WP5 Access and Interoperability Policies
Task 5.3 Data curation

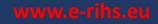York | 2 April 2020

Holly Wright | Archaeology Data Service

Task 5.3 Data curation

Addresses issues concerning data curation for heritage science, e.g. data quality assurance, the data life-cycle, data management and preservation. It defines policies to be adopted and provides guidelines for researchers, e.g. for the creation of data plans within research projects.

D5.3 Data curation policy

- Reviews issues concerning data curation for heritage science.

- Provides a policy framework to be implemented by E-RIHS, but designed to be of use to all those with interests in data within the heritage science domain.

- Follows the framework provided by the FAIR principles (Findable, Accessible, Interoperable and Re-usable) but interprets them in the context of heritage science.

- Examples are given with reference to a substantial appendix which covers a broad range of heritage science data types.

Findability

1. E-RIHS repositories will need to assign persistent identifiers to datasets, and E-RIHS users should reference these PIDs in their research outputs.

2. E-RIHS should provide information about best practices in data citation to the heritage science research community and repositories, allowing users to easily cite the data, e.g. by using a standardised button which says 'How to cite this dataset'.

3. E-RIHS users should register for an ORCID.

4. E-RIHS should build communities to develop relevant metadata schemas and standards for heritage science.

Accessibility

5. E-RIHS should ensure Heritage Science research data is easily accessible and retrievable with well-defined access conditions, using standardised communication protocols.

6. E-RIHS should work to create and sustain appropriate E-RIHS repositories.

7. E-RIHS repositories should obtain appropriate certifications.

8. E-RIHS researchers should consider legal requirements, discipline-specific policies and ethics protocols when applicable.

Accessibility

9. E-RIHS researchers should work to make their data Open Access whenever possible.

10. If data cannot be made Open Access, the metadata should be, which at least allows data discovery.

11. E-RIHS repositories should make (meta)data publicly accessible and harvestable by e.g. search engines, vastly improving accessibility.

12. E-RIHS should use standardised protocols to enable greater interoperability.

Accessibility

13. E-RIHS should maintain and publish a registry of protocol endpoints as part of DIGILAB.

14. E-RIHS should support new and developing repositories and provide best practice guidance to ensure they take the form most optimal for re-use within the E-RIHS data ecosystem.

15. E-RIHS should support repositories that make data associated with publications more accessible.

Interoperability

16. E-RIHS should support interoperability standards that are both human and machine-readable.

17. E-RIHS should promote active standards-based user development communities for heritage science

18. E-RIHS should publish the metadata models in use by the heritage science community as part of the resources in DIGILAB.

19. E-RIHS should document the technical specifications of metadata models, including defining the classes and properties, including those which are mandatory and recommended.
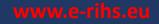
Interoperability

20. All data files held in E-RIHS repositories should be in an open, international, standardised file format to ensure long-term interoperability in terms of usability, accessibility and sustainability.

21. Researchers should use preferred formats which are recommended by E-RIHS and are independent of specific software, developers or vendors.

Re-usability

22. E-RIHS researchers should ensure heritage science research data is ready for future research and future processing.

23. E-RIHS researchers and laboratories should ensure research data is systematically documented.

24. E-RIHS researchers and laboratories should ensure they maintain adequate version control for research data.

25. E-RIHS researchers should follow a precise and consistent file naming convention.

Re-usability

26. E-RIHS repositories should develop guidelines recommending standardised preferred file formats that are widely used in the Heritage Science community.

27. E-RIHS repositories should develop metadata requirements that include information about provenance of samples, name of the laboratory, methodology and equipment.

28. To permit the widest reuse possible of (meta)data, it should be clear who the (meta)data rights holder is and what license applies.
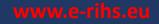
Re-usability

29. E-RIHS should adopt the Creative Commons licencing framework, and map other frameworks used within E-RIHS to it.

30. Metadata should be made available under a CC-0 licence by default.

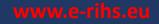31. Datasets should be made available under a CC-BY licence by default.

Data Management Planning

32. Completion of a data management plan should be a requirement for E-RIHS support for access to E-RIHS facilities.

33. E-RIHS should adopt the PARTHENOS DMP template as the default.

Appendix Workflow

Overview/Planning

- Purpose - why is the data being created? Are there limitations to the approach or subject?

- Expected reuse/intended audience - are the intended outputs limited or restricted in any way?

- Needs (dissemination/preservation) - are there existing requirements to share or preserve the data?

- Preparation - are there documentation and procedures relating to location, data collection, equipment, and testing?

Appendix Workflow

Collection/Creation

- Raw data - identify the 'raw data' alongside any initial (in-device or manual) data cleaning that is undertaken.

- Format options - what formats are available at these early stages. Are these native or open formats, standards-compliant, etc.?

- Equipment settings - is there documentation for the settings or environments used in data collection or processing? Are these files self-generated or require manual documentation? Do documented protocols exist?

Appendix Workflow
Processing (post-acquisition)

- Policy and protocols - as above during the collection phase, are there policy or protocol documents available for the post-acquisition phase that may help users understand and reuse data?

- Intermediary datasets - are there intermediate or transitory sets of data (i.e. between the raw and final dataset) that should be saved?

- Other versions of data - are there multiple outputs for different purposes (dissemination, preview, graphical, etc.)?

Appendix Workflow
Long-term Curation

- Files / Formats - is the final data suitable for long-term preservation (format)?

- Data Selection - has the final dataset had working or draft files removed?

- Structure - is there a meaningful structure to the final dataset? Are relationships between files (raw, derived, etc.) clear? Is the relevant documentation included?

Spectroscopy and Material Analysis

- X-ray Crystalline Powder Diffraction (XRD and XRPD)
- X-ray Fluorescence Spectroscopy (XRF)
- Raman Spectroscopy
- Infrared Spectroscopy
- Thermal Analysis

Microscopy

- Optical, Fluorescence and Metallographic Microscopy
- Particle Analysis
- Confocal Laser Microscopy (CLSM)
- Scanning Electron Microscopy (SEM)

Dating Methods

- Potassium Argon

Biomolecular Methods

- Palaeoproteomics (ZooMS; proteomics)
- Isotope ratio mass spectroscopy (IRMS)

Synchrotron Methods

- X-ray absorption spectroscopy (XAS)
- X-ray Crystalline Powder Diffraction (XRD and XRPD)
- X-ray absorption near-edge structures (XANES)
- X-ray Fluorescence Spectroscopy (XRF)

E-RIHS

EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

website	email	contact
www.e-rihs.eu	co@e-rihs.eu	holly.wright@york.ac.uk

E-RIHS was officially adopted by the ESFRI roadmap in December 2016