

Conservation Data Policy Primer and Template



THE INSTITUTE OF CONSERVATION



INTERNATIONAL INSTITUTE
FOR CONSERVATION OF
HISTORIC AND ARTISTIC WORKS



american
institute for
conservation

**Preserving Cultural
Heritage**

Contributions by: Sara Crofts, Anni Mantyniemi, Sarah Stannage, Kristen St.John, Athanasios Velios, Eryl Wentworth and others.

Data policy primer

Introduction

The use of data which is suitable for querying and analysis benefits many domains of practice and research. In conservation such data is not widely available. This policy primer and policy template assist with planning for organisations to make such data widely available. This policy primer and template is not intended to be restrictive or to set out a single set of required actions; instead it should be used to agree on an aspirational path for the future which aligns with the strategy and priorities of the organisation.

Linked Conservation Data

Linked Conservation Data (LCD) is a network of partners exploring best practices and tools for making conservation documentation records accessible. This includes the use of technologies collectively referred to as Linked Data. Linked Data allows the integration of records across different organisations. This integration allows answering questions which are currently impossible to answer by querying individual datasets. They range from current trends on treatments, to decision-making for environmental control, to long-term departmental planning. This document has been compiled as part of the work within the Linked Conservation Data consortium. The use of this document does not require contribution to the consortium. The work of the consortium has been funded by the Arts and Humanities Research Council.

What is a Data Policy for conservation?

It is a document outlining objectives for handling data produced during conservation work. The spirit of the document is to explain how an organisation undertaking conservation work can distribute the resulting conservation records for reuse by relevant audiences. The data policy explains the principles, organisational steps and roles to achieve this.

What is this Data Policy template?

It is a document template which assists policy makers in producing a Data Policy document for conservation in their organisation. The policy template outlines areas for consideration when producing the policy. It also recommends the adoption of best practices as discussed in the LCD consortium. The use of the template for the production of a policy does not require specific tools or subscription to specific services. Adopters of the policy are free to customise it as they see suitable.

The role of professional bodies

The International Institute for Conservation of Historic and Artistic Works (IIC), the Institute of Conservation (Icon) in the UK and the American Institute of Conservation (US) have contributed to the development of this document and support its aim for guiding organisations towards open and good quality documentation records.

Challenges and Opportunities

Best practices

With the increasing variety of the quality of data which is shared online, it is appropriate that discussions around best practices take place within organisations and the profession.

Open data

Progression of practice and thinking in the domain of conservation is augmented by sharing conservation data openly. This means sharing without licensing restrictions for use and reuse of data. This is not widely observed in conservation and this document champions sharing data openly while respecting confidentiality and copyright. This policy template allows excluding datasets when sensitive material is documented or where permission for sharing has not been granted.

Benefits from sharing documentation

Sharing good quality conservation data requires for data to be produced during conservation work. The additional effort of producing rich data is balanced by the benefit of consuming data produced by others and by informing decisions for our own conservation work based on previous observations.

Technical tools

Sharing conservation data requires technical infrastructure. Organisations may be in different stages of establishing such infrastructure. Good practices for documentation are essential **prior** to that infrastructure being in place. This will enable its gradual delivery and ensure that it is fit for purpose. Organisations which lack the resources for implementing such infrastructure could seek partnerships with larger organisations for hosting their datasets thus forming a network of support for sharing conservation data. Such collaborations may require legal agreements between organisations. Sharing data with open licenses may help reduce the effort needed for such agreements.

How to use the template?

The template should be customised by removing and adding points in the various sections. The “Principles” section is critical in describing the nature of the shared data and the idea of openness should be preserved where possible. The template can also be used “as is” when replacing the token words between “<” and “>” and considering the optional sentences between “[“ and “]”.

Questions and Answers

What is data?

Section 2 of the template (Definitions) includes a short description of what *data* can be considered. Anything that provides an insight or documents activity in conservation could be considered as data. For the purposes of this policy, common types of conservation data include treatment reports, assessments, and surveys; scientific data, and environmental or other overall data on collections.

In practice data needs to be encoded in a format that can be transferred and transformed using software. Examples of data include:

- 10 °C (measurement of a dimension such as temperature)
- paper (object description, such as its material)
- 10 January 2005 (important dates such as the date that an object entered the conservation studio)
- condition map / drawing (topography and type of damage of an object)

What is metadata?

Metadata (A) is data which is often used to describe other data (B). In conservation it may not be possible to publish B, but it may be possible to publish A so that interested parties can locate B. For example, the individual measurements comprising an XRF spectrum are data and they may not be possible to be published as individual measurements. The type of analysis and the sampled area are metadata which is enough to discover the fact that this analysis has been done. Section 4 of the template (Principles) includes a list of types of data covered by the policy. Metadata can be specified there.

Why share?

Conservation documentation records represent the observations and descriptions made by conservators during conservation work. As such they represent knowledge. By sharing conservation data it is possible to provide access to such knowledge to colleagues and also have access to knowledge provided by them. This allows more informed decision making for future conservation work and enriches our understanding of collections.

What technical tools or resources are needed for this?

Specifying the tools for sharing records using Linked Data principles should be done in collaboration with the technical team of the organisation. In many cases it is likely that there will be no capacity to support new tools for sharing conservation records, but existing tools with some modifications may be adequate.

In general, sharing records using Linked Data principles requires:

- A system to issue and maintain identifiers, i.e. a method for being able to point to individual pieces of data shared. This will allow anyone to link to your specific database entries once they are shared. For more information see: <https://www.ligatus.org.uk/lcd/faqs>
- Structured records, i.e. records held in a database as separate fields or entities and not as free text. This will allow easier matching of corresponding data in fields across databases. Free text can accompany structured records.
- Vocabulary platform, i.e. an application which allows tracking of the vocabulary terms used in the documentation records. Such a system can help with matching terms from different databases which refer to the same concept or characteristic. For more information on vocabularies see: <https://www.ligatus.org.uk/lcd/output/193>
- Publication platform / repository, i.e. an application accessible on the Web where your records can be published. See next question.

Where do I share or publish this data to?

Section 4 of the template indicates where data can be shared. Sharing data depends on the existing or planned infrastructure of your organisation. If your organisation has a managed space or digital repository which is accessible online then it makes sense to use that for publishing conservation records. Records published as Linked Data can be shared as downloadable files. They can also be shared through purposely built software called Linked

Data Platforms which allow sophisticated ways of querying data. The Linked Conservation Data consortium has worked on a pilot (<https://lcd.researchspace.org>) which demonstrates the use of a Linked Data Platform for sharing conservation records.

I do not have permission to share data

Section 4 of the template (Principles) can be used to specify the types of data to be shared. Specify the types of data for which you have permission to share. This may be metadata of a record, parts of a record, or a subset of records from a collection, etc.

If you do not currently have permission to share any data, then you can specify which future datasets will be possible to share. You can also consider specifying the intended audience. For example, if you cannot share with people outside your organisation, the audience may be the employees of the organisation. You can specify the audience in section 1 (Introduction).

If you cannot share with the recommended licenses (<CC0 or CC-BY>), you can replace those with stricter licenses in section 4 (Principles).

I want to share data but there is no technical support in my organisation

Section 3 of the template (Context) can be used to specify partnerships with other organisations which have the technical capacity to share conservation data. It is envisaged that large organisations which have such infrastructure can support smaller organisations as documentation hubs. If such partnerships are not possible then you can engage with the LCD consortium for technical support and adopt external systems for sharing data such as data repositories which come with minimal or no cost.

I want to share data and have the technical support but there is no time to do it

Section 4 of the template (Principles) can be used to specify how much of the data can be shared. There is no expectation for datasets to be complete, but sharing some data is better than not sharing at all. When specifying the data to be shared, you can include data from specific projects for which there are resources. You can also express that data sharing will be an integral part of future projects and try costing it as part of the overall budget.

Am I liable for the quality of data?

Section 9 of the template (Disclaimer) includes a statement of no liability for users of the data, clarifying that it is offered “as is”. The quality of data will vary depending on circumstances. The spirit of the document is to encourage sharing of data rather than imposing penalties for “low” quality data. It is widely understood that data quality can always improve and that no conservation organisations hold only “good quality” data.

Data policy template

Policy number	
Summary	
Who this policy applies to	
Author/policy contact	
Approved by/approval level	
Date effective	
Last reviewed or updated	

1. Introduction

This data policy specifies how conservation records should be produced and/or published by the <organisation>.

The core objective of this policy is to set a time-frame for producing and/or publishing conservation data and inform planning.

The policy applies to the <conservation department> of the <organisation>. The staff of the <conservation department> are responsible for implementing the policy.

The <conservation department> will plan the publication of conservation data [in collaboration with <other departments>] [of <other organisations>] and create new or adapt current workflows to produce and consume conservation data.

The audiences that this policy refers to are <members of the public or organisation employees>.

2. Definitions

Audiences: groups of people who benefit from the activities of the <organisation> and include: [students, conservators, general public, ...].

Data: records relevant to conservation work including observations, descriptions of conservation activity, plans for future conservation work, analysis results, environment measurements and others. These records should be in a suitable format, such as a database, which enables automatic transformation using software tools for the intended result.

Linked Data: Sharing data online following recommended guidelines by the World Wide Web Consortium for the Semantic Web including the use of the Resource Description Framework.

Open data: Data that can be used without licensing restrictions.

External authorities: controlled vocabularies and lists of records which are broadly accepted as accurate and can be used as reference points (e.g. the Getty Arts & Architecture Thesaurus)

Two ISO standards are particularly relevant to this policy:

- ISO 25964 on the interoperability of thesauri
(<https://www.niso.org/schemas/iso25964>)
- ISO 21127 on ontological modelling of cultural heritage data
(<https://www.iso.org/standard/57832.html>) which is also known as the CIDOC-CRM.

The technologies described as Linked Data (<https://www.w3.org/standards/semanticweb/data>) and the relevant initiatives and recommendations of the World Wide Web Consortium are core to the implementation of this policy.

Further definitions and introductory material can be found as Frequently Asked Questions on the Linked Conservation Data project website: <https://www.ligatus.org.uk/lcd/faqs> .

3. Context

The <organisation> is committed to offering high quality data to its audiences through the <organisation's> website.

This policy is needed to:

- a) maximise the value of conservation data for the <organisation> and its audiences, i.e. ensure that small changes in practice have large impact in data quality
- b) articulate the kind of access to conservation data that the <organisation's> audiences should expect to have
- c) set a timeline for offering such access
- d) plan infrastructure and training to enable such access

This policy depends on:

- a) Technical tools: The production, management and publication of conservation records depends on the availability of appropriate technical tools. The <conservation department> will work closely with the <technical department> to provision for such tools.
- b) Culture and training: The production of good quality conservation records depends on good practice on behalf of the <conservation department> staff members. The <conservation department> will provide training courses on the value of Open Data and the use of best practices in producing data.

Following an internal review of requirements and an assessment of the resources needed, we concluded that [the <organisation> has internal capacity to implement tools for sharing conservation data within the next decade.] [the <organisation> does not have internal capacity to implement tools for sharing conservation data. Therefore we are partnering with <other organisation> which can act as a host of our conservation datasets.]

The <organisation> is committed to offering training to conservation staff on best practices for sharing conservation data and disseminating [through <the organisation's social media accounts>] the availability of conservation data as public and open datasets.

4. Principles






This policy should be implemented by <2030> for newly produced records [and legacy records] so that time conservation documentation in the <organisation> will be produced and published according to the following principles:

- Conservation data should be published using the <CC0 or CC-BY> license. This means that there are <no restrictions> to the use of this data.
- Conservation data will be assessed taking into account legal, ethical and security considerations prior to publication and sensitive data should not be included. The <organisation> cannot publish data for which it does not hold copyright or license or

permission to publish.  This may include reasons of:

ownership, when conservation work was undertaken by external parties to the <organisation> or when all data produced by conservation work is restricted by agreements with external parties

cultural ownership, when permission for sharing material is required from groups such as communities, nations and users

- Conservation data should be published at the latest <5 years>  after its production, thus allowing the producer of the data to publish work based on this data first.
- Conservation data should include attribution to the agent (staff member or team) who produced it and the date of publication. This means that the provenance of the data is published alongside it.
- Conservation data should be published using open formats on the <organization's or partner organization or consortium> website and/or institutional repository. When this is not practically feasible (e.g. when there are no open formats available or an institution is committed to other formats), then metadata should be published using an open format[<and, or> data in a proprietary format should be made available]. 
- Where possible conservation data should point to external authority files to disambiguate records. This should include:
 - the publication of local vocabularies used in <the organisation> alongside their matching synonyms from external authority files where available, 
 - the mapping of  local database schemas to the CIDOC-CRM ontology and its extensions. 

Datasets to be published and revised on a regular basis include: 

- [data about the technology and condition of collection items]
- [data about past treatments alongside assessments of success]
- [data about proposals for new treatments]
- [data about environmental monitoring of relevant spaces]
- [data about scientific analysis or other technical work]
- [metadata that indicates treatment and/or analysis has occurred in lieu of full data from the treatment]
- [etc.]

[Success for the implementation of this policy will be assessed based on the following metrics:

- [number of projects sharing datasets,]
- [scale of legacy conservation records being shared,]
- [percentage of staff members able to share data,]

- etc.]

5. Organisational approach

By adopting this policy the <organisation> will:


- contribute to the ongoing research and practice in the discipline of conservation and inform conservation work in the <organisation>
- champion transparency in safe-keeping and conservation work in <the organisation's> collections
- engage with audiences by delivering rich conservation data

This policy forms part of and aligns with the <organisation>'s wider strategy for:

- pioneering in the digital realm by sharing collection information online
- enabling engagement of new audiences with rich content
- support the development of collections through expert staff knowledge
- promote the history and culture of the <region>

6. Related policies, procedures, processes, forms, guidelines and other resources

This policy forms part of a broader collection of policies by the <organisation> which support our mission for greater use of our collections and broader engagement with our audiences, which include:

- <related policies of the organisation> 

In implementing this policy we are consulting with a number of external partners and resources, including:

- [the Linked Conservation Data Network with its a) terminology guidelines, b) data modelling recommendations and c) training material,]
- [the SPECTRUM standard,]
- [the ResearchSpace and Sinopia platforms for Linked Data,]
- [the Linked Art Network with its software tools,]
- [the CIDOC-CRM special interest group,]
- [the Open Data Institute guidelines on Open Data policies,]
- etc.

7. Roles and responsibilities

This policy specifies that members of the <conservation department> should be producing

and publishing data according to the principles outlined in section 4. They will be led by the <conservation documentation lead>. Members of the <digital team> should be responsible for maintaining the infrastructure to allow production and publication of data.

The <head of the conservation department> and the <head of digital> should be responsible for setting plans to implement the policy. In particular the <head of the conservation department> will be responsible for staff training while the <head of digital> will be responsible for establishing the required infrastructure.

The <conservation documentation lead> will be responsible for supporting the published datasets and answering queries by audiences about the datasets. The <conservation documentation lead> will also be responsible for liaising with the relevant communities to maintain best practices and inform this policy revisions.

The policy is formally approved and ratified by the <board of trustees> of the <organisation>.

8. Managing implementation and review

Implementation of this policy is the responsibility of the <head of the conservation department> who may assign this task to other members of staff. Implementation decisions rely on regular review of similar work in other organisations and are based on the experience gained by projects such as the Linked Conservation Data Pilot (<https://www.ligatus.org.uk/lcd>).

This policy is subject to an <annual> review, assessment and revision. The <head of the conservation department> is responsible for undertaking this task in collaboration with the <head of digital> and in consultation with staff members of the <conservation department> and primarily the <conservation documentation lead>. The <head of the conservation department> finalises the revised policy document before ratification by the <board of trustees>.

9. Disclaimer

The data are provided “as is,” without warranty of any kind. The <organisation> is not liable for any costs incurred because of the use of the data.

10. Feedback

This policy has been prepared by <name(s) of policy author(s)> based on a template provided by the Linked Conservation Data Network. This policy has been ratified by the <body which ratified the policy>. The <organisation> welcomes feedback on its policies.