

Meeting Report

SciELO 25: A reflection on technologies enabling accessibility

Erik Schultes^{a,b,*}

^a*Senior Researcher, Leiden Academic Center for Drug Research, Leiden, The Netherlands*

^b*FAIR Implementation Lead, GO FAIR Foundation, Leiden, The Netherlands*

Abstract. In September 2023, prominent voices in Open Science and FAIR Data gathered in São Paulo, Brazil to celebrate the 25th anniversary of the Scientific Electronic Library Online (SciELO). Presentations and workshops were held that reviewed progress in the previous decades and scoped priorities for the time to come. The celebration was organised around the ideas of Open Science with Impact, Diversity, Equity, Inclusion and Accessibility (IDEIA). Here, I give a brief report on discussions around Accessibility.

Keywords: Open science, FAIR guiding principles, FAIR digital objects, Nanopublications, Equitable access

From September 25–29, 2024, the Scientific Electronic Library Online (SciELO) celebrated 25 years of service. From the earliest days of the Internet, SciELO envisioned a library covering Brazilian scientific journals that was made “electronic”, and as such could minimize the cost of scholarly knowledge dissemination (anticipating Open Science) and providing real-time services such as machine-readable metadata and automatic language translation services. Over the years, the SciELO project has helped to develop a common methodology for the preparation, storage, dissemination and evaluation of scientific literature in electronic format across Latin America. The 25th anniversary event adopted a theme of “Open Science with Impact, Diversity, Equity, Inclusion and Accessibility (IDEIA)”. Highlights of the workshop covering each of the thematic areas can be viewed on line [1].

I found myself among the panel devoted to Interoperability and Accessibility [2]. This included ideas around Open Access, but also to the practical accessibility of scholarly content crossing language boundaries, for people who may be visually impaired, for people with dyslexia or access specifically for machines (such as the emerging class of Large Language Models) that consume scientific information. There was also discussion around the imperative that some classes of data, no matter how valuable they may be for research purposes will nonetheless be restricted and will never be openly accessible (examples include patient medical records). They may be, however, accessible to researchers under well-defined conditions that adhere, for example, to the laws and regulations operating within defined jurisdictions.

Following along the SciELO 25 program, I was inspired by the unfolding discussions recounting a quarter century of historical developments where key technologies have opened the scholarly mind to more

*E-mail: eriks@gofair.foundation. ORCID: <https://orcid.org/0000-0001-8888-635X>.

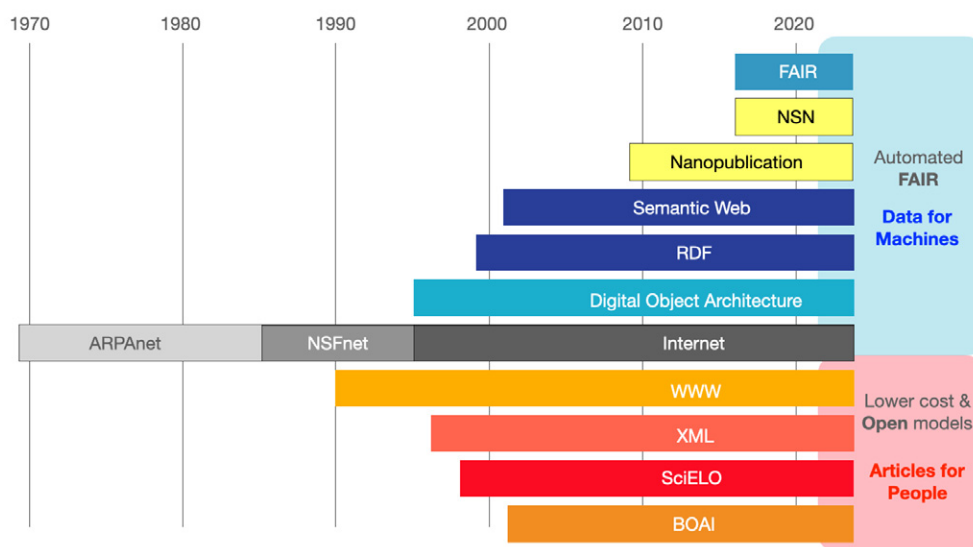


Fig. 1. A historical timeline depicting the some of the technology trends most relevant to the 25 year history of SciELO. The center bar (grey) represents the three stages of the development of the Internet (technology allowing for network interoperability). The Internet, driven the core technology TCP/IP had become by 1995 a mature infrastructure supporting digital information exchange. The bars below the Internet represent the emergence of enabling technologies (the World Wide Web and Extensible Markup Language) and social movements (the Scientific Electronic Library Online and the Budapest Open Access Initiative) that promoted open access to traditional articles. The bars above the Internet represent technologies (described in the text) for enabling semantically rich, machine-actionable operations supporting the automated Findability, Accessibility, Interoperation and ultimate Reuse of data. Whereas the lower bars represent efforts to make narrative articles more broadly accessible to human scholars, the upper bars represent efforts to make data more broadly accessible to machine assistants.

equitable modes of access in academic publishing. In preparing my own presentation [3], I decided to make these historical developments more explicit, and show how they relate to the mission of SciELO (Fig. 1). This overview began with the Internet and World Wide Web, but included the most recent technological developments around the FAIR Principles [4] and machine-actionability that they engender.

Now familiar and ubiquitous, 25 years ago (in 1998) the Internet was still an emerging technology and “Google” had yet to become a synonym of search. Universal and instantaneous file sharing and electronic mail (email) were still novelties, although academic scholars were among the early users. Like electricity, the Internet today is now taken for granted, available at reasonable costs, and powering a diverse array of digital applications.

The technology behind the Internet has its origins in a series of federally funded research projects predating SciELO, beginning in the 1960’s. Version one of the legendary ARPAnet came online in 1969 and over the course of two decades, helped to develop the core technology that drives the Internet today (Transmission Control Protocol/Internet Protocol or TCP/IP). Beginning in 1985, the National Science Foundation launched a project using TCP/IP to connect computer networks that had emerged independently among US academic centres. The resulting “NSFnet” was so highly successful that even industry joined, not necessarily for research purposes, but for commerce. By 1992, the private sector participants had come to outweigh the public sector users and it was decided to privatise the publicly funded NSFnet. The aim was to transition the network from a government funded project to a self-sustaining infrastructure

like telephone networks or electrical grids. This government-to-industry handover (completed by 1995) is typically taken as the “birth” of the Internet that we know today [5] (Fig. 1, center bar).

With this core infrastructure coming into place, Tim Berners-Lee had, by 1990, built the main components of the World Wide Web (WWW) [6]. The WWW made it easy to link to documents (i.e., web pages) and reflected the spirit of the Internet in its minimal requirements and its bottom-up, voluntary adoption. Another technology development having relevance for SciELO’s mission is the emergence (1996) of the Extensible Markup Language (XML) [7]. XML became a facilitating technology for the standardisation of metadata describing research articles, and a workhorse for distributing that metadata on the WWW. Taken together, the Internet, the WWW, and XML all contributed to the lowering costs associated with academic publishing, and created new possibilities for open models that could make research articles more available for scholarly exchange and academic collaboration (Fig. 1, lower bars). It is a testament to the vision of Abel Packer and the co-founders of SciELO that they should have recognized, only a few years into the burgeoning Internet, that enormous transformative potential and were bold enough to commit to action. By late 2001, “a small but lively meeting convened” in Budapest by the Open Society Institute to codify the vision, and launch the Budapest Open Access Initiative (BOAI). Participants collaborated on strategies and actions that would accelerate the transition making research articles in all academic fields freely available to readers on the Internet [8].

As SciELO and the BOAI were signaling the emergence of the open access movement, the architects of the Internet had begun to shift their attention towards technologies that would support higher-level “networking” of data and data operations. The interest was motivated by the possibility that both the complexity and the volume of information on the Internet would exceed human ability to track and control. It was anticipated that as more information became available on the Internet, humans would need the assistance of automated, computational services that would find, access, combine and reuse that information. To put it more strongly, without assistance of computers, humans would soon drown in their own data overload.

As early as 1995 (the same year that the Internet was born) Robert Kahn (co-inventor of TCP/IP) proposed a framework for “distributed digital object services” [9]. In this Distributed Object Architecture, services would operate on digital objects (like, for example, articles or datasets) using standardised information structures that make use of persistent identifiers and minimal metadata descriptions. The vision, like the Internet, was grand: an autonomous technology that would interconnect decentralised data and services into a single virtual database. The digital object would be for data interoperation what TCP/IP is for network interoperation.

While work on the technical specifications and early prototypes of the Distributed Object Architecture continued, the WWW had independently inspired a vision to make data more machine-readable. The so-called Semantic Web would extend WWW resources to interconnect data (for example, the content that appears within in web pages). This, along with corresponding query languages, would allow the machine to not only find, but to process knowledge in ways that humans would find meaningful and useful. An important extension towards this aim was the development of the Resource Description Framework (RDF). RDF expresses unitary thoughts (relations) in subject-predicate-object combinations (triples) where the subject, the predicate, and the object are “resources” on the WWW. By resources, we mean that they are “on line” and have persistent identifiers. The World Wide Web Consortium (W3C) adopted RDF as a recommendation in 1999 [10]. Curiously, although the Distributed Object Architecture and the Semantic Web shared many of the same objectives and design elements, they continued to develop mostly in isolation.

By 2008, Jan Velterop and Barend Mons recognised in actual practice the information overload anticipated by Kahn in 1995. Keeping track of increasingly large datasets via narrative articles was becoming increasingly difficult for reasons that had technical (e.g., broken links) and equitability (e.g., pay walls) dimensions. Their solution was to decompose the traditionally large data collections (like articles linked to supplemental data) into smaller, more precisely managed, publishable entities [11]. In its most fundamental formulation, individual semantic assertions (data) would be traceable as stand-alone publications and would be as accessible on the WWW as webpages and other content had become. These minimal publishable units were called “nanopublications” reflecting their diminutive size. As nanopublications, scientific and other scholarly information would then be available for analytical purposes unencumbered by diverse, idiosyncratic, and complicated academic publishing environments. To do this, each semantic RDF triple (i.e., a single assertion about the world) would require provenance metadata (minimally author, time/date) to make them citable and to preserve in the scientific record the contributions of scholars, research performing organisations, funding agencies, etc. The provenance would also be presented in RDF making the nanopublication entirely machine-readable. By 2009 an early nanopublication schema had been designed [12]. By 2014, the first nanopublications representing real-world data were created [13]. By 2016, a collective of academic developers had built the first decentralised network running services supporting various aspects of nanopublication (authoring, publishing, storing, search, access) [14]. The Nanopublication Server Network (NSN) now contains more than 40 services, and millions of nanopublications covering a multitude of domains [15]. It should be noted that nanopublications and the NSN, as it exists today, is radically open (open access content run on open source services).

Although the Digital Object Architecture, Semantic Web and Nanopublications have realised many of the visions for machine actionability of data, like the open movement, their uptake by academic publishers has been cautious and slow. In 2014 the data overload problem was addressed directly by a broad collection of stakeholders in a Lorentz Workshop [16]. Participants included experts in all the technologies described above and after two years of deliberations published the modern form of the FAIR Guiding Principles [4]. Together, the 15 one-liners of the FAIR Principles provided a proscription for operations necessary to make data **F**indable, **A**ccessible under well defined conditions, automatically **I**nteroperable and **R**eusable. In short, to automate as much as possible F, A, I and R. The FAIR Principles are notable in that they address many of the aspects important to SciELO such as helping to provide the larger context for accessibility (i.e., well-defined conditions be they open or restricted) and emphasizing the importance of machine-actionability that would permit high fidelity format conversion making scholarly content accessible to a broader range of human and machine scholars [17,18] (Fig. 1, upper bars).

SciELO has been a leading organization in the last quarter century of the open access and open science movements. In parallel to this are the long-term trends, seemingly inevitable, leading towards machine-actionability. With many of the leaders in open science and FAIR present in São Paulo, we looked forward to the next 25 years of SciELO helping to make access to knowledge more equitable and thus maximizing the collaborations of scholars, and to help humans and machines work together in the preservation and extension of the scholarly record.

References

- [1] SciELO 25, Open Science with IDEIA, São Paulo, September 25–29, 2023, <https://25.scielo.org/>. Internet Archive link: <https://web.archive.org/web/20240527055836>.
- [2] Standards, methods, techniques, practices. Interoperability and Accessibility. Video: <https://www.youtube.com/watch?v=yDJFH2aLh6Q&t=4714s>.

- [3] E. Schultes, "Standards, methods, techniques, practices: Interoperability and Accessibility - Academic publishing using FAIR Digital Objects", <https://doi.org/10.48331/scielo.XUI6Z9>, SciELO, V1.
- [4] M.D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data* **3** (2016), 160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18), Erratum in: *Sci Data* **6**(1) (2019), 6. PMID: 26978244; PMCID: PMC4792175.
- [5] E. Schultes, *FAIR Digital Objects for Academic Publishers*, 2024, pp. 15–21. <https://content.iospress.com/articles/information-services-and-use/isu230227>.
- [6] History of the Web, World Wide Web Foundation, <https://webfoundation.org/about/vision/history-of-the-web/>.
- [7] Brief History of XML, Web Reference, <https://webreference.com/xml/#>.
- [8] Background on the Budapest Open Access Initiative, <https://www.budapestopenaccessinitiative.org/faq/>.
- [9] R. Kahn and R. Wilensky, A framework for distributed digital object services, *International Journal on Digital Libraries* **6**(2) (2006), 115–123. doi:[10.1007/s00799-005-0128-x](https://doi.org/10.1007/s00799-005-0128-x).
- [10] W3C Issues Recommendation for Resource Description Framework (RDF), <https://www.w3.org/press-releases/1999/rdf/>.
- [11] B. Mons and J. Velterop, Nano-Publication in the e-science era, *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*, 523, 2009, https://www.researchgate.net/publication/228675568_Nano-Publication_in_the_e-science_era/citation/download.
- [12] P. Groth, A. Gibson and J. Velterop, *The Anatomy of a Nanopublication*, 2010, pp. 51–56. <https://content.iospress.com/articles/information-services-and-use/isu613>.
- [13] M. Lizio, J. Harshbarger, H. Shimoji et al., Gateways to the FANTOM5 promoter level mammalian expression atlas, *Genome Biol* **16** (2015), 22. doi:[10.1186/s13059-014-0560-6](https://doi.org/10.1186/s13059-014-0560-6).
- [14] T. Kuhn, C. Chichester, M. Krauthammer, N. Queralt-Rosinach, R. Verborgh, G. Giannakopoulos et al., Decentralized provenance-aware publishing with nanopublications, *PeerJ Computer Science* **2** (2016), e78. doi:[10.7717/peerj-cs.78](https://doi.org/10.7717/peerj-cs.78).
- [15] T. Kuhn, R. Taelman, V. Emonet, H. Antonatos, S. Soiland-Reyes and M. Dumontier, Semantic micro-contributions with decentralized nanopublication services, *PeerJ Comput Sci* **7** (2021), e387. doi:[10.7717/peerj-cs.387](https://doi.org/10.7717/peerj-cs.387), PMID: 33817033; PMCID: PMC7959648.
- [16] Lorentz Center Workshop, 22–26 January 2024, The Road to FAIR and Equitable Science, <https://www.lorentzcenter.nl/the-road-to-fair-and-equitable-science.html>.
- [17] K. Bazargan, "Standards, methods, techniques, practices: Interoperability and Accessibility - XML: Why it's the most important format", <https://doi.org/10.48331/scielo.DDVS6A>, SciELO, V1.
- [18] L.S.A. Cabral, "Padrões, métodos, técnicas, práticas: Interoperabilidade e Acessibilidade - Acessibilidade atitudinal na produção e difusão do conhecimento científico", <https://doi.org/10.48331/scielo.CMBXL6>, SciELO, V1.