LETTER

Challenges in the Maintenance of FAIR-Compliant Paediatric Data-Sharing Repositories

Claudia Pansieri*,†, Anando Sen‡, Annunziata D'ercole§, Ronald Cornet‖,¶, Victoria Hedley‡, Rebecca Leary‡, Alexandre Betourne**, Sanchita Bhattacharya††, Scott Shaunessy‡‡, Ramona L. Walls**, Fedele Bonifazi†,§ and Mariagrazia Felisi*,†

To meet the growing need for strong clinical evidence on a global scale, both public and private sectors have invested to standardise health data elements and achieve greater connectivity and interoperability of health data systems. This is enabled by FAIRification (making Findable, Accessible, Interoperable, and Reusable) of clinical trial data increasingly recognised as a crucial step in enhancing the value and utility of clinical data across the research community. This is even more relevant in the paediatric and rare disease field where data remains highly fragmented in terms of data collection practices, ontologies, and clinical reporting standards, and often locked in silos with diverse formats and standards.

To meet the growing need for strong clinical evidence on a global scale, both the public and private sectors have invested to standardise health data elements and achieve greater connectivity and interoperability of health data systems.[1] This is enabled by FAIRification (making Findable, Accessible, Interoperable, and Reusable) of clinical trial data increasingly recognised as a crucial step in enhancing the value and utility of clinical data across the research community.[2] This is even more relevant in the paediatric and rare disease field, where data remain highly fragmented in terms of collection practices, ontologies, and clinical reporting standards, and are often locked in silos with diverse formats and standards.

Being a FAIR-compliant data-sharing repository (DSR) involves committing to practices that ensure data is in a standardized format, accessible, and useful for current and future research. A range of technological tools, guidelines, and ontologies are emerging to facilitate FAIRification. Examples include metadata standards (e.g., schema.org) and repositories that offer open data sharing solutions. However, implementing these tools within clinical research workflows is not yet standardized or universally adopted and requires substantial effort and investment.[3]

To understand both the added value and the challenges in the development and maintenance of paediatric and FAIR-compliant DSRs, members of the connect4children (c4c) data standardization teams conducted interviews with representatives from three major DSRs in 2022: Clinical Study Data Request (CSDR), Immunology Database and Analysis Portal (ImmPort), and the Rare Disease Cures Accelerator – Data and Analytics Platform (RDCA-DAP).[4–6]

The two main takeaways from the interviews were:

(i) Having more unstandardized data available is better than having less fully standardized data.
(ii) Standardization requires substantial resources, including skilled personnel, time, and financial investment.

Two key challenges in maintaining a FAIR-compliant DSR were also identified, as outlined below:

* Consorzio per Valutazioni Biologiche e Farmacologiche (CVBF), Via Luigi Porta, 14, 27100 Pavia PV, IT

† TEDDY European Network of Excellence for Paediatric Research, Pavia, IT

‡ John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research Institute, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, NE1 3BZ, UK

§ Fondazione per la Ricerca Farmacologica Gianni Benzi Onlus, Via Giulio Petroni 91/D, 70124 Bari BA, IT

‖ Amsterdam UMC – University of Amsterdam, Medical Informatics, Meibergdreef 9, Amsterdam, NL

¶ Amsterdam Public Health, Methodology & Digital Health, Amsterdam, NL

** Rare Disease Cures Accelerator – Data Analysis Platform (RDCA-DAP), Tucson, US

†† Institute for Computational Health Sciences, University of California, San Francisco, CA 94158, US

‡‡ Anaqua-ClinicalStudyDataRequest (CSDR), Boston, MA 02116, US

Corresponding author: Claudia Pansieri (cpansieri@cvbf.net)

(i) Data originates from various sources and is provided in various formats. As a consequence, further efforts are needed for the conversion to modern standards such as Fast Healthcare Interoperability Resources (FHIR)[7] or the Observational Medical Outcomes Partnership (OMOP) Common Data Model before data can be shared and reused.[8]

(ii) Continuous monitoring and updating of data handling procedures need to be compliant with regulations and other local data protection laws (e.g., in the US, the HIPAA **Health Insurance Portability and Accountability Act**).

The full results of the interviews are reported in **Table 1**.

The interviews highlight the critical elements for maximizing the value of repositories, particularly in specialized contexts such as paediatrics and rare diseases. They underscore the importance of ongoing multidisciplinary collaboration to ensure the effective maintenance and evolution of these repositories. Such

Table 1: Interview's report.

Item	ImmPort	RDCA-DAP	CSDR
Major challenges associated with maintaining a repository.	**HIPAA (Health Insurance Portability and Accountability Act) compliance**, deidentification, determining **level of curation.**	Funders provide **capital to setup but not to maintain**. Keeping up with **new data formats**.	Researchers asking for **data which is no longer available.**
Importance of interoperability at institution.	**Absolutely required**. Enhance collaboration, streamline research flow, scalablity and democratization.	**Yes, it is a core principle**. Communication with other platforms (e.g., Vivli) in the hope of joining forces.	**Would like everything interoperable**. Some standardization using SAS.
Thoughts on FAIR principles.	**We are FAIR – as much as possible**. Ambiguous names and units are problematic.	**We are 25% there towards being FAIR.** Challenges with contributors collecting non-standard data.	**FAIR compliant but depends on interpretation**. Would like data to be understood by both humans and machines. Attempt to make data sharing agreement simple.
Costs related to being FAIR compliant.	Cost is directly proportional to types and volumes of data.	More costly than to just put data on the platform but to make the data useful, FAIR is a must.	It would if we took it to the highest possible level.
Steps taken to implement FAIR principles.	Implementing metadata standards, developing user-friendly interface for data search, collaborating with other data portals for improving data discoverability.	**Global unique identifiers, standardizing to OMOP**, adding ontologies, create customized data catalogues.	**Connecting with other repositories** like Vivli.
More data in a less standardised form or less data that is all standardised.	**Standardized data is rewarding for downstream secondary data reuse.**	**More data is always better**.	**Would prefer all data**.
Preferred format for data.	Have several templates (tab-delimited) with mandatory and optional columns.	**Machine-readable data – JSON**. Excel converted to CSV and then parsed.	Not really applicable as it really depends on data providers.
Additional challenges with paediatric data.	No restrictions on searching. **Not aware of additional challenges**.	**Paediatric patients move through multiple health care systems**. No paediatric strategy. Plenty of paediatric data within platform.	**No major challenges unless the diseases are rare**.
Thoughts on how to obtain more data from rare disease patients.	No answer.	**Mission of platform to get rare disease data but** must respect privacy and patient wishes.	Data providers can choose not to share **rare-disease data as they may be identifiable**.
Interest in academic third-party partnerships.	Would be great to get high value disease-specific datasets that would benefit the community for data reuse.	**Would be great** to have more people come to us to share data.	**Absolutely open to it**.

collaborative efforts are essential for transforming repositories into powerful tools that drive more efficient and impactful scientific discoveries.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Conceptualization: FMG, HV, CR, PC. Formal analysis: AS. Investigation: SA, PC, CR, VH. Methodology: SA. Writing: PC. Review and Editing: All.

All authors have read and agreed to the published version of the manuscript.

References

1. **Franklin JB, Marra C, Abebe KZ, Butte AJ, Cook DJ, Esserman L, Fleisher LA, Grossman CI, Kass NE, Krumholz HM, Rowan K, Abernethy AP, and JAMA Summit on Clinical Trials Participants.** Modernizing the Data Infrastructure for Clinical Research to Meet Evolving Demands for Evidence. *JAMA.* 2024;332(16):1378–1385. DOI: https://doi.org/10.1001/jama.2024.0268

2. **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.** The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016 Mar 15;3:160018. DOI: https://doi.org/10.1038/sdata.2016.18

3. **Felisi M, Bonifazi F, Toma M, Pansieri C, Leary R, Hedley V, Cornet R, Reggiardo G, Landi A, D'Ercole A, Malik S, Nally S, Sen A, Palmeri A, Bonifazi D, Ceci A.** Mapping of data-sharing repositories for paediatric clinical research—A rapid review. *Data* 2024;9(4):59. DOI: https://doi.org/10.3390/data9040059

4. **Clinical Study Data Request (CSDR).** Homepage. https://www.clinicalstudydatarequest.com/

5. **Immunology Database and Analysis Portal (ImmPort).** Homepage. https://www.immport.org/shared/home

6. **Rare Disease Cures Accelerator – Data and Analytics Platform (RDCA-DAP®).** Homepage. https://portal.rdca.c-path.org

7. **Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D.** The Fast Health Interoperability Resources (FHIR) Standard: Systematic literature review of implementations, applications, challenges and opportunities. *JMIR Medical Informatics.* 2021;9(7):e21929. DOI: https://doi.org/10.2196/21929

8. **Observational Health Data Sciences and Informatics.** Standardized Data: The OMOP Common Data Model. https://www.ohdsi.org/data-standardization/

Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS