

# Open Science, Open Data and Open Access infrastructure: interview with Tullio Basaglia



CERN [1] is the European Organization for Nuclear Research, one of the world's largest and most respected centres for scientific research in particle physics. Founded in 1954, the CERN laboratory in Geneva is one of Europe's first joint ventures and now has 23 member states.

The CERN Scientific Information Service (SIS) [2] is the name of the group which includes, in addition to the archive and the library, the sections dealing with INSPIRE digital library and Open Access (OA). The CERN library collections support the scientific research conducted by the CERN Laboratory. The Open Science (OS) section develops strategies for negotiating read and publish agreements with publishers and plans to develop training modules aimed at authors who wish to publish their articles and research data

in OA mode.

Tullio Basaglia has been a librarian at CERN's SIS since 1997 and has been the Chair of the library and bookshop at SIS since 2009. He is an international librarian and has been president of Association of International Librarians and Information Specialists of Geneva since 2010. Tullio Basaglia is now retired and we wanted to ask him to describe the library's impact on the OA, OS and Open Data research infrastructure:

Q1. How has the scholarly communication at CERN evolved over the past two decades?

We can identify three interrelated phenomena in the evolution of scholarly communication at CERN in the period 2004–2024.

## *The development and establishment of Open Access-related policies*

CERN's OA publishing policy [3] was published in 2014 and slightly updated twice in 2017 and 2021, without significant modification of the core principles of the original document. The Open Data policy for the Large Hadron Collider (LHC) experiments [4] was approved in 2020, after a fairly long period devoted to the development of the Open Data Portal [5], for educational and further analysis purposes. The policy itself resulted from intensive consultation with the LHC experiments' community. Finally, the OS policy [6], released in 2022, was the fruit of a lengthy process of consultation with a variety of protagonists of OS, namely, the communities involved with the development of Open Source Software, Open Hardware and Open Educational Resources.



---

### *The emergence of a new Open Access scholarly publishing paradigm*

Scholarly Consortium for OA Publishing in Particle Physics (SCOAP3) [7], began operations in 2014, after a long period where CERN collected pledges of potential institutional members, devised the governance of the consortium and launched negotiations with publishers, a collective effort spearheaded by CERN that gave birth to a worldwide partnership ensuring that more than 90% of the literature published yearly in high energy physics is made freely accessible to anyone.

### *The creation of two Open Access services for the storage and retrieval of scholarly literature*

Zenodo [8], active since 2013 and powered by an open-source CERN-developed software, Invenio [9], that provides an outlet for free submission and perpetual archival of collections of articles, software and data sets, as well as for community-created and managed collections. INSPIRE-HEP [10], an Invenio-based knowledge hub serving the global particle physics community that fully replaced SPIRES-HEP in 2012 and underwent a technical update in 2020, aiming among others at a stronger involvement of the users' community in the effort of bibliographic records curation:

- Q2. What role has the Open Access movement played in CERN's scholarly communication strategy, and how has the institution contributed to the promotion of Open Access principles and practices?

The OA movement saw CERN as one of its early members. The serial price crisis of the nineties accelerated CERN's decision to join the OA movement, a decision in line with the OA culture inspiring scholarly communication within the particle physics community. The creation of ArXiv [11], founded by a particle physicist to archive and exchange preprints in his research domain, even predated the launch of the World Wide Web and paved the way to a pioneering experiment in scholarly publishing, the launch of *Journal of High Energy Physics* [12] in 1997, a peer-reviewed, fully OA journal managed by members of the community, that CERN actively promoted right from its start, against the scepticism of the commercial publishers and of parts of the researchers' community:

- Q3. Can you discuss any specific initiatives or programs implemented by CERN's library to support Open Access and facilitate access to research outputs produced by CERN researchers and collaborators?

Two categories of initiatives can be mentioned in this regard: one concerns repositories and the other relates to the removal of barriers to scholarly literature.

CERN IT Department developed Invenio, an open-source software framework for large-scale digital repositories. This software powers CERN's institutional repository, CERN Document Server (CDS) [13], the library catalogue [14], INSPIRE [15] an OA digital library for the field of high energy physics and Zenodo, an open repository that allows researchers to deposit research papers, data sets, research software, reports and any other research-related digital artifacts. CDS and Zenodo are also operated by CERN, while INSPIRE is jointly managed by an international collaboration.

SCOAP3, the Sponsoring Consortium for OA Publishing in particle physics, where libraries, funding agencies and research centres are represented, has converted key journals in the field of high-energy physics to OA and supports OA publishing in these journals at no cost for authors:

- Q4. How has the CERN library adapted its services and resources to meet the changing needs and preferences of researchers in the digital age, particularly in terms of providing access to electronic journals, databases and other online resources?

The CERN library has moved to e-only subscriptions for most of journal titles as early as 2008 and has steadily increased the amount of e-books in its collections. Today, the ratio in the library catalogue is about two e-books to one paper book. All library services (loan requests, renewals, ILLs, acquisition/subscription proposals, digitization, etc.) are accessible online:

- Q5. What challenges has the CERN library encountered in managing and preserving digital scholarly content, and how has the institution addressed issues related to long-term access and archiving?

According to CERN rules, all articles produced using CERN research facilities should be submitted in electronic form to CERN's institutional repository, the CDS. The main challenge in ensuring full coverage of CERN's scientific output consists in the high degree of "decentralization" in the production of scholarly content, which sometimes results in research groups and collaborations failing to follow the guidelines.

As a result and in spite of the efforts invested in the establishment of procedures [16] – an internal policy document stipulates the preservation and archival of paper artifacts – in the past, not all scientific content was submitted to CDS. It should be noted that CERN does not yet have a general policy for the perpetual preservation and archival of e-documents. Finally, a centralized budget for retrospective and systematic digitization of paper documents would be needed:

- Q6. In what ways has the CERN library supported interdisciplinary collaboration and knowledge exchange among researchers working in diverse scientific domains, both within CERN and with external partners?

Zenodo is an open repository that is interdisciplinary by design. Submitters' communities are welcome to create their own collections of objects related to projects, institutions, domains or conferences.

CERN is a member of the EIROforum [17], a collaboration among some of the key European research laboratories (ESA, EMBL, European XFEL, ESO, ESRF, ILL and Eurofusion), aiming at facilitating discussions on issues of common interest and relevant to research and development. The SIS is represented in two EIROforum Working Groups dealing, respectively with library and OS-related topics:

- Q7. How has the development and diffusion of the Web, precisely from CERN, influenced academic communication within the particle physics community and how has the CERN library responded to these developments?

The Web provided a new infrastructure for the distribution of prepublications, a crucial vehicle of scholarly communication within the worldwide particle physics community. This open-source, community-developed infrastructure paved the way for the creation of OA repositories, where a community-driven model for peer review emerged.

A pioneering example of such a development is provided by ArXiv.

In an early phase, CDS tended to massively import metadata of particle physics preprints from ArXiv, aiming at an exhaustive coverage of high-energy physics literature worldwide. In a later phase, a differentiation of functions between INSPIRE and CDS was introduced, whereby INSPIRE aims at exhaustivity, while CDS focuses on CERN output:

- Q8. Can you discuss any collaborative projects or partnerships that the CERN library has been involved in to promote innovation and best practices in scholarly communication, both at CERN and globally?

On the impulse of the SIS, CERN has signed the CoARA agreement [18] as one of the early signatories. CERN is committed to support the reform of research assessment both promoting the establishment of a CERN-wide policy and by providing tools and services to facilitate the adoption of new practices.

The CERN-UNESCO school on digital libraries [19] is an example of a partnership aiming at training African university personnel to become familiar with the principles of OA and OS, and at providing them with practical solutions for setting up and running digital libraries:

- Q9. What strategies does the CERN library employ to promote research integrity and responsible conduct of research among its users, particularly in relation to issues such as plagiarism, authorship disputes and data sharing?

According to CERN's OS policy:

"CERN is committed to ensuring the integrity of research. In order to facilitate the reuse of its research products, CERN provides infrastructures to accommodate the scale and complexity of its research outputs."

The publication committees of CERN Departments and scientific collaborations are responsible to carry out investigations to ensure that the principles of research integrity are respected:

- Q10. Looking ahead, what do you see as the most pressing opportunities and challenges for the future of scholarly communication at CERN, and how do you envision the role of the library evolving to address these challenges?

In my view, the main challenge will consist in developing innovative measures of research impact and in establishing assessment policies inspired by these new methods.

In terms of opportunities, I would mention the move away from subscription model to a new paradigm to make research output openly accessible, whereby libraries would pay fees for editorial services, commensurate with their scientific output.

A new opportunity has opened up with the SCOAP3 for Books initiative, which converted 113 books (monographs and textbooks) to OA [20]. A committee of experts is responsible for the selection of frontlist titles, to be made available in OA mode.

**Tullio Basaglia**

*CERN, Geneva, Switzerland, and*

**Anna Maria Tammaro**

*Department of Information Engineering, University of Parma, Parma, Italy*

## Notes

1. CERN website: <https://home.cern/>
2. CERN Scientific Information Service web site: <https://library.cern/>
3. Open Access publishing policy <https://cds.cern.ch/record/1955574>

4. Open Data policy for LHC <https://cds.cern.ch/record/2745133>
5. Open Data Portal <https://opendata.cern.ch/>
6. Open Science policy <https://cds.cern.ch/record/2835057>
7. SCOAP3 <https://scoap3.org>
8. Zenodo <https://zenodo.org>
9. Invenio <https://inveniosoftware.org/>
10. INSPIRE-HEP <https://inspirehep.net/>
11. ArXiv <https://arxiv.org>
12. Journal of High Energy Physics <https://jhep.sissa.it/>
13. CERN Document Server (CDS) <https://cds.cern.ch>
14. Library Catalogue <https://catalogue.library.cern>
15. INSPIRE <https://inspirehep.net>
16. <https://cds.cern.ch/record/1202773>
17. EIROforum [www.eiroforum.org/](http://www.eiroforum.org/)
18. COARA <https://openscience.cern/research-assessment>
19. CERN UNESCO school <https://school-digilib.web.cern.ch/home>
20. <https://scoap3.org/scoap3-books/>