CHORUS Forum: Making FAIR'S Interoperability and Reusability Data Goals Possible Q&A and Chat Thread, 17 June 2022

Q & A

1. It doesn't help to frame or view the four FAIR functions as mutually exclusive strengths or capabilities because one of them, viz interoperability, has a lot to do with how the other three perform. Interoperability is also one, which we need to parse more carefully because it is potentially loaded with many facets that define its strength. For instance, we can define interoperability at the following levels: data, software, machine, services, semantics, etc. How do you address interoperability without defining it at greater granularity? *Bhaskar Ramachandran*

Answer: these are not mutually exclusive and it is an accident of implementation that the findability and accessibility aspects have been treated as separable. In large part that is historic by accident and not the fundamental principle. Ingrid D., also agrees that it is not mutually exclusive and finds FAIR really isn't anything new, especially if you look at the digital preservation community who has been working on this for a very long time. We also see funders calling out for assessments as they want to know how FAIR data are on the one hand and on the other hand, we want to make sure we don't exclude communities because FAIR is a journey. She believes we should take it integral and one-step at a time and be able to translate the steps taken into other areas so we are not reinventing. We should not look at the four aspects of FAIR separately. *Christopher Marcum*

2. May I ask if there is any guideline for the research of interoperability? *Shuai Wang (VU# he/him)*

Answer: Shelley S. noted that Sarah Nusser has done research on reusability and to some extent interoperability. Shuai Wang offered to do some research and write a guideline for this. Sarah N. (ISU, UVA, she/her) shared the following - This is the NASEM Roundtable's broader work, which spawned Helios: https://www.nationalacademies.org/our-work/roundtable-on-aligning-incentives-for-open-science. *Shelley Stall*

Christopher Marcum also mentioned the NSF's FEROS platform has some capacity for funding research in this area.

3. One challenge to FAIR data is the traditional scholarly publishing system that does not adequately recognize data as a research product, doesn't consistently require data citation or depositing. How/do the US federal government (Chris) and European Union (Ingrid) intend to work with publishers? e.g., mapping semantic artifacts doesn't happen without publisher buy in. Can you provide any more thoughts or detail?) *Amy Koshoffer (she/her)*

Answer: Research just published by three Croatian researchers who approached 1800 researchers who published in journals in the field of epidemiology with a data availability policy, where 120 replied and out of that hardly any data that were reusable came of it so that shows the magnitude of the problem. Changing that whole cultural system in science is maybe even a bigger issue. *Ingrid Dillo*
Christopher Marcum answered that it's not just data, it's ALL research products.

4. Is EOSC collaborating at all with Open Science Framework? Seems like some overlap & mutual interests. *Amy Koshoffer (she/her)*

Answer: This link might useful https://www.eoscsecretariat.eu/eosc-and-international-initiatives. *Shelley Stall*

5. Do you think that the lack of data sharing could be a result of poor data management skills? Should we as a community focus on teaching and showing the importance of data management skills? A group working on this: https://www.nsf.gov/awardsearch/showAward?AWD_ID=2126334 *JD Campbell (SoyBase.org)*

Answer: That is definitely a large part of the problem. In Europe a lot of work has been done in this area. See e.g., https://www.fairsfair.eu/fair-adoption-handbook-report-good-practices-fair-competence-education *Shelley Stall*

6. We know the bulk of our research teams are international and growing and more complex, and it really takes more than one discipline to solve these complex problems. Thinking about the researcher themselves, how do we start and walk my ways towards these common challenges? *Shelley Stall*

Answer: That is likely an internal issue as the supply and demand don't meet automatically. With the enormous pressure researchers are under so if it is not directly related to their work, there is no time to really dive into it with them. The main solution coming out of RDA is to make sure that we have smaller entities. At the national level or small regional level, we need to reach out to the researcher. And it is up to the groups like RDA to make that translation or be the intermediary in bringing it to the researchers. *Ingrid Dillo*

This also taps into the deeper issue of culture tenure in the academy and that research products that are adjacent to data sharing are not rewarded in the same way as publications and individual projects and grants. And work done by Greg Tenenbaum and others on trying to change the incentive structure at the university is a conversation all the funders need to be having. *Christopher Marcum*

7. There are also issues around equity and inclusion and there is a real worry that if we make science more open, make data more available in a way that is reusable potentially, that if we do this poorly, that we will cause more barriers for equity and inclusion. Are those being discussed and if so, what is your approach? *Shelley Stall*

Answer: There is real concern about equity and inclusion and to the culture of attribution. There is a distinction that should be made clear is that the proprietorship aspect of data products that are fully funded research projects and the equity issues are more about cultural challenge than natural policy challenges. *Christopher Marcum.* There is also disparity in disciplines with some disciplines getting huge amounts of funding and others much less and in doing so we've made a value statement. Together with funders and publishers requirements need to also come support for researchers to comply with the requirements; whether that support be with awareness, training, and technical infrastructure, and the element of recognition. *Ingrid Dillo*

Chat

Shelley Stall: F-UGI Tool: https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool

Shelley Stall: Analysis/review of F-UGI: Devaraju, A., & Huber, R. (2021). An automated solution for measuring the progress toward FAIR research data. Patterns, 2(11), 100370. https://doi.org/10.1016/j.patter.2021.100370

Caroline Coward - NASA JPL: Without interoperability and reusability, we're left with "Data AF".

Mark Parsons: see also https://www.heliosopen.org. A consortium of universities trying to look beyond the paper.

Ramapriyan (Rama): Parsons et al - https://eos.org/opinions/credit-where-credit-is-due

Amy Koshoffer (she/her): some good data DEIA resources: https://weallcount.com/the-data-process/ and the CARE Principles for Indigenous Data Governance https://www.gida-global.org/care. Also Make Data Count https://makedatacount.org/

Ixchel Faniel: In talking about changing culture and practices, part of that would be weaving those changes into a discipline's curriculum and teaching the "care for"/document their data at the point of creation.

Caroline Coward - NASA JPL: If we limit the heavy labor to machine learning systems, that's fine. But as soon as we transition into artificial intelligence, we need to infuse an ethical framework into our algorithm development. Ethical AI is a critical next step for the IR in FAIR.

Paul Guinnessy, Physics Today AIP: To follow up on Caroline's point, this book is pretty good at giving the general ethical issues regarding machine learning/AI algorithm and what constraints should be put on them. https://bookshop.org/books/human-centered-ai/9780192845290.

Mark Parsons: see also http://datafeminism.io

Caroline Coward - NASA JPL: Standard DMPs, should this be part of FAIR 2.0, and must include accessibility statements at least, if not full blown DEIA statements.

Bhaskar Ramachandran: Another useful resource from a 2022 NASEM Workshop: AI/ML to Advance Earth System Science: https://www.nationalacademies.org/our-work/machine-learning-and-artificial-intelligence-to-advance-earth-system-science-opportunities-and-challenges---a-workshop

Taner Sen: Perhaps another undermentioned point is the importance of curation. Creating better algorithms to use machines for interoperability and reusability is critical, but the algorithms are never perfect, and even after automated machine-processing, a certain level of curation by domain experts is necessary. The research funding though primarily focuses on hypothesis-driven research or developing novel algorithms, and manual curatorial work ends up under-funded and under-appreciated.

Sarah Nusser (ISU, UVA, she/her): Fantastic talks all around! The ideas discussed in the second section are so promising and get at the heart of problems that small-group researchers face. Thank you!!

Caroline Coward - NASA JPL: Hard to cite data in the literature?? Is your data being cited? Join the Data Citation Community of Practice: https://data.agu.org/DataCitationCoP/information/