

Converge or collide? Making sense of a plethora of open data standards in healthcare: an editorial

Guy Tsafnat, Rachel Dunscombe, Davera Gabriel, Grahame Grieve, Christian Reich

Submitted to: Journal of Medical Internet Research
on: December 23, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
Supplementary Files.....	11
Figures	12
Figure 1.....	13



Converge or collide? Making sense of a plethora of open data standards in healthcare: an editorial

Guy Tsafnat^{1, 2, 3} PhD, FAIDH; Rachel Dunscombe^{4, 5*}; Davera Gabriel^{1, 6*} RNC, FHL7, FAMIA; Grahame Grieve^{7, 8*}; Christian Reich^{9, 10, 11} MD, BSc

¹OHDSI FHIR & OMOP Collaboration Working Group NA US

²Evidentli Pty Ltd Surry Hills AU

³Macquarie University Macquarie Park AU

⁴openEHR International St. Helens GB

⁵Imperial College London London GB

⁶School of Medicine Johns Hopkins University Baltimore US

⁷Health Level 7 International Ann Arbor US

⁸Health Intersections Pty Ltd Melbourne AU

⁹Odysseus Data Services Cambridge US

¹⁰OHDSI FHIR & OMOP Collaboration Working Group New York City US

*these authors contributed equally

Corresponding Author:

Guy Tsafnat PhD, FAIDH

Evidentli Pty Ltd

50 Holt St.

Suite 516

Surry Hills

AU

Abstract

The digital health sector presents unique and diverse data challenges that manifest as a fractured ecosystem, and often fails to deliver effective collaboration among interconnected entities. The consequences include inconsistencies between services rendered and payments, waste and most importantly, suboptimal care. Interoperable data is hailed as a near-future solution to many of these challenges and have been so for decades. Paradoxically, interoperability efforts have themselves been fractured and inconsistent, resulting in a plethora of incompatible interoperability standards, despite widespread acknowledgement that fewer standards would provide better interoperability. This paper presents a typology of healthcare data requirements and describes the challenges and opportunities of open data standards in healthcare.

Recognizing that different data standards represent different points of view and respond to different needs, and that no single standard would necessarily be able to meet all the requirements of all healthcare systems, we distinguish three domains of healthcare data with their own unique characteristics and challenges, and outline high-level design requirements. We distinguish between requirements that are common across all domains, and those that are specific to each domain.

(JMIR Preprints 23/12/2023:55779)

DOI: <https://doi.org/10.2196/preprints.55779>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http](#)

Original Manuscript

Preprint
JMIR Publications

Converge or collide? Making sense of a plethora of open data standards in healthcare

Guy Tsafnat^{1,2,3,*}, Rachel Dunscombe^{4,5,#}, Davera Gabriel^{3,6,#},
Grahame Grieve^{7,8,#}, Christian Reich^{3,9,10,#}

¹ Evidentli, Sydney, Australia

² Macquarie University, Centre for Health Informatics, Sydney, New South Wales Australia

³ OHDSI FHIR & OMOP Collaboration Working Group

⁴ openEHR International, St. Helens, England, UK

⁵ Imperial College London, London, England, UK

⁶ Johns Hopkins University, school of medicine, Baltimore, Maryland, USA

⁷ Health Level 7 International (HL7), Ann Arbor, Michigan, USA

⁸ Health Intersections, Melbourne, Victoria, Australia

⁹ Odysseus Data Services, Cambridge, Massachusetts, USA

¹⁰ Observational Health Data Sciences and Informatics, New York, New York, USA

* Corresponding author

These authors contributed equally and are listed alphabetically

Introduction

The digital health sector presents unique and diverse data challenges that manifest as a fractured ecosystem, and often fails to deliver effective collaboration among interconnected entities.¹ The consequences include inconsistencies between services rendered and payments, waste and most importantly, suboptimal care. Interoperable data is hailed as a near-future solution to many of these challenges and have been so for decades. Paradoxically, interoperability efforts have themselves been fractured and inconsistent, resulting in a plethora of incompatible interoperability standards, despite widespread acknowledgement that fewer standards would provide better interoperability.^{2,3} This paper presents a typology of healthcare data requirements and describes the challenges and opportunities of open data standards in healthcare.

Recognizing that different data standards represent different points of view and respond to different needs, and that no single standard would necessarily be able to meet all the requirements of all healthcare systems, we distinguish three domains of healthcare data with their own unique characteristics and challenges, and outline high-level design requirements. We distinguish between requirements that are common across all domains, and those that are specific to each domain.

Three Domains of Healthcare Data

Clinical care and administration

Healthcare data are generated as documentation of clinical service delivery for a variety of purposes including continuity of care and billing. A repository of clinical records containing the complete patient history of every patient is required in order to deliver continuous and consistent care in an efficient and effective manner. As clinical care sometimes lacks a strong evidence base from which best practices can be gleaned, care documentation should be able to address a broad range of data types, including “one off” events reflecting unforeseen patient circumstances.⁴

The main focus of standards for EMR interoperability are to protect users from vendor lock-in, in which the cost of changing to a better system outweighs the benefits expected from the switch. An electronic healthcare record system that uses an open standard can replace another that uses the same standard, far more easily, and with less disruption to clinical practice, than two incompatible systems where data transformation from one system to the next, represents the lions’ share of the cost.

Data exchange

Records of care provided are used in communication between different actors. For example, continuity of care depends on multiple healthcare providers sharing information about the patient, effective billing requires sharing information between providers and payers. It is often critical that information is passed between systems immediately to inform urgent decisions. Standards for the transmission of healthcare data should be able to be arranged in small quantities that include all and only the relevant information.

The main focus of data exchange standards is fidelity to the operational healthcare processes. For example, a focus is on capturing and faithfully representing the terminology used in the local processes, whatever that is, as opposed to restricting use to a common terminology.

Longitudinal analysis

The main focus of longitudinal analysis are patterns, trends and their predictors. A practical consideration is that such patterns tend to be subtle, and hence, data from multiple systems and organizations is required to be aggregated and/or compared. in order to elucidate reliable phenomena. As an empirical discipline, Medicine incrementally improved using inferences made from big data. Moreover, medical evidence should produce predictable outcomes across multiple settings and for all patients.

Analytical methods use generalizations such as grouping patients with similar characteristics such as age. “One off” events and subtle differences in clinical practice are inconsequential or even harmful to population statistics. For example, as age is often measured in years and stratified to a larger age bracket, the exact birth date of each patient can be substituted by an approximate date without changing the results of the analysis. Another common reason to lose fidelity is patient privacy and prevention of unauthorized re-identification of individuals from analytic datasets.

Practicality, feasibility and community

The technical suitability of a standard to the problem it addresses is important, but digital health and data interoperability are applied disciplines. Practical and socio-technical considerations, such as available skill sets are at least as important, if not more important, than technical superiority. While no standard is perfect, there are major differences in what standards can do “out of the box”, versus how much customization is required before they can be put to use. To provide data interoperability in practice, standards have to find the right balance between customizability and rigidity. The United States Core Data for Interoperability (USCDI) is a prime example of such balance.⁵ These core data

element definitions are regulated for use in the US, but can also be thought of as a mid-point between customizations needed to be made at every site, and rigidity with regards to specific data elements.

Community

Data standards are only as useful as their adoption by collaborating organizations and so standards that are in widespread use in the community are more likely to be beneficial to organizations seeking to join collaboration networks that use these standards. Engagement from communities that not only use a standard, but to also contribute and improve it over time, differentiates legacy standards from more sustainable ones. An actively engaged community will, over time, evolve what starts as an infantile standard into a superior one. An engaged and vibrant community is therefore a major advantage for the longevity of the data standards it uses.⁶ Open standards have a clear advantage over proprietary ones, as they are available for contribution from a larger audience, and are tested in more diverse environments. Conversely, it is usually easier to govern the development of proprietary standards which can lead to shorter improvement cycles.⁷

Terminologies

A common aspect of all data standards, regardless of application, is the use of common terminologies. Many standards rely on terminologies that are developed independently. Terminologies can be flat or ontological, specific to one or a handful of domains or more general, proprietary or open. Depending on the intended use of the standard and available resources, the inclusion policy for terminologies in a standard may differ. Standards designed for continuity of care may allow the use of external terminologies provided that they are available to all parties providing care, whereas standards used for data mining may restrict their use to only pre-vetted terminologies. As is the case of data standards, terminologies benefit from active communities to keep them up to date. As evident in differing design criteria, a truly interoperable, agile healthcare system that provides the best possible clinical care, implements multiple standards. For example, standard records from a repository of complete patient histories, may be transformed to a standard for transmission, and transformed again when entered into a data warehouse for analysis. Systematic harmonization between standards would reduce or eliminate information loss in the transformation processes.⁸

Emerging standards

A paradox of healthcare interoperability is the existence of a large number of standards, with significant overlap among them. Each may be more or less equally suited to its intended use, so selecting a standard for implementation can be difficult. Harmonization from all standards to all others is impractical. The interoperability paradox is that more standards lead to less interoperability. We therefore propose that the health informatics community converge on just three standards. We have chosen standards that are open, are backed by active communities, and have been tried and tested in their respective areas.

Clinical care and Administration

openEHR⁹ provides a common data structure for health IT systems including, but not limited to, electronic medical records and patient administration systems. The standard uses a centralized database with a tree structure for efficient retrieval of a complete patient record. The schema of the database is patient-centric, clinical, and extensible with reusable archetypes and templates. openEHR is increasingly preferred over proprietary formats as multiple vendors that offer commercial systems based on openEHR means a very expensive process for switching vendors can be greatly simplified.

Data exchange

Health Level 7's (HL7) Fast Healthcare Interoperability Resources (FHIR) ¹⁰ is an open standard particularly suited for data transfer. At the center of the FHIR specification is a RESTful API that provides real world healthcare services such as 'record patient details', 'prescribe a medication', 'suggest the correct treatment for a patient', or 'make an appointment with your doctor'. Its resource structure contains only information relevant to a particular API call so it can be transmitted efficiently. Specific minimum data sets of FHIR (called Cores) are required in the USA with other countries currently developing additional Core requirements. Using FHIR allows different entities in that country streamlines continuity of care by exchanging information between systems as well as between organizations.

Longitudinal analysis

Observational Health Data Sciences and Informatics' (OHDSI) Observational Medical Outcomes Partnership Common Data Model (OMOP) ¹¹ is an open standard designed for longitudinal interrogation of clinical records. Its use of a patient-centric relational model makes it suited for complex queries across many patients' records to discover patterns and trends using statistical and machine learning techniques.

As the requirements for each data standard mean that one standard cannot replace the others, harmonization projects are led by community volunteers to provide standard transformations from one standard to another. For example, the collaboration between OHDSI and HL7 manifests as a common working group with weekly meetings with the aim of producing an implementation guide for algorithms that convert data from FHIR to OMOP and from OMOP to FHIR. Similar projects for harmonizing openEHR with OMOP, and FHIR with openEHR, provide a standard architecture for effective data interoperability and reuse that will create a virtuous cycle of growing community participation, improving the standards and simplifying harmonization. Such volunteer groups are highly beneficial, not just to the standards they are dedicated to, but to all healthcare systems globally.

Which standard to use?

Most users will easily choose whether their needs' focus is operational care delivery or analytics, and identify which systems are involved. Figure 1 provides a high level key. They would then find it easy to choose whether they should use openEHR, OMOP or FHIR, and whether they need to invest in Extract-Transform-Load processes to change data representations from one to another. An interoperable health system would use openEHR to collect data, FHIR to transmit data between systems and organizations, and OMOP to find insights in the data.

It is possible for transformations from one standard to another to be "lossy". For example:

- fidelity may be lost when converting from a specific term (e.g. 'fracture of the fourth metatarsal on the left foot') to a more general one (e.g. 'fracture of the fourth metatarsal'),
- information can be lost when information captured in one standard (e.g. the time drug administration was witnessed in FHIR) has no place in another standard (e.g. OMOP),
- relationships may be lost when information has to be split (e.g. a diagnosis and a drug prescription), or
- accuracy may be lost when converting to a standard that requires information not found in the other (e.g. the time a drug was stopped is required for every exposure in OMOP and may need to be estimated whenever it is not present in the dataset).

Consideration should be given to minimizing such losses. These approaches can utilize tacit and/or implicit knowledge about the dataset, external knowledge or metadata.¹²

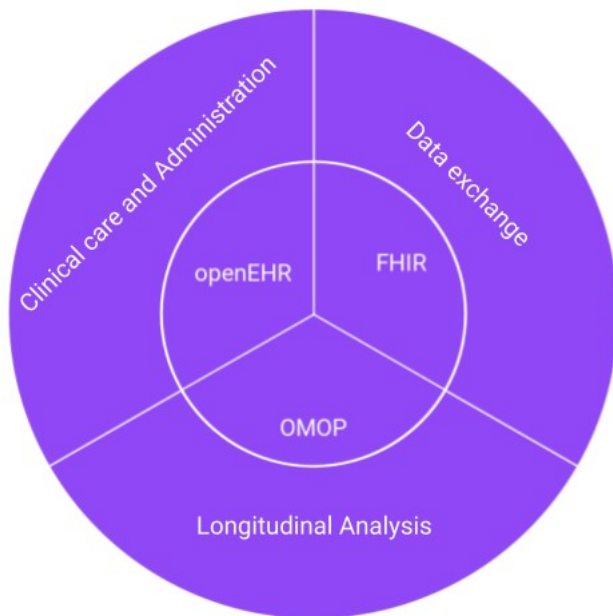


Figure 1: Alignment between domain (outer circle) and open standard (inner circle).

Conclusion

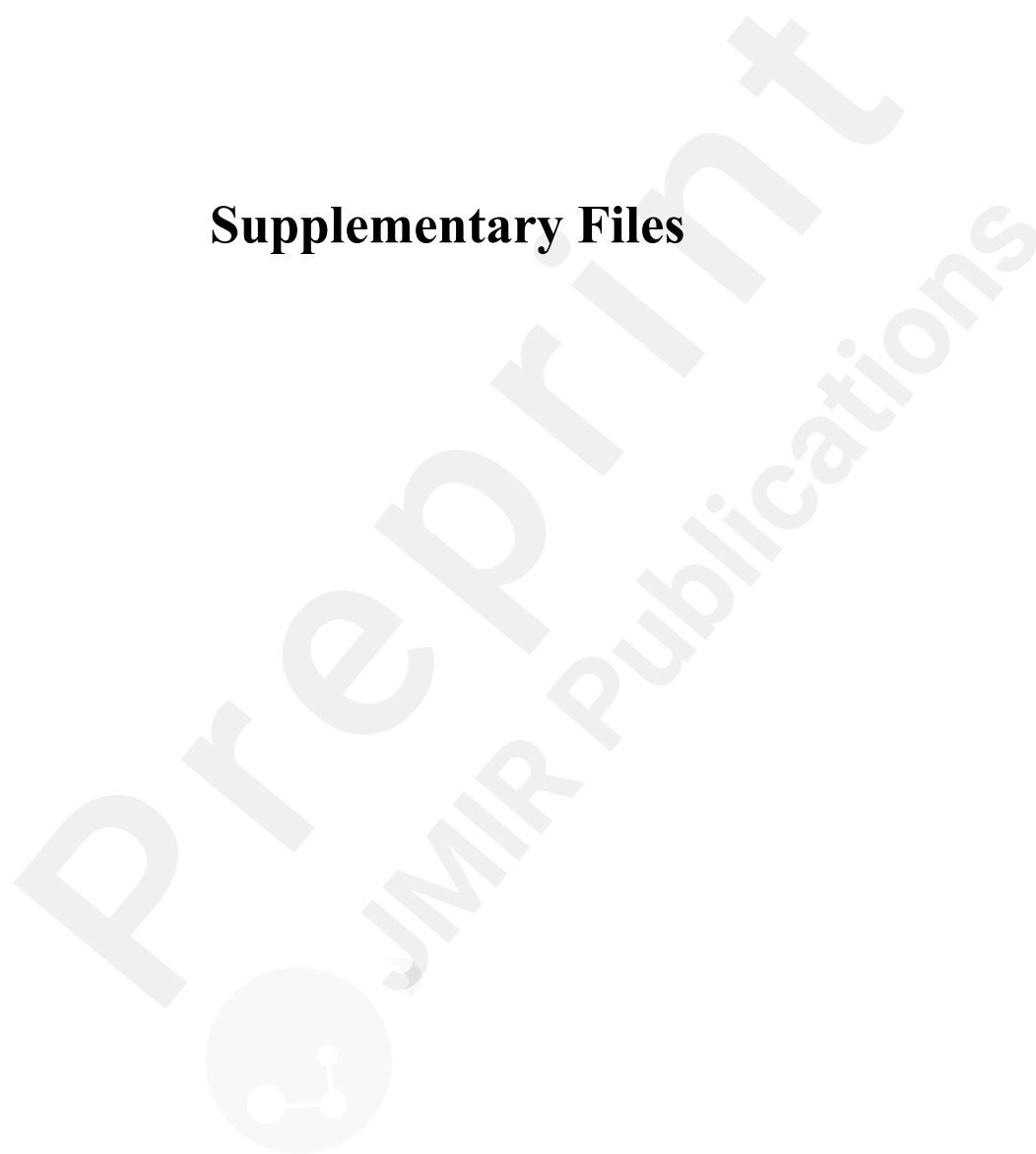
Open standards, backed by engaged communities, hold an advantage over proprietary ones. They enable broader contributions and testing, fostering continual improvement. However, achieving true interoperability requires harmonization among standards, and a pathway for migration from other standards and proprietary models. Three open standards are particularly suited to their respective healthcare domains: openEHR for clinical care and administration, FHIR for data exchange and OMOP for longitudinal analysis. The most important aspect of every standard is its community and all three have active and growing communities that continue to use and improve these standards.

References

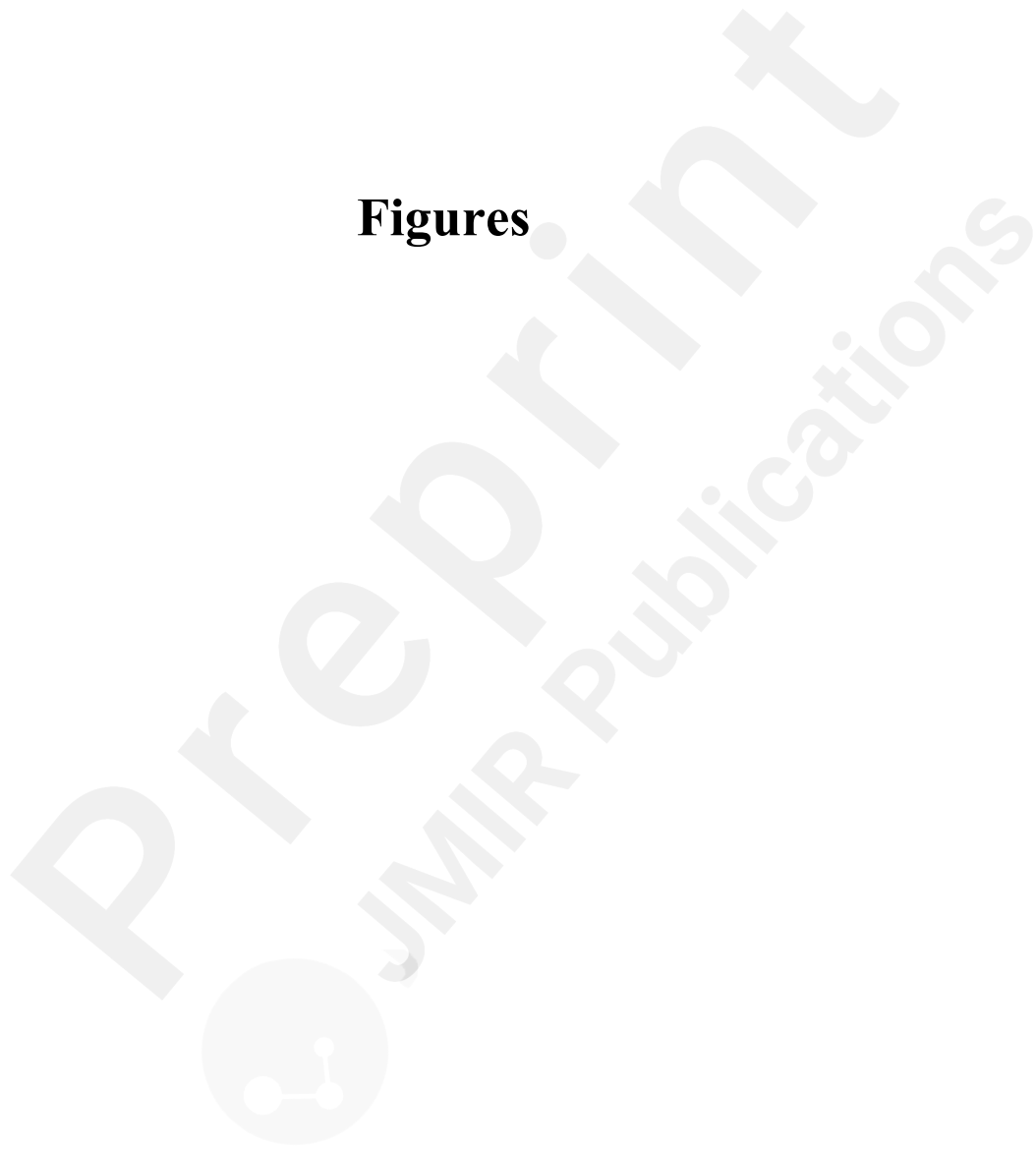
1. Coiera E. *Guide to Health Informatics*. CRC Press; 2015. doi:10.1201/b13617
2. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. *npj Digital Med*. 2019;2:79. doi:10.1038/s41746-019-0158-1
3. Ahmadi N, Zoch M, Kelbert P, et al. Methods used in the development of common data models for health data: scoping review. *JMIR Med Inform*. 2023;11:e45116. doi:10.2196/45116
4. Handler T, Holtmeier R, Metzger J, Overhage M, Taylor S, Underwood C. HIMSS Electronic Health Record Definitional Model. *Health Information Management Systems Society*. 2003;2003:1-8.
5. Technology Office of the National Coordinator for Health Information. United States Core Data for Interoperability (USCDI). 2018.
6. Coiera E. The standard problem. *J Am Med Inform Assoc*. August 31, 2023. doi:10.1093/jamia/ocad176
7. Reynolds CJ, Wyatt JC. Open source, open standards, and health care information systems. *J Med Internet Res*. 2011;13(1):e24. doi:10.2196/jmir.1521

8. de Mello BH, Rigo SJ, da Costa CA, et al. Semantic interoperability in health records standards: a systematic literature review. *Health Technol (Berl)*. 2022;12(2):255-272. doi:10.1007/s12553-022-00639-w
9. Kalra D, Beale T, Heard S. The openEHR Foundation. *Stud Health Technol Inform*. 2005;115:153-173.
10. Benson T, Grieve G. Principles of FHIR. In: *Principles of Health Interoperability: SNOMED CT, HL7 and FHIR*. Health information technology standards. Springer International Publishing; 2016:329-348. doi:10.1007/978-3-319-30370-3_18
11. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-578.
12. Bönisch C, Kesztyüs D, Kesztyüs T. Harvesting metadata in clinical care: a crosswalk between FHIR, OMOP, CDISC and openEHR metadata. *Sci Data*. 2022;9(1):659. doi:10.1038/s41597-022-01792-7

Supplementary Files



Figures



Alignment between domain (outer circle) and open standard (inner circle).

