FAIRing the radiation science commons

Balázs Madas[1,1], *Ulrike* Kulka[2], *Bernd* Grosche[2], *Mandy* Birschwilks[2], *Gayle* Woloschak[3], *Shin* Saigusa[4], *Soile* Tapio[5], *Michael* Gruenberger[6], *Paul* Schofield[6]

[1]Environmental Physics Department, MTA Centre for Energy Research
H-1121 Budapest, Konkoly-Thege Miklós út 29-33.
[2]Department Radiation Protection and Health Federal Office for Radiation Protection (BfS)
Ingolstaedter Landstrasse 1 85764 Oberschleissheim Germany
[3]Departments of Radiation Oncology, Radiology, and Cell and Molecular Biology, Robert E. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University 303 E. Chicago Ave., Ward 13-002, Chicago, IL 60611
[4]Institute of Radiation Epidemiology, Radiation Effects Association, 1-9-16 Maruishi Dai-ni Building 5F, Kajicho, Chiyoda-ku, Tokyo, 101-0044, Japan
[5]Helmholtz Zentrum München , Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) , Ingolstädter Landstr. 1 , 85764 Neuherberg , Germany
[6]Dept of Physiology Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

The FAIR principles

The primary data produced in the course of publicly-funded science represents a common asset for society as much as the analysed and interpreted results. Recent years have seen a unanimous agreement that such data and discoveries should be as accessible as possible by other scientists and the members of society in order to extract the maximum value from that investment. The concept of the science commons is well established and legal economic and social aspects of the commons are the subjects of intensive interest and examination [1]. Open access publication has been the focus of much attention and most investigators are now familiar with mandated open access publication mechanisms and requirements. However the public availability of primary data particularly requires the establishment of rules of governance and a sharing infrastructure. Through sharing the value of research commons are enhanced through the data cycle with iterated use and recontribution of new or modified data and materials creating an amplification effect that enhances the value and ultimately sustainability of the research commons.

The governance and infrastructure needed to support a research commons are two critical arms of successful scientific enterprise. The FAIR guidelines for data sharing were established in 2015 as a consequence of extensive consultation and apply to the community through consent and consensus [2]. The basic rules of FAIR are articulated in 15 areas summarised under the four letters of FAIR -Findable, Accessible, Interoperable, Reusable. The implications of the FAIR guidelines are that data should be discoverable and accessible by a human or by machine, that it should have sufficient metadata to be understandable and implementable and critically that the originator of the data should not be involved in the

[1] Corresponding author: balazs.madas@energia.mta.hu

decision as to whom it is made available. The FAIR principles do not preclude licensing or reasonable charges for access, so FAIR does not necessarily mean free or free of constraint over use, but that data should be accessible under reasonable conditions and in fact most of the data that concerns us lies within the pre-competitive space in any case. Major funding agencies such as the Wellcome Trust and the European Commission [3], and the NIH are now also trialing a FAIR data commons policy [4].

Data sharing in radiation science

Databases and repositories are the essential infrastructure for the research commons and require coordinated development and sustainable funding [5, 6]. In 2009 the STORE database was first set up under European Commission funding to encourage public data sharing and reuse in the domain of radiation biology. It was developed through successive grants and was opened to public use in 2014. STORE provides a sharing infrastructure for all kinds of data types and domains, ranging from epidemiology and human cohort data to 'omics and cytogenetics. Increasingly STORE is being used by large distributed projects to coordinate and archive primary and derivative data which is then used for support of publications. STORE mints persistent digital object identifiers and accession IDs which use a namespace formally registered with identifiers.org at the EBI. Similarly registered with the FAIR sharing initiative and r3data [7] STORE is a well recognised and accepted data repository. The database is physically located at the Bundesamt fuer Strahlenschutz in Munich and has the full security of a German Federal core data service. The BfS has undertaken to maintain the database indefinitely which means that data will be secure and accessible for the foreseeable future. Currently STORE contains around 3000 data objects across a wide range of data types; the number are increasing rapidly. The aim of STORE is to promote open access and reuse of data, as well as the archiving of at-risk or legacy data, thus promoting and enhancing the scientific commons. Consequently deposition and access to data are free to individual investigators and to funding agencies. STORE is available on http://www.storedb.org and users may sign up using their ORCID ID.

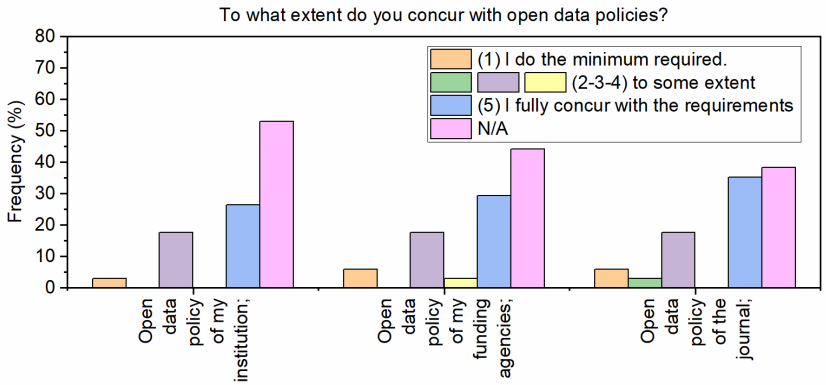

Figure 1. Concurrence of the respondent to open data policies.

Our experience of working with the radiation science community during the development of STORE has raised questions concerning the level of training in research data management, the culture of data sharing and various perceived disincentives for sharing [8]. This conflict with the development of policies of funding agencies and, increasingly, journals is a source of significant concern and we consider recommendations to address this problem through training, particularly of early-career scientists.

References

1. A. Mishra and T. Bubela, *OMICS: A Journal of Integrative Biology,* **18**, 254 (2014).
2 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, *et al.*, *Sci Data,* **3**, 160018, (2016).
3. H2020 Guidelines on FAIR Data Management:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
4. https://commonfund.nih.gov/commons
5. P. N. Schofield, J. Eppig, E. Huala, M. H. de Angelis, M. Harvey, D. Davidson*, et al.*, *Science,* 330, 592, (2010).
6. S. A. Sansone, P. Cruse, and M. Thorley, *Sci Data,* **5**, 180027, (2018).
7. https://www.re3data.org/about
8. B. G Madas, P. N. Schofield, Survey on data management in radiation protection research arXiv:1805.05463 [cs.CY] Accepted for publication in *Radiat Prot Dosim*.