



## **Implementation of the CoreTrustSeal**

The CoreTrustSeal board hereby confirms that the Trusted Digital repository ARCHE complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.

The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on March 19, 2018.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

## Assessment Information

Guidelines Version:	2017-2019   November 10, 2016
Guidelines Information Booklet:	<a href="#">CTS Requirements 2017-2019</a>
All Guidelines Documentation:	<a href="#">Documentation</a>
Repository:	ARCHE
Seal Acquiry Date:	Mar. 19, 2018
For the latest version of the awarded CoreTrustSeal for this repository:	<a href="https://www.coretrustseal.org/why-certification/certified-repositories/">https://www.coretrustseal.org/why-certification/certified-repositories/</a>
Previously Acquired Seals:	None
This repository is owned by:	<b>ACDH</b>

T +43 1 51581-2200  
E [acdh@oeaw.ac.at](mailto:acdh@oeaw.ac.at)

# Assessment

## 0. Context

### Applicant Entry

*Self-assessment statement:*

#### **Brief Description of the Repository's Designated Community**

[ARCHE](#) [1] is one of the central services of the Austrian Centre for Digital Humanities at the Academy of Sciences in Vienna ([ACDH-OeAW](#) [2]) and offers stable hosting for digital research data for the Austrian humanities community [3]. ACDH-OeAW was founded with the intention to foster the change towards the digital paradigm in the humanities. ARCHE welcomes data from all humanities researchers, be they based at the Academy, or elsewhere at the national and international level. While its predecessor, CLARIN Centre Vienna / Language Resources Portal was dedicated to digital language resources, ARCHE offers services open to a broader range of disciplines.

ARCHE is mainly meant to accommodate resources concerning Austria, that is both those that have been collected or created in Austria, and those pertaining to a geographical area or historical period that is of interest to Austrian scholars. However, we do not categorically exclude resources without direct relation to Austria.

The [collection policy](#) [4] details the types of data the repository is willing to accept and store.

#### **Level of Curation Performed. Select all relevant types from:**

A and B.

At present we do not execute file format conversion or extensive enhancement of documentation, but assist data depositors in doing so. We plan to fully implement Level C curation by the end of 2018. It is planned to offer automated format conversion via dedicated microservices for common formats and data types (e.g. convert JPEGs into TIFF). Enhancement of documentation will be done by a qualified curator in close cooperation with the data provider. ARCHE will not edit the content of deposited data at any time.

**Outsource Partners. If applicable, please list them.**

The services offered by ACDH-OeAW, and especially ARCHE receive extensive and dedicated support from reliable partners within the Academy and beyond. The computing centre of the academy ([ARZ](#) [5]) ensures sufficient hardware capacities for storage, computing power, and networking. Since its [foundation in 1976](#) [6], it is the sole responsibility of the ARZ to provide a number of computing services such as networking and server hosting, but also Email, network shares, or cloud services to all the institutes and departments of the Academy. ACDH-OeAW and ARZ have regular intensive exchange to resolve any issues arising from the cooperation, supported by a professional ticket tracking system. ARZ is currently elaborating a policy document that details the strategy and agenda of ARZ within the Academy for the upcoming years. This document will be approved by the Academy's governing bodies and is planned to be released in 2018.

In the context of the [EUDAT2020](#) [7] project, [CLARIN ERIC](#) [8], in which ACDH-OeAW takes part, established a cooperation with the Max Planck Computing and Data Facility ([MPCDF](#) [9]) in Garching, Germany. This allows for a regular off-site replication of the data managed by ACDH-OeAW, currently via the [B2SAFE](#) [10] service provided by MPCDF for the purpose of disaster protection. A letter of Understanding between MPCDF and the Service Providers is being worked out and will be signed in the first half of 2018.

CLARIN also mediates the provision of Persistent Identifiers (PIDs) from the [ePIC](#) [11] consortium. In addition, our deep involvement in the two pan-European projects of ERIC, [CLARIN](#) [12] and [DARIAH](#) [13], ensures a firm embedment of all the institute's activities and the sharing of state-of-the-art expertise and best practices with the European Digital Humanities community at large.

**Other Relevant Information**

ARCHE will pursue the principles of Open Access and Open Data. It encourages data depositors to use open licences, like CC-BY and CC-BY-SA, adhere to rules for good scientific practice, and apply the [FAIR Data Principles](#) [14].

ACDH-OeAW is a contributor of The European Strategy Forum on Research Infrastructures ([ESFRI](#) [15]) in the domain of linguistics and Digital Humanities. In fact, it is part of two ESFRI initiatives (CLARIN and DARIAH) and acts as their national coordinator while actively collaborating with other partners within several working groups.

The metadata of linguistic data is regularly harvested by the Virtual Language Observatory, part of CLARIN. Additionally, ARCHE will provide well documented APIs, which includes an OAI-PMH endpoint for external

service providers, such as Europeana.

- [1] (<https://arche.acdh.oeaw.ac.at>; main page of repository; 13.10.2017)
- [2] (<https://www.oeaw.ac.at/acdh/>; main page of ACDH; 13.10.2017)
- [3] (<https://arche.acdh.oeaw.ac.at/browser/about-service>; About ARCHE; 20.02.2018)
- [4] (<https://arche.acdh.oeaw.ac.at/browser/collection-policy>; collection policy of ARCHE; 13.10.2017)
- [5] (<http://www.oeaw.ac.at/arz>; computing centre of the academy; 13.10.2017)
- [6] (<https://www.oeaw.ac.at/arz/ueber-das-arz/ueber-das-arz/>; About page of ARZ; 20.02.2018)
- [7] (<https://www.eudat.eu/>; page of EUDAT; 20.02.2018)
- [8] (<https://www.clarin.eu/>; project page CLARIN; 13.10.2017)
- [9] (<http://www.mpcdf.mpg.de/>; page of MPCDF; 13.10.2017)
- [10] (<https://www.eudat.eu/catalogue/B2SAFE>; about B2SAFE; 20.02.2018)
- [11] (<http://www.pidconsortium.eu/>; pid consortium; 13.10.2017)
- [12] (<http://clarin.eu/>; CLARIN; 13.10.2017)
- [13] (<http://dariah.eu/>; DARIAH; 13.10.2017)
- [14] (<https://www.force11.org/group/fairgroup/fairprinciples>; FAIR Data Principles; 13.10.2017)
- [15] ([https://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri); ESFRI; 13.10.2017)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 1. Mission/Scope

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE is the successor of CLARIN Centre Vienna / Language Resources Portal (CCV/LRP) and is jointly funded by the Austrian Academy of Sciences and the Austrian Federal Ministry of Science, Research and Economy. The mission of CCV/LRP was to provide easy and sustainable access to digital language resources and provide depositing services for language resources created in Austria.

ARCHE extends the scope of CCV/LRP as it aims to offer an advanced and reliable data management and depositing service open to a broader range of humanities fields in Austria [1]. We recognise the challenges of ensuring sustainability, long-term preservation and accessibility in the face of the dramatically increasing amount and the rapidly changing nature of data. Therefore we aim for the continuous pursuit of technological and scientific development and the implementation of internationally accepted standards.

ARCHE is a central part of ACDH-OeAW's mission to foster the change towards the digital paradigm in the humanities [2]. ACDH-OeAW has undergone a dramatic development in the last two years, evolving from the predecessor Institute for Corpus Linguistics and Text Technology (ICLTT) with its focus on language data to an institutional and national centre of expertise in digital humanities. ACDH-OeAW pursues a dual agenda of conducting digitally enabled research and providing technical expertise and support to the research communities at the Academy, on the national, and international levels.

[1] (<https://arche.acdh.oeaw.ac.at/browser/about-service>; ARCHE Mission; 20.02.2018)

[2] (<https://www.oeaw.ac.at/acdh/about/mission/>; ACDH Mission; 13.10.2017 and <https://www.oeaw.ac.at/acdh/about/history/>; ACDH History; 13.10.2017)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## 2. Licenses

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

In principle, [CC0](#) [1] applies for all metadata in [ARCHE](#) [2]. The data provider retains all intellectual property rights to their data [3]. The depositor must grant distribution rights to ARCHE by signing a [Deposition Agreement](#) [4] and choose an access model (public, academic or restricted) and an applicable licence [5]. The depositors are encouraged to choose from standard and open licences [6] (such as Creative Commons or dedicated software licences). For restricted data sets, explicit permission from the depositor will be required in which case a login and explicit electronic signing of the licence is required.

The [Terms of Use](#) [7] and resource specific licences apply both for the use of the repository and the access to the resources. By using the resources, the user agrees to comply with the disciplinary and ethical norms as specified in [The European Code of Conduct for Research Integrity](#) promulgated by ALLEA (ALL European Academies) [8].

The access mode and licensing of every resource is clearly displayed to the users. For resources with limited access, like academic or restricted, users will be required to authenticate with their institutional account via Federated Login [9] (Shibboleth [10]) or to acquire a separate account.

In case of non-compliance with the terms and regulations, the user can be excluded from accessing the resources and general legal consequences according to national and international laws are applicable.

- [1] (<https://creativecommons.org/share-your-work/public-domain/cc0/>; description of CC0; 13.10.2017)
- [2] (see §6, d in <https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)
- [3] (see §5, n in <https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)
- [4] (<https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)
- [5] (see § 3 in <https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)
- [6] (<https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017 and <https://arche.oeaw.ac.at/browser/faq#faq-chose-licence>; FAQ about licences; 13.10.2017)
- [7] (<https://arche.acdh.oeaw.ac.at/browser/terms-of-use>; Terms of Use of ARCHE; 13.10.2017)
- [8] (<http://www.allea.org/allea-publishes-revised-edition-european-code-conduct-research-integrity/>; Code of

Conduct for Research Integrity; 13.10.2017)

[9] (<https://arche.acdh.oeaw.ac.at/browser/faq#faq-shibboleth>; FAQ in ARCHE about Shibboleth; 13.10.2017)

[10] (<https://www.shibboleth.net/>; Shibboleth home page ; 13.10.2017)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



### 3. Continuity of access

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

ARCHE is a core service provided by the Austrian Centre for Digital Humanities (ACDH-OeAW), an institute at the Austrian Academy of Sciences in Vienna. Continuity of access for the medium-term is secured by the funding of the institute. This funding is a combination of a base state financing via the global budget of the Academy, a dedicated start-up financing for the build-up of the centre by the National Endowment running for five years from 2015, and a mix of national and international third-party financed projects, either conducted directly by the institute or by a number of external cooperation partners.

To ensure the long-term perspective of ARCHE, negotiations with the legal department, the heads of the Academy and the directors of the individual institutes of the Academy are underway. They will result in an official declaration/commitment by the Academy, formalising the role of ARCHE as a central permanent service provided by the Academy for its own humanities institutes and the broader humanities community in Austria. This declaration will be underpinned by a business plan accompanied by a strategic paper for the years 2018-2021. These documents are currently in preparation and scheduled to be finished and officially released by mid 2018.

ACDH-OeAW is the national coordinator of CLARIAH-AT, a consortium of partners involved in the two large Digital Humanities European Research Infrastructure Consortia (ERICs) CLARIN and DARIAH, and as such it acted as coordinating instance for elaborating the [national strategy for Digital Humanities in Austria](#) [1] in 2015. One of the central goals of the strategy is the long-term preservation of research data (Leitlinie 4). A measure proposed to achieve this goal is the establishment of a national repository federation. The idea of the repository federation is to ensure long-term continuity of access to data hosted by individual partners, which include University of Graz (host of the repository [GAMS](#) [2]) and University of Vienna (host of the repository [PHAIDRA](#) [3]), by exchanging expertise, sharing technologies, and interlinking repository resources. The long-term goal is to reach an agreement between individual federation partners, according to which the other partners would step in as a fall-back repository in case one of the partners ceases to exist. Implementation of the measures supporting guideline 4 is part of the agenda for the CLARIAH-AT consortium for the upcoming three year period.

[1] (<https://doi.org/10.1553/DH-AUSTRIA-STRATEGIE-2015>; DH-Austria-Strategie; 20.02.2018)

[2] (<http://gams.uni-graz.at/context:gams> ; GAMS - University of Graz; 13.10.2017)

[3] (<https://phaidra.univie.ac.at/>; PHAIDRA - University of Vienna; 13.10.2017)

#### Reviewer Entry

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 4. Confidentiality/Ethics

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Every submission of resources is handled by the repository management team and is dealt with in direct communication with the depositor to ensure that the deposited resources meet the requirements and sensitive information is kept out, obfuscated or protected with the help of special access modes. Licensing, IPR and issues regarding ethical norms are cleared before the resources are accepted for deposition [1].

For example resources containing personal information have to be deposited in anonymised form, except for cases of explicit consent of the involved persons. For resources containing geographical information about endangered archaeological find-places the exact geographical coordinates have to be obfuscated or left out.

The depositor must sign an [agreement](#) [2] acknowledging that he has the right to deposit the data, allows ARCHE to disseminate the data according to chosen access modes, and has considered and taken care of any legal or ethical issues.

[1] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process>; Description of deposition process in ARCHE; 13.10.2017)

[2] (<https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 5. Organizational infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE is hosted by the [ACDH-OeAW](#) [1], which is part of the [Austrian Academy of Sciences \(ÖAW\)](#) [2], a major non-university research institution in Austria. ACDH-OeAW is financed by a base state financing via the global budget of the Academy, a dedicated start-up financing for the build-up of the centre by the National Endowment running from 2015 for five years and a mix of national and international third-party project fundings. IT resources are provided by the computing centre of the academy.

Currently ACDH-OeAW employs over fifty persons, who have a background in a wide range of disciplines, including Natural Language Processing, Literature, Linguistics, History, Archaeology, Social Sciences, and Anthropology. They are organised in five [core units](#) [3] dedicated to broad task areas, each having a long-term agenda, refined into individual measures and tasks. The core units coordinate in regular meetings and are managed by a head who reports to the ACDH-OeAW institute director. Coordination between the core units is ensured through day to day communication as well as bi-weekly heads meetings.

By the end of 2017, dedicated task forces were established to bring together members from different core units and work on a particular task. Each task force has a specification, an agenda to be fulfilled by 2021, holds regular meetings, is coordinated by a designated member, and reports to its assigned head of the associated core unit. One of these task forces, devoted solely to data preservation in general and ARCHE in particular, is concerned with discussing current issues and upcoming developments, communicating changes in the system or workflows, and making initial decisions, which are then reported to and agreed on with the head and in more prominent cases with the institute director.

Currently, five staff members, an equivalent of 3,5 FTEs, are solely dedicated to the development and the curation of ARCHE. Of these, 2 FTEs are concerned with the technical development of ARCHE, while 1,5 FTEs are filled by data archiving specialists. These positions are funded until 2021. Efforts are being made to ensure the continuity of the service by officially anchoring ARCHE as the central archiving service of the Academy for research data in the humanities, accompanied by a dedicated business plan, as described in Requirement 3.

Financing for ARCHE also covers advanced training and intensive participation in international conferences, meetings, and specialist trainings. This is detailed in Requirement 6.

- [1] (<https://www.oeaw.ac.at/acdh/acdh-home/>; page of ACDH ; 13.10.2017)  
[2] (<https://www.oeaw.ac.at/oesterreichische-akademie-der-wissenschaften/>; page of ÖAW ; 13.10.2017)  
[3] (<https://www.oeaw.ac.at/acdh/about/core-units/>; Organisation of core units in ACDH; 20.02.2018)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 6. Expert guidance

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

In terms of in-house guidance, the Austrian Center for Digital Humanities employs a plethora of interdisciplinary experts who are not necessarily involved with the development and service of ARCHE but offer constant information sharing in official and unofficial meetings or events. During regular meetings, such as the ACDH's fortnightly "Research Lunch", ARCHE staff has the opportunity to exchange knowledge with other ACDH peers involved in other digital humanities projects. For a list of the current ACDH team please visit: [https://www.oeaw.ac.at/acdh/team/current-team/ \[1\]](https://www.oeaw.ac.at/acdh/team/current-team/).

As a successor of CCV/LRP, ARCHE has an organical bond with the Austrian infrastructure projects CLARIN-AT and DARIAH-AT and the stable institutional context of the Austrian Academy of Sciences. These connections allow us to gather continuous feedback and consultation from experts of various fields such as linguistics, lexicography, information and computer science.

Furthermore, ACDH intends to establish an external flow of information with similar centres and institutions around the world. In November 2016, ACDH joined a mailing list bringing together members from different archival institutions, including ADS, IANUS, and DANS, in order to exchange experiences and ask other digital archivists for their advice when needed.

Funding for ARCHE also covers advanced training and intensive participation in international conferences and meetings. In addition ACDH organises many knowledge transfer events (e.g. Tool Galleries, ACDH Lectures) with regular participation of external experts. One of the most informative events that attracted a large group of participants has been the TEI Conference 2016 ([http://tei2016.acdh.oeaw.ac.at \[2\]](http://tei2016.acdh.oeaw.ac.at)) which was successfully organised by ACDH.

ACDH is strongly committed to continuing and increasing the knowledge transfer between in-house, external and international experts which surely benefits ARCHE in every respect.

- [1] (<https://www.oeaw.ac.at/acdh/team/current-team/>; team members of ACDH ; 13.10.2017)  
[2] (<http://tei2016.acdh.oeaw.ac.at>; TEI conference 2016; 13.10.2017)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 7. Data integrity and authenticity

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE recognises its responsibility to ensure the integrity and the authenticity of data in direct communication and cooperation with the depositor.

Our guidelines “[Before Submission](#)” [1], “[Curation Step](#)” [2] and “[Filenames, Formats and Metadata](#)” [3] contain fundamental information for both depositors and repository staff to make sure that the data is properly prepared and documentation of the data is complete. A dedicated command line application (“repo-file-checker”) was developed to help automate this process [4]. Additionally detailed descriptions and procedures for repository staff are documented in the internal wiki system, which is described in more detail in Requirement 12.

We utilise MD5 checksums to verify the integrity of resources stored in the repository. The procedure is performed when new data is added to the repository and on a weekly basis in order to ensure that no data was changed unintentionally and every time a backup is created.

As the core of our system, the repository management software [Fedora Commons 4](#) [5] offers built-in capabilities to ensure the data integrity through version control system. As a principle, all resources in our repository must have metadata. Once in the repository, the Fedora Commons 4 version control system ensures data authenticity by monitoring changes to the data and its metadata.

If new versions of resources become available, they are identified by a timestamp, allowing to refer consistently and unequivocally to every version, which is identified by a persistent identifier. Linking to previous versions is also anchored in the [metadata with dedicated properties](#) (acdh:isNewVersionOf and acdh:isPreviousVersionOf) [6]. In case rich metadata is available for resources in a dedicated format (CMDI, EAD, LIDO), which cannot be losslessly encoded in the current metadata schema, the corresponding metadata records are stored as separate binary objects and linked to the resources with a dedicated metadata property (acdh:hasMetadata). There is a two-way link between the resource and the metadata record in the repository browser, allowing to access either. The OAI-PMH can expose any of these records through the OAI-PMH protocol.



In order to apply custom business logic, ensure the validity of the ingested data, make sure that the resources and the metadata fulfill our requirements, and check that the required metadata fields are filled with correct values, a single point of access to Fedora was developed: [the doorkeeper](#) [7].

By end of 2018, we plan to fully support PREMIS metadata for tracking provenance information to strengthen data integrity and make all changes to data transparent to the users. A subset of the PREMIS elements are already supported (and used) out of the box by Fedora Commons 4 [8]. By expanding these and based on the PREMIS module for [Islandora](#) [9], a service will be implemented to display PREMIS metadata on demand.

Currently, ARCHE does not offer self depositing, which is why data depositors have to get in contact with ARCHE (either via e-mail or telephone) before depositing. A first proof of the depositors' identity is their institutional e-mail address, e.g. for staff members of the ÖAW. Usually at least one face to face meeting is arranged, especially with first time depositors, during which problems with identity may be detected. Depositing in ARCHE also requires a signed [Deposition Agreement](#) [10] which also collects the contact information of the depositor.

- [1] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process#before-submission> ; Description of how to prepare data for submission; 13.10.2017)
- [2] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process#curation-step>; Description of curation; 13.10.2017)
- [3] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata>; Requirements for filenames, formats and metadata; 13.10.2017)
- [4] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup in ARCHE; 13.10.2017 and <https://github.com/acdh-oeaw/repo-file-checker>; Code of repo-file-checker on GitHub ; 13.10.2017)
- [5] (<http://fedorarepository.org/>; Fedora repository software; 13.10.2017)
- [6] (<https://github.com/acdh-oeaw/repo-schema>; Metadata schema for ARCHE on GitHub; 13.10.2017)
- [7] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup in ARCHE with mention of doorkeeper; 13.10.2017)
- [8] (<https://wiki.duraspace.org/display/FF/Design+-+PREMIS+Event+Service>; Description of PREMIS related metadata in Fedora Commons ; 20.02.2018)
- [9] ([https://github.com/Islandora/islandora\\_premis](https://github.com/Islandora/islandora_premis); Use of PREMIS metadata in Islandora; 20.02.2018)
- [10] (<https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement for ARCHE; 20.02.2018)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 8. Appraisal

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE declares in its [Collection Policy](#) [1] that it functions as a digital data hosting service for the humanities in Austria. Consequently all humanities fields, including modern languages, classical languages, linguistics, literature, history, jurisprudence, philosophy, archaeology, comparative religion, ethics, criticism and theory of the arts all fall into our fields of relevance.

In these relevant fields of the humanities, our repository accepts a wide range of research data such as digital texts, lexical resources, tabular data, databases, images, file collections like GIS, 3D or CAD, multimedia files such as sounds and videos, websites and social media data, relevant digital tools, software and scripts.

The resources and data must meet our requirements and conform to the suitable formats and standards, which are defined on our "[Before Submission](#)" [2], "[Curation Step](#)" [3] and "[Filenames, Formats, and Metadata](#)" [4] guidelines.

The primary mechanism of quality control checks to ensure the completeness and understandability of data takes effect within the tight collaboration between the depositor and the ARCHE curators. Our data curators investigate the provided metadata to verify the consistency, understandability and completeness of data. If any issues arise, the data curator contacts the depositor and offers further assistance. If the research data doesn't fall into our field of relevance then we still offer support and share our knowledge and contacts with the aim of finding the most appropriate repository for the data in question.

[1] (<https://arche.acdh.oeaw.ac.at/browser/collection-policy>; Description of how to prepare data for submission; 13.10.2017)

[2] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process#before-submission>; Description of how to prepare data for submission; 13.10.2017)

[3] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process#curation-step>; Description of curation; 13.10.2017)

[4] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata>; Requirements for filenames, formats and metadata; 13.10.2017)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 9. Documented storage procedures

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE recognises its responsibility to monitor any changes which may impact the sustainability and long-term accessibility of its data. Therefore we continuously monitor technological and scientific developments and the implementation of such through our technical and organisational strategies.

Our repository is based on the well-established Fedora Commons 4 repository management platform which performs fine-grained versioning of digital objects (on the level of individual data streams) and allows for exporting, archiving and migrating the digital objects in XML-based formats. The actual repository system is only accessible to designated repository administrators. All other interactions with the system are managed via our [dedicated doorkeeper component](#) [1].

To ensure well-defined processes, the procedures in place are documented in the internal wiki system. Individual issues which arise, as well as each instance of a deposition process are administered via an internal issue tracking system, which supports the workflow by keeping track of the status and involved persons for each task.

A monitoring system based on the open source solution Icinga is employed to continuously check the availability of all services run by ACDH-OeAW, notifying the system administrators and the head of the technical core unit in case of any irregularities.

The primary server storage is a [RAID-6](#) [2] configuration that makes it possible to sustain read and write operations in the presence of up to two concurrent disk failures. Every night, the live data stored on the repository production server is copied up to the computing centre of the academy (ARZ) [NetApp](#) [3] production storage, of which numerous snapshots are stored on ARZ NetApp backup system in a separate location. Additionally the data is encrypted on our side and then copied to a long-term storage in the computing centre run by the Max Planck Computing and Data Facility (MPCDF) in Garching. The storage service provided by MPCDF replicates the data to a remote partner institution for high reliability and integrity.

ARCHE checks backups for integrity via MD5 checksums to verify the integrity of data. Recovery measures and steps are not only documented, but also tested to ensure swift recovery in case of emergency on the side of ARCHE.

Additional techniques for risk management are not yet set up. Risk analysis regarding data preservation will be accounted for in the Preservation Policy currently under development (see Requirement 10). Future development of ARCHE will also consider issues and topics mentioned in [DRAMBORA](#) [4].

- [1] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup in ARCHE ; 13.10.2017)
- [2] ([https://en.wikipedia.org/wiki/Standard\\_RAID\\_levels#RAID\\_6](https://en.wikipedia.org/wiki/Standard_RAID_levels#RAID_6); Description of RAID-6 on Wikipedia; 13.10.2017)
- [3] (<http://www.netapp.com/>; NetApp web page; 13.10.2017)
- [4] (<http://www.repositoryaudit.eu/>; DRAMBORA web page ; 20.02.2018)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 10. Preservation plan

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

As its primary preservation strategy, ARCHE performs migration of formats as opposed to providing software emulation. It aims to establish a high level of transparency with its depositors and users. Thus the deposition agreement and other relevant informative sections of our website highlight our responsibilities and our rights to copy, transform, store and provide access to the deposited items. All the actions relevant to preservation are specified in our documentation.

Further improvements and expansions in regard to long-term preservation strategies are intended to be worked on, implemented, and formalised in a dedicated Preservation Policy due to be released by the end of the first quarter of 2018. The policy will include essential documentation about archiving methods and processes (e.g. file formats, storage procedures, backup strategies, and migration procedures). Several relevant international best practices will be consulted, including the [Data Preservation Handbook \(by the Data Preservation Coalition\)](#) [1], the [Guides to Good Practice \(by Archaeology Data Service\)](#) [2], the [IT-Empfehlungen \(by IANUS\)](#) [3], the [Standard Recommendations \(by CLARIN\)](#) [4], the [Digital Preservation Strategies \(by the British Library\)](#) [5], documentation by the [Open Preservation Foundation](#) [6], and information about [preservation provided by the Library of Congress](#) [7].

Essential parts have already been put into practice. The [OAIS model](#) [8] is tightly integrated in ARCHE via the OAIS-compliant open-source repository system Fedora Commons 4. The [list of accepted and preferred formats](#) [9] is a crucial piece of information for depositors and curators to decide upon which formats to deposit and accept. Currently, format conversion is not normally executed by ARCHE, but data depositors are advised to do so if necessary. The [Deposition Agreement](#) [10] clearly states both depositors' and the repository's rights and obligations, in regard to deposition, curation and dissemination. The [Deposition Process](#) [11] explicates all relevant steps for deposition and curation to ensure sustainable archiving of data.

[1] (<http://dpconline.org/handbook/contents>; Data Preservation Handbook by Data Preservation Coalition; 13.10.2017)

[2] ([http://guides.archaeologydataservice.ac.uk/g2gp/ArchivalStrat\\_1-0](http://guides.archaeologydataservice.ac.uk/g2gp/ArchivalStrat_1-0); Guides to Good Practice by ADS; 13.10.2017)

[3] (<https://www.ianus-fdz.de/it-empfehlungen/>; IT-Empfehlungen, IANUS; 13.10.2017)

[4] (<https://www.clarin.eu/content/standard-recommendations>; CLARIN Standard Recommendations; 13.10.2017)

[5] (<http://www.bl.uk/aboutus/stratpolprog/collectioncare/digitalpreservation/strategy/dpstrategy.html>; Digital preservation Strategies by British Library; 13.10.2017)

[6] (<http://openpreservation.org/>; Open Preservation Foundation; 13.10.2017)

- [7] (<http://www.loc.gov/preservation/>; Digital preservation at Library of Congress; 13.10.2017)  
[8] (<http://public.ccsds.org/publications/archive/650x0m2.pdf>; Description of OAIS model; 13.10.2017)  
[9] (<https://arce.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#formats>; Preferred and accepted formats for ARCHE; 13.10.2017)  
[10] (<https://arce.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 13.10.2017)  
[11] (<https://arce.acdh.oeaw.ac.at/browser/deposition-process>; Description of deposition process in ARCHE; 13.10.2017)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 11. Data quality

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE employs two mechanisms to evaluate and ensure the quality of data: automatic validation and manual data curation by experienced curators.

Automatic validation and checking is done upon ingest with a dedicated command line application, [repo-file-checker](#) [1], and the doorkeeper component, which also allows for the validation of metadata against the [ARCHE metadata schema](#) [2]. We plan to improve this process by creating a user friendly web interface which will help depositors to deliver data in standard formats and to provide the required metadata. Information about [mandatory and recommended metadata properties is available for depositors](#) [3]. The metadata set includes properties to establish relations to other publications and sources as well to related data. [4].

The metadata schema for ARCHE takes into account international standards like the Dublin Core Metadata Initiative [5] and [Component Metadata by CLARIN](#) [6]. It is generic and flexible enough to preserve the documentation of heterogeneous resources.

Manual quality checks are performed by data curators as soon as automatic checks have been performed successfully. Since ARCHE's designated community is multidisciplinary, it is necessary to work in close cooperation with the depositors. A data curator inspects the provided metadata to verify the consistency, understandability and completeness of the data. If any issues arise, the data curator contacts the depositor for further actions. As far as possible, ARCHE will assign the curation of the deposited data to staff with a matching disciplinary background.

[1] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup of ARCHE; 13.10.2017 and <https://github.com/acdh-oeaw/repo-file-checker>; Code for repo-file-checker on GitHub; 13.10.2017)

[2] (<https://github.com/acdh-oeaw/repo-schema>; Metadata schema of ARCHE on GitHub ; 13.10.2017)

[3] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#metadata>; Information about metadata for users; 13.10.2017)

[4] (See the section "Relations to other projects, collections or resources in the table in <https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#metadata>; Information about metadata for users; 13.10.2017)

[5] (<http://dublincore.org/documents/dcmi-terms/> ; DCMI terms; 13.10.2017 and



<http://dublincore.org/documents/dces>; Dublin Core element set; 13.10.2017)

[6] (<https://www.clarin.eu/content/component-metadata>; CLARIN Component Metadata; 13.10.2017)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 12. Workflows

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The data submission process is accompanied by a data curator who is in direct communication with the depositor throughout. In addition, the comprehensive deposition workflow is clearly documented on the [ARCHE website](#) [1].

After an initial review to determine if the data is suitable for the repository, the curators counsel the depositor to facilitate the submission of data in preferred formats and accompanied by sufficient descriptive, administrative and structural metadata. In our guidelines, there is a section about "[Filenames, Formats and Metadata](#)" [2] which provides all necessary information on the preferred formats. Information about the workflow leading up to and following the deposition can be found in the section "[Deposition Process](#)" [3].

During the deposition process, the curators make sure that the data formats are in compliance with the supported and recognised standards, the data is properly prepared, and the documentation is complete. Any changes to formats or to the documentation required are carried out in close interaction with the depositor. During the data ingestion, a persistent identifier is automatically assigned to every resource.

After automatic validation and checking (described in Requirement 11) and manual curation, the ingestion into the repository system is executed with the help of a script and its corresponding configuration file to ensure a reproducible and consistent workflow. Since the repository is open for a broad range of subjects, almost any data type is allowed. If previously unknown data types cannot be handled by the ingestion script, it is changed accordingly and the changes are documented in the internal wiki system and communicated to all team members.

If a depositor's data doesn't match our collection policy, we still offer support to identify and recommend the most appropriate repository for the data in question.

The internal wiki system is based on Redmine, an open source project management web application. It supports the management of multiple projects, two of which are solely dedicated to ARCHE (one for the technical development and the other for data curation). Each project has a separate wiki space containing descriptions of

workflows and procedures, which are administered and updated by the task force described in Requirement 5. Previous versions are saved by automatic versioning provided by Redmine.

For technical development issue and bug tracking as well as documenting, feature requests and the roadmap function are extensively used. An overall documentation for the system can be found in the wiki, which points to more extensive documentation on GitHub.

For data curation, a ticket template with sub issues representing the curation workflow has been set up. Each of this issues contains a short description of the task and points out the software and tools to be used. This template is copied and used for each new data collection that is ingested into ARCHE. If changes in the workflow occur, the template is changed accordingly.

Redmines offers custom user roles with different permissions, thus allowing for a clear assignment of editing rights for each user.

[1] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process>; Description of depositon process in ARCHE; 13.10.2017)

[2] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata>; Requirements for filenames, formats and metada for ARCHE; 13.10.2017)

[3] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process>; Description of depositon process in ARCHE; 13.10.2017)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### 13. Data discovery and identification

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

ARCHE offers an easy-to-use [web interface](#) [1], where users can browse and perform a wide variety of search actions to discover stored data. In principle, our repository requires metadata for every resource and all metadata is openly searchable via our web interface. Users can select predefined search filters (facets), enter their keywords. Additionally complex SPARQL search queries can be performed via a dedicated user interface.

Additionally the repository exposes a public [OAI-PMH endpoint](#) [2], which is already regularly harvested by the [Virtual Language Observatory \(VLO\)](#) [3], CLARIN's main metadata catalogue.

The repository systematically assigns [Handles](#) [4] to published resources to ensure persistent referenceability of digital objects, irrespective of their future location. Furthermore each resource page features a recommended citation. In case the published data with a Handle is withdrawn, a tombstone page will be created containing the metadata, which will be kept.

The web interface provides a basic user friendly view of selected metadata associated with the resources as well as an expert view displaying all available metadata. The metadata is also available as RDF/XML. The system is equipped with a [growing set of dissemination services](#) [5] for displaying specific data types, so that the users can view the resource directly and seamlessly. A typical example is a [TEI](#) [6] file (or any other XML) that can be transformed to HTML or PDF for viewing.

[1] (<https://arche.acdh.oeaw.ac.at/browser/>; ARCHE website with search interface on start page; 13.10.2017)

[2] (<https://arche.acdh.oeaw.ac.at/oai/>; OAI-PMH endpoint of ARCHE; 13.10.2017)

[3] (<http://vlo.clarin.eu>; Virtual Language Observatory; 13.10.2017)

[4] (<http://www.handle.net/>; Handle web page ; 13.10.2017)

[5] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup in ARCHE ; 13.10.2017)

[6] (<http://www.tei-c.org>; Description of TEI; 13.10.2017)

#### Reviewer Entry

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 14. Data reuse

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Long-term accessibility and understandability of data is one of the priorities of ARCHE. Thus, the sections "[Filenames, Formats and Metadata](#)" [1] and "[Deposition Process](#)" [2] of our web-platform contain information about mandatory and recommended metadata for the benefit of the users.

As stated in Requirement 11, the metadata schema for ARCHE was developed respecting international standards. Mandatory metadata is required to ensure the inner workings of the repository system, while recommended metadata was selected to increase the understandability of data. The metadata is encoded in RDF/XML and provided as CC0, which maximises the interoperability between ARCHE and the depositors and other service providers which harvest our metadata.

We strongly encourage our depositors to provide the resources in standard formats suitable for long term preservation and acknowledged by the respective international research communities. We also provide support in converting the data if this is necessary and feasible.

Formats ingested into the repository are carefully monitored for possible obsolescence and when a migration becomes necessary a bespoke migration plan will be tailored for the resources affected.

[1] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata>; Requirements for filenames, formats and metadata in ARCHE; 13.10.2017)

[2] (<https://arche.acdh.oeaw.ac.at/browser/deposition-process>; Description of deposition process in ARCHE; 13.10.2017)

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 15. Technical infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The repository implements the tasks and functions of the [Open Archival Information System reference model](#) [1]. The implementation is facilitated by the underlying repository platform Fedora Commons 4, which supports ingest of Submission Information Packages (SIP) and processing of Archival Information Packages (AIP).

All mandatory responsibilities described in the [OAIS reference model](#) [2] are already implemented in ARCHE. Negotiations for appropriate information that take place during the deposition process are based on the [Collection Policy](#) [3] and further auxiliary documentation detailing the requirements for deposition of research data. The [Deposition Agreement](#) [4], ensures that ARCHE obtains sufficient control of the information for long term preservation. The designated community is likewise specified in the Collection Policy. The independent understandability of data is ensured by our [metadata requirements](#) [5] and data quality checks as described in Requirement 11. The preservation procedures to follow are documented in the internal wiki system described in Requirement 12 and will be further formalised in the Preservation Policy (Requirement 10) and the Business Plan (Requirement 3). Lastly, the preserved information is already disseminated via ARCHE's web interface.

We maximise the quality of our data and the interoperability of our repository service by following the international and community standards and principles such as Handles, [OAI-PMH](#) [6], [Linked Data Platform](#) [7], Dublin Core Metadata Initiative [8] and [Component Metadata of CLARIN](#) [9].

Regarding data formats suitable for archiving, we primarily refer to the ["IT-Empfehlungen" by IANUS](#) [10], which in turn take into account several international standards and recommendations. For curation tasks like format validation or format conversion ARCHE relies on tools and services provided by renowned digital archiving institutions, as e. g. the tools collected in the [E-ARK project](#) [11] or by the [Open Preservation Foundation](#) [12], as well as [PRONOM](#) [13] as a file format registry. We strive to use widely adopted open software whenever possible.

The technical setup of the repository is detailed in a respective [user information page](#) [14] and many of its components are publicly available via [GitHub](#) [15]. The main components of the repository system are [Fedora Commons 4](#) [16] and [Drupal 8](#) [17], both community supported and wide-spread software solutions. Custom modules and functions are mainly implemented in [PHP 7](#) [18].

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



The embeddedness of ARCHE in the institutional context of ACDH-OeAW and the ÖAW allows for the continuous development of the repository according to relevant standards and practices. The use of a suitable hardware setup with sufficient capacity and measures against hardware failure is ensured through close interaction and consultation with the computing centre of the academy (ARZ). See Requirement 16 for further details of the technical setup.

The ARZ itself aims to respect renowned international and national standards for the hardware and services they provide, including ISO/IEC 27001, ISO/IEC 27002, and the Austrian manual for information security ([Österreichisches Informationssicherheitshandbuch](#) [19]). Currently, efforts are made at ARZ to create formal IT guidelines for the academy.

- [1] (<https://public.ccsds.org/Pubs/650x0m2.pdf>; OAIS model description; 13.10.2017)
- [2] (<https://public.ccsds.org/Pubs/650x0m2.pdf>; mandatory responsibilities in the OAIS model, page 3-1 ; 22.02.2018)
- [3] (<https://arche.acdh.oeaw.ac.at/browser/collection-policy>; Collection Policy of ARCHE; 22.02.2018)
- [4] (<https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>; Deposition Agreement of ARCHE; 22.02.2018)
- [5] (<https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#metadata>; Metadata requirements of ARCHE; 22.02.2018)
- [6] (<https://www.openarchives.org/pmh>; OAI-PMH; 13.10.2017)
- [7] (<https://www.w3.org/TR/ldp/>; Linked Data Platform ; 13.10.2017)
- [8] (<http://dublincore.org/documents/dcmi-terms/> ; Dublin Core Metadata Initiative; 13.10.2017 and <http://dublincore.org/documents/dces>; Dublin Core Metadata Element Set; 13.10.2017)
- [9] (<https://www.clarin.eu/content/component-metadata>; Component Metadata of CLARIN; 13.10.2017)
- [10] (<https://www.ianus-fdz.de/it-empfehlungen> ;IT-Empfehlungen by IANUS; 22.02.2018)
- [11] (<http://www.eark-project.com/resources/eark-tools>; Tool list provided by E-ARK project; 22.02.2018)
- [12] (<http://openpreservation.org/>; Open Preservation Foundation web page; 22.02.2018)
- [13] (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>; PRONOM file format registry ; 22.02.2018)
- [14] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup>; Description of technical setup of ARCHE; 13.10.2017)
- [15] (<https://github.com/acdh-oeaw?q=repo>; Code of ARCHE components on GitHub ; 13.10.2017)
- [16] (<http://fedorarepository.org/>; Fedora Commons web page; 22.02.2018)
- [17] (<https://www.drupal.org/8> ; Drupal 8 web page; 22.02.2018)
- [18] (<http://www.php.net/>; PHP web page; 22.02.2018)
- [19] (<https://www.sicherheitshandbuch.gv.at/>; Österreichisches Informationssicherheitshandbuch; 22.02.2018)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Send back to applicant for modification

*Comments:*

DISCUSS WITH ISABELLE

## 16. Security

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

There is a clear internal agreement on the division of roles between the ARCHE team at ACDH and computing centre of the academy (ARZ) with respect to the technical infrastructure, ARZ being responsible for the maintenance and provision of server hardware (computing and storage capacity), including physical security measures, network configuration, firewall setup and basic virtualisation layer.

Security is assured on different levels:

#### **a) Data backups**

Four levels of backup are used. On the first level daily snapshotting allows to recover past repository state, e.g. in case of unintentional data removal. On the second level the RAID-6 setup of the network-attached storage protects from hardware failures of up to two drives at the same time. On the third level live replication to an identical storage matrix in the second ACDH-OeAW data centre protects from major infrastructure failures making the primary data center unusable and from major hardware failures (e.g. a RAID controller failure). On the fourth level we are pushing a data backup to the [MPCDF](#) once a week which is then replicated within the [B2SAFE](#) infrastructure. The fourth level backup allows for major disaster recovery.

The first backup level is being managed by the ACDH, the second and third by the ARZ computing centre and the fourth again by the ACDH. The first and fourth level is used on demand while the second and third level are automatically monitored by ARZ.

All the levels were tested initially and now we are testing them twice a year.

#### **b) Live monitoring**

We use the Icinga software to provide live monitoring of all vital services. For the repository we are monitoring the state of all important software components (Fedora Commons, the Blazegraph triplestore, the webpage providing a GUI, the OAI endpoint, and the service providing resolution of Handles). In the same way we are monitoring system resources usage (available RAM, CPU and storage space). If one of the tests fails, a notification

is automatically sent to the system administrators.

#### **c) Failover**

Different failover procedures are prepared for different kind of failures.

Data storage failures are handled by levels two to four of the data backup strategy described in a). Failures related to level two don't cause any downtime because of the nature of the RAID. Issues related to level three can be handled within minutes by reconfiguring the virtualisation host to use the spare data matrix. Recovery procedures related to level four may take up to a few days.

Virtualisation host failures are handled by migrating all the software components to a different host. As all components are prepared as Docker containers with their source kept in git repositories (both locally and on GitHub) and the data storage is network-attached, migration to another virtualisation host is straightforward and takes as long as physically copying the software code and the data from the backup location. The procedure of setting up the whole software stack within a Docker-enabled environment is described within the documentation of the repository in our internal Wiki system described in Requirement 12.

#### **d) Network security**

Network security is assured by ARZ. ARZ provides a firewall between the Internet and data centres network and enforces strict routing rules. Only incoming HTTPS connections are accepted from external networks and from the ACDH local network only HTTPS and SSH connections are routed. ARCHE's VM is using its own firewall accepting only HTTPS and SSH connections. All connections are encrypted and the available cipher configuration is updated according to the current [CertBot](#) [1] settings.

#### **e) Physical security**

Physical security of the hardware is assured by ARZ. Both data centres being in use are placed in dedicated locked rooms, with only authorised personnel able to access.

#### **f) Software lifecycle**

The ARZ data centre uses VMware ESXi 6.0 virtualisation software which is supported until 12-03-2020 and a migration to ESXi 6.5 is planned this year to provide support until 15-11-2023. Virtual machines' operating system is CentOS 7 with support provided until 30-06-2024. For the Docker updates we rely on CentOS package updates. All other software components (Fedora Commons, Blazegraph, PHP, Drupal) are continuously updated as new releases become available.

### **g) Authorisation**

Write/modify access to the repository is restricted to the administrators. The read access can be flexibly granted to particular users or to the public using access control rules provided by the [Fedora Commons WebAC module](#) [2]. Authorisation is provided by Fedora Commons and authentication is provided by the Doorkeeper (see Requirement 9). Authentication methods include local login & password as well as Shibboleth-based federated login (AAI) coupled with the [eduGAIN](#) [3] federation.

Documentation of the setup is partially publicly available via the respective GitHub repositories and on ARCHE's web page. Documentation of the full setup and procedures is contained and maintained in the internal wiki system described in Requirement 12.

Besides the storage procedures [4] in place and documentation of recovery procedures ARCHE is not using specific risk analysis tools at this time. Risk analysis regarding data preservation will be made explicit in the Preservation Policy currently under development (see Requirement 10) and possible risks for ARCHE as a whole will be acknowledged in the future Business Plan (see Requirement 3), with consideration of relevant standards as e.g. DRAMBORA.

[1] (<https://certbot.eff.org/>; CertBot web page; 22.02.2018)

[2] (<https://wiki.duraspace.org/display/FEDORA4x/WebAC+Authorization+Delegate>; Fedora Commons WebAC module; 22.02.2018)

[3] ([https://www.geant.org/Services/Trust\\_identity\\_and\\_security/eduGAIN](https://www.geant.org/Services/Trust_identity_and_security/eduGAIN); eduGAIN web page ; 22.02.2018)

[4] (<https://arche.acdh.oeaw.ac.at/browser/technical-setup#storage-procedures>; Storage Procedures in ARCHE; 22.02.2018)

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **17. Comments/feedback**

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### **Applicant Entry**

*Statement of Compliance:*

0. N/A: Not Applicable.

*Self-assessment statement:*

Some of the requirements have a great overlap and it might be advisable to revise them in this regard.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*