

# 1. Data collection and documentation

## 1.1 What data will you collect, observe, generate or re-use?

During the lifetime of the project images of depictions of animals from different cultures (from Ancient Egypt, Classical Greece, Medieval Europe, Contemporary Asia) and times will be collected, described and analysed. The art works are part of museums and collections from all over the world. The data will be collected within the virtual research environment XY according to a pre-defined data model and the data can be easily exported in various file formats which are suitable for long-term preservation (tiff, jpg2000, mp4, xml, csv) at the end of the project. A rough estimation of the size of this dataset is 250 GB. The dataset will be enriched by links to authority records (such as pleiades, geonames, and iDAI.gazetteer), URIs of existing data compilations, standardised names, time periods. Where possible, links to the artwork in museum or collection databases will be added.

Textual data will be compiled from original sources (Egyptian, Greek, Latin), from other datasets, and from secondary literature.

Furthermore, three already compiled datasets will be reused:

Name of dataset	Type(s)	File format	File size
Dataset 1: Greek depictions	text, image	tiff, jpg, pdf	50 GB
Dataset 2: Ancient Egyptian textual sources	text	pdf, csv, xml	10 GB
Dataset 3: Medieval animal depictions	text, image	csv, jpg	15 GB

Concerning other existing very recent and thus partially unpublished data which we would like to (re)use, we are currently contacting several institutions and the PIs of research projects to clarify the conditions for collaboration and a (re)use of their data. We seek agreements based on the *Reuse agreement template between Cultural Heritage Institutions and researchers* (<https://hal.science/halshs-03367459v2>). Agreements with two institutions have already been achieved:

- Institution 1: All data and images will carry a CC0 or CC BY license and we are allowed to include the data in our archiving dataset.
- Institution 2: We will receive access to the data, images and videos prior to publication, but are not allowed to include the data in our archiving dataset. Articles referencing these data may only be published after the data have been published.

Research results will be published in journal articles and by the end of the project the research data itself will be available in repository XY. Data created by the project will be made public with a CC-BY (<https://creativecommons.org/licenses/by/4.0/deed.en>) license. For the images and videos the licenses will vary, depending on the policy of the respective owning institution. The following procedure is intended:

- If an institution agrees that we include images or videos of one of its objects in our dataset in the repository, the images / videos will be visible. Respective copyright statements and license information will be attached. If the institution provides images or videos already with persistent identifiers on an own IIF compatible server, however, we will show and reference them from there, but not archive them as part of our own data to avoid duplication and unnecessary storage costs. If no persistent identifiers are provided by the institution, the image / video will be part of our archived dataset.
- If an institution does not agree that we include any media files of objects in its collection in our dataset in the repository, we will use the media files only during research, but they will not be part of the archived data. If the institution possesses an object database which can be accessed openly, we will add a link to the object in the database.

We estimate that about 80% of the objects can be either shown directly or be easily accessed at the owning institution.

## 1.2 How will the data be collected, observed or generated

Data will be collected

- from the literature
- from published datasets, museum databases and other discipline-specific databases

Data will be generated

- by defining an appropriate data model
- Data generation itself will happen within the virtual research environment XY according to the defined data model

For the place names links to gazetteers or authority databases such as Pleiades, the Getty Thesaurus of Geographic Names (TGN), iDAI.gazetteer or geonames will be added to cover also alternate names, spellings and transliterations. Similarly, institutions will be linked to authority databases. Data which we will reuse will be harmonised and enriched by additional data - e.g. named entities, linked open data, controlled vocabularies - according to common standards to ensure interoperability and transparency of data provenance.

*If you use the DaSCH Service Platform already during the project phase you may add:*

“The project will make use of the generic data management portal of DaSCH with basic functionalities. It allows project members to store and manage research data in accordance with the prior-defined, project-specific data model.”

### **1.3 What documentation and metadata will you provide with the data?**

The data collected will be published as Open Research Data (ORD) in compliance with the FAIR and CARE principles. All collected data will be accompanied by at least the following minimal set of metadata: storage institution or storage location, culture, time period, genre. This minimal set is extended by further metadata such as inventory number or author/artist, if known. At the end of the project, the main project data will be archived in *repository XY*. Please see *the webpage of your repository for details concerning metadata and add a few sentences here*.

*In the case that your repository of choice is DaSCH, you may e.g. write:*

“The data collected will be published as Open Research Data (ORD) in compliance with the FAIR and CARE principles. Digital publication formats aim at human and machine oriented re-use. At the end of the project, the main project data will be archived in DaSCH (Swiss National Data and Service Center for the Humanities) where it can be accessed and queried directly by humans as well as by machines via an API. As a FAIR repository, DaSCH provides metadata about the project itself using a standard metadata schema (<https://meta.dasch.swiss/>). DaSCH provides ARK persistent identifiers (PID) for the project, as well as for each single object within the dataset which allows for precise citations. The DaSCH data model will contain mappings to standard ontologies such as CIDOC CRM and its extensions to ensure interoperability for machines. To ensure that the data are as understandable as possible for secondary users and thus to foster reuse of the data, a documentation of the methodology and of the data model will be compiled which will explain each property.”

## 2. Ethics, legal and security issues

### 2.1 How will ethical issues be addressed and handled?

This point does not apply to the data created in connection with this research project, since only depictions of animals in art are examined, and not remains such as Egyptian animal mummies which might raise ethical issues.

### 2.2 How will data access and security be managed?

The project does not generate sensitive data.

Data access and security during the ongoing project will be managed and guaranteed by the IT Services of the University XY. Data access and security after the end of the project will be managed and guaranteed by repository XY. *Please see the webpage of your repository for details concerning data access and security and add a few sentences here.*

*In the case that your repository of choice is DaSCH, you may e.g. write:*

“The DaSCH Service Platform implements an authentication and permission system if necessary. Users can be granted access to data according to their credentials. Versioning and historization ensure both security and traceability of modifications. However, we expect to be able to publish the vast majority of the data as open research data with a [*give intended license here*]. Thus, no login and storage of personal data will be necessary to access the data. Concerning data security standards, DaSCH data is currently hosted by the University of Basel which is in the process of implementing a solution which follows closely the [NIST Cyber Security Framework](#).”

### 2.3 How will you handle copyright and Intellectual Property Rights issues

The owner of the data will be the University XY, copyright and intellectual property right stays with the project members. The data transferred to *repository XY* is released under a Creative Commons license (*give intended license here*). Concerning the third-party data which we intend to (re)use we are currently contacting the responsible institutions to detail the conditions by using the Reuse agreement template between Cultural Heritage Institutions and researchers as a guideline (<https://hal.science/halshs-03367459v2>).

## 3. Data storage and preservation

### 3.1 How will your data be stored and backed-up during the research?

**Case 1. If you use some kind of virtual research environment which is run on a server**

I will use the virtual research environment XY during the research which is run on a virtual machine which is provided by the IT services of the University XY. *Please contact your local data stewards or IT services to learn about management of servers, frequency of backups, protection against data loss.*

*If you use the DaSCH Service Platform as virtual research environment during your research, you may e.g. write:*

“The DaSCH Service Platform is operated in an environment that is strongly protected against data loss. The servers are provided by the University of Basel. The data is backed up every night and kept at different geographical locations and on varying storage technologies (disk and tape).”

**Case 2. If you use storage provided by your institution**

All data is stored on UniXY's servers, which provide large-storage capacity. These servers have backup measures such as daily automatic backups at 30-day rolling intervals with the possibility of data recovery. In addition, I will store my data on a daily basis on the server and on an external hard disk. The server is also accessible when I am on field trips via a VPN-client.

### 3.2 What is your data preservation plan?

For long-term data storage, the project's main research data will be deposited at *repository XY*. *Please see the webpage of your repository for details concerning information about how and how long data will be preserved and add a few sentences here.*

*In the case that your repository of choice is DaSCH, you may e.g. write:*

“For long-term data storage, the project's main research data will be deposited at DaSCH. DaSCH's infrastructure is based on open standards such as the Resource Description Framework (RDF), the Web Ontology Language (OWL) and the query language SPARQL. These are important Semantic Web technologies that allow for linked (open) data. According to its mandate, DaSCH will ensure preservation and long-term availability of the data in human and machine-readable form.”

## 4. Data sharing and reuse

### 4.1 How and where will the data be shared?

- ***In which repository do you plan to share your data?***
- ***How will potential users find out about your data?***

The project's main research data will be available in repository XY. *Please describe how your data can be found and viewed there, see the repository's website for respective information.*

*In the case that your repository of choice is DaSCH, you may e.g. write:*

"The project's main research data will be available at DaSCH. The DaSCH Service Platform provides a graphic user interface to search and view resources of all projects that are in the repository. No login is needed to view and discover the data. In addition, data is searchable based on a RESTful API. The data model is provided in both human and machine-readable forms using Linked Open Data standards (RDF/RDFS) and can be exported in standard JSON-LD. The metadata of the project will be available on the DaSCH metadata server which will be harvested by Connectome (SWITCH) and SWISSUBase."

### 4.2 Are there any necessary limitations to protect sensitive data?

No sensitive data is collected. Data will be shared at the time when the relevant scientific publications are published.

### 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.

Yes (checkbox)

### 4.4 I will choose digital repositories maintained by a non-profit organisation.

Yes (checkbox)