

Implementation and dissemination plan for data-envelopes

Authors: Mari Wigham, Maria Eskevich

Please refer to the [Glossary](#) for explanation of the terms used in this text

Dataset documentation such as data-envelopes[1] and Datasheets for DCH[2] make it easier for dataset users to find, assess and reuse datasets in a responsible manner. In discussions with both data providers and dataset users, all expressed enthusiasm for this kind of rich, interoperable dataset documentation [3].



Fig 1: Without context, it is hard to find the right datasets

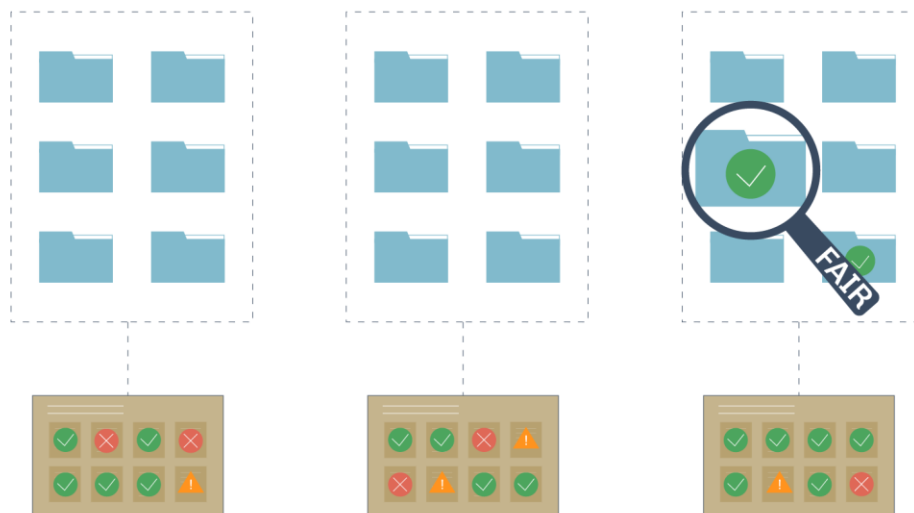


Fig 2: With context, it is easier to find, assess and use appropriate data

In this document we set out an implementation plan for dataset documentation in the humanities and cultural heritage communities, during the next five years. This plan is a result of the [‘Accessing Context’](#) project, a small-scale project of one year that focused on the practices of two large archives and a national research institute. Within this project, during work with the partners and during workshops with a wider range of organisations (universities, research institutes, large archives, libraries, and museums offering datasets for research), it became clear that terminology, practices and motivations related to dataset documentation differ significantly between types of organisation. We expect that there will be further variation among the organisation types who fell outside the reach of the project. This variation is discussed further in the section ‘Types of organisation’.

Due to this variation, creation of an implementation plan that covers all types of organisation in both the heritage and humanities communities at once appears not feasible and potentially very confusing. However, we can make important progress focusing on selected types of organisation and thus gain experience that will help other types of organisation later. This implementation plan is based on the experiences of the project and lays out the important next steps for future projects, which will have to be conducted in collaboration with large archives, research institutes and data hubs. The plan also indicates where there is room for activities that would stimulate potential extension to more organisation types.

This plan is intended to encourage discussion and guide the creation of (pilot) projects and collaborations. It is not a blueprint for how to document the context of datasets.

As 'Accessing Context' used a template of dataset contextualisation called a 'data-envelope' this document is written using this term. However, considering the high potential for interoperability between the different templates, it is equally valid for other types of rich dataset documentation, such as datasheets for digital cultural heritage.

We first introduce some types of 'data providers' organisations. Next, we list the steps to be taken during the coming five years to bring dataset contextualisation into wider use. Finally, we discuss how we will raise awareness and spread knowledge about dataset contextualisation in our dissemination plan.

See the document 'Towards a vision for implementation of dataset contextualisation' for details of the roles, policies and workflows we envisage for the implementation of dataset contextualisation. This implementation plan works towards the realisation of these roles, policies and workflows.

Types of organisation

In the (cultural) [heritage and humanities communities](#), different organisations act as data providers. Here, we discuss these types and how the variations between them (as we observed during the project or as we surmise based on the information we have) affect their use of dataset documentation. This variation helps to explain why it is not possible to have a 'one-size-fits-all' implementation plan for dataset documentation. Neither the list of types nor the variations are exhaustive; part of the implementation plan is to investigate this further. Note that there is large variation within each type, for example some research organisations actively share their datasets while others do not regard the output of their research as a 'dataset'.

Research organisations that share data

A research organisation exists primarily to conduct research. Increasingly, research organisations are stimulated to share their data and thus support further research, for example as a requirement by funders. Research organisations have, in general, an excellent awareness of good scientific practice for producing their datasets, which is usually documented in published papers, and of the needs of scientific researchers. There is often less awareness of the needs of dataset users and how to support (re)use. As dataset publication is not their primary goal, it can be hard to allocate resources to dataset documentation and motivate researchers to create documentation. Research organisations publish their datasets in a wide range of ways (e.g. Zenodo, DANS datastition), each of which has its own requirements regarding dataset description. Rich metadata such as that specified in datasheets and data-envelopes can be deposited now as part of a dataset, but is not integrated as data that could be used as an additional search facet of those platforms.

Contact points: Huygens Institute, one of the project partners, is a research organisation, and other research organisations such as universities and national research institutes were consulted during the project workshops.

Heritage organisations that support research

Heritage organisations exist primarily to preserve heritage, but as part of making this heritage available, many larger heritage organisations specifically offer their data to researchers. This includes support for research, such as creating custom platforms, making data available via APIs, or collaborating in research projects. Heritage organisations typically possess a wealth of information about their datasets and how they were produced but, unlike research organisations, there is not a traditional avenue for publishing this in detail. Collection descriptions tend to focus on what is in the dataset and how users can work with it. Heritage organisations are often not so familiar with research practices and the needs of researchers, in particular dataset users. As heritage collections are the focus of heritage organisations, we have noticed more willingness to create dataset documentation for this specific target group.

Contact points: Sound & Vision and the City Archives of Amsterdam, both project partners, are heritage organisations that support research. Other heritage organisations supporting research were consulted during the project workshops.

General heritage organisations

General heritage organisations that do not offer research services exist primarily to make their preserved collections available to the general public, for example museums. They have little to no awareness of the needs of researchers or other dataset users, while their collections may be of great value for research. They generally focus more on enabling their collections to be searched and browsed, than to be used as datasets. Particularly for small organisations, there may be few or no resources for dataset documentation, and little scope for training people to create it.

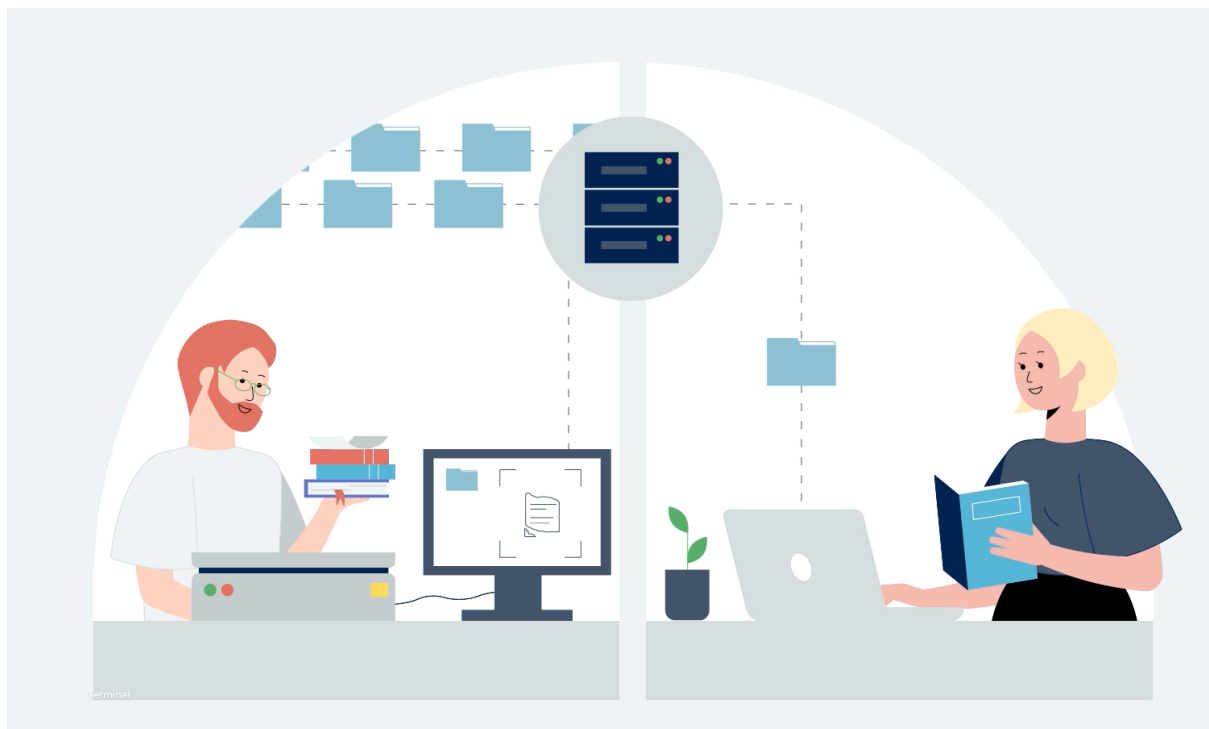
In the Netherlands, heritage organisations have access to the knowledge, tools and services of the [NDE](#) (Netwerk Digitaal Erfgoed – Dutch Digital Heritage Network). The NDE promotes making digital access to heritage and history available to everyone. They assist organisations in taking steps towards publishing heritage datasets and making them FAIR (findable, accessible, interoperable and reusable). The dataset description required by the NDE to publish datasets in their register forms the essential core of dataset documentation. Rich metadata such as that specified in datasheets and data-envelopes is not yet supported, but the NDE encourages the delivery of up-to-date and rich dataset descriptions [4].

Contact points: This type of heritage organisation was not consulted during the project. The NDE has much experience with these organisations and so is a vital partner both for understanding these organisations and for assisting in contacting them.

Other organisations providing data

Other organisations can provide data that is useful to the heritage and humanities communities. Think for example of census data, geographical coordinates, or vocabularies. Their needs and available resources are yet to be defined.

Implementation plan



In this implementation plan we describe the steps to be taken towards the realisation of data-envelopes in practice within the cultural heritage field and humanities. Financing and partners will be sought to realise this implementation plan. (To find partners for implementation, outreach and dissemination are essential. These are described in the Dissemination section).

The implementation plan covers the creation and exchange of data-envelopes, the formulation of best practices, the development of templates and tools, and the extension to other types of organisation.

Data-envelope creation

The practice of data-envelopes creation is already occurring at a small scale within the Huygens Institute and serves as a testbed for making rich dataset documentation. The

next step is to make creation a structural part of data management practices within an organisation. This includes creating policies, fine-tuning the templates and tools used and training/supporting researchers and data managers to create data-envelopes.

At the same time, introduction of rich documentation into other organisations will ensure interoperability and larger impact. The experiences of the 'Accessing Context' project suggest that this can best start with an initial pilot. The organisation should select a small list of representative datasets, choose an initial template and editor, and allocate a small team to create data-envelopes for these datasets. Their experiences can inform the choices for policies, templates and tools. Partners of the 'Accessing Context' project are open to answering practical questions about setting up such a pilot while supporting one at the Huygens sister institute, the Meertens Institute in the upcoming years.

To optimise data-envelope exchange (next section) it is essential to already consider expected requirements at the stage of creating data-envelopes. For this reason, it would be helpful to include hubs as partners in these pilots, or to maintain close contacts with exchange work.

Data-envelope exchange

Data hubs for cultural heritage and humanities datasets exist within the Netherlands, for example the NDE Dataset Register, and internationally, for example Europeana. Some third parties that use the data from data hubs already recognise the need for richer dataset descriptions to handle aspects such as data ethics, see for example [5]. In Accessing Context, steps have been taken towards data-envelope exchange by mapping the data-envelope format to DCAT-3. Further work will be undertaken, together with partners, to develop a common central model for dataset documentation and means of mapping between formats to facilitate documentation exchange. To bring this work into practice, a pilot is needed that will implement this exchange for at least one data provider and one hub. This should include the following activities:

- Set up partnerships with hubs, e.g.
 - NDE
 - CLARIAH-NL (SSHOC_NL project)
 - DANS
 - E-RIHS
 - Europeana
- Investigate the possibility of partnership in and alignment with current initiatives, such as the planned pilot for using datasheets in Europeana Dataspace

- Set up a project with at least one data provider and one hub to pilot a data exchange between them, focusing on:
 - Conversion between formats
 - Exchange mechanisms
 - Integration of data-envelope information in hub
 - Tests with data users

Templates and tools

A template and editor exist for data-envelopes and are already used in practice. However, there is scope for improvement. For the template it would be beneficial to produce a standardised template that can be flexibly modified to suit the needs of different domains and organisations. This could potentially draw inspiration from the NDE approach to publishing datasets as linked data [6], whereby a generic model is combined with domain-specific models, to enable both integration and more in-depth exploration. In a similar way, the dataset description could combine a standardised base template with domain-specific profiles. For the editor, improvements are needed in user-friendliness and flexibility. A first version of a flexible, user-friendly editor for both datasheets and data-envelopes is in development within the SSHOC-NL project, and further work will be needed to improve this first version, for example by adding vocabulary support. Additional tools can also be helpful, for example to present data-envelopes in a flexible, user-friendly way, or to extract information required for data-envelopes from the dataset and existing documentation. Data-envelopes can be integrated into existing tools such as data provider metadata systems, making it easier for users. Finally, user feedback is essential to guiding development.

This is a summary of the required lines of actions:

- Templates
 - Flexible yet standardised template
 - Division between publicly available documentation and restricted internal documentation
 - Help/Manual and examples
- Editor tools
 - Flexibility, user-friendliness of the editor
 - Support of multilinguality when filling in the information
 - Selection and integration of vocabularies, as well as possibility to use multiple vocabularies
- Broader integration and uptake
 - Extraction of the information from available unstructured dataset descriptions
 - Integration in data provider systems
 - Flexible, user-friendly presentation

- Collecting user feedback

Best practices

Data providers need to implement data-envelope creation in the way that best suits their organisation. However, it can greatly help if experience in best practices and examples can be shared, to prevent data providers having to reinvent the wheel and to encourage harmonisation. Best practices can include:

- Workflows
 - Selection
 - Creation
 - Review (Quality control)
 - Publication
- People
 - Incentives
 - Roles
 - Training
 - Professional development
- Policies
 - Licensing
 - Citation
 - Data Ethics
 - Etc.

Best practices should be developed in projects and organisations using data-envelopes and shared on the [data-envelopes information site](#).

In addition to best practices, it may be useful to lay out a growth model for dataset documentation, with a series of steps that providers and hubs can take to gradually move towards implementing dataset documentation (similar to the star system for Linked Data).

Extension to other organisation types

As stated in the introduction, this implementation plan is based on experiences with two large archives and a national research institute. The section ‘Types of organisation’ introduced some organisational types and their relevant variations. To enable a broad range of types of organisation to successfully document their datasets, it is vital to extend the templates, tools and best practices to cover their needs. Note, however, that dataset documentation may not be feasible or desirable for some types of organisation, so it is important to maintain a balance of costs and benefits.

This work will include the following steps:

- Further identify (and prioritise) types of organisation
- Contact organisations of these types, directly or via networks such as NDE
- Assess the value of dataset documentation to the organisation type
- Set up collaborations
- Gather requirements
- Specify modifications needed to templates, tools and/or best practices
- Work to implement these modifications (where possible)
- Ensure that a clear introductory guide is created for each organisation type to help them on their way towards dataset documentation

Dissemination plan



Outreach and dissemination are essential to find partners for the implementation plan and to encourage awareness and uptake of data-envelopes. We aim to reach the following target groups:

- Data providers
- Data hubs
- Researchers
- Data managers
- Potential partners/funders

In addition to the dissemination activities already performed in the Accessing Context project, the following activities will be carried out before the end of the project:

- Publish the explainer video, on the [data-envelopes information site](#)

- Publish an overview of created data-envelopes on the information site
- Add guidelines and examples to the data-envelopes information site
- Publish this implementation plan on the data-envelopes information site
- Promote the project results and share the explainer video on social media such as LinkedIn
- Link to the results from the partners' websites

The following dissemination activities will occur after the end of the Accessing Context project:

- Publish a set of data-envelopes (on the data-envelopes site and additionally on a data repository such as Zenodo)
- Present results (and explainer video) via forums and community events such as the E-RIHS event
- Publish blogs on partners' websites
- Encourage the discussion of this work in academic conferences and journals (panel proposal at DHBenelux and paper at DHBenelux journal already submitted)
- After the release of the SSHOC-NL datasheets/data-envelopes editor, publish a blog together with CLARIAH-NL
- Further conversations with network

References

[1] - Mrinalini Luthra and Maria Eskevich. 2024. Data-envelopes for cultural heritage: Going beyond datasheets. In Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024, pages 52–65, Torino, Italia. ELRA and ICCL.

[2] - Alkemade, H., Candela, G., Claeysens, S., Colavizza, G., Eren, S., Freire, N., Irollo, A., Isaac, A., Lehmann, J., Neudecker, C., Osti, G., van Strien, D., & Wevers, M. (2025). Datasheets for Digital Cultural Heritage Datasets (Version 2). Zenodo.
<https://doi.org/10.5281/zenodo.15828222>

[3] – Summary of Data Contextualisation Workshops, 2025, <https://github.com/knaw-huc/data-envelopes/blob/main/Summary%20of%20Dataset%20Contextualisation%20workshops.pdf>

[4] - Netwerk Digitaal Erfgoed,. Van Data naar dienst – Visie op de ontwikkeling van verbonden erfgoeddiensten, November 2025, [10.5281/zenodo.17541399](https://doi.org/10.5281/zenodo.17541399)

[5] - Dienstplatform Datahub Koloniale Collecties deel 3: van NDE-infrastructuur naar dataspace, Gertjan Filarski, 2025, <https://www.p-322.com/nl/posts/20251127-dienstplatform-datahub-koloniale-collecties-deel-3-naar-dataspace/>

[6] - Netwerk Digitaal Erfgoed, Schema.org Application Profile for NDE (SCHEMA-AP-NDE), 14th January 2026, <https://docs.nde.nl/schema-profile/#generic-and-domain-data-models>