

Impact case study (REF3)

Institution: University of Oxford		
Unit of Assessment: 12 – Engineering		
Title of case study: The FAIR Principles and FAIRsharing: Transforming Data Policies, Supporting Reproducibility and Reusability World-Wide		
Period when the underpinning research was undertaken: 2010 – 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Prof. Susanna-Assunta Sansone	Associate Professor	2010 – present
Dr Philippe Rocca-Serra	Associate Member of Faculty	2010 – present
Dr Peter McQuilton	Project Coordinator	2015 – present
Period when the claimed impact occurred: July 2016 – December 2020		
Is this case study continued from a case study submitted in 2014? N		
1. Summary of the impact		
<p>Data-driven discovery requires the use of data at scale, and a transition from the current closed and silo-based approaches to research towards more networked scholarship. University of Oxford research to transform data management has resulted in the collaborative development of the FAIR Principles to ensure data is Findable, Accessible, Interoperable, and Reusable autonomously, with minimal human intervention. FAIR has de facto become a global norm for good data stewardship, a prerequisite for reproducibility, and has guided data policies, actions and professional practices in the public and private sectors. In their commitment to encouraging FAIR research data, meetings of international policy makers, such as G20 and G7, are publicly endorsing FAIR data; funding bodies are consolidating FAIR into their funding agreements; publishers have united behind FAIR as a way to promote and remain at the forefront of open research; and in the private sector FAIR is adopted and enshrined in policy in major biopharmas, libraries, and unions. Furthermore, the European Commission estimated the cost of not having FAIR Data for EU-28 at EUR10,200,000,000 (01-2019) a year. The adoption by these stakeholders of FAIR, the requirement for greater transparency and preservation of underpinning research that supports discoveries, are changing the culture of data science. Fundamentally, better data means better science.</p>		
2. Underpinning research		
<i>Work on data sharing policies, research data management and reproducibility</i>		
<p>Thanks to the amount of data that is increasingly available in the public domain we are starting to see the rise of scientific discoveries that are made using other people's data. However, the vast majority of data in the public domain are still not reusable, primarily because data are poorly described for third party use. This is placing strain on the mechanisms we currently have for peer review and quality control of the information that is shared. For scalable, effective and trustworthy data-driven science, we need new technological and social infrastructure to improve data management and stewardship, as well as cultural and policy changes.</p>		
<p>For over two decades Prof. Sansone and Dr. Rocca-Serra have worked with researchers, funders and other stakeholders, such as pharmas, services providers, and scholarly publishers to address the grand challenges related to information science and scholarly communications, where data quality and readiness for (re)use is a prerequisite for success. After joining the University of Oxford in 2010 they worked with representatives of major biomedical funding organizations in the UK, EU, USA, and Canada to centralize bioscience data sharing policies and drive their harmonization. Originally a blogging site, BioSharing was redesigned, informed by funders' requirements, and launched in 2011 as a prototype 'one stop shop' resource to map and interlink the description of</p>		

public Biomedical Sciences data repositories and community standards (essential for the unambiguous description, management and sharing of datasets) and explore their adoption by data policies.

In 2012 Prof. Sansone and Dr. Rocca-Serra reached a major milestone for another relevant activity: the ISA Framework. They led an international group of collaborators in a publication showcasing a set of open source tools and their wide application for data management, and a general-purpose **data representation model** built on the ‘Investigation’ (the project context), ‘Study’ (a unit of research), and ‘Assay’ (analytical measurement) metadata categories. The ISA framework helped researchers and curators to meet funders’ data sharing policies, by describing, sharing, and publishing biomedical data using community standards [R1]. Initiated and driven by Prof. Sansone and Dr. Rocca-Serra, the ISA Framework has been successfully embedded in a number of research infrastructures and adopted by a growing number of communities in different areas listed under the ISA Commons. These include global Biomedical Sciences data repositories, such as EMBL-EBI Metaboblights and NASA Genelab, and data journals, such as Springer Nature’s *Scientific Data*. An extension of the ISA model for nanotechnology applications has become a formal ASTM standard [R2]. In 2015 Dr. Peter McQuilton joined Prof. Sansone’s group and was appointed BioSharing project coordinator. The group then expanded its userbase beyond funders working with, and for, scholarly publishers, librarians, and services providers in the public and private sectors, adding content and functionalities and releasing an enhanced community service [R3].

Community drivers of digital transformation: launching the FAIR Principles and their indicators

The research and development activities on ISA and BioSharing placed Prof. Sansone and her group at the core of the international community movement for reusable and reproducible research. In 2015, as a part of a high-level group of 53 internationally recognized data leaders, Prof. Sansone’s group started to work on a set of guiding principles to ensure that contemporary data resources and scholarly output are **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable, in short: FAIR. Prof. Sansone and Dr. Rocca-Serra’s long-standing expertise and contribution is highlighted by the inclusion of both the BioSharing and ISA Framework as key FAIR-enabling exemplars. Published in 2016, the FAIR Principles describe characteristics and aspirations for systems and services to support the creation of valuable research data that could then be rigorously evaluated and extensively reused or reproduced, with appropriate credit, to the benefit of both creator and user [R4].

However, the FAIR Principles are aspirational, and their rapid adoption required the development of a qualitative and quantitative evaluation method to measure the level of FAIRness of the data. Prof. Sansone was one of the leaders of a small group (from the co-authors of the FAIR Principles, including Drs. Rocca-Serra and McQuilton) that paved the way to the definition of maturity indicators to measure FAIRness [R5] and demonstrated their implementation in the first automatic FAIR evaluation tool [R6].

3. References to the research

- [R1] Sansone, S.A., Rocca-Serra, P., Maguire, E., Trefethen, A., et al. “Toward interoperable bioscience data”. *Nat Genet.* 44, 121–126 (2012). doi:10.1038/ng.1054 (Journal article)
- [R2] Baker, N.A., [...] Rocca-Serra, P., Sansone, S.A., et al. “Standardizing data”. *Nat Nanotechnol.* Feb;8(2) (2013): doi:10.1038/nnano.2013.12 (Journal article)
- [R3] McQuilton, P., Gonzalez-Beltran, A., Rocca-Serra, P., Thurston, M., Lister, A., [...] Sansone, S.A., et al. “BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences”. *Database* Vol. 2016 (2016). doi:10.1093/database/baw075 (Journal article)
- [R4] Wilkinson, M., [...] Gonzalez-Beltran, A., Rocca-Serra, P., Sansone, S.A., et al. “The FAIR Guiding Principles for scientific data management and stewardship”. *Sci Data.* 3, 160018 (2016). doi:10.1038/sdata.2016.18 (Journal article) (*September 2020: accessed 147,000 times, cited 1,482 times Web of Science, cited 1,767 times CrossRef - putting it in the top 5% of all research outputs*)

- [R5] Wilkinson, M., Sansone, S.A., Schultes, E. et al. "A design framework and exemplar metrics for FAIRness". *Sci Data*. **5**, 180118 (2018). doi:10.1038/sdata.2018.118 (Journal article)
- [R6] Wilkinson, M., [...] Sansone, S.A., Rocca-Serra, P., McQuilton, P., "Evaluating FAIR maturity through a scalable, automated, community-governed framework". *Sci Data* **6**, 174 (2019). doi:10.1038/s41597-019-0184-5 (Journal article)

4. Details of the impact

The success of the FAIR Principles drew in more community interest in BioSharing, and started to receive national and international recognition. In response to this interest, in 2016 Wellcome commissioned Prof. Sansone and Dr. Rocca-Serra to evidence the status quo of community standards, the role of BioSharing in the nascent FAIR-enabling ecosystem, and the actions to support open research and reproducibility into disciplines beyond Biomedical Sciences [S1]. Now rebranded as FAIRsharing, this is a curated, informative, and educational resource that interlinks community standards to databases, repositories, and data policies. FAIRsharing guides consumers to discover, select, and use these resources with confidence, and guides producers to make their resource more discoverable, adoptable, and citeable. The FAIRsharing resource has become a recognized element in the research data life cycle for stakeholders working with data, and their endorsement brings to FAIRsharing hundreds of thousands of users [S2]. This uptake is being achieved by embedding FAIRsharing in discipline-specific communities and by expanding and developing new features and tools to access and use the content (for example, by connecting FAIRsharing to data management planning tools and systems for the transparent evaluations of FAIRness) [S3].

Policy Impact: The FAIR Principles: aligning global policy makers around common guidelines

The FAIR Principles have propelled the global debate about better data stewardship and transparent and reproducible data worldwide, and have become a mainstream reference for policy makers since their endorsement by global and intergovernmental leaders. For example:

- The G20 Leaders' Communique Hangzhou Summit (2016): "*We support appropriate efforts to promote open science and facilitate appropriate access to publicly funded research results on Findable, Accessible, Interoperable and Reusable (FAIR) Principles*" [S4].
- The G7 Expert Group on Open Science (2017): "*Research data management adheres to the FAIR principles whereby data is findable, accessible, interoperable, and reusable*" [S5].
- The Organisation for Economic Cooperation and Development (OECD) Committee for Scientific and Technological Policy (2018): "*The FAIR principles have de facto become an international norm, helping to guide policy actions*" [S6].

Policy Impact: funding bodies mandating FAIR data [S7]

- The European Commission (EC), as a part of the Horizon 2020 funding, began a flexible *Open Research Data Pilot* (2016) to "maximize access to and re-use of research data" and to help beneficiaries ensure they are applying the FAIR Principles. The FAIR Principles and FAIRsharing are also recommended in the *Model Grant Agreement* (2019), in the *Final Report and Action Plan from the European Commission Expert Group on FAIR Data: Turning FAIR into Reality* (2018) and *Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs* (2020). FAIR is also firmly placed in the *Horizon Europe 2021-2027 Draft Programme* (2019) in which open research data must be ensured: "in line with the principle 'as open as possible, as closed as necessary'; Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Re-usable) and Open Research Data".
- The UK Open Research Data Task Force built upon the principles in the *Concordat on Open Research Data* (2016) in the output *Realizing the Potential: Final Report of the Open Research Data Task Force* (2018), which already encouraged FAIR, and recognized that UKRI was in an ideal position to harmonize and lead the effort of implementing FAIR Principles across the UK. The associated *Open Research Data Task Force Case Study* (2018) also features FAIRsharing.

- Science Europe, an association representing the interests of 37 major public research performing and research funding organisations from 28 European countries, also adopted FAIR and recommends FAIRsharing in the core of its *Presenting a Framework for Discipline-specific Research Data Management* (2018).
- In the USA the *National Institutes of Health (NIH)'s Strategic Plan for Data Science* (2018) commits to adhering to the FAIR Principles to ensure a modern data-resource ecosystem.
- Australia's *Policy Statement on FAIR Access to Australia's Research Outputs* (2017) specifies that all publicly funded researchers and research organizations' policies and practices be FAIR in the *National Research and Innovation Framework* by 2020.

Policy Impact: publishers placing themselves at the forefront of data science by adopting FAIR [S8]

- Springer Nature, Wiley, Taylor and Francis, and Elsevier are among the major FAIR-supportive scholarly publishers using and adopting FAIRsharing and recommending it as the authoritative source of standards in their journals' guidance to authors. Since 2017 FAIRsharing has strengthened their data policy requirements that, in total, will affect 1,000s of journals by unifying the process in which editors select and recommend data repositories and community standards.
- Driven by FAIRsharing Prof. Sansone and her group, in collaboration with Datacite, are working with major journal publishers active in, and around, open data. This collaboration has been created to identify a common set of criteria for selecting and recommending data repositories and associated standards. Each of the publishers involved (PLOS, Springer Nature, F1000, Hindawi, Wiley, Taylor and Francis, Elsevier, EMBO Press, eLife, GigaScience, Cambridge University Press, and Oxford University Press) have published details of this FAIRsharing-standard and how it furthers their dedication to Open Science.
- The International Association of Scientific, Technical and Medical Publishers (STM, the global voice in publishing, which unites 150 members based in over 20 countries around the world) launched a Share, Link and Cite initiative to boost FAIR-compliant data sharing in 2020. STM states that it bases "[...] its activities on evidence-led, community-driven outputs such as: the FAIRsharing.org co-curated database of standards, policies and databases".

Policy Impact: adoption in the private sector, unions, libraries and associations [S9]

- Major biopharmas (including Bayer, Novartis, Hoffmann-LaRoche, and AstraZeneca) which also work in pre-competitive R&D initiatives around digital transformation, such as the global Allotrope Foundation and Pistoia Alliance, have developed and maintained a *FAIR Toolkit* to help companies implement FAIR principles in the life science industries.
- CODATA (part of the International Science Council non-governmental organization of 40 international scientific unions and associations and over 140 national and regional scientific organizations) created the *Beijing Declaration on Research Data* (2019) supporting international efforts to make research data open via FAIR.
- GO-FAIR, a bottom-up initiative of national governmental-funded offices in The Netherlands, France, Germany, USA, and Brazil fosters the coherent development of the *Internet of FAIR Data & Services (IFDS)* to standardise online research data to maximise discovery, accessibility, and usability. Since 2019 FAIRsharing has been a formal GO-FAIR Implementation Network (a community-led and self-governed consortium committed to define and creating tools for IFDS) in *FAIR StRePo: making Standards, Repositories, and Policies FAIR* (2019).
- The Research Data Alliance (RDA) was launched by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation. In 2015 Prof. Sansone established the RDA BioSharing Working Group (later FAIRsharing Working Group), in collaboration with the CODATA Executive Director and jointly with Force11 (a community of librarians, archivists, publishers and funders). In 2016 Dr. McQuilton joined as co-chair and since 2019 the FAIRsharing resource has been a flagship RDA-endorsed output, benefitting more than 10,000 members from 145 countries.

- Library associations in Europe, such as Liber, Europeana, and the International Library Carpentry, recommend FAIRsharing in their guidance and provide courses on its benefits.
- Discipline-specific communities, such as the American Geophysical Union (60,000 members in 137 countries) and the Federation of Academies of Sciences and Humanities (ALLEA, 50 academies from over 40 EU and non-EU countries) also recommend FAIRsharing. The pan-European biomedical ELIXIR (220 institutes in 23 nodes, including Oxford led by Prof. Sansone) recommends both FAIRsharing and the ISA Framework.

Economic Impact – Cost of not having FAIR Data for EU-28 has been estimated at EUR10,200,000,000 (01-2019) a year [S10]

FAIR has been calculated to be economically advantageous. In 2018 an EC report calculated the “annual cost of not having FAIR data to a minimum of EUR10,200,000,000 per year [01-2019]. The actual cost is likely to be much higher due to unquantifiable elements such as the value of improved research quality and other indirect positive spill-over effects of FAIR research data.” This estimation was based on three areas of the research data lifecycle:

1. Research activities – time spent, cost of storage, licence costs;
2. Further research opportunities – research retraction, double funding, interdisciplinarity;
3. Innovation – potential economic growth (% of GDP).

Furthermore, the report highlighted that the impact of FAIR on innovation alone “[...] could add another EUR16,000,000,000 to the minimum cost we estimated.” For context, regarding the scale of the costs of not implementing FAIR, “EUR10,200,000,000 per year is 78% of the Horizon 2020 budget per year and ~400%, of what the European Research Council and European research infrastructures receive combined”. These economic calculations were also accompanied by policy recommendations to capitalise on these savings.

5. Sources to corroborate the impact

[S1] Commissioned Wellcome article expanding Biosharing: Sansone, SA; Rocca-Serra, P (2016): *Review: Interoperability standards*. Wellcome Trust doi:10.6084/m9.figshare.4055496.v1

[S2] Sansone, *et al.* FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol*, 37, 358–367 (2019) doi:10.1038/s41587-019-0080-8

[S3] Publications corroborating expansion of FAIR and the development of tools to evaluate FAIRness of open data to a uniform standard, see: Wilkinson, M.D., *et al.* Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci Data* 6, 174 (2019). doi:10.1038/s41597-019-0184-5 and Clarke, D., *et al.*, "FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources", (preprint) (2019) doi:10.1101/657676

[S4] G20 Leaders' Communique Hangzhou Summit, supporting FAIR Open Science (2016)

[S5] G7 Expert Group on Open Science “Focus: Infrastructures for an optimal use of research data” statement ensuring FAIR in Data Management Plans (2017)

[S6] “OECD Science, Technology and Innovation Outlook” report reviewing key trends in STI policy in OECD countries and stating that FAIR is the norm for international policy on data (doi: 10.1787/25186167, 2018)

[S7] Select exemplars of policies from funding bodies mandating FAIR including: **a)** European Commission (2018), **b)** US National Institutes of Health (2018), and **c)** Australia's National Research (2017)

[S8] **a)** A memorandum (2019) and preprint paper (doi: 10.17605/OSF.IO/N9QJ7, 2019) evidencing FAIRsharing and journal representatives' commitment to create a common list of mandated standards and repositories, including: PLOS, Springer Nature, F1000, Hindawi, Wiley, Taylor and Francis, Elsevier, EMBO Press, eLife, GigaScience, Cambridge University Press, and Oxford University Press. **b)** STM's announcement on increased research data sharing and its effectiveness based on policies such as FAIRsharing (2020).

[S9] Select exemplars of policies from private sector mandating FAIR: **a)** biopharmas (Pistoia Alliance); **b)** Library Associations (Liber, 2017); **c)** discipline-specific communities (ALLEA, 2020)

[S10] European Commission, *Cost of not having FAIR Research Data: Cost-Benefit analysis for FAIR research data*, (2018) publication evidencing the cost of not having FAIR research data to the EU community and subsequent policy recommendations