Check for updates

STUDY PROTOCOL

Mapping the international ecosystem of national health data spaces. A scoping review protocol

[version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]

Lauren Maxwell

Heidelberger Institut für Global Health, Universitätsklinikum Heidelberg, Heidelberg, Baden-Württemberg, 69120, Germany

v1 **First published:** 06 Jun 2023, **3**:87
https://doi.org/10.12688/openreseurope.15982.1
Latest published: 06 Jun 2023, **3**:87
https://doi.org/10.12688/openreseurope.15982.1

Abstract

Background: The reuse of participant-level health data by public health and surveillance institutions, hospitals, doctors, and patients is an emerging priority for a number of national governments. Technical and semantic interoperability of health data ecosystems is important for detecting and responding to global health challenges, including emerging infectious diseases, antimicrobial resistance, and vaccine-preventable illnesses. In this scoping review, we will identify and describe health data ecosystems, spaces, clouds, and commons, national-level mechanisms for enabling the reuse of participant-level health data.

Methods and analysis: We will apply the Arskey and O'Malley scoping review approach to describe governance, content, and semantic and technical interoperability of data and metadata in national health data ecosystems. We selected a scoping rather than a systematic review methodology to provide a high-level analysis of the current state of health data ecosystems' implementation of the FAIR principles for data resources. The systematic search strategy was pilot tested and tailored for Ovid(Medline), CINAHL, and Web of Science. We will also conduct web scraping and consult stakeholders to identify additional health data ecosystems. Two reviewers will conduct the title-abstract and full-text screening and data charting independently. Discrepancies will be resolved by consensus, and results will be summarized in narrative form.

Ethics and dissemination: Ethical approval is not required for this scoping review of published studies and grey literature. The scoping review protocol was registered prior to initiating the search strategy. Study results will be submitted for publication in an Open Access journal.

Keywords

Scoping review, health data ecosystem, health data space, health data commons, data sharing, data reuse

Open Peer Review

Approval Status ? ✓ ✗

	1	2	3
version 1 06 Jun 2023	? view	✓ view	✗ view

1. **Mate Beštek**, IT University of Copenhagen, København, Denmark
2. **Rada Hussein**, Ludwig Boltzmann Gesellschaft, Vienna, Austria
3. **Sandra Geisler**, RWTH Aachen University, Aachen, Germany

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Horizon 2020 gateway.



This article is included in the Research Infrastructures collection.

Corresponding author: Lauren Maxwell (lauren.maxwell@uni-heidelberg.de)

Author roles: Maxwell L: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The drafting of this protocol was supported by the ReCoDID project, which is funded by the EU Horizon 2020 research and innovation programme (grant agreement 825746) and the Canadian Institutes of Health Research (CIHR) Institute of Genetics (grant agreement 01886-000).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Maxwell L. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Maxwell L. **Mapping the international ecosystem of national health data spaces. A scoping review protocol [version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]** Open Research Europe 2023, **3**:87 https://doi.org/10.12688/openreseurope.15982.1

First published: 06 Jun 2023, **3**:87 https://doi.org/10.12688/openreseurope.15982.1

Abbreviations

CODATA: Committee on Data; DMP: Data Management Plan; FAIR: findable, accessible, interoperable, reusable; FIP: FAIR Implementation Profile; LILACS: Literatura Latinoamericana y del Caribe en Ciencias de la Salud; MEDLINE: Medical Literature Analysis and Retrieval System Online; MeSH: Medical Subject Headings; PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols; VRE=virtual research environment

Introduction

National health data ecosystems, spaces, virtual research environments (VREs), or data commons have been or are being launched across a number of geographies to enable the reuse of participant-level health data. Health data ecosystems are especially important for enabling machine learning and related artificial intelligence (AI)-based approaches to health data analysis and can facilitate public health surveillance, observational clinical research, and the design and conduct of randomized controlled trials (RCTs) by enabling global analyses of health data.

How health data ecosystems implement the FAIR (findable, accessible, interoperable, reusable) principles[1] and data reuse-related governance, including legal and ethical concerns, has important consequences for cross-ecosystem semantic and technical interoperability and data access. When considering the adoption of data capture or exchange standards for discovery or descriptive metadata or participant-level data, using the same standards or standards that have established crosswalks can support cross-ecosystem interoperability. Similarly, interoperability, transparency, and machine actionability of ecosystem governance, including how consent for data or sample reuse is operationalized at the participant level, is important for enabling cross-ecosystem and cross-domain analyses. In Table 1, we provide an overview of the different levels of interoperability important to consider in cross-ecosystem analyses of participant-level health data.

FAIR implementation profile

Whether and how FAIR is implemented has consequences for how metadata and participant-level data can be reused. FAIR convergence, as when related stakeholders take the same or similar routes to implementing FAIR, can help foster interoperability in the health data space where there are a high number of standards for the capture or exchange of health data and for which there is limited cross-ontology interoperability[2]. Open collaboration on FAIR implementation also helps ensure that stakeholders reuse available resources and standards rather than developing new ones.

To enable FAIR convergence, stakeholders have to gain an in-depth understanding of how FAIR has been implemented in similar initiatives, which resources have been used, and why these choices have been made. The GO FAIR project[3] has developed machine actionable FAIR Implementation Profiles (FIPs)[4], which build on prior work to develop machine actionable data management plans (DMPs)[5], and which make visible the choices communities of practice make when implementing FAIR[6].

Objectives

In this project, we will use a scoping review approach to 1) identify health data ecosystems and 2) describe how these health data ecosystems have implemented FAIR. We will use this information to develop a draft health data ecosystem FIP which will be refined through a series of meetings with a diversity of stakeholders, including data engineers, ethicists, legal experts, and patients, that interact with health data ecosystems. The agreed-upon FIP will be completed by health data ecosystem stakeholders (e.g. project managers, infrastructure experts). The cross-ecosystem analysis of FIP results will

Table 1. Levels of interoperability across health data ecosystems.

Laws & policies	Data access laws
Governance	Guidelines & agreements for accessing data
Data flows	Process for submitting & accessing data
IT infrastructure	Interfaces between infrastructures (e.g., APIs for discovery, federation, analysis); storage; compute
Security measures	Authentication, authorization
Information	Shared definitions & content, interoperable participant-level data & metadata standards & shared content
Analytic outputs	Machine-readable outputs
KPIs	Shared metrics

KPI=key performance indicator; IT=information-technology

inform a roadmap for FAIR data convergence and addressing and aligning legal and ethical issues through machine-actionable, interoperable governance systems.

The objective of this scoping review is to identify and describe health data ecosystems with a focus on how FAIR ecosystems implement FAIR and address legal and ethical concerns. This scoping review is part of the Committee on Data for Science and Technology (CODATA) Health Data Case Study.

We will work with a representative stakeholder group to:

a. develop commonly agreed upon FAIR implementation profiles (FIP) and dedicated machine-readable metadata templates for health data ecosystems

b. clarify and align priorities for cross-ecosystem FAIR convergence

Methods

We will conduct a rapid systematic review to identify and describe ongoing and planned health data ecosystems with a focus on their implementation of FAIR and governance structures. The systematic review protocol will be created in keeping with the PRISMA-P guidelines[7], and the systematic review will be published in keeping with the 2020 PRISMA guidelines[8].

Stage I: Identifying the research question

The initial review questions were developed through consultation with the research team and the Committee on Data for Science and Technology (CODATA)[9] leadership. These questions will be refined with feedback from national health data ecosystem stakeholders.

Initial review questions

1. What are the different definitions for national health data ecosystems, spaces, commons, etc.?
2. What national health data ecosystems, spaces, commons exist or are planned?
3. Who are the relevant stakeholders, and what are their priorities for cross-ecosystem convergence?
4. Which stakeholders were engaged in the development of health data ecosystems, and how were they engaged?
5. What domains and types of health data are included?
6. How have health data ecosystems implemented FAIR?
7. How are decisions made about data reuse, and what is the decision-making process?
8. What are the similarities and differences in data flow and governance from data capture to data reuse?
9. How do different stakeholders define the "success" of health data ecosystems?

Stage II: Identifying relevant health data ecosystems, spaces, commons

We will systematically search Ovid (Medline), Web of Science, and Literatura Latinoamericana y del Caribe en Ciencias de la Salud (LILACS) to identify national health data ecosystems, spaces, commons, or other infrastructures that enable the reuse of participant-level health data. We will use a combination of MeSH and text terms tailored to each database and export citations to EndNote Version 7 for deduplication.

Search strategy

The search strategy includes MeSH and text terms related to data sharing. The search strategy has been tailored to each database and piloted. We will not limit the search by geography or language of publication and will search all databases from January 1, 2000 to the date the search is run. The searches will not be further restricted by location or language of publication. The search strategies are presented in Table 2.

Stage III: Study selection and eligibility criteria

Two independent reviewers will conduct title-abstract and full-text screening and data charting, which will be managed in COVIDENCE systematic review software[10]. Disagreements will be resolved through consensus. We will develop a web scraping algorithm to identify additional health data spaces. The list of country-level health data ecosystems identified through the search will be circulated to stakeholders to identify additional ecosystems through reverse snowball sampling.

Inclusion criteria

Health ecosystems or spaces and health data commons included in this review will be country-level systems that make available participant-level data and related metadata from health systems or research.

Exclusion criteria

Health ecosystems or spaces and health data commons that serve as metadata catalogues but do not provide some means of linking datasets to end users, either through allowing health data downloads, access in a secure VRE, or federated access without directly interacting with the data, will be excluded.

Stage IV: charting the data

We will develop and pilot a data extraction and charting form to facilitate the descriptive synthesis of our findings prior to beginning the charting process. Key data that will be extracted from included studies or reports are presented in Table 3.

Data extraction and data charting will be completed independently by two reviewers in COVIDENCE systematic review software[10].

Stage V: collating, summarizing and reporting results

Results of the data charting process will be presented in a narrative form. Results of the systematic search will be reported

Table 2. Database-specific search strategy.

	Ovid(Medline)	Web of Science	LILACS
1	((health care OR healthcare OR digital health OR health data OR healthcare data OR health care data) adj2 (space* OR ecosystem* OR cloud OR commons)).ti,ab	TI=((health care OR healthcare OR digital health OR health data OR healthcare data OR health care data) adj2 (space* OR ecosystem* OR cloud OR commons)) OR AB=((health care OR healthcare OR digital health OR health data OR healthcare data OR health care data) near/2 (space* OR ecosystem* OR cloud OR commons))	(ti:((health care OR healthcare OR digital health OR health data OR healthcare data OR health care data) near/2 (space* OR ecosystem* OR cloud OR commons)) AND (db:("LILACS"))
2	(Virtual research environment OR VRE).ti,ab	TI=(Virtual research environment OR VRE) OR AB=(Virtual research environment OR VRE)	(ti:(virtual research environment OR VRE) OR AB=(virtual research environment OR VRE) AND (db:("LILACS"))
3	1 or 2	1 or 2	1 or 2
4	exp animals/not humans.sh	TS=(animal model or animal* NOT human)	
5	3 not 4	3 not 4	
6	limit 6 to yr= "2000 -Current"	PY=(2000-2021)	DA:20$

LILACS=Literatura Latinoamericana y del Caribe en Ciencias de la Salud

in a PRISMA flow diagram. The data charting process will serve as the basis for the draft FIP.

Stage VI: stakeholder consultation & involvement

Stakeholder engagement is key to the conduct and potential impact of this scoping review exercise. We will leverage our work with CODATA, an international consortium focused on enabling cross-disciplinary collaborations and open science, to engage stakeholders in developing and implementing the data charting strategy. We will use findings from the scoping review to develop a draft FIP for health data ecosystems. We will interact with stakeholders in the health data ecosystem space to refine the list of key metadata. We will convene contacts in a series of working meetings with stakeholders to review:

1. the overall concept of a FIP and FAIR convergence
2. a draft FIP that includes key information on FAIR implementation and governance structures
3. revise FIP
4. complete FIP
5. review findings of the cross-FIP analysis
6. develop a roadmap for FAIR convergence within the national health data ecosystem space

We expect to hold six three-hour meetings over a 6-month period. We will develop a living network graph to present the international ecosystem of health data spaces that is tied to updates in each group's FIP.

Cross-domain alignment of FAIR

Health does not exist in a vacuum. Efforts to improve the adoption of FAIR in climate, ecology, agriculture, conflict, education, etc., have important implications for understanding the social determinants of health and the distribution of disease. Climate change is directly related to the emergence of novel pathogens and to epidemics of well-characterized non-communicable and communicable diseases. By understanding how health data ecosystems implement FAIR, CODATA will help highlight areas for cross-domain convergence (e.g. metadata ontologies). Moving towards FAIR, cross-domain metadata will also help enable the One Health approach[11], where animal and ecosystem health are considered alongside human health.

The key metadata identified by stakeholders in the health ecosystem space will be the basis for cross-domain metadata content and interoperability conversations facilitated by CODATA. We will conduct a final cross-domain stakeholder meeting with health data ecosystem implementers and WorldFAIR[12] to review opportunities for cross-domain FAIR convergence[2].

By documenting how health data ecosystems operationalize governance, including legal and ethical concerns, CODATA will facilitate conversations with stakeholders around opportunities to align approaches to interpreting shared concerns. Lastly, understanding what ontologies and tools or approaches are used by existing health data ecosystems will help health data ecosystems that are in the formative stages use existing tools and ontologies, which, in turn, facilitates semantic and technical interoperability across health data ecosystems.

Ethics and dissemination

This scoping review will summarize evidence from published studies and, as such, does not require ethical approval. We will publish the scoping review results in an open-access journal. We will work with stakeholders to produce a roadmap for FAIR convergence and the alignment of health data ecosystem governance. The network graph of FAIR implementation choices across ecosystems and the supporting data will be made publicly available as a website under a creative commons license. Our findings will be disseminated in multiple contexts to facilitate public engagement in the conversation around health data reuse across health data ecosystems.

Table 3. Draft items for data charting.

Domain	Items
Health data ecosystem	• Name • Years of development • Years of implementation • Development process • Description of resource, including classification (e.g. ecosystem, commons)
Content	• Health domains covered (e.g. demographic data, including race, education, income; vital signs; medicines administered; routine care; sexual and reproductive health; health outcomes; social outcomes) • Data source (e.g., health system/routine clinical care; research; biobank; census) • Data types (e.g., clinical, genetic, imaging) • Types of metadata collected • Types of populations included or excluded (e.g. indigenous, minors)
Data & metadata interoperability	• Metadata, participant-level data capture, data exchange formats applied
Technical approach to data submission and reuse	• How data submission takes place • How data reuse takes place • How these are described (i.e. is the description machine actionable)
Legislative environment	• Is ecosystem written into legislation? • Is the funding stream written into legislation? • Is there legislation around data reuse?
Governance of data reuse	• Governance mechanism • How governance mechanism described • How governance mechanism varies by type of end user (e.g. national government, local government, academic partner, industry) • Types of data that can be reused • Types of users • Types of use cases
Ethical concerns	• How and whether participant- or institutional-preferences for reuse (e.g. broad consent) is operationalized • How local or cultural concerns are considered • How benefit sharing with patients/participants, research teams, institutions, local communities is implemented and evaluated
Funding mechanism & sustainability	• Type of survey (cross-sectional, longitudinal) • How survey disseminated • How survey administered (self administer paper form, online survey) • How survey analyzed
Evaluation & review timeline & process	• How is reuse measured? • How is KT or CRT measured? • What other ways is success measured?

CRT=clinical research translation; KT=knowledge translation

Conclusion

This scoping review will provide a comprehensive overview of the rapidly developing ecosystem of national health data spaces and ecosystems. Information from this review will be used to identify challenges and opportunities building technical and semantic interoperability, shared implementation of the FAIR principles, and alignment of governance structures within the national health data ecosystem space and beyond.

Study status

The current study status and anticipated timeline are presented in the GANTT chart in Table 4.

Table 4. GANTT Chart.

	2023									2024	
	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb+
Pilot test MeSH and text term strategy											
Pilot test web scraping search strategy											
Implement systematic search and web search											
Identify representatives from stakeholder groups											
Convene small stakeholder group to refine domains and items for data extraction											
Screen title-abstract and full texts of publications identified through systematic search of electronic databases											
Conduct web scraping and search of grey literature sources											
Identify health data ecosystem contacts for FAIR, policy, legal, technical, and KT-domains											
Conduct reverse snowball sampling where identified health data ecosystems are asked to identify additional ecosystems					SE	SE	SE	SE	SE	SE	SE
Complete initial data charting											
Introduce larger group health data ecosystem stakeholders to FIP tool					SE						
Draft FIP tool with data charting findings											
Refine draft FIP with feedback from health data space developers, technical, policy, and ethical stakeholders						SE					
Complete initial FIP tool with health data ecosystem stakeholders							SE				
Create network graph to make health data spaces FAIR implementation choices observable									D		
Draft and submit findings for publication										D	
Collaborate with stakeholder group to create a health data space interoperability roadmap								SE	SE	SE	D
Disseminate findings to patient advocate groups, policy makers, funders										D	
Host cross-disciplinary alignment meeting										SE	

* D=dissemination activity; FAIR=findable, accessible, interoperable, reusable; FIP=FAIR implementation profile; KCT=knowledge and clinical practice translation; SE=stakeholder engagement

Ethics and consent

Ethical approval and consent were not required.

Data availability

No data are associated with this article.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data.* 2016; **3**(1): 160018.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Sustkova HP, Hettne KM, Wittenburg P, *et al.*: **FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources.** *Data Intelligence.* 2020; **2**(1–2): 158–70.
Publisher Full Text
3. **GO FAIR Initiative.** (accessed December 10 2021).
Reference Source
4. Schultes E, Magagna B, Hettne KM, *et al.*: **Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence.** Cham: Springer International

Publishing; 2020; 138–47.
Reference Source

5. Pergl R, Hooft R, Suchánek M, *et al.*: **"Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning.** 2019; **18**(1): 59.
Publisher Full Text

6. **FAIR Implementation Profile.** (accessed September 14 2022).
Reference Source

7. Moher D, Shamseer L, Clarke M, *et al.*: **Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.** *Syst Rev.* 2015; **4**(1): 1.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Page MJ, Moher D, Bossuyt PM, *et al.*: **PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.** *BMJ* 2021; **372**: n160.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. **CODATA's Mission.** 2022; (accessed September 22 2022).
Reference Source

10. Covidence: **Covidence Systematic Review Software.** Melbourne, Australia; 2017.

11. Lebov J, Grieger K, Womack D, *et al.*: **A framework for One Health research.** *One Health.* 2017; **3**: 44–50.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. **WorldFAIR: Global cooperation on FAIR data policy and practice – a major two-year project starts today!** 2022.
Reference Source

Open Peer Review

Current Peer Review Status: ? ✓ ✗

Version 1

Reviewer Report 19 October 2023

https://doi.org/10.21956/openreseurope.17256.r34861

© **2023 Geisler S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✗

Sandra Geisler
RWTH Aachen University, Aachen, Germany

The article describes the conduction of a study in the field of health data ecosystems. Particularly the aspects of how FAIR has been implemented in existing health data ecosystems are the aim of the study.

My assessment was not quite positive because of the following major concerns and comments:

- Crucial definitions are missing. It is not clear how the authors define a (health) data ecosystem or a (health) data space.
- Methods are not sufficiently described. For example, it is not clear, what is a Arskey and O'Malley scoping review and it is also not described in the report. It is also not clear what is a PRISMA flow diagram,
- The selection of search engines was not reasoned. Why these and not others?
- The authors claim that they will not filter the papers regarding language. But will they translate papers from all possible languages?
- The text is written very lengthy and many sentences are far to complicated and long.
- The description of the aims in the beginning is very abstract and not getting very concrete / stays only at the surface. For example, it is not clear what FAIR convergence means exactly.
- Table 2 are not levels for me. In my opinion they are very rough and high level aspects of data ecosystems. How these have been assembled remains unclear.
- The content of Table 3 is not sufficiently explained
- The search terms are very broad (e.g., cloud, healthcare, commons). Why are the search terms for the three databases different? Why is FAIR not contained?

Is the rationale for, and objectives of, the study clearly described?
No

Is the study design appropriate for the research question?
Partly

Are sufficient details of the methods provided to allow replication by others?
No

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Data ecosystems, data spaces, FAIR data, RDM

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 07 September 2023

https://doi.org/10.21956/openreseurope.17256.r34870

© 2023 Hussein R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rada Hussein
Institute for Digital Health and Prevention, Ludwig Boltzmann Gesellschaft, Vienna, Vienna, Austria

The scoping review aims to identify health data ecosystems and describe how existing health data ecosystems have implemented FAIR. As a result, a commonly FAIR implementation profile (FIP) will be developed. This FIP will be discussed and refined with various stakeholders, including, WorldFAIR cross-domain experts, to clarify and align priorities for cross-ecosystem FAIR convergence.

The protocol is clear and well-designed. However, in Table 1, the level of interoperability across health data ecosystems can be organized and or grouped to match the common levels of interoperability, namely: semantic, structural, functional, and technical interoperability.

The planned work mainly addresses the technical and governance aspects of how these health data ecosystems have implemented FAIR. I suggest (if applicable within the scope of the ReCoDID project) to consider the other social factors that affect the “success” of health data ecosystems (as mentioned under the initial review question #9), e.g., public engagement in addition to capacity

building of researchers/healthcare workforce to engage with health data. In this way, the results can be widened to create the required synergy with the recommendations of the current call for action, like turning FAIR into reality in Europe, and how to expand to international levels. On the other hand, that may accelerate the adoption of the available FAIR Cookbooks on a wider scale.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health Informatics, EHDS, Healthcare standards and interoperability

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 September 2023

https://doi.org/10.21956/openreseurope.17256.r34025

© 2023 Beštek M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Mate Beštek
IT University of Copenhagen, København, Denmark

The paper describes a study protocol that would perform a scoping review for concepts similar or linked to health data spaces including health data ecosystems, spaces, clouds, commons, and national-level mechanisms that enable reuse of participant-level health data.

In that sense, the study is interesting and timely as there are a lot of changes anticipated with the new European health data spaces legislation.

The study introduces Arskey and O'Malley's scoping review approach as the one to be used for conducting evidence synthesis. However, the paper then further uses both concepts of a systematic review and scoping review interchangeably. Therefore I would suggest that the author introduce a more precise use of the correct approach that will eventually be used in the study. I also attach a citation to a paper by Munn et.al [1] that can perhaps help with choosing a scoping

review VS a systematic review.

The objectives of the study are defined as 1) identification of health data ecosystems and 2) description of how these health data ecosystems have implemented FAIR. At this point, I have not been clear about the role of the other concepts used in the abstract, namely, spaces, clouds, commons and national-level concepts for enabling the reuse of participant-level health data. Further on, the concept of commons is used simultaneously with health data commons later in the paper. Health data commons is a subtype of commons and has a different set of guiding principles or design principles in comparison to the traditional commons. I do feel that health data commons might be more appropriate in this paper. But I would also suggest noting the semantic commons which do try to address the semantic interoperability aspect of health data commons in several publications.

In Stage I the initial review questions are stated. I was not able to clarify if the questions would be enough to identify approaches to governing health data that focus on the data design phase in comparison to data curation approaches where we assume the existence of data in whatever form and then different methods are used to align ontologies that specify the schema of the datasets.

Further on, in Stage II, the concept of infrastructures is used as a synonym for health data ecosystems, spaces and commons. The concept of infrastructures is used in information systems literature and is an important concept. I wonder, however, if information infrastructures were pointed to. In any case, I would kindly ask for the author to introduce the correct concept in the beginning so it does not confuse the reader.

Within the same section, Stege II, Table 2 is referenced. I would suggest a short description of the parameters in the table to improve clarity.

In Section III, reverse snowball sampling is introduced. I would suggest introducing the method at the beginning of the paper to make the methods clearer.

In this section, the health data commons are introduced. As I commented at the beginning of this review, I suggest clarifying the correct commons and what it entails.

In Stage VI a few more concepts are introduced. The living network graph and stakeholder engagement.

Also, within the concept of cross-domain alignment of FAIR, the focus is on metadata that will be used for defining the alignment. At this point, I understand the need for such alignment, but I also think about the underlying FHIR governance principles which are not yet clear. I would benefit from a more concise description.

Overall, I will greatly anticipate the results of the study and look forward to reading the final version in a journal as anticipated by this paper.

References

1. Munn Z, Peters MDJ, Stern C, Tufanaru C, et al.: Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach.*BMC Med Res Methodol*. 2018; **18** (1): 143 PubMed Abstract | Publisher Full Text

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Partly

Are sufficient details of the methods provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health data platforms, ecosystems, spaces and commons. Interoperability, especially semantic interoperability and the governance approaches that enable design time interoperability in comparison to data curation approaches that are usually used after the data already exists.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.