



# The Heritage Data Reuse Charter: from principles to research workflows

Erzsébet Tóth-Czifra, Laurent Romary

## ► To cite this version:

Erzsébet Tóth-Czifra, Laurent Romary. The Heritage Data Reuse Charter: from principles to research workflows. 2020. halshs-02475692

**HAL Id: halshs-02475692**

**<https://shs.hal.science/halshs-02475692>**

Preprint submitted on 27 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# The Heritage Data Reuse Charter: from principles to research workflows

Erzsébet Tóth-Czifra; Laurent Romary

**Summary:** There is a growing need to establish domain- or discipline-specific approaches to research data sharing workflows. A defining feature of data and data workflows in the arts and humanities domain is their dependence on cultural heritage sources hosted and curated in museums, libraries, galleries and archives. A major difficulty when scholars interact with heritage data is that the nature of the cooperation between researchers and Cultural Heritage Institutions (henceforth CHIs) is often constrained by structural and legal challenges but even more by uncertainties as to the expectations of both parties. The Heritage Data Reuse Charter aims to address these by designing a common environment that will enable all the relevant actors to work together to connect and improve access to heritage data and make transactions related to the scholarly use of cultural heritage data more visible and transparent.

As a first step, a wide range of stakeholders on the Cultural Heritage and research sector agreed upon a set of generic principles, summarized in the [Mission Statement](#) of the Charter, that can serve as a baseline governing the interactions between CHIs, researchers and data centers. This was followed by a long and thorough validation process related to these principles through surveys<sup>1</sup> and workshops<sup>2</sup>. As a second step, we now put forward a questionnaire template tool that helps researchers and CHIs to translate the 6 core principles into specific research project settings. It contains questions about access to data, provenance information, preferred citation standards, hosting responsibilities etc. on the basis of which the parties can arrive at mutual reuse agreements that could serve as a starting point for a FAIR-by-construction data management, right from the project planning/application phase. The questionnaire template and the resulting mutual agreements can be flexibly applied to projects of different scale and in platform-independent ways. Institutions can embed them into their own exchange protocols while researchers can add them to their Data Management Plans. As such, they can show evidence for responsible and fair conduct of cultural heritage data, and fair (but also FAIR) research data management practices that are based on partnership with the holding institution.

---

<sup>1</sup> Dorian Seillier, Anne Baillot, Marie Puren, Charles Riondet. Survey on researchers requirements and practices towards Cultural Heritage institutions: Documentation and analysis. [Technical Report] Inria Paris. 2017. [\(hal-01562860\)](#)

<sup>2</sup> See e.g. in The Hague on 09.04.2018: <https://libereurope.eu/events/feedback-workshop-reuse-of-cultural-heritage-charter/>; at the Bibliothekartag 2018 conference in Berlin: <https://www.bibliothekartag.de/archives/2018/bibliothekartag2018.de/programm/programm-3/index.html>; and at the DARIAH Annual Event 2019: [https://dariah-ae-2019.sciencesconf.org/data/pages/AE2019\\_BookOfAbstracts\\_1.pdf](https://dariah-ae-2019.sciencesconf.org/data/pages/AE2019_BookOfAbstracts_1.pdf)

# I. The need for domain-specific approaches in research data management

Due to the increasingly data-driven nature of arts and humanities research and also to the growing pressure of European open data policy mandates, there is a strong demand for domain-specific approaches to research data management. We need to explore pathways to data sharing that are meaningful and relevant for arts and humanities research practices.

Clearly, the generic research data management guidelines do not always align well with the cultural, conceptual and epistemological complexity of research data in the arts and humanities and the many entailments of this complexity such as:

- Data comes in the humanities in a wide variety of source types, formats and corpus sizes. The word 'data' itself is hardly used and mostly replaced by the notion of *primary source*;
- Researchers usually lack know-how as to how to deal with the various dimensions of data management: documentation, hosting, identification or re-use condition are not part of the education curricula in the humanities;
- A fundamental difference between the epistemic cultures of hard sciences and arts and humanities is that in the arts and humanities the wide range of scholarly information referred to as cultural heritage data are not autonomous products of research projects but are deeply embedded in the memory of the institutions (museums, libraries, archives) that preserve, curate and (co)produce them;
- These institutions are not only data providers as ownership of heritage data is inherently shared between them, the researcher communities, the public and the people and cultures that give rise to the objects in question.
- Access to and the digital availability of cultural heritage as the primary condition of research in the majority of humanities disciplines that defines the reusability and accessibility of scholarship built on them.

At DARIAH, we are sensitive that if we are facing a community mandate or an EU mandate to share our research data, then the infrastructure needs to put something in place to ease to make it easier for our communities. We find it important to start our advocacy work on research data management with strong integration of the cultural heritage sector and support collaboration between Cultural Heritage Institutions and researchers on different scales.<sup>3</sup>

---

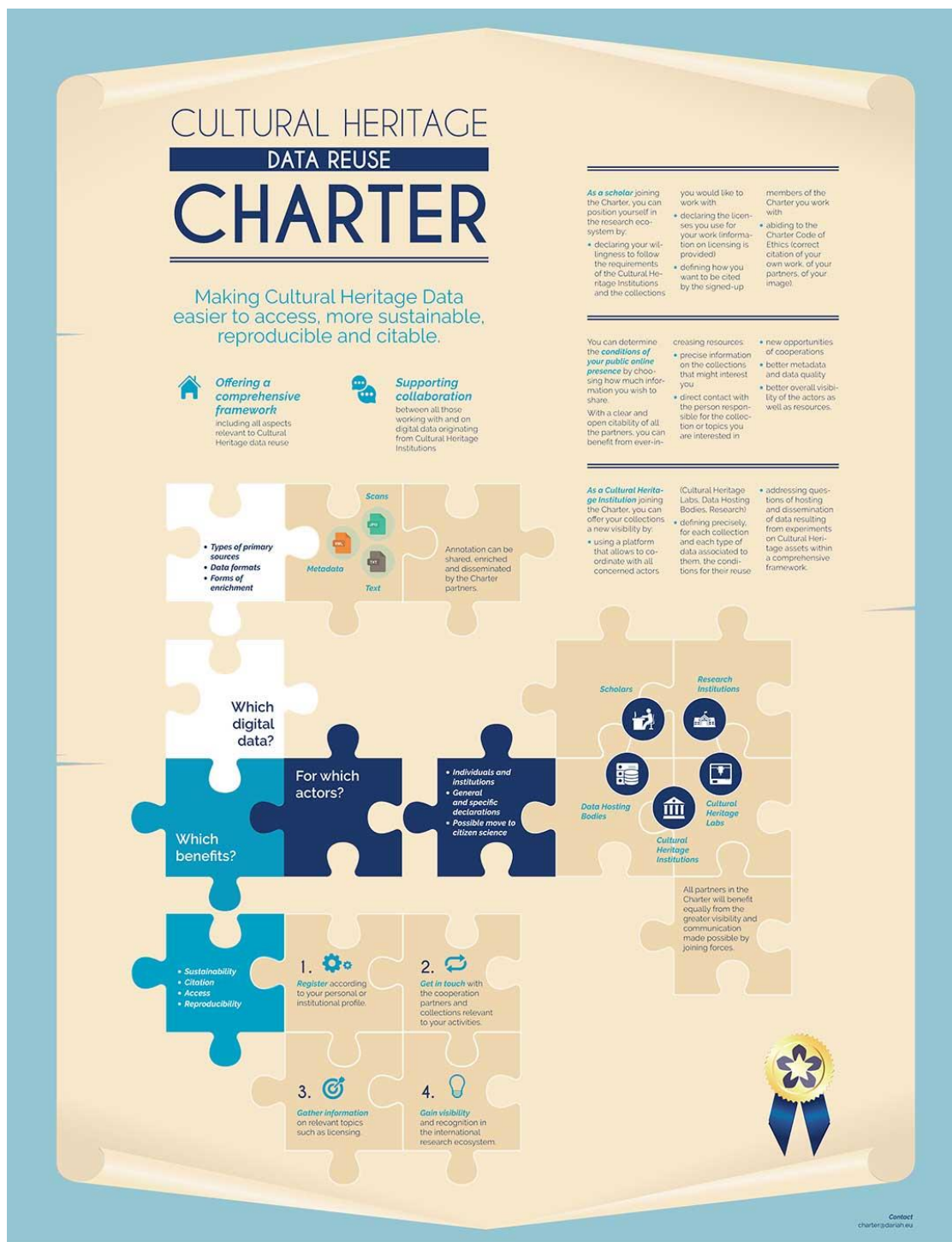
<sup>3</sup> See for instance our [Open data guide](#) for humanists in which, among others, we give practical **checklist** to arts and humanities scholars **with the questions to ask from the curator** of the collection of his/her interest during their first visit (virtual or physical) to the library, archive or museum. On a different scope, recently we published a [collection of best practices to facilitate cooperation between humanities researchers and Cultural Heritage Institutions](#).

## II. The Heritage Data Reuse Charter: a framework to facilitate the handshake between scholars and Cultural Heritage Institutions

The [Heritage Data Reuse Charter](#) aims to address the above mentioned defining characteristic of arts and humanities data and data workflows, namely, its crucial dependence on the availability of Cultural Heritage resources.

A major difficulty when scholars interact with heritage data is that the nature of the cooperation between researchers and Cultural Heritage Institutions (henceforth CHIs) is often constrained by structural and legal challenges. The fact that CHIs and academia operate in different legal and institutional silos has many consequences in the exchange mechanisms across them. The uncertainties around reuse conditions, the translation of the intellectual property rights into scholarly activities, agreeing upon hosting responsibilities and the sustainable preservation of the enrichments scholars contribute to the cultural heritage materials or assuring mutual acknowledgement - these are all cornerstone issues that all too often remain unclear in the course of their interactions. The lack of a clear and comprehensive framework that could serve as a general baseline for such interactions are acknowledged as a recurrent problem that affect the working conditions of all the involved stakeholders and also put serious limitations of accessibility and downstream reuse of the scholarship built on the primary sources.

This recognition led several European organizations such as [APEF](#), [CLARIN](#), [Europeana](#), [E-RIHS](#) with the support of European projects such as [Iperion-CH](#) and [PARTHENOS](#) to come together and join forces under the governance of [DARIAH](#) to set up principles and mechanisms for improving the conditions for the use and re-use of cultural heritage data issued by cultural heritage institutions and studied and enriched by researchers. Our shared aim is to design a common environment that will enable all the relevant actors to connect and improve together access to heritage data and make transactions related to the scholarly use of cultural heritage data more visible and transparent.



**Fig. 1. The concept of the Cultural Heritage Data Reuse Charter.**

## II. 1. The vision behind the of the Heritage Data Reuse Charter: the Mission Statement

The first step of this activity concerned the definition and validation of a set of generic principles, summarized in the [Mission Statement](#) of this Charter, that can serve as a baseline governing the interactions between CHIs, researchers and data centres. These are: Reciprocity, Interoperability, Citability, Openness, Stewardship and Trustworthiness. These principles are fully compliant with and map onto the FAIR principles and can be taken as their optimization for cultural heritage data exchange settings. Key elements of the future environment would include primarily conditions of access and reuse (e.g.license information) for each collection or object registered in the Heritage Data Reuse Charter as well as contact information to the person responsible for specific datasets. Signing the Charter would allow all stakeholders to declare their commitment to the reuse conditions expressed for each collection or object registered in the Charter and thus increase the transparency and visibility of their own workflows.

In the last 2 years, a long validation process related to these principles has taken place through surveys<sup>4</sup> and workshops. The community feedback shows strong and univocal endorsement of the six principles from all stakeholder groups, the comments we received are mainly asking for more details on how these high-level principles can be realized in the day-to-day interactions between institutions, researchers and infrastructure providers.

## II. 2. Translation of the principles to research workflows

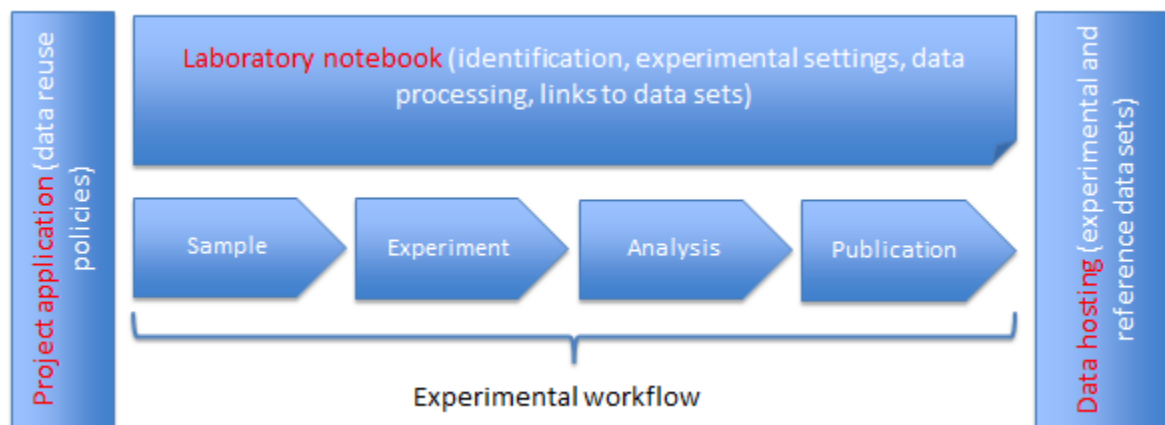
Having established the values that must underpin the increased fluidity of data exchange within the cultural research sphere, we now seek to translate them into a set of actions that can realise this vision.

Creating a common environment to support smooth end-to-end communication between CHIs and researchers will be an essential component of the FAIR data ecosystem in the arts and humanities domain. Such a trusted framework would be a key step in achieving a better alignment of data creation and curation with downstream reuse.

Below you can find a model of such an environment that encapsulates a schematic research life cycle within a political, editorial and technical environment that will ensure that the resulting data sets are generated and managed in an optimal manner. In the following, we will focus on and describe only its first building block, the project application system.

---

<sup>4</sup> See e.g. Dorian Seillier, Anne Baillot, Marie Puren, Charles Riondet. Survey on researchers requirements and practices towards Cultural Heritage institutions: Documentation and analysis. [Technical Report] Inria Paris. 2017. ([hal-01562860](#))



**Fig. 2.** Model workflow for setting up a proper data management environment for scholarly workflows involving cultural heritage resources. Source: Laurent Romary.

Clearly articulating intentions, legal aspects/rights and roles of responsibility in the initial stages of project planning is crucial since in most cases the access conditions to the primary sources will define the reusability of and access to the final research outputs. Keeping this in mind, in the project application phase the Heritage Data Reuse Charter would allow both CHIs, researchers and, if relevant, other infrastructure providers such as research facilities, data centers or data repositories, to clarify their goals at the beginning, to specify their exchange mechanisms, and to clear their expectations in terms of citation and attribution standards, or hosting responsibilities.

The questionnaire below in **Appendix 1** would serve as a tool to guide and structure the conversations and to build mutual agreements on these crucial issues between the parties. It is broken down along six core principles of the Mission Statement into questions that allow each party to explain how they are planning to interact with the data, what they intend to do in practice to abide by the principles, what are their constraints (e.g. on size, formats, documentation, access conditions) and their own expectations towards the other stakeholders during the transaction. These commitments, signed by the relevant stakeholders/partners (the research team, the CHI and, if relevant, an infrastructure provider third party) will be articulated around the following elements:

- Commitment concerning the re-use possibilities associated to source materials, future data sets and research publications built on them (Reciprocity, Openness)
- Commitment to an agreed set of standards in terms of data sharing formats, processes and protocols (Interoperability)
- Commitment to hosting and maintenance responsibilities (involving, if needed third party repositories) and best efforts to enabling the interconnectedness of source data (e.g. via PIDs or linked data protocols<sup>9</sup>, their enrichments and the resulting publications (Stewardship)
- Commitment to keep the richest possible track of documentation and provenance information (Trustworthiness)

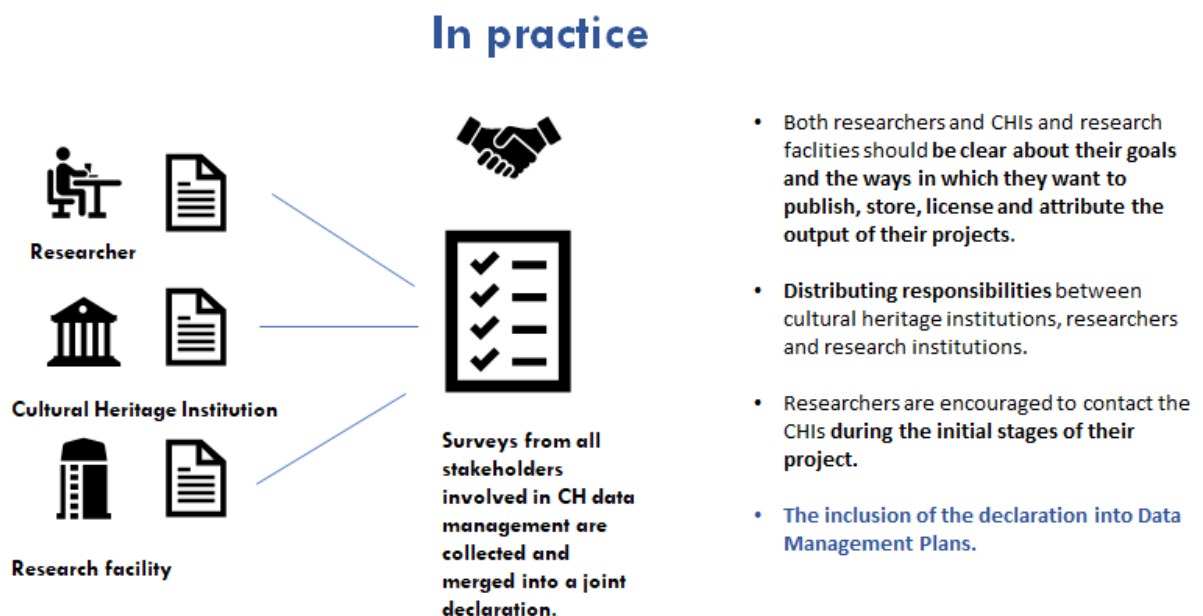


## II. 3. Integration of the Reuse Agreements to exchange protocols (Data Management Plans, project application platforms or the institutions' own exchange protocols)

In practice, the survey below and the mutual agreements between research teams and CHIs can be flexibly applied in platform-independent ways.

Institutions who sign the Charter could use it (and expect to use such surveys) in their own exchange protocols. Another direction of future developments is to set up a platform dedicated to such exchanges. Our previous pilot experiments with the redesign of the ScienceCall platform for this purpose suggest that it could include traceability mechanisms through the entire life cycle of Cultural heritage data.

On the other hand, researchers are encouraged to contact the CHIs during the initial stages of their project in order to explain their plans and figure details of transaction together. This mutual declaration can later be a powerful component in their Data Management Plans as it shows evidence for responsible and fair conduct of cultural heritage data, and fair (but also FAIR) research data management practices that are based on partnership with the holding institution.



**Fig. 3.** Enclosing mutual reuse agreements between CHIs and researchers to Data Management Plans.

As enclosing a Research Data Management Plan to grant applications is becoming a more and more common requirement among research funders, we need to raise the funders' awareness to the fact that such bi- or trilateral agreements and data reuse declarations among researchers,



CHIs and infrastructure providers are crucial domain-specific components of FAIR data management.

## Annex 1. Charter template between CHIs and researchers to guide mutual reuse agreements in the project planning phase

### Reciprocity

*« Both Cultural Heritage Institutions and Researchers agree to share content and knowledge equally with each other, making use of data centers and research infrastructure »*

- **Researcher**

Are you planning to communicate your research results to the institution that gave you access to the primary sources as well as the equipment providers, enabling them to enrich their original resources?

If so, could you please provide us with contact information to the corresponding person?

- **Heritage Institution**

Under what conditions would you allow researchers to access the collection in question?

Please specify any potential legal and/or technical issues involved in allowing access to your resources.

How would you like to benefit from the research results obtained using your sources, including potential corrections or enrichments regarding your initial data?

Are there any specific enrichments you are especially interested in?

Are there any mechanisms at your disposal to facilitate these exchanges (mail, online form, an open API etc.)?

- **Research Facility / Equipment (if applicable)**

Would you like to be kept informed about the research results? In which form (specific message, publication, etc.)?

Are there any specific enrichments you are especially interested in?

Are there any mechanisms at your disposal to facilitate these exchanges (mail, online form, etc.)?

## Interoperability

*« Cultural Heritage Data will be made accessible in a form that facilitates reuse of the data for research. Formats should work and be interoperable for both scholars and CHIs. »*

- Researcher

Are you planning to share your data and/or metadata in standardized formats, processes and protocols? If so, in which ones?

- Heritage Institution

Are you planning to share your data and/or metadata in standardized formats, processes and protocols? If so, in which ones?

- Research Facility

Do the data you provided rely on a standardized (and open if possible) and/or documented format? If not, is there a reliable documentation on the format(s) being used?

## Citability

*« Cultural Heritage data and any resulting research need to be fully citable to increase their visibility and impact. Relevant data citation standards should be applied »*

- Researcher

Do you have a recommended citation model for your research data? Are you willing to cite the institutions / equipment providers that gave you access to the data / services / tools needed for your work?

- Heritage Institution

Do you have a recommended citation model for the experimental material / sources that you curate? Are you willing to cite the other stakeholders (researchers and equipment providers) who conducted research using your sources?

- Research Facility

Do you have a recommended citation model for the experimental material / sources that you possess? Are you willing to cite the other stakeholders (researchers and equipment providers) that conducted research using your sources?

## Openness

*« Cultural Heritage data should be shared under an open license whenever possible, taking into account the existing copyright and any restrictions due to national legislation and privacy issues »*

- Researcher

In which ways are you planning to share your results, data visualizations and any derived data?

Please specify the license (and the licensing framework) under which you are planning to share your research results. Are you planning to use an open license (such as Creative Commons under CC-BY)?

- Heritage Institution

Which licenses do you use to disseminate your data? Are there any particular constraints regarding the dissemination of the research data obtained using your material (sensitive data, intellectual property, etc.)? Would you recommend using specific licenses to disseminate these data?

- Research Facility

What is your policy regarding data obtained using your equipment?  
Would Do you recommend using specific licenses?

## Stewardship

*« Long-time preservation, persistence, accessibility and legibility of cultural heritage data should be a priority »*

- Researcher

Where are you planning to store your data to ensure their long-term preservation? If so, what infrastructure will you use? (This might be a repository or other appropriate solution.)

- Heritage Institution

Do you provide a data storage service for the research results or are you associated with any reliable and public host?

- [Research Facility](#)

Do you provide a data storage service for the research results or are you associated with any reliable and public host?

## Trustworthiness

*« The provenance of Cultural Heritage data and any consequent research should be clear, up to date, openly available and therefore trustworthy*

**Commitment:** *Cultural Heritage Institutions, Researchers and Research infrastructures will put best effort into documenting the provenance of the data that they share, including records of any relevant materials, equipment, techniques, procedures and protocols used. They will attribute any resources used and flag any issues of authenticity or missing data and will also alert to any defamatory use of Cultural Heritage data.»*

- [Researcher](#)

Are you planning to document all those who participated in the data production process in order to ensure its traceability?

- [Heritage Institution](#)

Will you consent to be all your materials used for a research project listed and cited under its provenance metadata?

- [Research Facility](#)

Will you consent to be all your materials used for a research project listed and cited under its provenance metadata?