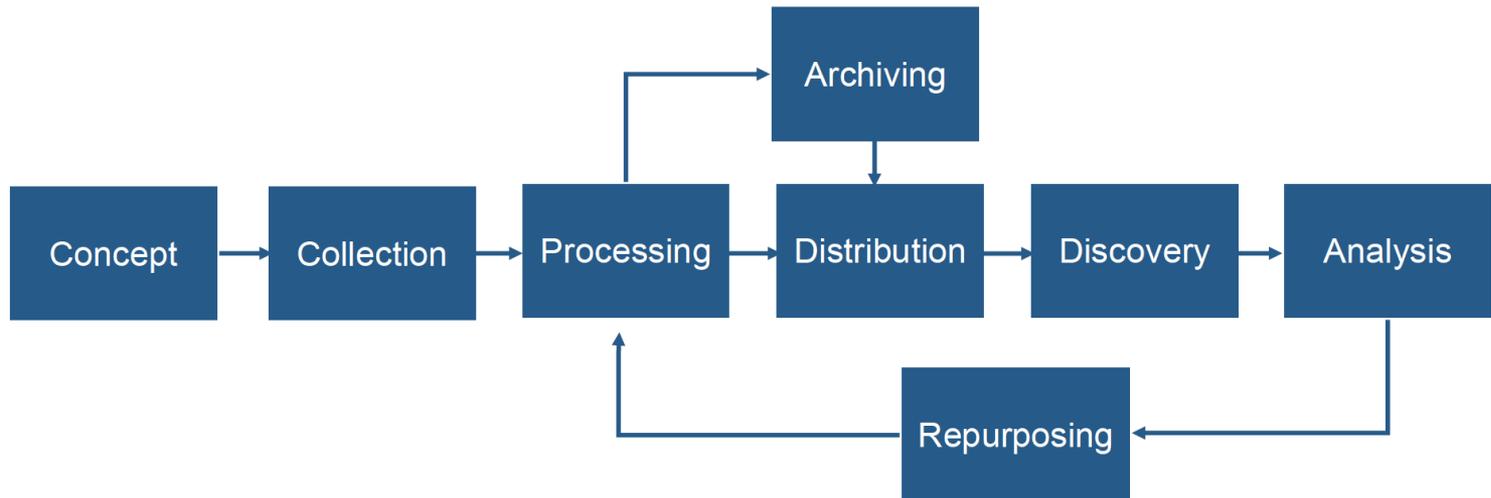
SSHOC Partners



FAIR principles

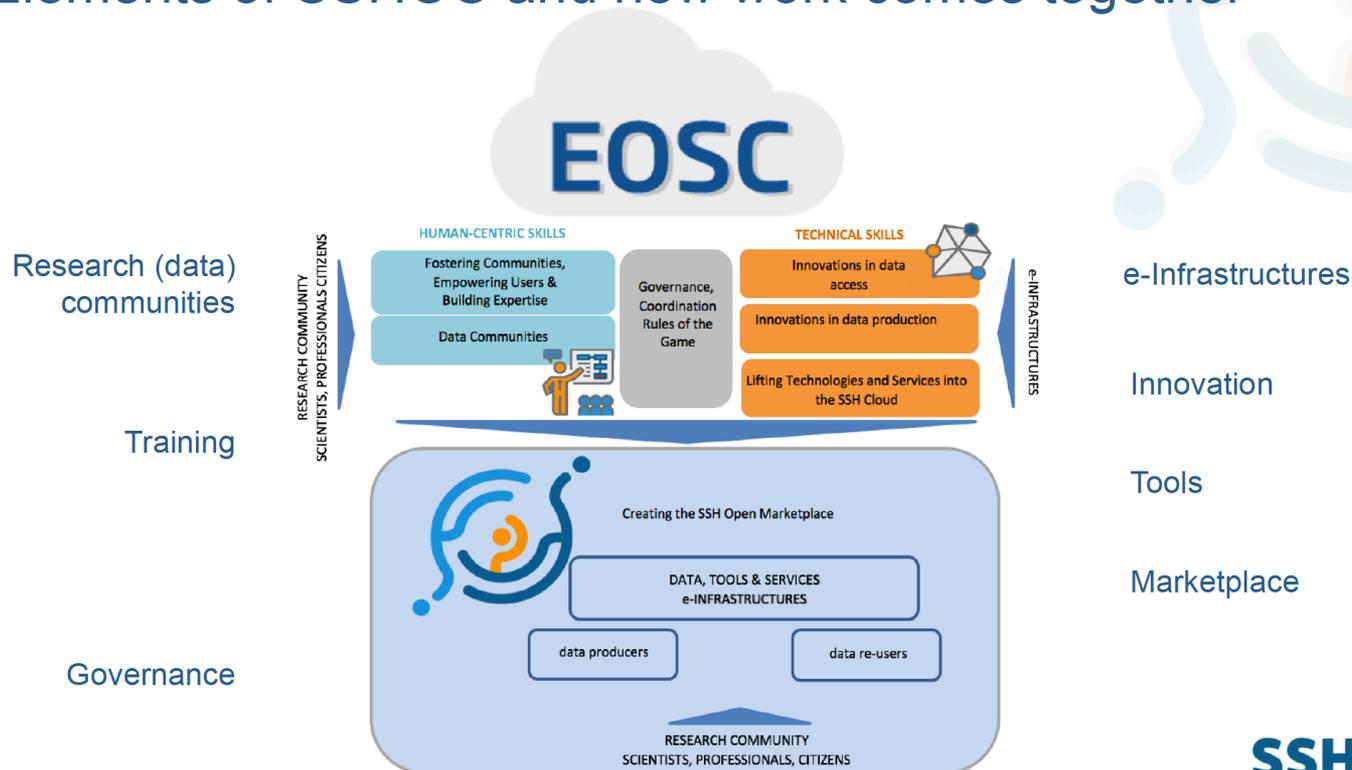
SSHOC consortium covers the whole data cycle

🌀 Following the FAIR Principles



SSHOC and EOSC

Elements of SSHOC and how work comes together



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

- **Task 5.6 Issues in providing Open Data in Heritage Science and Archaeology**

Complex Heritage Science datasets present special accessibility and interoperability issues.

Archaeological and heritage data may provide the only archival record of heritage which is destroyed in the course of the research. In different European countries there are varying approaches to open access for archaeological data reflecting different legal protection systems, sensitivities surrounding site location, and different attitudes to data collection e.g. metal-detecting by members of the public.

This task will enable progress in data sharing in this domain and will be applicable to other domains handling diverse interdisciplinary datasets.

ARIADNE_{plus}

- Home
- About
- Consortium
- Portal
- Community
- Resources
- Transnational access
- News
- Events
- Features



ARIADNEplus Kick-off meeting

READ THE NEWS

- Catalog
- Services
- About



Deputy Coordinator
Lead archaeological partner

ARIADNE

All fields Search for resources in the Ariadne catalog ...

Welcome
Explore the digital resources and services that ARIADNE has brought together from across Europe for archaeological research, learning and teaching.

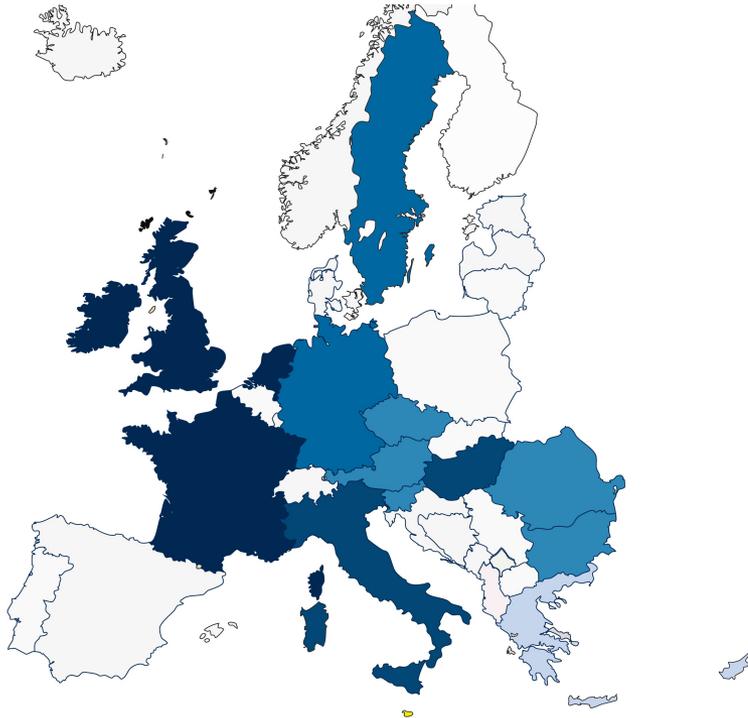
Browse the Catalog

- Where: [Map of Europe]
- When: [Line graph showing data over time]
- What: pits (earthworks), churches (buildings), kilns, cairns, houses, forts, ditches, graves, burials, farms, farmhouses, cargo vessels

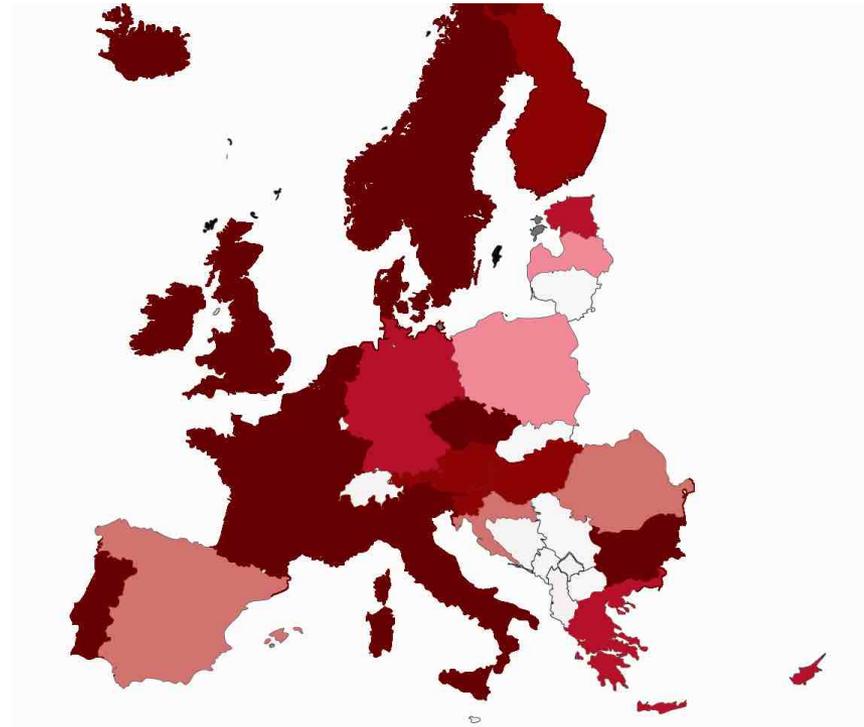
http://archaeologydataservice.ac.uk

Extending geographically

- ARIADNE:
- 24 partners; 18 countries

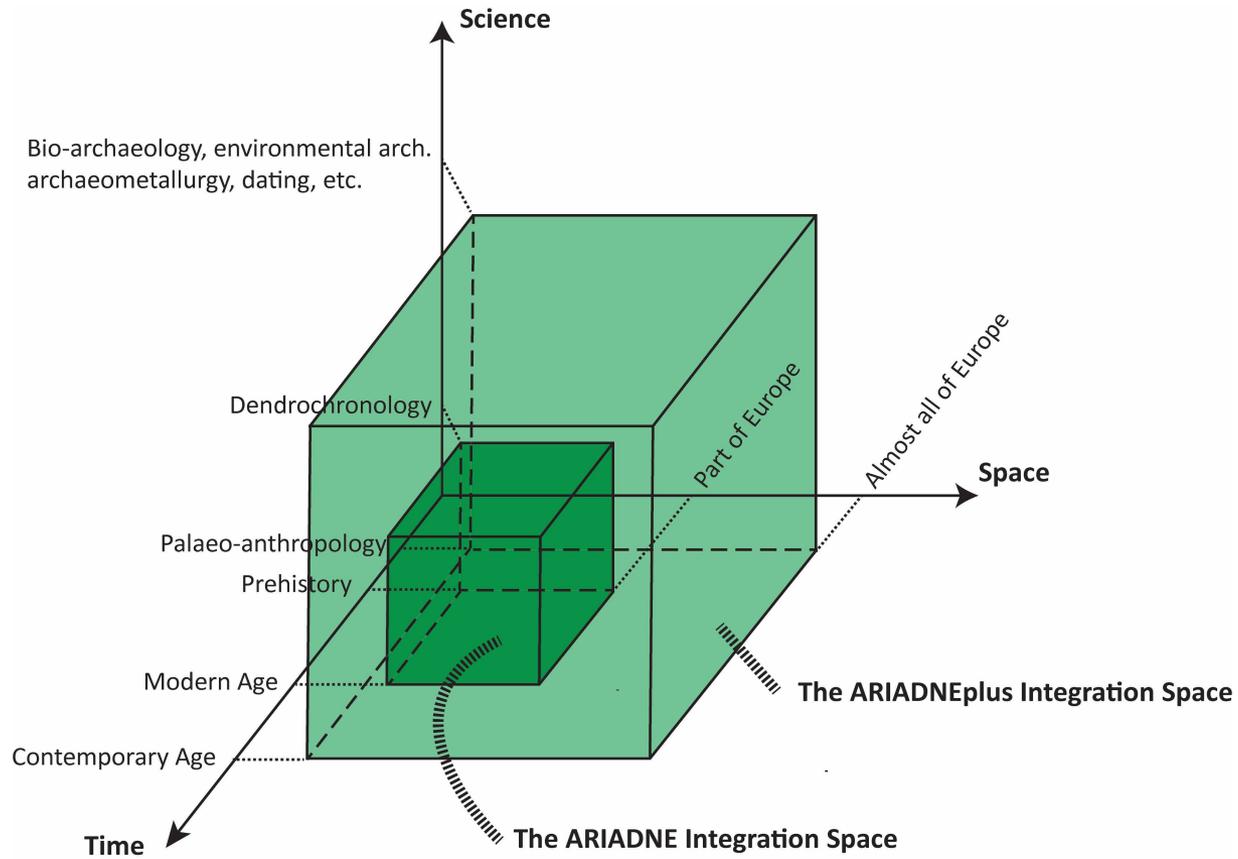


- ARIADNEplus:
- 41 partners; 27 countries



Plus: Argentina, Japan & USA

Extending thematically



ARIADNEplus special interest groups

<u>Paleo-anthropology</u> (CENIEH, Spain)	<u>Remote Sensing</u> (ZRC-SAZU, Slovenia)
<u>Bio-archaeology and Ancient DNA</u> (FORTH, Crete)	<u>Standing Structures</u> (LNEC, Portugal)
<u>Archaeological finds made by general public</u> (Aarhus University, Denmark)	<u>Spatio-temporal data</u> (ARUP-CAS, Czech Republic)
<u>Environmental Archaeology</u> (Umea University, Sweden : SEAD)	<u>Maritime and underwater archaeology</u> (DGPC, Portugal)
<u>Inorganic Materials Study</u> (INFN, Italy)	<u>Archaeological fieldwork</u> (INRAP, France)
<u>Field Survey</u> (University of Groningen, Netherlands)	<u>Inscriptions</u> (University of Barcelona, Spain)
<u>Burials</u> (OAEW, Austria)	<u>Dating</u> (INFN, Italy)

ARIADNEplus and FAIR data

- Good practice in archaeological data management
- ARIADNEplus policy support tools
 - A DMP flexible template & domain protocol
 - A policy wizard explaining the main principles concerning archaeological data management
 - A standardization wizard, documenting major standards in the archaeological domain as well as authority files such as thesauri, reference collections, gazetteers etc
- Providing guidelines and support on repository creation and management
- Providing guidelines and support on repository quality control
- Managing FAIRness of archaeological data and IPR
- Training on FAIR Data Management



Thank you for listening

E-RIHS PP is a Horizon 2020 project funded by the European Commission under Grant Agreement n. 739503



SSHOC is a Horizon 2020 project funded by the European Commission under Grant Agreement n. 823782



ARIADNEplus is a Horizon 2020 project funded by the European Commission under Grant Agreement n. 823914.



The views and opinions expressed in this presentation are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

Contact:

julian.richards@york.ac.uk

Twitter: @julian62523002