

# Rucio

## Streamlining Complex Data Management Across Distributed Networks

### Highlights

#### Efficient data Handling

Rucio efficiently manages and organizes massive volumes of data across distributed computing environments, facilitating seamless access and retrieval.

#### Automated Replication

Implements robust data replication strategies to optimize storage utilization and ensure data redundancy across multiple locations.

#### Scalable Architecture

Designed to handle data on the scale of exabytes and billions of files, making it ideal for large-scale scientific projects.

#### Keywords

Data management, distributed computing, large-scale data, data replication, open-source software



### Description

Rucio is an open-source data management framework tailored to handle large volumes of scientific data across distributed computing environments. Originally developed for the high-energy physics community at CERN, it is adept at managing data from petabytes to exabytes, making it applicable across various research domains. Rucio manages now over an exabyte of data in a single deployment for the ATLAS experiment. It organizes data into objects linked with extensive metadata, facilitating advanced indexing and querying capabilities for efficient data management. It automates key processes such as data replication, recovery, and access, adhering to user-defined policies on data lifecycle, security measures, and archival rules.

The framework is scalable and adaptable, integrating smoothly with different storage systems and data transfer protocols. It supports extensions through plugins to meet specific needs or integrate with other technological ecosystems. Rucio prioritizes security with robust authentication and authorization mechanisms that comply with global standards, ensuring that data access is restricted to authorized users. By optimizing geographic data distribution, Rucio enhances accessibility and reduces latency, essential for fast data retrieval necessary for data-intensive applications.



# Rucio

## Advantages

### Data Localization and Accessibility

Rucio optimizes data placement across global storage to enhance computational efficiency and reduce latency. By aligning data storage with access patterns and computational demands, it ensures rapid data retrieval, improving system performance and responsiveness.

### Monitoring and Diagnostics

Rucio's monitoring tools provide real-time insights into data transfers, storage health, and user activity, enabling efficient management and troubleshooting. This suite of tools helps maintain system reliability and supports informed decision-making about data management.

### FAIR Data Principles

Rucio enhances the Findability, Accessibility, Interoperability, and Reusability of scientific data. It assigns unique identifiers and detailed metadata for easy discovery. Using open protocols, Rucio ensures secure access with robust authentication. It supports accessible languages for seamless data integration and interoperability. Rucio also provides rich data descriptions and provenance, optimizing data reuse across scientific domains.

For further information and licensing requests, please reach out to the CERN Venture Connect Program Manager.

### CVC Programme Manager

Linn Kretzschmar

[Linn.kretzschmar@cern.ch](mailto:Linn.kretzschmar@cern.ch)

### Research Contact

Martin Barisits

[Martin.Barisits@cern.ch](mailto:Martin.Barisits@cern.ch)

## Applications

**Need for distributed data, efficient data handling, automated replication and scalable architecture, such as:**

- **Scientific Research:** Manage data in large-scale physics experiments, genomics research, and climate modeling.
- **AI and Machine Learning:** Supports the data needs of AI research by managing training datasets distributed across various nodes.
- **Big Data Analytics:** Enhance industries like finance, healthcare and medical research (imaging, genomics).
- **Climate Science and Environmental Monitoring:** Manage data from satellite observations and climate models, and monitor various sensors and locations.

The applications listed above are examples of where Rucio could be utilized. The versatility of the technology encourages further exploration.

## Website

<https://rucio.cern.ch>

## Publications

BARISITS, M., BEERMANN, T., BERGHAUS, F. ET AL. RUCIO: SCIENTIFIC DATA MANAGEMENT. COMPUT SOFTW BIG SCI 3, 11 (2019). [HTTPS://DOI.ORG/10.1007/S41781-019-0026-3](https://doi.org/10.1007/s41781-019-0026-3)

Large User Base

Manages Over an Exabyte Data in a Single Deployment  
(ATLAS experiment)