Advancing Large-Scale Scientific Collaborations with Rucio

Hugo González Labrador[1,*] and *Martin* Barisits[2,**]

[1]CERN, IT Department, 1 Esplanade de Particules, Geneva, Switzerland
[2]CERN, EP Department, 1 Esplanade de Particules, Geneva, Switzerland

Abstract. This article provides an overview of the data deluge observed and projected in the next decade in high-energy physics and related sciences, and how Rucio, a mature and modular scientific data management platform used by various scientific collaborations, is employed to manage these extensive datasets. The paper highlights Rucio's key features, current deployments, and its role in facilitating scientific innovation across diverse domains

1 A data deluge in science

Modern scientific experiments, particularly those involving complex instruments, generate or plan to generate unprecedented volumes of scientific data. Figure 1 shows that since the beginning of the LHC Run3 in July 2022, CERN experiments have stored more data on tape media (from detectors, simulation and analysis), than in the entire history of the Laboratory (CERN just celebrated its 70th anniversary in 2024). This amounts to over 500 Petabytes of data from all experiments in the last 30 months. CERN HL-LHC project (commonly known as HiLumi) projects a 10-fold increase[1, 2] in the amount of data volume that will need to be collected by the end of the decade. Such data explosion is not unique to CERN; others major scientific experiments will be ready by the end of the decade, further exacerbating the challenge posed by managing vast amounts of data. SKA[1][3], KM3NET[2][4] and DUNE[3][5] are a few examples of such data intensive experiments.

The combined data production of some of these experiments: SKA (700 PB/year[3]), HL-LHC (over 500 PB/year[1, 2]) and DUNE (over 100 PB/year[5]) is projected to hit the 1 Exabyte/year mark by the next decade.

*e-mail: hugo.gonzalez.labrador@cern.ch
**e-mail: martin.barisits@cern.ch

[1]The Square Kilometre Array (SKA) is an intergovernmental international future radio telescope project
[2]KM3NeT is a research infrastructure housing a future next generation neutrino telescope.
[3]The Deep Underground Neutrino Experiment is a future neutrino experiment

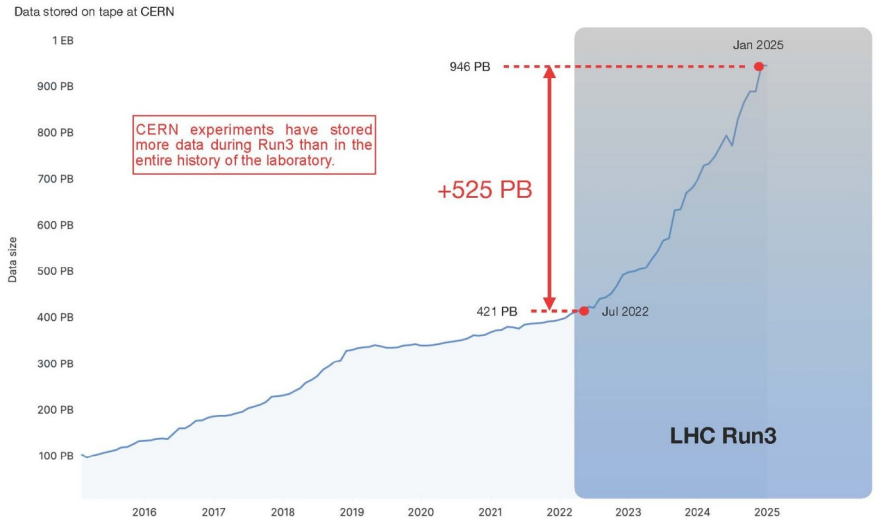

Figure 1. Historical representation of data volume (in Petabytes) stored on tape at CERN

2 Bound by the speed of light

WLCG[6] stands for the Worldwide LHC Computing Grid. WLCG combines about 1.4 million computer cores and 1.5 exabytes of storage from over 170 data centers in 42 countries. This distributed computing infrastructure allows more than 12,000 physicists around the world to connect near-real-time to the LHC data and the computing power to process it. Figure 2 shows the onion representation of the WLCG federation.

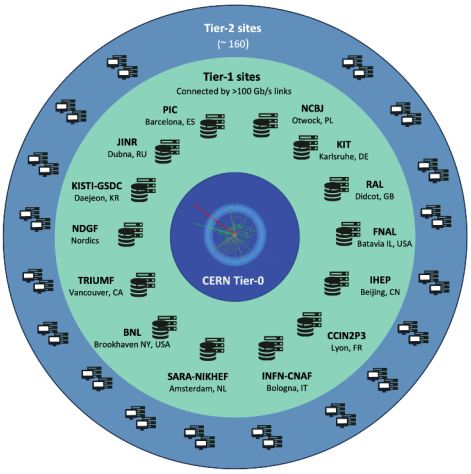

Figure 2. Logical representation of WLCG sites

These sites are usually interconnected by fiber links, usually part of the LHCONE[7] routing network. For example, a physicist geographically located at CERN can access the re-

sources with very little latency (CERN's inside data centre latency is around 0.25ms). However, if physicists in the US would like to access the resources online, the latency to CERN could be as high as 125ms (400 times slower) because of thousands of miles between US and CERN. Therefore, such latency cannot be eliminated completely simply because we are bound by the speed of light. There are mechanisms to hide latency for WAN data access, such as having local caching mechanisms like XCache[8] or using WAN-optimized data access protocols, such as XROOTD[9]. Despite these ongoing efforts, data transfers within WLCG happen 24/7, with rates peaking at 260 GB/s[10, 11].

Some common transfer methods in computing such as Secure Copy Protocol (scp) and FTP[4] may been used in the past to transfer small quantities of data, however, given all the potential network issues (i.e, router failures, ISP (Internet Service Provider) blocking connections from certain countries, data corrupted in transit, ...), a robust alternative is needed to manage the data distribution at scale.

3 Rucio

Rucio[12] is a mature and modular scientific data management platform designed to catalog and handle the distribution of vast amounts of data across multiple data centers. Rucio uses a global single namespace to store associated metadata (such as calibration data, data campaign, etc ...). Rucio is location-aware and proficiently manages data within heterogeneous distributed environments (HPCs[5], data lakes, archiving facilities, ...), encompassing creation, location tracking, transfer, deletion, annotation, and access control. Besides that, Rucio orchestrates data flows through the implementation of both low-level and high-level policies, which ultimately trigger the copy of data to multiple of these locations, each operating under distinct administrative domains, by relying on a transfer tool to orchestrate and monitoring the successful transfer of the data. The two most common transfer tools that Rucio utilizes are FTS[13] and globus[14].

Rucio is a free and open-source software solution licensed under Apache v2.0. Rucio fosters an open, community-driven[15, 16] development process. Figure 3 shows that in the last 5 years, more than 200 people have contributed to the project. Furthermore, there is a stable contribution trend line, with over 10 monthly authors (more than 60 annually), ensuring the continuity of the project and further adoption by other institutions. The CERN IT department is one of the new contributors to the Rucio community. CERN IT contributes with software development person-power to the core Rucio project while prototyping a new class of service that may be adopted by the small and medium experiments at CERN, ensuring a coherent vertical integration with existing CERN core storage solutions (CTA, EOS, FTS).

3.1 Core Features and Advantages of Rucio

In a nutshell, Rucio answers the five Ws for data management: **Who** created the data and who moved the data? **What** data is cataloged? **When** was the data created, moved or deleted? **Where** is the data? **Why** do we have this data? The answers to these questions are provided by the two fundamental features of Rucio:

1. **Automation of Data Management Operations**: Rucio reduces the need for manual intervention by allowing users to specify desired data placement outcomes rather than

[4]FTP stands for File Transfer Protocol, a standard network protocol used to transfer computer files between a client and a server on a computer network.

[5]High Performance Computing

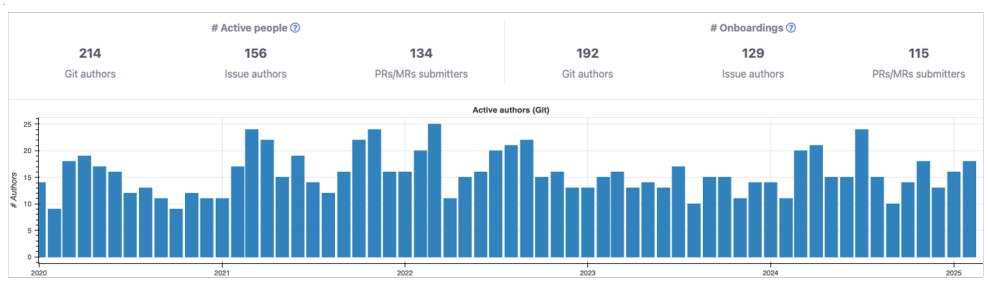

Figure 3. Number of contributors in the last 5 years and historical evolution over time.

explicit procedures. For example, users can request "Three copies of this dataset, distributed across multiple continents, with at least one copy on tape" or "One copy of this file anywhere, as long as it resides on a high-performance disk." This method of emphasizing the declarative model allows to have replication rules that governs the data distribution, including:

- Dynamic addition and removal of rules, with authorization mechanisms
- Rule evaluation engine optimizes data transfers and deletions
- Capability to lock data against deletion for specified durations
- Dynamic creation of cached replicas based on usage patterns
- Integration with workflow systems for automated rule generation

2. **Abstraction through a Scalable Metadata Catalog:** Rucio provides a scalable catalog that abstracts the underlying storage protocols (e.g., S3, WebDAV[17], XROOTD[9]) and WLCG storage technologies (e.g., dCache[18], EOS[19], CTA[20], Storm[21]), offering a unified data access layer. This indirection allows a data centre to switch the underlying storage technology without impacting the user community, allowing to introduce new technologies and re-tune the storage characteristics and protocols without having to impact the user workflows.

Figure 5 depicts the interaction of and end-user with Rucio and the shielded abstraction to the underlying storage infrastructure.

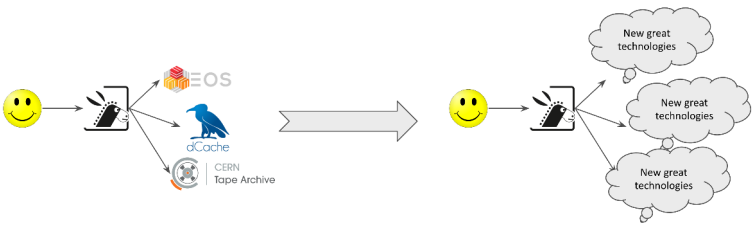

Figure 4. End users, Rucio and the underlying storage infrastructure.

4 Deployment Landscape: Rucio in Scientific Collaborations

Rucio has been adopted by over 20 scientific collaborations and has emerged as the *de-facto* standard for scientific data management in high-energy physics (HEP) and related disciplines. It is currently employed by CERN experiments such as ATLAS[22], CMS[23], and the Neutrino Platform (DUNE[24], ICARUS[25]), as well as other collaborations including Belle II[26, 27], LIGO/VIRGO[27], the Cherenkov Telescope Array Observatory (CTAO)[28], and the Vera Rubin Observatory[25]. Furthermore, Rucio is being integrated into projects such as the SKA[27], KM3NeT[4], IHEP[29], and XENONnT[27]. Collectively, all these Rucio sites manage over 2 Exabytes of data across its various deployments.

Figure 5. Non-exhaustive representation of communities evaluating and using Rucio.

5 Catalyst for Innovation

Rucio, as the platform sitting between the end-users and high-level applications and the storage systems, serves as an enabling technology in numerous scientific activities. Some of these include initiatives such as: the integration between Rucio and Sense[30], DaFab: Extending Rucio for Enhanced Earth Observation Data Management[31], data discovery, analysis, and reproducibility in Virtual Research Environments[32] or extending Rucio capabilities with OpenData[33]. An extended description of some of these activities is provided below.

5.1 DaFab

The DaFab project is an initiative aiming to enhance the exploitation of Copernicus[34] data through advanced AI and High-Performance Computing (HPC) technologies. Especially, DaFab seeks to improve the timeliness, accuracy, and accessibility of EO (Earth Observation) data. Rucio is a key component in the DaFab computing infrastructure[31] and complement the EO data store with metadata filtering. Figure 6 shows the usage of Rucio as connecting platform between the end-user and the EO feature extraction system.

The objective of the DaFab metadata developments in Rucio are to expand Rucio's metadata functionality to store, manage, and search heavily structured metadata, such as metadata in the STAC[35] format. While the first use-case of this development is the Earth Observation domain, the features are highly generic and can be exploited by any Rucio community.

5.2 Streamlining open data with Rucio

This initiative is a funded EU project under the OSCARS[6] umbrella. The growing need for interdisciplinary research, particularly in multi-messenger science, has highlighted the

[6]OSCARS is a macro-European project bringing together world-class European Research Infrastructures (RIs) in the ESFRI road-map to foster the uptake of Open Science in Europe.

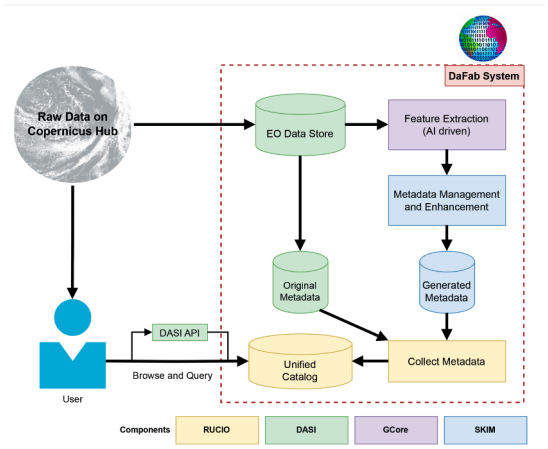

Figure 6. Rucio in the DaFab computing architecture.

importance of sharing scientific data across institutions performing complementary research (infrared, gravitational waves, etc.). However, institutions often rely on archaic or error-prone methods for data sharing. Moreover, given that Rucio does not recognize open data as a main data type, experiments need to copy this data to other systems to make them FAIR[7], incurring extra costs for the storage of the copy of the data. The main objective of this project is thus to introduce native support in Rucio to manage open data, where embargo and public access policies can be defined, so a copy of the data is not needed to make data open. Figure 7 highlights the workflow from data curation to data publication.

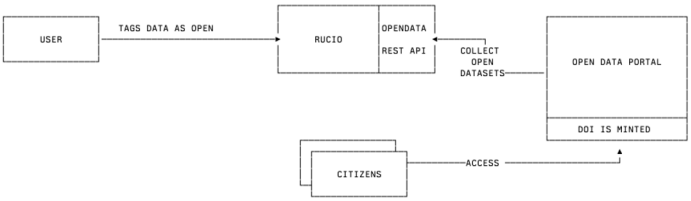

Figure 7. Rucio and open data portal integrations

6 Challenges and Future Directions

The Rucio project faces several challenges and opportunities for the near-future development:

- **Authentication Mechanisms:** The current authentication infrastructure relies on a combination of x509 certificates and WLCG tokens[36]. The proposed evolution is to use WLCG tokens, a derived token schema from the standard OIDC[8]. Current works involve optimizing the minting of these tokens and ensuring valid lifetimes for the execution of

[7]FAIR data is data which meets the principles of findability, accessibility, interoperability, and re-usability.
[8]OIDC is an identity authentication protocol that works with OAuth 2.0 to standardize the process for authenticating and authorizing users when they sign in to access digital services

downstream Rucio actions, such as data tranfers, which happen asynchronously and may be queued for hours if not days.

- **Integration with Workflow Management Systems:** For large-scale deployments, automating the transfer of data to the closest computing facility where users launch their jobs can significantly enhance efficiency. The DIRAC interware[37] is a promising candidate for this integration, given its proven track record with Belle II[26] and current evaluation at IHEP[29].

7 Conclusion

This article provides an update over previous bibliography[27, 38]. Rucio is an open, reliable, and efficient data management system that supports some of the existing and future world's largest scientific experiments. Rucio is maintained by a stable community that is continuously extending it to meet the evolving needs and requirements of the scientific community. A strong collaboration link is established between physics and other neighboring sciences, actively participating in the Rucio ecosystem, helping advancing the project to new horizons.

References

[1] A. Di Girolamo, Z. Marshal (ATLAS), Tech. rep., CERN, Geneva (2022), https://cds.cern.ch/record/2802918
[2] C.O. Software, Computing, Tech. rep., CERN, Geneva (2022), https://cds.cern.ch/record/2815292
[3] Ska data rates, https://indico.cern.ch/event/1349135/contributions/5832985/attachments/2833776/4951647/Breen%20-%20LHCOPN-LHCONE.pdf
[4] M. Bouwhuis, V. Cecchini, K. Graf, F.S. Greus, F. Schnabel, Toward full implementation of the km3net research infrastructure, EU Cordis platform (2023).
[5] T.D. collaboration, Dune offline computing conceptual design report (2022), 2210.15665, https://arxiv.org/abs/2210.15665
[6] I. Bird, F. Carminati, R. Mount, B. Panzer-Steindel, J. Harvey, I. Fisk, B. Kersevan, P. Clarke, M. Girone, P. Buncic et al., Tech. rep., CERN (2014)
[7] E. Martelli, S. Stancu, Lhcopn and lhcone: status and future evolution, in *Journal of Physics: Conference Series* (IOP Publishing, 2015), Vol. 664, p. 052025
[8] A. Hanushevsky, H. Ito, M. Lassnig, R. Popescu, A. De Silva, M. Simon, R. Gardner, V. Garonne, J. De Stefano, I. Vukotic et al., Xcache in the ATLAS distributed computing environment, in *EPJ Web of Conferences* (EDP Sciences, 2019), Vol. 214, p. 04008
[9] A. Dorigo, P. Elmer, F. Furano, A. Hanushevsky, Xrootd-a highly scalable architecture for data access, WSEAS Transactions on Computers **1**, 348 (2005).
[10] Wlcg website, https://home.web.cern.ch/science/computing/grid
[11] Wlcg transfers, https://monit-grafana.cern.ch/d/ZqU5ugjMz/fts-status-board?orgId=20&viewPanel=111
[12] M. Barisits, T. Beermann, F. Berghaus, B. Bockelman, J. Bogado, D. Cameron, D. Christidis, D. Ciangottini, G. Dimitrov, M. Elsing et al., Rucio: Scientific data management, Computing and Software for Big Science **3**, 1 (2019).
[13] E. Karavakis, A. Manzi, M.A. Rios, O. Keeble, C.G. Cabot, M. Simon, M. Patrascoiu, A. Angelogiannopoulos, FTS improvements for LHC Run-3 and beyond, in *EPJ Web of Conferences* (EDP Sciences, 2020), Vol. 245, p. 04016

[14] I. Foster, C. Kesselman, The globus toolkit, The grid: blueprint for a new computing infrastructure pp. 259–278 (1999).
[15] Rucio github, https://github.com/rucio/
[16] Rucio open source community report, https://cauldron.io/project/8934
[17] E.J. Whitehead, M. Wiggins, Webdav: Ieft standard for collaborative authoring on the web, IEEE Internet Computing **2**, 34 (1998).
[18] G. Behrmann, P. Fuhrmann, M. Grønager, J. Kleist, A distributed storage system with dCache, in *Journal of Physics: Conference Series* (IOP Publishing, 2008), Vol. 119, p. 062014
[19] A.J. Peters, E.A. Sindrilaru, G. Adde, EOS as the present and future solution for data storage at CERN, in *Journal of Physics: Conference Series* (IOP Publishing, 2015), Vol. 664, p. 042042
[20] E. Cano, V. Bahyl, C. Caffy, G. Cancio, M. Davis, O. Keeble, V. Kotlyar, J. Leduc, S. Murray, CERN Tape Archive: a distributed, reliable and scalable scheduling system, in *EPJ Web of Conferences* (EDP Sciences, 2021), Vol. 251, p. 02037
[21] A. Ghiselli, L. Magnoni, R. Zappi, Storm: A manager for storage resource in grid, Nuovo Cimento. C **32**, 273 (2009).
[22] V. Garonne, R. Vigne, G. Stewart, M. Barisits, M. Lassnig, C. Serfon, L. Goossens, A. Nairz, A. Collaboration et al., Rucio–The next generation of large scale distributed system for ATLAS Data Management, in *Journal of Physics: Conference Series* (IOP Publishing, 2014), Vol. 513, p. 042021
[23] E. Vaandering, Transitioning CMS to Rucio Data Managment, in *EPJ Web of Conferences* (EDP Sciences, 2020), Vol. 245, p. 04033
[24] S.C. Timm, Tech. rep., Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States) (2021)
[25] B. White, D. Lee, Y. Guo, Tech. rep., Fermi National Accelerator Laboratory (FNAL), Batavia, IL (United States) (2022)
[26] C. Serfon, R. Mashinistov, J.S. De Stefano, M.H. Villanueva, H. Ito, Y. Kato, P. Laycock, H. Miyake, I. Ueda, Integration of rucio in belle ii, in *EPJ Web of Conferences* (EDP Sciences, 2021), Vol. 251, p. 02057
[27] M. Lassnig et al., Tech. rep., XENON. Technical report, ATL-SOFT-PROC-2020-017, CERN, Geneva (2020)
[28] L. Arrabito, J. Bregeon, P. Maeght, M. Sanguillon, A. Tsaregorodtsev, The Cherenkov Telescope Array production system proto-type for large-scale data processing and simulations, in *EPJ Web of Conferences* (EDP Sciences, 2021), Vol. 251, p. 02029
[29] X. Zhang, X. Zhang, H. Li, Y. Bi, H. Hu, H. Wang, Distributed Data Management System at IHEP, in *International Symposium on Grids and Clouds (ISGC)* (2023), Vol. 19
[30] F. Würthwein, J. Guiang, A. Arora, D. Davila, J. Graham, D. Mishin, T. Hutton, I. Sfiligoi, H. Newman, J. Balcas et al., Managed Network Services for Exascale Data Movement Across Large Global Scientific Collaborations, in *2022 4th Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)* (IEEE, 2022), pp. 16–19
[31] Harnessing the power of rucio for earth data observation, https://www.dafab-ai.eu/blog/2024/07/01/harnessing-the-power-of-rucio/
[32] E. Gazzarrini, E. Garcia, D. Gosein, A.V. Moya, A. Kounelis, X. Espinal, The virtual research environment: towards a comprehensive analysis platform, arXiv preprint arXiv:2305.10166 (2023).
[33] Streamlining opendata policies in rucio, https://oscars-project.eu/projects/streamlining-open-data-policies-rucio-data-management-platform

[34] S. Jutz, M.P. Milagro-Perez, Copernicus: the european earth observation programme, Revista de Teledetección pp. V–XI (2020).

[35] Spatiotemporal asset catalog (stac) specification, https://stacspec.org

[36] Dack, Thomas, Agostini, Federica, Basney, Jim, Cornwall, Linda, De Stefano Jr, John Steven, Dykstra, Dave, Giacomini, Francesco, Litmaath, Maarten, Miccoli, Roberta, Sallé, Mischa et al., Wlcg transition from x.509 to tokens. status, plans, and timeline, EPJ Web of Conf. **295**, 04054 (2024). 10.1051/epjconf/202429504054

[37] F. Stagni, A. Tsaregorodtsev, A. Sailer, C. Haen, The DIRAC interware: current, upcoming and planned capabilities and technologies, in *EPJ Web of Conferences* (EDP Sciences, 2020), Vol. 245, p. 03035

[38] T. Barisits, Martin et Beermann, D. Cameron, J.A. Clark, R. Di Maria, G.G. Fronzé, I. Johnson, M. Lassnig, C. Serfon, E.W. Vaandering, Experience with Rucio in the wider HEP community, in *EPJ Web of Conferences* (EDP Sciences, 2021), Vol. 251, p. 02030