Perspectives on Data Sharing and the New NIH Policy from the European Union

Jan G. Bjaalie[1,2] Carole Goble[3,4] Susanna-Assunta Sansone[5,6,4] Richard Nakamura[7,8] Maryann E. Martone[9]

[1]Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway,

[2]EBRAINS, Human Brain Project, European Union , Europe,

[3]Department of Computer Science, University of Manchester, Manchester, England, United Kingdom,

[4]ELIXIR-UK, United Kingdom,

[5]Department of Engineering Science, University of Oxford, Oxford, England, United Kingdom,

[6]Oxford e-Research Center, University of Oxford, Oxford, England, United Kingdom,

[7]Former Center for Scientific Review, National Institute of Health, Bethesda, Maryland, United States of America,

[8]Former National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, United States of America,

[9]Department of Neurosciences, University of California San Diego, La Jolla, California, United States of America

ABSTRACT

The U.S. National Institutes of Health sponsored a meeting in April 2021 to consider how to increase the scientific culture of data sharing. The editors of this special section on the meeting asked three prominent scientists to provide perspectives from Europe on data sharing and management and about the possibilities for global cooperation.

Keywords: Data management and sharing plans, DSM, FAIR, citation, data sharing, privacy, intellectual property, database

Media Summary

The April 2021 National Institutes of Health (NIH) workshop necessarily focused on implementation and impact of the new NIH Data Management and Sharing Policy on biomedical science in the United States. Among the speakers were Drs. Carole Goble, Susanna-Assunta Sansone, and Jan Bjaalie, with long pedigrees in the areas of open science, data management and sharing, and representing two large, well-established data sharing efforts in the European Union: ELIXIR (Goble and Sansone) and the European Human Brain Project (HBP; Bjaalie). ELIXIR is an intergovernmental organization that brings together life science resources from across Europe, including databases, software tools, training materials, cloud storage, and supercomputers. ELIXIR's goal is to coordinate these resources so that they form a single infrastructure. HBP is a large-scale EU project to put in place cutting-edge research infrastructure to enable scientific and industrial researchers to advance knowledge in the fields of neuroscience, computing, and brain-related medicine. The HBP led to the creation of EBRAINs, a digital infrastructure to support neuroscience and brain-inspired technology.

ELIXIR and the HBP have been underway for several years and provide valuable insights and experience from which to draw as we move forward in the United States. We asked our colleagues to provide some reflections on the workshop and perspectives on data sharing in the European Union versus the United States.

This conversation took place asynchronously throughout September 2021. Revisions were made in December 2021. **Richard Nakamura** (Former Director of the Center for Scientific Review at the National Institutes of Health) and **Maryann Martone** (Department of Neurosciences, University of California, San Diego) sent written questions to the following participants:

Jan G. Bjaalie, MD PhD, Professor of Neuroscience, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway; Infrastructure Development Director of the EU Human Brain Project; EBRAINS Data services leader

Professor Carole Goble, CBE FREng FBCS CITP, Professor of Computer Science, Department of Computer Science, University of Manchester, Manchester, England, United Kingdom; Joint Head of Node ELIXIR-UK, United Kingdom

Susanna-Assunta Sansone, PhD, Professor of Data Readiness, Associate Director of the Oxford e-Research Centre and University-wide Academic Lead for Research Practice, Department of Engineering Science, University of Oxford, Oxford, England, United Kingdom; ELIXIR Interoperability Platform ExCo; ELIXIR-UK, United Kingdom.

The workshop in April generated a lot of excitement and great conversations around the ambitious goal of the workshop: "Changing the culture around data management and sharing." After a few months' reflection, what were your main takeaways from the meeting?

Carole Goble (CG): There is a change in temperature in the community, from lip service to or ignoring of RDM (research data management) and hostility to sharing to at least acceptance and possibly enthusiasm! A big takeaway is that FAIR (findability, accessibility, interoperability, and reusability; Wilkinson et al., 2016) takes a village; from the different stakeholders ranging from infrastructure to policy to the fact that we cannot delegate data management and sharing to the individual. We all are responsible, we all have a role to play, and we all must be held to account. Cultural change has to be everywhere. Another point is that cultural change is expensive. FAIR is not Free. I am not sure the sustainability aspects were fully developed in the workshop.

Susanna-Assunta Sansone (SS): The FAIR brand has made RDM more glamorous. Whilst before it was seen as a (boring) service, nowadays good RDM is on the agenda of every stakeholder, including publishers and funders. Via their FAIR-driven data policies, RDM has become more visible to researchers, who are progressively learning to embed RDM practices in their work. These are small but essential steps that are driving cultural changes. The main takeaway from the meeting is that FAIR and good RDM are team sports, and we have just started. The road ahead of us is long. Having major funders and organizations that help us turn these aspirations into practices, as shown by the workshop, is essential.

Jan G. Bjaalie (JGB): The workshop demonstrated that the well-established FAIR Principles are increasingly embraced by the broader scientific community. It also discussed many of the concrete steps required to make data FAIR and what the (not insignificant) implications are for the individual researchers. While many valuable and practical solutions have been developed over the recent years, we are far from having truly smart and scalable solutions for all kinds of data sharing. It will be important to build on the growing number of success stories.

How do you compare the policies and approaches being used in the United States versus the European Union? What have you found to be most effective? Ineffective?

CG: The European Commission has already mandated data management plans (DMPs) at proposal time for its new Horizon Europe (2022) work plan and offers advice on DMP, reproducibility, and open research practices, including the public publishing of DMPs. What is interesting is that we now hear of final DMPs being rejected for H2020 projects (the previous program), so the European Commission is actually reviewing those. Which is essential if we are to get real changes in behavior.

At the national level, policies vary even within subdivisions in one funding agency, but we see efforts to really harmonize. A consistent policy is important—no researcher just funds their research from one agency, and having different policies from different agencies for pooled research is confusing at best.

SS: In the European Union there is a major investment in RDM and FAIR data by the European Commission, via the European Open Science Cloud (EOSC) program across all disciplines, and in partnership with the private sector, via the Innovative Medicine Initiative in the biomedical area, which has clear RDM requirements (Innovative Medicine Initiative, 2018). Perhaps what we do not need is too many 'coordination projects' that often become echo chambers. For example, in the EOSC ecosystem there is a layer cake of FAIR coordination projects, working groups, and task forces: the higher you go, the further away you are from the researchers. Although coordination is important, it is clear that solving the current profusion of FAIR-focused coordination projects simply by adding another layer of coordination might not be the best solution. While an extra layer may seem to offer a sweet way of bringing various ingredients together, the result can be a gooey mess.

JGB: The European Commission has funded numerous large research projects with a strong data-sharing component. Some are parts of the European Union and nationally funded European research infrastructures. Others represent the starting point for new research infrastructures. Europe's policy on research infrastructures may have some advantages in terms of creating sustainable mechanisms for FAIR sharing of data and tools for analysis. It remains to be seen over the coming years if this promising trend will be further strengthened, for example, through the new European Open Science Cloud (EOSC). I have the same perspective on EOSC as expressed by Susanna: it may go well if EOSC and other comparable initiatives will succeed in connecting strongly to the levels where the actions are taking place.

What additional barriers are there around international data sharing that were not addressed?

CG: The ethics of sharing and the need for contributor credit and access were addressed to some extent, but the big issue of legal governance was pretty lightly skimmed over. Organizational and legal interoperability are way more tricky than technical or semantic interoperability and can really block progress. We didn't really talk a lot on licensing either. The table on page 4 of the EOSC Interoperability Framework (European Commission et al., 2021) summarizes some of the issues. I am working on a project right now that needs cancer data to be shared between the United States of America and the United Kingdom, and really the technical barriers are not the hard bit—it's the governance, and it isn't clear what best practice is or where to go for how-tos.

SS: Any international practice requires an all-hands-on deck approach. If we can overcome the technical and motivational challenges, then we are left with the legal and ethical ones, as Carole says. The first two categories are rooted in well-known issues with health information systems and standards, for which solutions exist. However, it is harder to get funding to connect systems and bridge standards; few are willing to sponsor the 'data pipeline.' For the last two categories, the solutions lie in an international dialogue aimed at generating consensus on data policies and instruments for sharing data. But where there is a will (and a need) there's a way!

JGB: For understandable reasons, the challenges related to the sharing of data across countries and continents were not discussed in great depth. Data originating from human subjects are the most challenging to share, for regulatory as well as ethical reasons. Some of the concrete examples shown at the workshop, however, have served as a source of inspiration for a task force of the International Brain Initiative that recently developed a set of recommendations for international data governance (Eke et al., 2022).

Have you started to see tangible impacts of the EU policies' approaches on data management and/or data sharing?

CG: European Commission projects are required to submit DMPs during their reporting, and we hear that now these are being rejected at review—this means that they are being taken seriously, and grant holders and their universities are taking it seriously.

SS: First, any data policy is only as good as its implementation; second, data policies need to go hand in hand with practice and funding. Hopefully, the results of these investments in the European Union will soon be evident.

JGB: One of the most evident examples of how EU policies influences data sharing is seen in the [EBRAINS Data and Knowledge services](). This service for FAIR sharing of neuroscience data is so far mostly populated by the research produced by the EU Human Brain Project, largely because the data sharing was a strict requirement in this project. Since the service was introduced in 2018, more than 1,500 researchers have contributed nearly 1,300 datasets. An increasing number of investigators from other projects and also from outside Europe are starting to use this service.

Given that the NIH funds projects outside of the United States, do you see any conflicts arising for NIH-funded projects in the European Union?

JGB: We should continue to build on the historical traditions for trans-Atlantic collaborations. Data management and sharing will certainly benefit from having researchers from both the United States and Europe in the same consortia.

What is the single biggest challenge we face in trying to change the culture, based on your experiences?

CG: Getting the 'FAIR is a Village' (Borgman & Bourne, 2022) to work. Strategic policy without *at the same time* addressing metrics, processes, and organizations at the operational level leads to stress at best and failure at worst. FAIR sharing needs trusted research environments and repositories to use for the sharing to make it feasible.

SS: Culture is often defined as shared everyday habits. Therefore, changing culture requires scientists to change their existing habits and bring good RDM into their daily habits. Preferably, we should create a career path for a new class of scientists who are RDM professionals. Good RDM and data sharing will then not be seen just as services, or second-class activities, but become a recognized research and development subject in all sectors.

JGB: The single biggest challenge is, in my view, the incentives system. It is of great help that funders and institutions have started to introduce requirements for data management and sharing. It is also promising that early stage researchers are beginning to realize that transparency in the field of data (and also tools) management and sharing will help them build a solid platform for their future career. But there is a danger that

the ongoing change will be too slow and that much of what is delivered by science will continue to have problems with reproducibility and replicability for many years to come.

Closing Comments

The United States is joining the European Union and many other countries in the push toward open and FAIR science built on a foundation of open and FAIR data and code. The workshop and these reflections highlight that are many opportunities to learn from existing efforts for what works and what does not as the United States moves forward, but also opportunities to join forces to tackle some of the persistent challenges limiting progress. As data sharing becomes the norm within geopolitical borders, increasing attention will have to be paid to sharing across them as well. Such work will benefit from international organizations like the Global Alliance for Genomic Health, the International Neuroinformatics Coordinating Facility, and the International Brain Initiative, which bring together experts across the world to work on issues such as standards and international data governance. Much needs to be done, yet current progress in Europe and elsewhere, along with U.S. efforts, will provide strong examples of the benefits of open science and global sharing. International cooperation can help us build on these steps to accelerate the transformation of biomedicine to be open and FAIR.

Disclosure Statement

Richard Nakamura, Jan G. Bjaalie, Carole Goble, and Susanna-Assunta Sansone have no financial or non-financial disclosures to share for this article. Maryann Martone is a founder and has equity interest in SciCrunch Inc, a tech start up that provides tools and services in support of rigor and reproducibility.

References

Eke, D. O., Bernard, A., Bjaalie, J. G., Chavarriaga, R., Hanakawa, T., Hannan, A. J., Hill, S. L., Martone, M. E., McMahon, A., Ruebel, O., Crook, S., Thiels, E., & Pestilli, F. (2022). International data governance for neuroscience. *Neuron 110*(4), 600–612. https://doi.org/10.1016/j.neuron.2021.11.017

European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K. (2021). EOSC interoperability framework: Report from the EOSC Executive Board Working Groups FAIR and Architecture, Publications Office. https://doi.org/10.2777/620649

Horizon Europe. (2022). *Programme guide: Version 2.0.* European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

Innovative Medicine Initiative. (2018, November 22). Open access and data management for projects. https://www.imi.europa.eu/resources-projects/open-access-and-data-management-projects

National Academies of Sciences, Engineering, and Medicine. (2021, April 28–29). Changing the culture of data management and sharing: A workshop. https://www.nationalacademies.org/event/04-29-2021/changing-the-culture-of-data-management-and-sharing-a-workshop

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, M., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, Article 160018. https://doi.org/10.1038/sdata.2016.18