

ESMRMB annual meeting roundtable discussion: “Challenges and solutions in data sharing—an MRI perspective”

- [Hermien E. Kan](#),
- [Stephan Heunis](#),
- [Stuart Taylor](#),
- [Adina Wagner](#) &
- [Tonya White](#)

Introduction

As part of the program of the upcoming 2021 Annual meeting of the European Society for magnetic Resonance in Medicine and Biology (ESMRMB), the program meeting committee challenged various scientists involved in data sharing of MRI datasets to join a roundtable discussion.

With the increasing attention to data sharing and open science, there is a growing need to make data publicly available. Both funders and scientists play active roles in this effort. There are now guidelines on how to make data Findable, Accessible, Interoperable and Reusable (FAIR) (<https://www.go-fair.org/>), and many funders require datasets and published works to be openly accessible. While it might seem obvious that data which have been generated using public resources should be available to the community, there are challenges associated with sharing data, and some of these are specific to MRI. For the roundtable discussion, we have invited contributions from different scientists involved in open science and data sharing and have asked them to reflect on various aspects of data sharing of MRI data.

The scientists have a wide range of backgrounds and work in different Universities or University Medical Centres:

- Stephan Heunis, PhD—Institute of Neuroscience and Medicine, Research Centre Juelich, Germany; real-time fMRI, methodology, data quality, open science community building.
- Stuart Taylor, MD, FRCR—Centre for Medical Imaging, University College London, United Kingdom; large-scale prospective multicentre trials of imaging technology in bowel and oncological imaging.
- Adina Wagner, Msc—Institute of Neuroscience and Medicine, Research Centre Juelich, Germany; open science advocate and research data management enthusiast, development of tools and resources for reproducible, open, and FAIR neuroscience.

- Tonya White, PhD—Child & Adolescent Psychiatry and Radiology & Nuclear Medicine, Erasmus University Medical Centre, Sophia Children’s Hospital, The Netherlands; mapping typical neurodevelopment and neurodevelopment of individuals with, or at high-risk for developing serious psychiatric disorders.
- The scientist panelists will be joined by a law representative from the European Commission during the roundtable discussion.

In preparation for the roundtable discussion, we have invited representatives to present a short pitch on three questions which we share with the community in the following sections.

1. What are the main challenges and solutions in data sharing of MRI data from the:

a. ethical perspective.

b. legal perspective.

c. technical perspective.

2. Which technological developments have enabled the evolution of data sharing the last 10 years and which ones are directly needed for the future?

What are the main challenges and solutions in data sharing of MRI data?

Challenges

All participants in this debate highlight the significant challenge in the inherent contradiction between, on the one hand, prioritizing the open sharing of research data for transparency, reproducibility, and being essential for developing artificial intelligence (from the ethical perspective of serving the public) and, on the other hand, complying with information governance and GDPR principles.

From an ethical perspective, the critical element is to assure that adequate consent has been obtained. With collections of template consent forms, initiatives like the Open Brain Consent [1] assist in this requirement for data sharing for current and future studies, but not for sharing data from older studies with insufficient consent. Obtaining individual participant consent for use of large volumes of retrospective data is often impractical or even impossible, often preventing this legacy data from being openly shared. An important aspect that was highlighted by multiple participants in the debate is data ownership. While some panel members argue that the individual is actually the

true owner of their data, there seem to be many mixed views and many stakeholders claim to be the 'true owners' of the data. Under the GDPR, informed consent needs to provide much greater detail in how the data will be used and who will have access to the data, as is the depth of anonymization, for example sharing of completely deidentified, pseudonymized or fully identifiable patient data. Not giving participants the opportunity to make a decision as to whether they want to have their data shared or not (i.e., by not including the option into the consent process) can be perceived as a solution, but this raises ethical issues. These considerations impact on required permissions and processes for sharing. It was also mentioned that ethical considerations go beyond data sharing and should consider if data are representative of age and ethnicity for example, so outputs are applicable to the wider patient population. Some panel members mentioned that participant involvement in data sharing processes is imperative and shared data must be proportionate to the aims of the research, while this is not highlighted by others. All panel members agree that safeguards must be in place and that data protection impact assessments consider mitigations and inform risk management.

From a legal perspective, challenges that were mentioned include that the academic Institutions who support research and data collection are given considerable responsibility to safeguard and protect human subject data. Data sharing may make these institutions more vulnerable to legal actions against them. If a university was a closed unit without data sharing and data held only within computers lacking any access to the outside world, there would be the most minimal risk of a data leak. While this would protect the university, it would impede science as data sharing is clearly enhancing scientific discovery.

From a technical perspective, even if technical solutions seem to be around the corner, there still are challenges. Those highlighted by the roundtable panelists can be broadly classified into issues inherent to the *system* and issues inherent to the *data*. *System issues* encompass the greater context in which data are shared. For instance, a lack of sufficient expertise forming part of the full research pipeline, from ideation to design, to running and reporting on experiments, is a substantial challenge. The solution is clear, but implementing an academic system that values, promotes, and provides job security for the required diverse set of skills is the counter-challenge. In addition, while data standardization has been a significant advancement (e.g., the BIDS standard), its adoption still varies a lot both within and outside the field. *Data issues* are more directly related to the MRI images that are shared, and the technical systems used for this. Storage demands, especially of today's larger or longitudinal acquisitions can exceed available infrastructure and transfers become non-trivial. In addition, preventing data leaks in an era where nefarious types are hacking databases is a

challenge. Ensuring that no data leaks occur is very important, as major data leaks could harm research efforts and research subjects. Pooling datasets requires harmonization of data elements (especially epidemiologic data). In addition, beyond interoperability of data, it is useful to increase the interoperability of *tools* along the tool chain of data acquisition, storage, sharing and processing. Finally, there has been considerable work in meta-analyses, but approaches using mega-analyses are more difficult as a result of data sharing issues.

Solutions

Possible solutions for accessing large volumes of MRI data include.

- central permissions processes;
- national and international data sharing repositories or digital innovation hubs—group—or site-specific infrastructure with permission-based access to protected data, standard analysis pipelines, report publishing, or collaboration;
- and patient consent opt out schemes.

International data sharing may be enabled by organizations such as the Global Digital health partnership. The concept of decentralized data access and computation was mentioned by many: infrastructure and tools that allow data to be accessed and analyzed remotely without actually “seeing” the data. Technical solutions like ENIGMA computing [2] (“bringing-compute-to-data”) allow this solution, where data never leave the host-infrastructure. Scalable infrastructure solutions (e.g., XNAT) and clients that allow encrypted file transport, partial (e.g., single file) or disk-space aware (e.g., with a separation of file content and file metadata) file retrieval (e.g., git-annex, DataLad) can mitigate this. To improve the validity and usefulness of metadata (cf. FAIR-principles [3]), open (domain-)standards (e.g., ReproIn) are important. If data structures and descriptors are standardized, there is a positive domino effect of interoperability that allows the development of automated tools for data validation, quality checking, processing, visualization, and of course, hosting on a fit-for-purpose server infrastructure, all of which promote data sharing and collaboration. When this metadata is used by processing software, invalid metadata can be identified and corrected early on. Open source enables interoperability of all involved tools.

Which technological developments have enabled the evolution of data sharing the last 10 years and which ones are directly needed for the future?

(Neuro)science shaped the emergence of open hosting infrastructure for data sharing (e.g., OpenNeuro), domain standards (e.g., BIDS [4]) and increasing technological adaptation to them (e.g., BIDSApps, cloud-computing infrastructure like brainlife.io), and it started to give academics credit to

data sharing (SciData's "Data descriptors"). Domain-agnostic tools for data versioning and transport (e.g., DataLad) allow to streamline retrieval, publication, and updating of data. In addition, algorithms are now emerging that allow for data to be analyzed without leaving the site, or cloud-based solutions that adhere to legal rules (i.e., GDPR) that allow researchers to access multiple data sets. Finally, the enhancement of parallel computing and continued improvement in storage ability has been critical for large data sets.

In the future, to bridge the gap between tools used across the life cycle of imaging data further, several panel members mention that more harmonization and interoperability between the different technologies involved in data sharing, such as hosting and storage of data, authentication and data-usage-agreements and transport are needed. Data harmonization will require improved metadata descriptors that allow close-to-complete interpretation of the data: improved linking and tagging of data sources, versions, computational history, collaborators and more to define the data in its larger context; and more tools that can interpret this machine-readable information and generate human-readable information to make the data findable and accessible.

SWOT analysis

We next challenged the participants to provide a SWOT analysis. The combined SWOT table is outlined as follows:

Strengths

- Universal imaging "language"-Dicom.
- Health care systems afford large volumes of rich MRI datasets available with genetic and histopathological correlates.
- High potential for academic-academic and academic-commercial collaboration.
- Advancing scientific discovery through responsible reuse and aggregation of existing resources.
- Enabling wider participation in the scientific community worldwide.
- Increased confidence in scientific findings via greater transparency.
- Reproducibility of research outputs (if "FAIR").
- Higher return-on-investment (ROI) in science.

Weaknesses

- Lack of standardization of MRI data between platforms and manufacturers.
- Requirement for time intensive image annotation.
- Lack of reference standard or ground truth.

- Open data are much less reusable without quality control, useful descriptors/metadata or without a useful structure results in data that are not reusable.
- Data need maintenance, sharing data once does not suffice.
- GDPR compliant hosting infrastructure for imaging data yet lacking.
- Sharing for the sake of sharing without diligently considering the risks involved.

Opportunities

- Improvements in MRI scanner efficiency (workflow and image acquisition).
- Improved and more efficient image interpretation.
- Development of new quantitative MRI biomarkers.
- Tools, standards and infrastructure for decentralized and privacy-aware data management, sharing and computing.
- Training and embedding of data experts and the next generation of scientists into secure positions as part of the full research pipeline in academia.
- Scientific credit and visibility for researchers/labs.
- General advancement of the field.

Threats

- Complex regulatory framework for data collation and sharing.
- Tension between academic and commercial entities.
- Complexity and non-uniformity of data extraction pipelines between institutions.
- Data abuse/privacy breaches.
- Data rot.
- Time-vs-reward trade-off (yet too little recognition for invested time).
- Our current collective ignorance about how shared data could be used in future (intentionally or not) for nefarious purposes or marginalization.

Summary

The entirety of the collated inputs of all roundtable participants on data sharing of MRI data highlighted that the ethical, legal, and technical challenges in MRI are very interconnected. As a result, while isolated solutions that tackle single “categories” of challenges are important, they are likely to be insufficient to “fix” data sharing at large. Going forward, a more “holistic” view and understanding of the problem is the only way to foster working solutions.

References

1. [con/open-brain-consent 1.1.0 | Zenodo. https://www.zenodo.org/record/4264941#.YQKPWEQzYkx](https://www.zenodo.org/record/4264941#.YQKPWEQzYkx)

2. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, Baune BT, Bertolín S, Bralten J, Bruin WB, Bülow R, Chen J, Chye Y, Dannlowski U, de Kovel CGF, Donohoe G, Eyer LT, Faraone SV, Favre P, Filippi CA, Frodl T, Garijo D, Gil Y, Grabe HJ, Grasby KL, Hajek T, Han LKM, Hatton SN, Hilbert K, Ho TC, Holleran L, Homuth G, Hosten N, Houenou J, Ivanov I, Jia T, Kelly S, Klein M, Kwon JS, Laansma MA, Leerssen J, Lueken U, Nunes A, Neill JO, Opel N, Piras F, Piras F, Postema MC, Pozzi E, Shatokhina N, Soriano-Mas C, Spalletta G, Sun D, Teumer A, Tilot AK, Tozzi L, van der Merwe C, Van Someren EJW, van Wingen GA, Völzke H, Walton E, Wang L, Winkler AM, Wittfeld K, Wright MJ, Yun JY, Zhang G, Zhang-James Y, Adhikari BM, Agartz I, Aghajani M, Aleman A, Althoff RR, Altmann A, Andreassen OA, Baron DA, Bartnik-Olson BL, Marie Bas-Hoogendam J, Baskin-Sommers AR, Bearden CE, Berner LA, Boedhoe PSW, Brouwer RM, Buitelaar JK, Caeyenberghs K, Cecil CAM, Cohen RA, Cole JH, Conrod PJ, De Brito SA, de Zwarte SMC, Dennis EL, Desrivieres S, Dima D, Ehrlich S, Esopenko C, Fairchild G, Fisher SE, Fouche JP, Francks C, Frangou S, Franke B, Garavan HP, Glahn DC, Groenewold NA, Gurholt TP, Gutman BA, Hahn T, Harding IH, Hernaus D, Hibar DP, Hillary FG, Hoogman M, Hulshoff Pol HE, Jalbrzikowski M, Karkashadze GA, Klapwijk ET, Knickmeyer RC, Kochunov P, Koerte IK, Kong XZ, Liew SL, Lin AP, Logue MW, Luders E, Macciardi F, Mackey S, Mayer AR, McDonald CR, McMahon AB, Medland SE, Modinos G, Morey RA, Mueller SC, Mukherjee P, Namazova-Baranova L, Nir TM, Olsen A, Paschou P, Pine DS, Pizzagalli F, Rentería ME, Rohrer JD, Sämann PG, Schmaal L, Schumann G, Shiroishi MS, Sisodiya SM, Smit DJA, Søndersby IE, Stein DJ, Stein JL, Tahmasian M, Tate DF, Turner JA, van den Heuvel OA, van der Wee NJA, van der Werf YD, van Erp TGM, van Haren NEM, van Rooij D, van Velzen LS, Veer IM, Veltman DJ, Villalon-Reina JE, Walter H, Whelan CD, Wilde EA, Zarei M, Zelman V (2020) ENIGMA Consortium. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry* 10(1):100. PMID: 32198361; PMCID: PMC7083923. <https://doi-org.libproxy.ucl.ac.uk/10.1038/s41398-020-0705-1>
3. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:160018. Erratum in: *Sci Data*. 2019 Mar 19;6(1):6. PMID: 26978244; PMCID: PMC4792175. <https://doi-org.libproxy.ucl.ac.uk/10.1038/sdata.2016.18>

4. The Brain Imaging Data Structure (BIDS) Specification |

Zenodo. <https://www.zenodo.org/record/3890439#.YQKQOkQzYkw>
