

Aruna Object Storage

[[RDC Integration](#)] [[Overview](#)] [[Getting started](#)] [[User Guide](#)] [[Developer Guide](#)] [[AOS infrastructure](#)] [[AOS data structure](#)] [[References](#)]



Aruna Object Storage (AOS) is a modern distributed storage platform designed to meet the increasing demand for effective data management and storage of scientific data. It is the central storage of the [Research Data Commons](#) (RDC) cloud layer and the data foundation for the upper layers. It is a cloud-native, scalable system with an API and a S3-compatible interface. It allows resource organization into Objects, Datasets, Collections and Projects. Additionally, it provides an event-driven architecture which enables automation, data validation and improves accessibility and reproducibility of scientific results. AOS is open-source and available at <https://aruna-storage.org>.



Status: **PRODUCTIVE**

Current Version: v1.0.x; Version v2.0.x (planned October, 2nd 2023)

Weblink: <https://aruna-storage.org>

Target group:

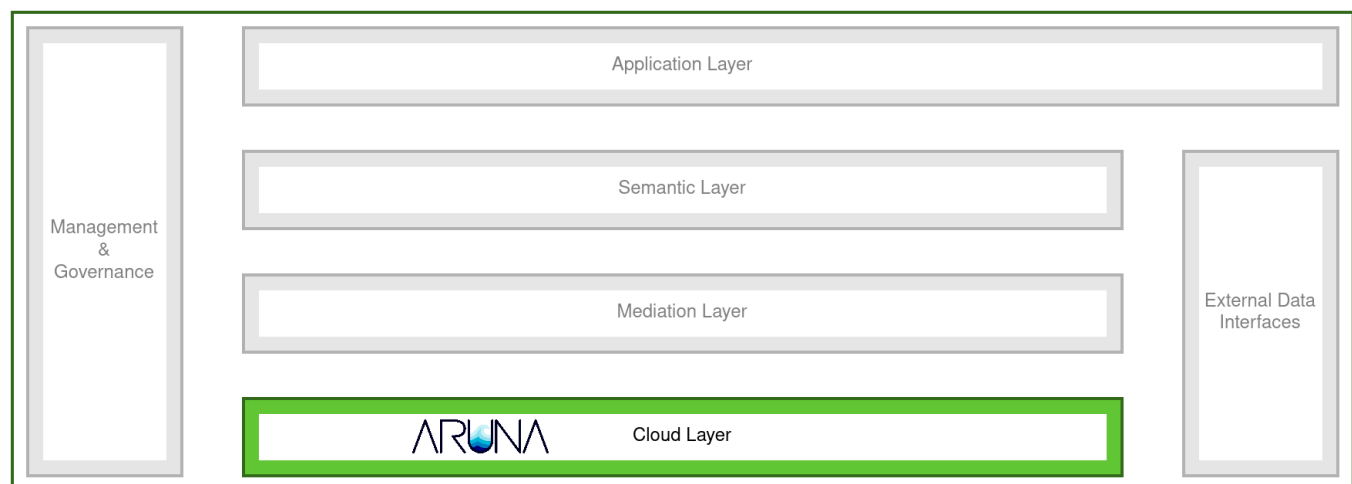
Keywords: data management, storage, FAIR, multi-cloud, cloud-native, data mesh

RDC Integration: integrated or connected

Product owner:

RDC Integration

The Aruna Object Storage (AOS) is part of the cloud layer of the RDC. It is an object storage designed as open source storage platform that allows scientists to store, annotate and share their data according to the FAIR principles.



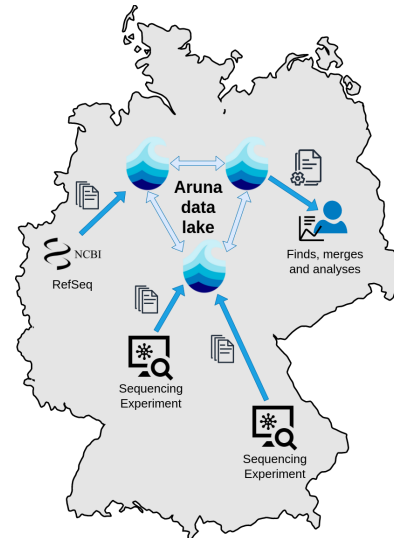
Overview

AOS is a fast, secure and geo-redundant data storage. It offers a sophisticated metadata management according to the FAIR principles. It builds the foundation for RDCs mediation and semantic layer and handles all stored data objects secure, and data-agnostically.

AOS key features are:

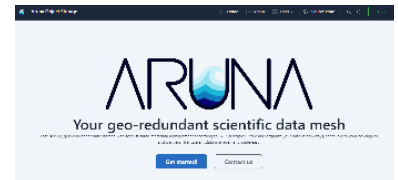
- Data storage - Easily store large volumes of arbitrary data.
- Metadata enrichment - Enrich and group a user's data and metadata together for easier access and better organization.
- Secure Storage - Secure data storage from the moment of upload.
- FAIR - All data is managed according to the FAIR principles.
- Easy Sharing - Flexible data sharing made easy with flexible presigned URLs.
- Effortless Migration - Effortless migration from other cloud storage providers such as AWS (Amazon Web Services) thanks to the S3 compatible interface.

Storing data in localized, domain-specific data silos has limited use for collaboration, reuse and data analysis. AOS offers significant benefits for research data in NFDI4Biodiversity. These benefits include improved collaboration, compliance with FAIR principles, scalability, robust data security and seamless integration with existing systems and workflow tools. Ultimately, this facilitates scientific progress and enhances the quality of our research.



Getting started

AOS is located at <https://aruna-storage.org>. Users can log in there. Currently, the AAI of the GWDG is used for this purpose, which requires a user account at the GWDG, the DFN or at LifeScience AAI. Nevertheless, additional identity providers are possible. Thus, login via an SSO of NFDI4Biodiversity (and other NFDIs) will be supported when the service is established. After the AOS account has been activated, the user can create a project. Further users can then be activated for this project to enable data exchange and joint processing. The project can then be filled with data either via the API or via the S3 interface.



User Guide

Basically, AOS is intended as a data backend for the RDC. For this reason, very few end users will use AOS directly. Data import, verification, transformation and processing is basically possible via the services in the mediation layer. This also ensures the consistency of the data. Users and services can be informed about changes to individual data objects or even entire projects via the AOS notification service and can thus react to these changes.

Developer Guide

The current documentation for using AOS is linked from the AOS home page at <https://aruna-storage.org>. This contains a complete description of the API. AOS consists of five main components: AOS Server, AOS Proxy, AOS API (and its S3 interface), AOS CLI and AOS Notification System. Of these components, the AOS team installs and maintains the servers and associated databases. AOS proxies can then be installed at various locations, which then communicate with the servers in each case. The actual data traffic from and to the storage backend then takes place via the AOS proxies. The interaction between a client and the proxies/servers takes place via the AOS API. To reduce the entry barrier, there is a command line interface called AOS CLI, which encapsulates API calls. Moreover, an S3 interface was implemented, since many software packages already support data storage via S3 as industry standard. Finally, the AOS notification system will soon be released to allow immediate response to changes in the AOS. This can be, for example, a data verification that is automatically initiated when a data upload is complete.

AOS infrastructure

The main component of AOS is a distributed database system. It synchronizes all data between several computers at different locations and thus generates fail-safety via this redundancy. This database is regularly backed up. The actual data can also be synchronized across multiple sites to provide redundancy. Nevertheless, all data will also be stored at one location in a redundant system. Due to the fact that data cannot be overwritten, but new versions of the data are then created, in combination with the redundant data storage at multiple levels, no backup of the data is currently performed. An implementation at a later date is currently being discussed.

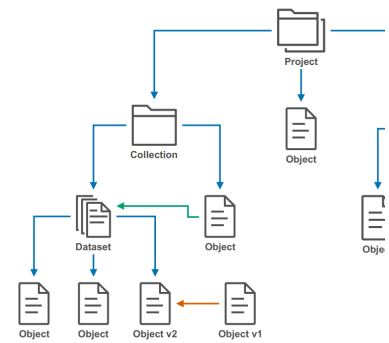
AOS data structure

AOS organizes data in Version 1.x into Projects, Collections, Object Groups, and Objects, starting with version 2.x the data structure will be even more flexible and are organized into Projects, Collections, Datasets, and Objects with a more flexible relation model.

[blocked URL](#)

UML diagram of the Aruna Object Storage data structure in Version v1.0.

x



UML diagram of the Aruna Object Storage data structure starting in Version v2.0. All resources form a directed acyclic graph of belongs to relationships (blue) with Projects as roots and Objects as leaves. Resources can also describe horizontal version relationships (orange), data/metadata relationships (yellow) or even custom user-defined relationships (green).

References

Links:

- Dokumentation and Aruna start page: <https://aruna-storage.org>
- Source-Code: <https://github.com/ArunaStorage>