

**30-07-2024**

## **Deliverable 8.2:**

# **AARC TREE Data Management Plan**

Publication Date	29-07-2024
Due Date	31-07-2024
Authors:	D. Groep (Nikhef), L.Florio (NORDUNet),
Version:	1.0
Document Code:	AARC-TREE D8.1
Publishing Organisation:	NORDUNet
DOI:	N.A.

### **Abstract**

This document describes the AARC TREE Data Management Plan.

### **Copyright**

© Members of the AARC community.

This work is licensed under a Creative Commons Attribution CC-BY 3.0 Licence.





# Table of Contents

<b>1. Data Summary</b>	<b>3</b>
2. FAIR data	3
2.1 Making data findable, including provisions for metadata	4
2.2. Making data accessible	4
2.3. Making data interoperable	5
2.4. Increase data re-use	5
3. Other research outputs	6
4. Allocation of resources	6
5. Data security	6
6. Ethics	6
7. Other issues	6
<b>References</b>	<b>7</b>

# 1. Data Summary

The AARC Technical Revision to Enhance Effectiveness (AARC TREE) project defines common strategies for the development and deployment of authentication and authorisation infrastructures (AAs) for Research and e-Infrastructures, to improve access and sharing of scientific resources and to improve interoperability among research infrastructure communities across the thematic areas. It pursues that goal by defining an architecture and a series of accompanying policy and good practice documents, through community engagement activities, structured and unstructured requirements analysis in the community, and by outreach. The project builds upon the existing body of documents and requirements, and augments this to achieve enhanced effectiveness for the European (and global) Research Infrastructures.

The primary outputs of the project are textual documents and white papers, in some cases supported by structured infrastructure and community survey results (questionnaires to collect infrastructure requirements) where no existing sources of community requirements are available.

Wherever possible, the project re-uses existing requirements collections and survey data, including those generated by the 'FIM4R' activity (Federated Identity Management for Research), and the structured infrastructure surveys performed in the context of the European Open Science Cloud service provider catalogues. These requirements and infrastructure survey results provide complementary input for the AARC TREE architecture definition, both reducing effort in the project (since such requirements do not need to be collected again) and ensure better community participation (since the communities will not be repeatedly asked the same questions, and therefore can focus on the enhancements needed in AARC TREE).

Documentary information is provided in the form of human-readable documents, including references to related documents and architectural and good practice guidelines. These are structured in the AARC Document Series (normative and information 'guidelines'), maintained for the benefit of the AARC Community by the AARC Engagement Group for Infrastructures (AEGIS).

Due to the nature of the documents and infrastructure surveys, the data volume is limited and not expected to exceed O(100 Mbyte) in total. Any structured infrastructure survey data will be provided in tabular formats and include complementary documentation in the form of deliverables or white papers. The re-used documentary and survey data is of a similar size, and consistent mainly of documents and limited tabular (infrastructure survey) data, available under open (creative commons) licences.

The AARC Guidelines provide the key outputs of the project, and will be foundational to AAI systems architects, implementers and deployers in the Research and e-Infrastructures in Europe and globally.

## 2. FAIR data

## 2.1 Making data findable, including provisions for metadata

The enhanced AARC Blueprint Architecture and Guidelines, as well as any structured infrastructure survey results, will be assigned a DOI persistent identifier as part of the repository deposition process. All depositions will be accompanied by at least Dublin Core meta-data. Where relevant, structured relationships between guidelines will be expressed through semantic association meta-data annexed to the submission in the repository. Any infrastructure survey results will be accompanied by a description of their semantics.

Given the lack of an ontology for AAI architectures or AAI infrastructure surveys, descriptive keywords will be used to improve findability of the guidelines. They will also be associated with the AARC Community in the chosen repository (Zenodo), so that all the AARC guidelines can be identified as a collection.

The selected repositories (specifically Zenodo) are indexed by OpenAIRE and searchable by common Internet search engines.

## 2.2. Making data accessible

### Repository:

Guidelines and data will be deposited in the Zenodo repository. This repository is well-known, appropriately managed, and already home to the existing suite of prior AARC Guidelines, hence choosing Zenodo for deposition improves findability and continuity for the generated documents and data.

Longevity of the repository is ensured by the European Organisation for Nuclear Research (CERN), which has a 70+ years track record in persistently storing data and making data openly accessible. In addition, CERN is also a member of the AARC TREE project consortium. The Zenodo repository has a current lifetime linked to that of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least. Zenodo commits to both FAIR and Plan-S principles (<https://about.zenodo.org/principles/>)

Zenodo ensures assignment of a resolvable DOI to all submissions, and ensures at least Dublin Core meta-data is associated with the deposition.

In addition, copies of the contextual data and references to the development history of the guidelines will be made publicly available through the AARC Community web site, which is persistently maintained by AEGIS and hosted in the GEANT infrastructure.

### Data:

All AARC TREE guidelines and documents are available under an Open Creative Commons (CC) licence. For new documents and data, this will be under a CC Attribution licence (CC-BY); other CC licences will be used when upstream source licences are required (in particular, some documents may be based on CC-BY-NC-SA enforcing Share-Alike requirements).

The infrastructure survey results do not include personal data (while retro-active identification of individuals responding to the survey may be possible based on the results, the survey results are representative of a broad (research) infrastructure community, and thus not reflective of the input or opinion of individuals). We expect this data to be public without any restrictions.

All documents and data will be available without further embargo periods after publication through standard access protocols (secured hypertext transport protocol) as enabled by the Zenodo repository. There are no restrictions on distribution, and persons accessing the data do not need to identify themselves.

#### **Metadata:**

All meta-data in Zenodo is publicly accessible without restrictions and is put in the public domain (CC-0). Meta-data is available as long as the repository exists, which is foreseen to be at least 20 years from deposition.

As the AARC TREE project does not produce software, no additional software documentation or source code is foreseen.

### **2.3. Making data interoperable**

There are no available vocabularies for AAI architecture and guidelines such as those generated in the AARC TREE project. Hence, only keywords will be provided alongside the documents. Y grouping documents in the AARC Community in the repository, and by assembling all guidelines (both from AARC TREE as well as past and future documents) in a structured document series that is managed by AEGIS on behalf of the community, we encourage both re-use and interpretability of the guidelines. Since guidelines do not have a qualifying relationship between each other, only full deprecation of guidelines by new guidelines can and will be expressed as qualified references in the meta-data in Zenodo. The Zenodo repository provides facilities to express qualified references.

### **2.4. Increase data re-use**

Documents and any (infrastructure survey) data will be freely available from the repositories, and use an open licence encouraging unencumbered re-use. Where document augment or re-use elements that were previously licensed under a Share-Alike or non-commercial Creative Commons licence, such a licence will be maintained on the derived works. Where relevant, source information and attribution will be provided with all documents.

Architecture and documents are intended and disseminated for immediate re-use by Research and e-Infrastructures in Europe and globally. Data will remain available after the end of the project.

The AARC Architecture and guidelines are peer reviewed through community consultation and the AEGIS review process, and represent community consensus.

## **3. Other research outputs**



There are no outputs foreseen for the project besides the AARC Architecture, the AARC guidelines, supporting documents and white papers, and meta-data, and the infrastructure survey results.

## **4. Allocation of resources**

The resources needed to make the AARC Blueprint Architecture (BPA), Guidelines, white papers and infrastructure surveys available and FAIR are included as part of the project activities and need not further specific funding.

Data management is the responsibility of the project-management work package and the project management team.

Long term preservation is ensured by deposition in the Zenodo repository for at least 20 years. Ancillary materials and the AARC Community presence is supported by AEGIS and its contributors and supporting projects. The AARC Community has, over the past decade, demonstrated to be able to maintain access to all information also in absence of specific AARC project funding. This model is set to continue during and after the AARC project completes.

## **5. Data security**

The data is preserved in the Zenodo repository and made available for at least 20 years.

## **6. Ethics**

There are no ethical issues relating to the AARC Architecture, guidelines, infrastructure requirements (survey) results, and other materials.

## **7. Other issues**

This data management plan is self-contained and constitutes the entire process for data management.



## References

- [AARC Projects]      <https://aarc-project.eu/about/documents/>
- [AEGIS]              <https://aarc-project.eu/about/aegis/>
- [FIM4R]              <https://fim4r.org/>