

D1.4 Data Management Plan

Action: AXIOM
Grant agreement no: 6455601
Partner: University of Applied Arts Vienna

Document version: 1
Date (DD.MM.YYYY): 15.09.2015

Outline:

As part of the action on open access to research data in Horizon 2020, the Data Management Plan (DMP) for the project AXIOM describes how research data quality, sharing and security of the research data will be accomplished. This paper targets the main questions as outlined in the *Guidelines on Data Management in Horizon 2020* (Version 16th of December, 2013), but also takes into account the *Data Management Plan: Eine Anleitung zur Erstellung von Data Management Plänen* Projekt e-Infrastructures Austria (Version 2.0, May 2015).

Authors:

Maximilian Gurrech, Sebastian Pichelhofer, Matthias Tarasiewicz

License:

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

DISSEMINATION

Publication of data and outcomes.

PRE-INGEST

Gathered data from various sources.

CLOUD COLLABORATION

Project management and collaboration systems.

INTRANET RAID

Backup on local raid storage.

STORAGE CLUSTER

Central project repository for partners and community.

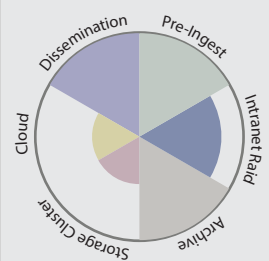
Long-term archival on LTO-6 tapes.

ARCHIVE



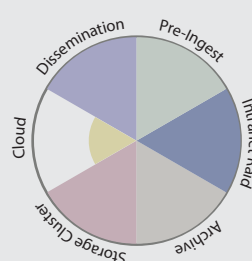
In the following **overview** a few data types are illustrated following the **6-tier Data Management** concept:

Video Footage



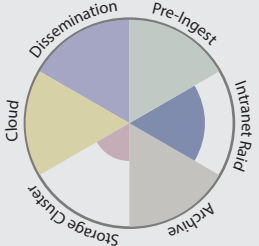
Raw footage is converted to archival format, sorted and named.
 Footage gets ingested to Intranet Raid for further editing.
 An archival-friendly copy gets transferred to the archive for longterm storage.
 Footage links are shared for collaborative editing through cloud.
 After completion results will be published through appropriate channels.

Photos



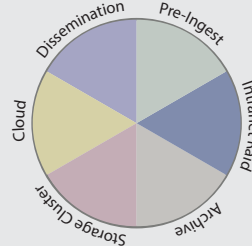
Raw data is titled, tagged and sorted out before ingesting to Intranet Raid for editing.
 The catalogue including adjustments and metadata is archived for longterm storage.
 Lower quality versions are uploaded to the Storage Cluster for universal access.
 Photos are used in publications and shared through appropriate channels.

Software



As developers are already maintaining source code on online repositories (github), a copy is pulled on regular intervals to the Intranet Raid.
 During a regular backup, concerned data is archived for longterm storage.
 Cloud Collaboration system can link tasks, bugs and requests to commits and codelines on github.
 Data will be disseminated and documented on project wiki and binary releases will be published on Storage Cluster.

General Data



During ingest data will be validated and curated for further use.
 Generally it will be copied from several sources to the Intranet Raid for internal use.
 During regular data collection it is archived onto LTO-6 tapes for longterm storage.
 The Storage Cluster includes data necessary for online collaboration and distribution.

a) What types of data will the project generate/collect?

The project generates different forms of data, which can be separated in the following sections:

Generic research data (process description and communication data)

- **Sources and resources:** collection of external references, studies, papers and open-access material from other research projects, individuals, as well as related contexts
- **Articles and commentary** on related topics, research and overview of the larger field of research, such as open source cinema, open hardware

Specific research data (outcomes)

- **Code** (software) in various formats
- **Technical and functional plans, drawings & 3D-Models (CAD)**

Documentation

- **Texts and articles:** documentation of processes and products of the main research
- **Photography** of processes, dissemination events, status
- **Video Communication:** updates for the (scientific) community, other developers as well as the interested public (title: apertus^o Team Talks).

Demonstrations

- Demonstration footage from the camera prototypes

Research Papers and articles

- Academic papers and articles about research outcomes and other aspects (not pure technical benefits) of Open Hardware

Title	Types of data	Dissemination
Generic research data	Various	Blog, Social Media, Wiki, Mailing List
Specific research data	Source code, technical plans and technical drawings	Phabricator, Wiki, Github
Documentation	Text, photo, video	Blog, Wiki, Social Media
Demonstrations	Video footage	Repository
Research papers and articles	Text and pdf	Blog, academic repositories

b) What standards will be used?

The project focusses mainly on open standards since this an integral part of the nature of the project. All outcomes, including documentation and research data produced, will be released under open licenses. Public release of all AXIOM project research data and results are published under a free licence (GNU GPLv3, CERN Open Hardware Licence 1.2, GNU Free Documentation License 1.3).

AXIOM will open everything from the beginning. All of the hardware (including optical and mechanical parts) and software produced in the course of the project, – including all knowledge/know-how generated during our research and development stage – will be made public and available on the Internet. It will be open for anyone to access without registration ('gold' open access).

List of data types with approximate sizes:

Data	Format/ Container	Description	Total size after successful finished EU project (approx.)	Open formats
Photo documentation	CR, ARW, CR2, DNG, NEF	Photographic images and documentation throughout the whole project	400 GB	DNG
Video Footage	MOV, AVI, MXF, DNG	Demo footage from AXIOM camera, documentation, guides, introduction videos, communication, events	24 TB	MOV, AVI, MFX, DNG
Technical Drawings	DWG, IPT, IAM, STP, STL, etc.	Mechanical components and assemblies as 3D CAD models created in various software tools	5 GB	STP
Finished Videos	MP4, MOV	Finished edited documentation and demonstration videos for publication/distribution	5 TB	h.264/ h.265
Illustrations, Graphic Designs	PSD, AI, EPS, PDF,	Drawings, illustrations for website and publications	10 GB	PDF, SVG
Animation Source Files	AEP	Illustration animations and motion graphics	1 TB	None
Texts and Source Code	Various	Documentation, articles, publications, software source code	1 GB	ODF, PDF, ASCII

c) How will the data management be implemented?

The data management in the project AXIOM consists of a 6-Tier system:

#	Tier	Implementation	Application
1	Pre-Ingest	Gathered data from various sources	Data gathering, data validation and selection
2	Intranet Raid	Ingest to backup: metadata and classification of data, immediate archival of high-priority data	Classification and description of data, adding of metadata, duplication on local storage.
3	Archival	Long-term archival on LTO-6 tapes (incremental as well as immediate)	Separation of data according to high-priority: instant archival of highly important data (data which is not redundantly stored yet).
4	Storage Cluster	Offsite server/storage cluster: preparation of data for internal project use	Transfer of relevant data to offsite storage cluster for availability to the whole consortium. Collection of external sources at the storage cluster. Backup of storage cluster through regular scheduled backups.
5	Cloud Collaboration	Project management and collaboration systems	Internal data stored in Google Drive, consortium documents and EU-relevant information in Phabricator, information for general public and larger team in apertus° wiki.
6	Dissemination	Publication of data and outcomes	External repositories (github, video hosting, social media, external photo storage, apertus° blog and wiki). See details in section b.

Used hardware for the implementation of the DMP

- Intranet RAID (local RAID-6 with 24 TB storage)
- Archival computer with LTO-6 tape drive (long-term archival)
- Storage Cluster with fileserver (research data sharing throughout the team, public availability of research data)

Targeting Data Degradation (data decay/data rot)

As “data degradation” is a key topic of the current discourse in long-term archival, we decided to use up-to-date filesystems to secure our data. In the case of file-level-corruption, we decided to work with ZFS and Btrfs file-systems, which implement integrity-checking and self-repair algorithms to prevent data rot.

d) How will this data be exploited and/or shared/made accessible for verification and re-use?

According to the 6-Tier system, data will be made accessible according to the scope of the tier:

1	Pre-Ingest	Availability only to responsible team members
2	Intranet Raid	Local availability at University of Applied Arts Vienna
3	Archival	Local availability at University of Applied Arts Vienna
4	Storage Cluster	Availability to whole consortium through fileserver (ftp/http). Selected parts will be made accessible to the larger audience via a public repository software.
5	Cloud collaboration	Internal data stored in Google Drive, consortium documents and EU-relevant information in Phabricator, information for general public and larger team in apertus° wiki.
6	Dissemination	General availability according to the external context.

Backup strategy for external sources

External source	Max. size	Backup to
Github	30 GB	Incremental pull-backup from archival computer to LTO tape. Additional redundancy through the nature of git-system; Additional daily pull on Storage Cluster.
Developer FTP	10 GB	Incremental pull-backup from archival computer to LTO tape. Mirror on Storage Cluster.
Google Drive	20 GB	Incremental pull-backup from archival computer to LTO tape. Mirror on Storage Cluster.
apertus° Wiki	15 GB	Incremental pull-backup from archival computer to LTO tape. Mirror on Storage Cluster.
apertus° Blog	2 GB	Incremental pull-backup from archival computer to LTO tape. Mirror on Storage Cluster.
Phabricator	2 GB	Incremental pull-backup from archival computer to LTO tape. Mirror on Storage Cluster.

Public Repository for research data

A public and self-hosted repository for all the research data will be made available during the project timeframe and is guaranteed to be hosted after the Horizon2020 project by the Artistic Bokeh Initiative. Details for the software used and implementation specifics will be outlined in Version 2 of the DMP.

e) How will this data be curated and preserved?

The described policy reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results. Data is curated by team members at the University of Applied Arts Vienna and will be preserved using current state-of-technology in data backup and data availability. Mission-critical data is always stored at least with dual-redundancy at all times, together with the long-term archival procedures data is available on tripple-redundancy level.

#	Tier	Curating	Preservation strategy
1	Pre-ingest	Immediate duplication on Intranet Raid. Selection by WP8 responsible team members.	Dual redundancy on Intranet Raid and source media.
2	Intranet Raid	Definition of initial folder structure and selection by WP8 responsible team members. Selection of mission-critical data for immediate archival by WP8 responsible members.	Additional redundancy via direct archival (to LTO) for mission-critical data. Weekly incremental backup to LTO tape.
3	Archival	Archival and long-term storage: LTO-6 Tape drive with redundant storage.	Weekly incremental backups, redundant archival of all research data to LTO-6 (storage of physical tapes on two different sites).
4	Storage Cluster	Selection by WP8 responsible team members	RAID-6 with ZFS file-system (trippel redundancy). Backups to LTO tape drive.
5	Cloud Collaboration	Selection by whole consortium, responsibility at the project lead.	Pull-backups to Storage cluster and to incremental backup to LTO tape drive.
6	Dissemination	External, outsourced storage	Pull-backups to Storage cluster and to incremental backup to LTO tape drive. Storage of dissemination results at Google Drive (Cloud Collaboration) as well as publishing to Phaidra system (University long-term database).

f) Dissemination of data and project outcomes

apertus° Wiki	Internal Hardware/Software documentation	http://wiki.apertus.org
apertus° Website	Primary dissemination outlet via project website	http://apertus.org
External video hosting service	Finished edited documentation and demonstration videos for publication/distribution.	http://youtube.com http://vimeo.com
External photo hosting service	Edited photos hosted for maximized availability	http://flickr.com https://commons.wikimedia.org
Social Media	Generic information and texts for wider audience.	http://facebook.com http://twitter.com http://plus.google.com
Academia.edu	Research papers and articles.	http://academia.edu
Phaidra	Selected documentation, texts and research content for long-term archival in Phaidra (P ermanent H osting, A rchiving and I ndexing of D igital R esources and A ssets)	https://phaidra.bibliothek.uni-ak.ac.at/
Cern Open Hardware Repository	A place on the web for electronics designers at experimental physics facilities to collaborate on open hardware designs, much in the philosophy of the free software movement.	http://www.ohwr.org
Github	Source code, source files for collaborative development	http://github.com

Literature list/sources

- Guidelines on Data Management in Horizon 2020, (Version 1.0, 16th of December, 2013)
- Data Management Plan: Eine Anleitung zur Erstellung von Data Management Plänen Projekt e-Infrastructures Austria (Version 2.0, May 2015)