Information Services & Use 44 (2024) 15–21
DOI 10.3233/ISU-230227
IOS Press

Meeting Report

FAIR digital objects for academic publishers

Erik Schultes[a,b,*]
[a]*GO FAIR Foundation, Leiden, The Netherlands*
[b]*Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands*

Abstract. For 200 years, collective scholarly knowledge was advanced by reporting new findings in the form of narrative text that is rich in logic, pinned to data, hedged with cautious nuance, and yielding novel claims. Authors' narratives evolved over the years into the now familiar academic research article, whose form has radiated into thousands of specialized intellectual niches (i.e., journal titles). In the last decades the corpus of collective scholarly knowledge (both narrative text and published data) has come to exceed human comprehension and challenges the ability of researchers, even those working in narrowly defined disciplines, to keep up. As a response, a wide range of abstracting and indexing services emerged and were among the first to push toward "electronic" publishing. By now, articles are routinely made available in digital repositories, but still the content and the form remain bound to human readers while the powerful information processing capabilities of machines, which ought to assist the researcher, are marginalized to the mundane calculation of impact scores. Today, the long-form narrative and the lack of standards in the academic publishing industry make the bulk of this information notoriously difficult for the machine process and reuse in scientific applications. As such, the classical research article has become increasingly untenable as a meaningful unit of intellectual progress. Since 2016, the FAIR Principles have provided guidance on how to compose data, including information contained in narrative text, to make them machine actionable. Recent developments in FAIR Digital Objects are now being exploited in academic publishing to expose FAIR information at the source, avoiding cumbersome text mining, and making exact meaning available to machine assistants supporting human scholars. Here I describe some of these long-term developments and how they relate to a new Sage/IOS Press journal called FAIR Connect. These trends signal an inevitable movement towards the FAIRifcation of scholarship, and hint at key issues that will impact the practice and business of academic publishing.

Keywords: FAIR Digital Objects, FAIR Guiding Principles, APIN, History of the Internet, Nanopublication, Data Stewardship, FAIR Connect

1. The long awaited FAIR digital object

In a world as digitally connected as ours, it is sometimes difficult to remember that computers were not always able to communicate with each other. In fact, this feature that seems so obvious now and that we take as a given, was at one point an engineering grand challenge. The solution to this problem, recounted below, is part of long-term trends toward information exchange and data interoperability. It is instructive to revisit this "ancient history" as it holds many parallels for the immediate future of academic publishing and scholarship in general.

In 2019, the former program director of the NSFnet, George Strawn, reminded us of the 50th anniversary of the Internet [1]. By this, Strawn meant that version one of the ARPAnet had begun operation in 1969.

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

Only four years later, the core technology that makes "interoperable networks" possible was invented by Robert Kahn and Vint Cerf. The engineering challenge at that time was to somehow interconnect early computer networks which had little standardisation in software or hardware. How could numerous networks be connected without having to create all pair-wise adaptors, or forcing every network to rebuild from scratch according to a single standard? Although it is today a problem hardly worth thinking about, it had by then become a serious research program in the Defense Advanced Research Projects Agency and culminated with version two of the ARPAnet (1983–1989). The brilliant solution was the Transmission Control Protocol/Internet Protocol (TCP/IP), a common and minimal standard that all networks could connect to with relatively minor modifications. In this way, network interoperability could be established among a diverse assemblage of home-grown technologies. Very different implementations (e.g., networks made of copper wires versus networks made of microwaves) could, with TCP/IP, seamlessly interconnect, behaving as a single, larger, virtual network. The approach proved to be economical and technically scalable (reaching hundreds of nodes) with benefits that would come to surprise even the engineers who built it.

For the next 10 years (1985–1995) Strawn had various responsibilities for the NSFnet, a federally funded project to extend the TCP/IP network to the computer networks at US Colleges and Universities and linking them to emerging supercomputer centers (constituting by then, already thousands of nodes). Curiously, Strawn discovered in 1991 that this US government funded academic research network had apparently become so useful that it was dominated by private industry members. It was then agreed in 1992 that the NSFnet, top-heavy with commerce, would better be managed by the private sector (accomplished in 1995), giving birth to what we now know as the Internet (soon thereafter reaching millions, then billions of nodes). Today TCP/IP continues to route our daily, and now global, data exchanges, from web browsing to music streaming, to social medial and emails to video calls. In 2004, Kahn and Cerf would win the highest honor in computer science, the Turing Award for their accomplishments [2].

Remarkably, in 1995, the same year that the commercial Internet was born, Kahn along with Robert Wilensky articulated another seminal vision: *A framework for distributed digital object services* [3]. Even at this early stage, Kahn and Wilensky were anticipating a time when the amount of information on the Internet would become so large, and so complex, that it would escape human ability to manage it. It was feared that more and more, data and their meaningful relations would be lost in plain sight (recall, at this time the now widely used Google Search Engine did not yet exist). This impending scalability problem could only be solved if computer systems could be enlisted to help humans in managing the information becoming available on the Internet.

Kahn and Wilensky proposed a new infrastructure that would be "open in its architecture and which supports a large and extensible class of distributed digital information services". These services would act on the "basic entities to be found in such a system, in which information in the form of digital objects is stored, accessed, disseminated, and managed". In essence, Kahn and Wilensky were proposing for information, what Kahn and Cerf had accomplished for networks: a technology solution that would interconnect, in a decentralized manner, data and services into a seamless virtual database. The digital object would be for data interoperation, what TCP/IP had become for network interoperation.

The vision behind this Digital Object Architecture percolated among technical experts for decades [4]. During that time, retrieval and access solutions for specialised problems permitted the discussion around the Digital Object Architecture to remain exploratory and theoretical. However, by the early 21st Century, the information overload anticipated by Kahn and Wilensky had begun to manifest and this set into motion a renewed interest in how data might be more automatically interconnected.

As early as 2010, biologists already feeling the weight of data overload, and convinced of the value of the semantic web and linked data, proposed a minimal schema (and its representation in the machine-readable Resource Description Framework) to make individual subject-predicate-object combinations (i.e., semantic assertions) become stand-alone publications, complete with provenance and Globally Unique, Persistent and Resolvable Identifiers (GUPRIs). Biological data such as gene-disease associations or protein-protein interactions, increasingly produced in large volumes in automated laboratories, could thus be more effectively captured, exchanged, and reused. In turn, authors could receive fine-grained credit for individual datum, and ambiguities that plague free-text and ordinary spreadsheet formats could be altogether avoided. This approach was named *nanopublication*, reflecting the idea that what is to be published are single, meaningful, atomic assertions about the world [5,6]. In the following years, several early attempts to publish complex biomedical knowledge as large collections of nanopublications had been attempted [7,8]. By 2016, a dedicated group of researchers created a decentralized network of services that could exploit the nanopublication format [9,10]. These included services to help people compose and publish nanopublications by providing public authentication keys, data and time stamps, licenses (by default open) and sophisticated search capabilities. Additional services were created to for minting blockchain-like "Trusty URIs" [11] where hash functions applied to the assertion and provenance information ensure unforgeable and uncorruptible GUPRIs for all nanopublications. The voluntary, bottom up Nanopublication Server Network demonstrated that researchers could "publish, retrieve, verify, and recombine datasets of nanopublications in a reliable and trustworthy manner" Its creators even proposed, independently of the ongoing discussions around digital objects but entirely consistent with them, "that this architecture could be used as a low-level data publication layer to serve the Semantic Web in general".

In another example where data overload elicited renewed interest in machine-actionability, a Lorentz Center workshop was held in Leiden, in 2014, entitled *Jointly Designing a Data FAIRport* [12]. The meeting was attended by a diverse cross section of data scientists, publishers, and funders from both the public and private sectors. Each came to the workshop looking to better characterise the problem of data overload and to scope possible solutions. Among the participants was George Strawn. Two years after the initial workshop, and after an extended period of community consultation, a commentary was published in *Scientific Data* that proposed a more focused and systematic approach to machine-actionably of data [13]. The vision was that of an entirely new infrastructure placing a "specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals". An innocuous figure item, "Box 2" annunciated the FAIR Guiding Principles - 15 "one liners" that captured the computational behaviors thought to be essential for automated Findability, Accessibility, Interoperability and Reuse of data [14].

The commentary and the FAIR Principles were immediately well received and embraced by the international stakeholder community, especially among policy makers, funders, and publishers. Currently the commentary is cited on average 7 times per day and according to Google Scholar, has by now accumulated over 11,000 citations.

Starting in 2020, the first discussions arose between communities interested in translating the FAIR Principles into practical implementations. This included GO FAIR [15] and communities that had ongoing discussions around digital objects supported primarily in the Research Data Alliance [16,17]. It became clear that the progress being made in FAIR seemed to have something critical to offer to the ongoing discussions around digital objects [18–20]. By October 2022, the First International Conference on FAIR Digital Objects was held in Leiden with keynotes provided by both Robert Kahn and George Strawn [21,22].

In its current form a FAIR Digital Object (FDO), is a self-describing, machine-actionable unit of (digital) information. It tells machine agents "What it is, what can be done with it, and what users are allowed to do with it". In keeping with the FAIR Principles, all the various components of the FDO have GUPRIs which makes them visible to machines, the first step towards their automated interpretation. The essence of the FDO is a minimal (and hopefully soon standardized) metadata record, with two essential elements: a type of description of the object and the object location. Each type of FDO has a metadata schema and vocabulary appropriate for it. FDOs can be thought of as minimal, standardized, machine-readable metadata tags that can be used to describe any resource (including non-digital entities such as, for example, butterfly specimens in a museum). With a minimal system of standardised self-description, it becomes possible to then build services that can operationalize the F, A, I and R functions, much as Kahn and Wilensky envisioned in 1995. The present grand challenge in the FDO community is to define the technical specifications around the minimal standards and build performant services that run FDOs [23].

Curiously however, in the First International Conference on FAIR Digital Objects mentioned above, collaborative work was reported at the GO FAIR Foundation that suggested a nanopublication assertion, when constructed in a particular manner, would be a close approximation to, if not be conformant with, the emerging FDO specification [24,25]. Specifically, nanopublications with assertions that explicitly give resource types and locations, along with the GUPRI services provided on the existing Nanopublication Server Network, already provide an example of the open architecture supporting: "a large and extensible class of distributed digital information services" envisioned in the Digital Object Architecture. Although simple and intriguing in hindsight, the idea that a certain class of nanopublications might be instances of FDOs came as a sudden surprise during very practical work using nanopublications to represent community-specific FAIR Implementation Profiles [26]. If the close correspondence of the nanopublication and FDO specifications is shown to hold, then we come to the remarkable conclusion that the community has already been "doing" FDOs for most of the last decade and that nanopublications are currently among the most technologically mature and widely used examples of FDOs.

2. APIN and FAIR Connect

Given the convergent technology trends between FDOs and nanopublications, the academic publishers broadly construed, be they society publishers, preprint platforms or commercial houses, were called to action by GO FAIR in 2020, to develop and promote best practices for "publishing for machines". The initiative was called the Academic Publishers Implementation Network (APIN), and it proposed the collective, pro-active formulation of protocols and standards that would publish all scientific research material, from data and code to the narrative text, as FDOs [27]. Of course, in lieu of finalised, and widely endorsed specifications for FDOs, APIN would proceed with its own version of a minimal, open standard. This bottom-up approach taken by APIN is inspired by developments in the early Internet where researchers and engineers had been encouraged to make progress by the mantra "rough consensus, running code".

By September 2022, IOS Press and the GO FAIR Foundation took the first concrete steps in response to the APIN call, and founded *FAIR Connect*, a journal devoted to the professionalisation of FAIR Data Stewardship [28]. FAIR Connect is an ISSN-registered, Sage/IOS Press journal as well as a web platform [29] that is managed by the FAIR Connect Foundation [30,31]. Although, in as far as FAIR Connect is devoted to the acceleration of FAIR in general, and thus has clear overlap with the FAIR ambitions of

APIN, the primary relevance of FAIR Connect here is not the journal content *per se*, but the journals mode of operation.

Realizing that data stewards around the world often confront common problems that are typically solved in isolation and without knowledge of previously created solutions, FAIR Connect aims to publish their creations as nanopublication-based FDOs. More specifically, in FAIR Connect, authors create "articles" by filling out nanopublication templates that help structure and focus the content to guarantee machine-readability [32–35]. Once this information is published as nanopublication-based FDOs, it can be automatically transcribed into human-readable prose in any human language.

FAIR Connect articles provide descriptions of so-called FAIR Supporting Resources (FSRs) which are explicitly defined in a FAIR ontology [36]. FAIR Connect FSR types are limited in number and include FAIR Data Policies; FAIR Data Stewardship Plan Templates; FAIR Implementation Profiles; FAIR Enabling Resources (having 12 sub-types); Data Steward Professional Profiles; FAIR Data Stewardship Events; FAIR Practices; FAIR Supporting Services; and formalized, short-form articles describing published FSRs. As FDOs, FSRs enjoy fine-grained search, explicit access points, zero ambiguity and automated interoperation. Hence, in FAIR Connect, Sage/IOS Press has accepted the primary challenge of APIN "to 'publish for machines' in alignment with the FAIR principles".

Although the FAIR Connect platform is intended to offer streamlined capability to publish and retrieve FSR nanopublications, the content is itself open and available to anyone via the Nanopublication Server Network. Indeed, as has always been the case, any organisation is welcomed to launch and run services on the open and decentralized Nanopublication Server Network. While others are encouraged to also build dedicated search capabilities fit to (their) purpose, FAIR Connect nonetheless aspires to diamond open access (free to publish, free to read) via the custodianship of the FAIR Connect Foundation.

The benefits of the FDO approach are immediately apparent. First, data stewards practicing their craft have a near-real time communication platform for exchanging resources that they create and manage. This will mitigate the rampant and needless "reinvention of the wheel". This could lead to significant cost savings and accelerated resource FAIRification worldwide. Second, the reuse of FSRs can be tracked in the emerging collection of domain-specific FAIR Implementation Profiles, themselves represented as nanopublications. This allows data stewards publishing at FAIR Connect to also receive recognition and credit for their contributions that would otherwise be invisible. Third, FSR nanopublications created in FAIR Connect are subject to a rapid and lightweight process of editorially controlled peer-review that is itself mediated by nanopublications [37]. Part of this review process includes the qualification and endorsement of FSRs according to explicit criteria that may be set forth by any third-party. These qualifications are issued, again as nanopublication-based FDOs and can be used in the search for FSRs having desired qualifications. These qualifications may come from recognised expert communities, funding organizations or publishers and will help drive FAIR convergence to high-quality and widely used FAIR implementations.

Taken together, FAIR Connect extends a long-term trend in information technology towards increasing interoperability of information, and leverages the recent developments in the FAIR Principles, FAIR Digital Objects and the Nanopublication Server Network to create a fully machine-actionable academic publication platform supporting the proceedings of professional data stewardship. As an APIN initiative, the FAIR Connect platform itself becomes an exemplar of lightweight and open tools that can help academic publishers to realistically transition to FDOs.

Acknowledgements

Many thanks to George Strawn, Louis Ter Meer, Einar Fredriksson, and Barbara Magagna for reading early drafts and generously offering corrections and thoughtful contributions to the final text.

References

[1] M. Hildreth and N. Meyers, Final Report: FAIR Hackathon Workshop for Mathematical and Physical Sciences Research Communities, Michael Hildreth, 2020. https://doi.org/10.7274/r0-rwpp-as13.
[2] Chronological listing of A. M. Turing Award winners. https://amturing.acm.org/byyear.cfm.
[3] 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.
[4] For example, see Cordra-managed digital objects at https://www.cordra.org/documentation/api/doip.html.
[5] B. Mons and J. Velterop, Nano-publication in the e-science era. Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), Washington, DC, 2009. https://www.semanticscholar.org/paper/Nano-Publication-in-the-e-science-era-Mons-Velterop/195c7924d58e729c1c66ace76a0bab06a856afd5.
[6] P. Groth, A. Gibson and J. Velterop, The anatomy of a nanopublication, *Information Services & Use* **30**: (2010), 51–56. doi:10.3233/isu-2010-0613.
[7] M. Lizio, J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin et al., Gateways to the FANTOM5 promoter level mammalian expression atlas, *Genome Biology* **16**(1) (2015). doi:10.1186/s13059-014-0560-6.
[8] Conference: Proceedings of 8th International Conference on Semantic Web Applications and Tools for Life Sciences. At: Cambridge, UK, Volume: 1546, January 2016. https://www.researchgate.net/publication/299388453_Finding_novel_associations_across_domains_using_linked_data_a_case_study_on_genetic_variants_disrupting_transcription_start_sites.
[9] T. Kuhn, R. Taelman, V. Emonet, H. Antonatos, S. Soiland-Reyes and M. Dumontier, Semantic micro-contributions with decentralized nanopublication services, *PeerJ Computer Science* **7** (2021), e387. doi:10.7717/peerj-cs.387.
[10] 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.
[11] T. Kuhn and M. Dumontier, Trusty URIs: Verifiable, immutable, and permanent digital artifacts for linked data. in: *The Semantic Web: Trends and Challenges. ESWC 2014*, V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab and A. Tordai (eds), Lecture Notes in Computer Science, Vol. 8465, Springer, Cham, 2014 doi:10.1007/978-3-319-07443-6_27.
[12] Lorentz Center Workshop, 13–16 January 2014, Jointly designing a data FAIRPORT. https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html.
[13] M. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., The FAIR guiding principles for scientific data management and stewardship, *Scientific Data* **3**(1) (2016). doi:10.1038/sdata.2016.18.
[14] Interpreting FAIR, GO FAIR Foundation webpage. https://www.gofair.foundation/interpretation.
[15] The FAIR Digital Objects (FDO) Forum Implementation Network. https://www.go-fair.org/implementation-networks/overview/fair-digital-objects-forum/.
[16] Group of European Data Experts in RDA. https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda.
[17] P. Wittenburg, Moving Forward on Data Infrastructure Technology Convergence: GEDE Workshop, Paris, 28–29 October 2019; Paris, France, 2019. Available online: https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects/Paris-FDO-workshop (accessed on 1 March 2020).
[18] E. Schultes and P. Wittenburg, FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure, 2019. https://doi.org/10.1007/978-3-030-23584-0_1.
[19] L.O. Bonino da Silva Santos, FAIR Digital Object Framework Documentation, 2021. https://fairdigitalobject framework.org/.
[20] FAIR DO Group. Joint Statement on FAIR Digital Object Framework, 2020. https://docs.google.com/document/d/11FmDxgncy-LynQqTlvxFProW-i5Il7JBFtp7ELyztlg/edit.
[21] First International Conference on FAIR Digital Objects, Edited by Tina Loo, Francisco Andres Rivera Quiroz, Wassim Deirieh, 2023. https://doi.org/10.3897/rio.coll.190.
[22] Leiden Declaration on FAIR Digital Objects, 2022. https://www.fdo2022.org/programme/leiden-declaration-fdo.
[23] FAIR Digital Objects Forum. 2020. https://fairdo.org.
[24] E.A. Schultes, B. Magagna, T. Kuhn, M. Suchánek, L.O. Bonino da Silva Santos and B. Mons, The Comparative Anatomy of Nanopublications and FAIR Digital Objects, *Research Ideas and Outcomes* **8** (2022), e94150. doi:10.3897/rio.8.e94150.

[25] E.A. Schultes, Are Nanopubs FDOs? Presentation given to the FDO Forum June 30, 2023. https://osf.io/qakjp.
[26] E. Schultes, B. Magagna, K.M. Hettne, R. Pergl, M. Suchánek and T. Kuhn, *Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence,* Lecture Notes in Computer Science, 2020, pp. 138–147. doi:10.1007/978-3-030-65847-2_13.
[27] J. Velterop and E. Schultes, *'An Academic Publishers' GO FAIR Implementation Network (APIN),* 2020, pp. 333–341. doi:10.3233/ISU-200102.
[28] IOS Press and the GO FAIR Foundation Announce a New Joint Initiative: FAIR Connect, September 2, 2022. https://www.iospress.com/news/ios-press-and-the-go-fair-foundation-announce-a-new-joint-initiative-fair-connect.
[29] FAIR Connect Hub. https://fairconnect.pro.
[30] E. Schultes, Presentation: When papers become FDOs, Academic Publishers Europe 2023, January 9–10, Berlin. https://osf.io/pjb7q.
[31] E. Schultes, Presentation: Academic publishing using FAIR Digital Objects, SciELO 25, September 25–29, 2023, Sao Paulo. https://osf.io/hg9we.
[32] C.-I. Bucur and T. Kuhn, 'Special Issue on Semantic Publishing with Formalization Papers'. 1 Jan. 2022: 1–9. https://content.iospress.com/journals/data-science/5/1.
[33] T. Kuhn, The Future of Science Publishing - Personal Views. IOS Press. IOS Press 35 Anniversary meeting. 2022. URL: https://www.iospress.com/ios-press-35.
[34] C. McNamara, Blog: The Future of Science Publishing – Focus on Nanopublications and Formalization Papers, March 28, 2022. https://labs.iospress.com/news-blog/future-science-publishing-focus-on-nanopublications.
[35] T. Kuhn, Blog: Welcome to the New Era of Scientific Publishing, August 31, 2022. https://labs.iospress.com/news-blog/welcome-new-era-scientific-publishing.
[36] E. Schultes, T. Kuhn and B. Magagna, FAIR Implementation Profile (FIP) Ontology. https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html.
[37] C. Bucur, T. Kuhn, D. Ceolin and J. van Ossenbruggen, Nanopublication-based semantic publishing and reviewing: a field study with formalization papers, *PeerJ Computer Science* **9** (2023), e1159. doi:10.7717/peerj-cs.1159.