Check for updates

STUDY PROTOCOL

What does good governance of federated biomedical data reuse look like? A scoping review protocol.

[version 1; peer review: awaiting peer review]

Lauren Maxwell [1,2], Melanie Stecher [3], Maria Panagiotopoulou [4], Monica Falk [5]

[1]Ecraid Foundation, Utrecht, Utrecht, 3584 BA, The Netherlands
[2]Heidelberger Institut für Global Health UniversitatsKlinikum Heidelberg, Heidelberg, Baden-Württemberg, Germany
[3]Department of Infection Control and Vaccines, The Norwegian Institute of Public Health, Oslo, Oslo, 0456, Norway
[4]European Clinical Research Infrastructure Network, Paris, Île-de-France, France
[5]Infection Control and Emergency Preparedness, The Norwegian Institute of Public Health, Oslo, Oslo, 0456, Norway

v1 **First published:** 25 Sep 2026, **6**:377
https://doi.org/10.12688/openreseurope.24344.1
Latest published: 25 Sep 2026, **6**:377
https://doi.org/10.12688/openreseurope.24344.1

Abstract

Background

Federated reuse of participant-level biomedical data enables public health institutions, healthcare providers, and researchers to analyse data without moving it from its source, providing access to a more diverse data ecosystem. Data governance determines whether this reuse is trustworthy, addressing the legal, ethical, and technical questions raised when decentralised analysis balances privacy with utility. Effective governance is central to protecting participant rights, ensuring benefit sharing, and supporting equity while enabling collaboration across data sources. This scoping review will identify and synthesise recommendations for data governance in the federated reuse of biomedical data, focusing on how approaches address participant rights, benefit-sharing, inclusivity, and legal and ethical compliance, and how they vary across reuse mechanisms.

Methods

We will apply the Arksey and O'Malley framework to map current recommendations and best practices and to identify gaps. A search strategy developed in line with the PRESS guidelines was piloted and tailored for Ovid (MEDLINE), EMBASE, and Web of Science, and covers the period from January 2000 to the search date, without language or geographic limits. Title and abstract screening will be conducted by

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

one reviewer with ASReview as the second reviewer; full-text screening by two independent reviewers; and data extraction by one reviewer, verified by a second. Stakeholder consultation will refine the charting framework and the interpretation of findings.

Results

As part of the review protocol, this article reports no results. The completed review will map governance models, chart how each addresses participant rights and benefit-sharing, and compare approaches across data types, technical architectures, geographies, and legal contexts.

Conclusions

The review will provide the first structured synthesis of data governance for federated reuse of biomedical data, informing efforts to build transparent, equitable, and scalable frameworks. Ethical approval is not required; the protocol was registered with Open Research Europe before searching, and results will be published open access.

Plain Language summary

Health research increasingly depends on combining information from many hospitals, studies, and countries. Moving sensitive patient data between these places raises privacy, legal, and trust concerns. One alternative is “federated” data reuse, where the analysis is sent to the data rather than the data being gathered in one place. The data stays where it is held, and only the results of the analysis are shared. This approach can help protect privacy while still allowing researchers to learn from large and diverse groups of people, which is especially valuable for rare diseases and for responding quickly to outbreaks.

For federated data reuse to be trustworthy, it requires strong data governance: clear rules on who can access the data, for what purposes, how participants’ rights are respected, and how the benefits of research are shared fairly. At present, these rules are scattered across technical, legal, and ethical sources and are described inconsistently.

This scoping review will gather and map the recommendations and practices for governing the federated reuse of health and biomedical data. We will search three research databases, screen the results, and summarise what is known about different governance models, how they protect participants, and how they comply with laws such as the GDPR. We will also consult people who design and use these systems. The findings are intended to help researchers, institutions, and policymakers build clearer, fairer, and more consistent governance for health data reuse in Europe and beyond.

Keywords
federated data reuse, data governance, scoping review, biomedical data reuse, benefit sharing, informed consent, General Data Protection Regulation

This article is included in the Horizon Europe gateway.

This article is included in the Research Ethics and Integrity collection.

Corresponding author: Lauren Maxwell (lauren.maxwell.us@gmail.com)

Author roles: Maxwell L: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Stecher M**: Methodology, Project Administration, Writing – Review & Editing; **Panagiotopoulou M**: Writing – Review & Editing; **Falk M**: Funding Acquisition, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project is funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101136531 (Coordination Mechanism for Cohorts and Trials [CoMeCT]). The funder and sponsor had no role in the development of this protocol.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Maxwell L *et al*. 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, Stecher M, Panagiotopoulou M and Falk M. **What does good governance of federated biomedical data reuse look like? A scoping review protocol. [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:377 https://doi.org/10.12688/openreseurope.24344.1

First published: 25 Sep 2026, **6**:377 https://doi.org/10.12688/openreseurope.24344.1

Abbreviations

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

What is federated data reuse?

Federated data reuse, also referred to as federated, distributed, or multiparty learning, enables data reuse without moving it from its original location or allowing data users to access it directly. This approach is particularly beneficial for sensitive data, such as healthcare records, as it allows collaborative model training across decentralised data silos while preserving data privacy.[1] A centralised or decentralised aggregator facilitates this process by integrating model updates that may vary by machine learning method (e.g., neural networks, decision trees, or gradient boosting). Federated learning is especially impactful in healthcare, enabling the development of artificial intelligence (AI) models for rare diseases, early disease detection, and medical surveillance without compromising data security. This method contrasts with other approaches, such as edge computing or swarm learning, which further decentralise computation.[2] Key governance challenges include ensuring data security through differential privacy, deciding between centralised and decentralised aggregation, and fostering collaboration among competing stakeholders. Addressing these issues is essential for building trust in federated reuse of biomedical data and samples.[3]

Federated learning offers numerous advantages over centralised data reuse. Keeping data at its source enhances privacy, strengthens privacy protection, supports compliance with regulations such as the General Data Protection Regulation (GDPR), and reduces the risk of breaches. It also optimises bandwidth and storage costs by minimising data transfer, making it scalable and suitable for real-time applications. Federated learning supports inclusivity by enabling participation from rural areas with limited connectivity and leveraging diverse, siloed datasets to improve model generalizability.[1] Empirical evidence suggests that federated approaches can achieve performance comparable to that of centralised models, particularly in rare diseases and public health surveillance. However, federated learning poses challenges, including security risks arising from decentralised data processing and the need for reliable systems that continue to function during node failures or connectivity disruptions. Initial setup and ongoing maintenance are resource-intensive and require skilled personnel, which can be a limiting factor.[1] Additionally, federated learning is constrained by certain algorithms, such as tree-based models, and must manage heterogeneous computing and storage requirements. These benefits and challenges highlight the importance of robust governance frameworks to maximise potential while addressing limitations.

Federated data reuse for research and public health responses to (re) emerging pathogens

Federated reuse of participant-level biomedical data offers significant advantages for responding to emerging pathogens by balancing rapid data sharing with privacy, security, and scalability concerns. Enabling in situ analysis minimises the transfer of sensitive data and protects patient privacy. It supports rapid response by facilitating distributed data analysis across institutions and countries, without delays caused by data transfer agreements, ensuring timely insights during public health emergencies.[4] Federated systems also address compliance with data protection regulations such as the GDPR while enabling international collaboration.

Incorporating datasets from diverse populations and healthcare systems enhances the representativeness and depth of analyses on pathogen characteristics, transmission dynamics, and treatment responses. Scalability supports the integration and analysis of large, multi-source datasets without the logistical complexities of centralisation, which is critical during large-scale outbreaks. Continuous real-time data updates from local sources provide dynamic insights for monitoring pathogen spread and evaluating interventions. By reducing the risk of data misuse and promoting equitable access, particularly in low-resource settings, federated reuse is an essential and ethical approach to addressing the challenges posed by emerging pathogens.

Technical approaches to federated analysis

Federated analysis of participant-level biomedical data encompasses several technical approaches, each addressing unique challenges in data sharing, privacy, and utility.[1] Swarm learning eliminates centralised servers by distributing learning coordination across multiple nodes, enhancing decentralisation and resilience, but it poses challenges for maintaining consistent privacy standards.[2] Device-based distributed learning trains models directly on smartphones and smartwatches, minimising data transfer, but faces limitations due to device heterogeneity and computational power. Software such as DataSHIELD and TensorFlow Federated enable privacy-preserving computations across data silos by

sharing model updates rather than raw data, thereby balancing privacy with analytic capabilities.[5] The OHDSI Network's approach applies standardised code across harmonised datasets to generate insights without exchanging participant-level data.[6] These methods reflect trade-offs between privacy and utility, employing strategies such as differential privacy, encryption, and secure computation. They also raise legal and ethical questions about consent, inclusivity, and governance, particularly when addressing sensitive or marginalised populations.

Variation in attitudes towards federated data reuse

Empirical research examining public attitudes toward federated reuse of participant-level biomedical data remains limited, with most evidence drawn from adjacent literatures on distributed research networks and general health data sharing. Reviews of this evidence indicate broad but qualified public support for sharing health data for research, contingent on trust, transparency, robust governance, and perceived fairness.[7,8] How participants themselves understand and evaluate federated reuse models remains underexplored, and addressing this gap is a central motivation for the present review.

What is data governance?

Data governance is the framework of policies, practices, and standards that ensures the ethical, secure, and effective management of data throughout its lifecycle. In the context of federated reuse of participant-level biomedical data, data governance can play a crucial role in balancing privacy, utility, and equity while addressing ethical and legal concerns. Effective data governance should ensure equitable benefit sharing by recognising the contributions of both researchers and research subjects. This includes returning research findings to participants, improving access to healthcare, and acknowledging researchers' intellectual contributions.[3] Ethical concerns, such as informed consent, inclusivity, and the prevention of data misuse or exploitation, must be addressed to foster trust and fairness. Legally, data governance frameworks must uphold participants' rights, including the ability to withdraw consent at any point in time, access their data, and understand its use, while ensuring compliance with regulations such as the GDPR and Health Insurance Portability and Accountability Act (HIPAA) to protect against potential harms.[3] By addressing these issues, data governance enables federated systems to leverage sensitive biomedical data to advance science and healthcare responsibly.

Data governance approaches

A range of governance frameworks has been described for federated data reuse, including decentralised Data Access Committees, centralised health data access boards such as that envisaged under the European Health Data Space, federated data governance networks, community-based models, public-private partnerships, and self-governance by data custodians.[3] These models differ in how they balance local autonomy, cross-site consistency, and participant involvement, and the appropriate choice depends on the federated initiative's specific goals, scale, and ethical considerations. Mapping and comparing these governance models is a focus of the scoping review described in this protocol.

Ethical concerns related to federated analysis of participant-level biomedical data

Consent Models for Federated Data Reuse

Consent is a foundational aspect of data governance in federated learning, with models such as broad, dynamic, and tiered consent commonly explored. Broad consent permits data to be used for a wide range of future research, whereas dynamic consent provides participants with ongoing control and updates on how and if their data can be used.[9] Federated data infrastructures pose unique challenges to these models, as participants may not fully understand how their data is analysed without leaving its original location. Effective consent models must balance participant autonomy with the operational needs of federated systems, ensuring clarity, flexibility, transparency, and compliance with ethical and legal standards. Federated data governance is not a replacement for informed consent.

Objective

We will use a scoping review approach to identify and summarise best practices for data governance in the context of federated data reuse, broadly defined as reusing the data without having direct access to the participant-level data, either through direct download or transfer or secure data access in a virtual or trusted research environment (VRE, TRE).

Methods

The scoping review protocol was drafted in accordance with the PRISMA-P guidelines and will be published in accordance with the 2020 PRISMA-ScR guidelines.[10] We will apply the Arksey and O'Malley scoping review framework[11] to map current recommendations and best practices on data governance for the reuse of federated biomedical data.

Registration. The protocol was registered with Open Research Europe before initiating the search.

Amendments. This protocol is not an amendment of a previously published protocol. Important amendments to the protocol will be documented in the Open Research Europe registration record, including the date, a description of the change, and the rationale. They will be reported in the final review.

Stage I: Identifying the research question

The initial review questions were developed through consultation with the research team and related stakeholders, including the eight health-related institutions engaged in the Coordination Mechanism for Cohorts and Trials (CoMeCT) project. The scoping review will cover how federated reuse of participant-level biomedical data is defined in current literature and the data governance frameworks that guide these approaches. It will examine variations in governance practices across data types, technical architectures, and legal-ethical contexts, and identify how participant rights, privacy, and benefit-sharing are addressed.

Initial review questions

- How is federated reuse of participant-level biomedical data defined in the current literature?
- What are the different types of data governance in use or recommended for federated reuse of health data?
- How do these recommendations and approaches differ by data type, mechanism for reuse of federated data, disease area, and population? What other factors are related to the differences in the recommended approach?
- How are data access and requests reviewed, approved and monitored in practice?
- What are the objectives of the data governance structures?
- What metrics are used to identify whether the governance structures are meeting their intended objectives?
- What are the recommended data governance models for the federated reuse of biomedical data?
- How are the rights of patients or research participants addressed in different approaches to federated data reuse?
- How is benefit-sharing with research producers and research subjects addressed in federated or swarm analyses?
- What are the legal and ethical concerns related to federated or swarm analysis in the health data space? What GDPR roles are generally assigned to different actors in federated data reuse?
- What approaches are described to support transparency, trust, sustainability, benefit sharing, and equity in federated governance models?
- How do the above differ by: technical approach to federated data reuse, including how the privacy/utility trade-off and data protection are addressed; data type; geography; population; data source (research, routine clinical care, surveillance, smart devices); or data permissions?

Stage II: Identifying relevant studies

Search strategy

We will systematically search Ovid (Medline), EMBASE, and Web of Science to identify national health data ecosystems, spaces, commons, or other infrastructures that enable the reuse of participant-level health data. The search uses a combination of MeSH and text terms tailored to each database (Table 1). We will export citations to EndNote Version 21 for deduplication.

The search strategy presented in Table 1 includes MeSH and text terms related to federated data reuse and data governance and was developed in accordance with the 2015 Peer Review of Electronic Search Strategies (PRESS) Guidelines.[13] The search strategies were developed by reviewing search strategies from related publications and through a literature review. They were reviewed by an information scientist at the Norwegian Institute of Public Health (NIPH). The search strategy was first developed for Ovid (Medline) and subsequently tailored and piloted for the other two

Table 1. Database-specific search strategies.

Ovid (Medline)	EMBASE	Web of Science
exp federated learning/or exp data sharing/	'federated learning'/exp OR 'data sharing'/exp	TS = ((federat* OR "federated learning" OR "privacy-preserving" OR "secure multiparty" OR "federated model" *OR"distributed research network*") NEAR/2 (learning OR analysis OR model* OR computation OR training OR inference))
((federat* or distribut* or decentralis* or decentraliz* or swarm or multiparty or "multi-party" or "multi-institutional" or "multi institutional" or "cross-institutional" or "privacy-preserving") adj3 (learning or reuse or "re-use" or computation or analysis or training or modelling or modelling or inference or approach* or method* or framework* or network* or research)).ti,ab,kf.	((federat* OR distribut* OR decentralis* OR decentraliz* OR swarm OR multiparty OR 'multi-party' OR 'multi-institutional' OR 'multi institutional' OR 'cross-institutional' OR 'privacy-preserving') NEAR/1 (learning OR reuse OR 're-use' OR computation OR analysis OR training OR modelling OR modelling OR inference OR approach* OR method* OR framework* OR network* OR research)):ti,ab	TS = ("trusted research environment" *OR" secure research environment*" OR "virtual research environment" *OR" trusted data environment*" OR "safe haven" *OR"data safe haven*" OR "secure data access" OR "controlled access environment*")
("analysis to data" or "analysis-to-data" or "code to data" or "code-to-data" or "model to data" or "model-to-data" or "data to model" or "data-to-model" or "distributed analytics" or "remote analysis" or "data localisation" or "data localization" or "federated access" or "federated analysis" or "federated infrastructure" *or" federated data" or" health data space*" or "data space" *or"personal data space*" or "personal health train" or PHT or "secondary use of health data" or "secondary use of biomedical data" or "policy analys" *or "data flow*" or "data sharing"). ti,ab,kf.	'analysis to data':ti,ab OR 'analysis-to-data':ti,ab OR 'code to data':ti,ab OR 'code-to-data':ti,ab OR 'model to data':ti,ab OR 'model-to-data':ti,ab OR 'data to model':ti,ab OR 'data-to-model':ti,ab OR 'distributed analytics':ti,ab OR 'remote analysis':ti,ab OR 'data localisation':ti,ab OR 'data localization':ti,ab OR 'federated access':ti,ab OR 'federated analysis':ti,ab OR 'federated infrastructure*':ti,ab OR 'federated data':ti,ab OR 'health data space*':ti,ab OR 'data space*':ti,ab OR 'personal data space*':ti,ab OR 'personal health train':ti,ab OR pht:ti,ab OR 'secondary use of health data':ti,ab OR 'secondary use of biomedical data':ti,ab OR 'policy analys*':ti,ab OR'data flow*':ti,ab OR 'data sharing':ti,ab	TS = (DataSHIELD OR OpenSAFELY OR "Open Secure Analytics for Electronic Health Records" OR OHDSI OR "Observational Health Data Sciences and Informatics" OR OMOP OR "Common Data Model" OR TriNetX OR i2b2 OR VODAN OR "Virus Outbreak Data Network" OR "Five Safes" OR "5 Safes" OR "Health Data Research UK" OR "HDR UK" OR Findata OR "European Health Data Space" OR EHDS OR BEACON OR "Health data hub" OR H3Africa OR "DS-I Africa" OR GA4GH OR "Global Alliance for Genomics and Health")
("trusted research environment" *or" secure research environment*" or "virtual research environment" *or" trusted data environment*" or "safe haven" *or"data safe haven*" or "secure data access" or "controlled access environment*"). ti,ab,kf.	'trusted research environment*': *ti,ab OR 'secure research environment*':ti,ab OR 'virtual research environment*':ti,ab OR 're*': ti,ab OR 'vre*':ti,ab OR 'trusted data environment*':ti,ab OR 'safe haven*': *ti,ab OR 'data safe haven*':ti,ab OR 'secure data access':ti,ab OR'controlled access environment*':ti,ab	#1 OR #2 OR #3
(DataSHIELD or OpenSAFELY or "Open Secure Analytics for Electronic Health Records" or OHDSI or "Observational Health Data Sciences and Informatics" or OMOP or "Common Data Model" or TriNetX or i2b2 or VODAN or "Virus Outbreak Data Network" or "Five Safes" or "5 Safes" or "Health Data Research UK" or "HDR UK" or Findata or "European Health Data Space" or EHDS or BEACON or	datashield:ti,ab OR opensafely: ti,ab OR 'open secure analytics for electronic health records':ti,ab OR ohdsi:ti,ab OR 'observational health data sciences and informatics':ti,ab OR omop:ti,ab OR 'common data model':ti,ab OR trinetx:ti,ab OR i2b2:ti,ab OR vodan:ti,ab OR 'virus outbreak data network':ti,ab OR 'five safes':ti,ab OR '5 safes':ti,ab OR 'health data research uk':ti,ab OR 'hdr uk':ti,ab	TS = ((health OR biomedical OR clinical) NEAR/3 (governance OR stewardship OR regulation OR policy OR privacy OR consent OR sharing OR access OR accountability OR ownership OR control OR transparency OR sovereignty OR custodian* OR controller* OR processor*))

Table 1. *Continued*

Ovid (Medline)	EMBASE	Web of Science
"Health data hub" or H3Africa or "DS-I Africa" or GA4GH or "Global Alliance for Genomics and Health" or "Trusted Research Environment*").ti,ab,kf.	'hdruk':ti,ab OR findata:ti,ab OR 'european health data space':ti,ab OR ehds:ti,ab OR beacon:ti,ab OR 'health data hub':ti,ab OR h3africa: ti,ab OR 'ds-i africa':ti,ab OR ga4gh: ti,ab OR 'global alliance for genomics and health':ti,ab	
1 or 2 or 3 or 4 or 5	#1 or #2 or #3 or #4 or #5	TS = ("data governance" OR "data steward*" OR "data custodian" OR "data trust*" OR "data commons" OR "data access committee" OR DAC OR "health data access bod*" OR HDAB OR "data protection officer" OR DPO OR "ethics committee*" OR "institutional review board" OR IRB OR "research ethics board*" OR REB OR "data sharing agreement" OR "data use agreement*" OR "data license" OR "governance framework*" OR "governance model" OR "governance structure*")
exp data protection/ or exp information management/ or exp medical ethics/ or exp informed consent/ or exp legislation/ or exp confidentiality/ or exp access to information/or exp health policy/	'data governance':ti,ab OR 'data steward':ti,ab OR 'data custodian': ti,ab OR 'data trust':ti,ab OR 'data commons':ti,ab OR 'data access committee':ti,ab OR dac:ti,ab OR 'health data access bod':ti,ab OR hdab:ti,ab OR 'data protection officer':ti,ab OR dpo:ti,ab OR 'ethics committee':ti,ab OR 'institutional review board':ti,ab OR irb:ti,ab OR 'research ethics board':ti,ab OR reb: ti,ab OR 'data sharing agreement': ti,ab OR 'data use agreement':ti,ab OR 'data license':ti,ab OR 'governance framework':ti,ab OR 'governance model':ti,ab OR'governance structure*':ti,ab	TS = ("informed consent" OR "broad consent" OR "dynamic consent" OR "tiered consent" OR "community consent" OR "opt-in" OR "opt-out" OR "data sovereignty" OR "benefit sharing" OR "benefit-sharing" OR "secondary use" OR "secondary data use" OR "data reuse" OR "data re-use" OR "social license" OR "social licence" OR "public trust" OR "patient rights" OR "participant rights" OR "data subject*" OR "right to withdraw" OR "withdrawal of consent")
((data or health or biomedical or clinical or research) adj3 (governance or stewardship or regulation or policy or policies or oversight or management or privacy or consent or sharing or access or accountability or ownership or control or transparency or sovereignty or custodian* or controller* or processor* or authority or authorities)).ti,ab,kf.	'informed consent':ti,ab OR 'broad consent':ti,ab OR 'dynamic consent':ti,ab OR 'tiered consent': ti,ab OR 'community consent':ti,ab OR 'opt-in':ti,ab OR 'opt-out':ti,ab OR 'data sovereignty':ti,ab OR 'benefit sharing':ti,ab OR 'benefit-sharing':ti,ab OR 'secondary use': ti,ab OR 'secondary data use':ti,ab OR 'data reuse':ti,ab OR 'data re-use':ti,ab OR 'social license':ti,ab OR 'social licence':ti,ab OR 'public trust':ti,ab OR 'patient rights':ti,ab OR'participant rights':ti,ab OR'data subject*':ti,ab OR'right to withdraw':ti,ab OR 'withdrawal of consent':ti,ab	TS = (GDPR OR "General Data Protection Regulation" OR HIPAA OR "Health Insurance Portability" OR "data protection law" OR "data protection regulation" OR "privacy law" OR "privacy regulation" OR "health data regulation" OR "legal framework" OR "regulatory framework" OR "ethical framework" OR compliance OR interoperability OR "FAIR data" OR "FAIR principles" OR "open data" OR "responsible data" OR "European Health Data Space" OR "health data space*")
("data governance" or "data steward" or "data custodian" or "data trust" or "data commons" or "data access committee" or DAC or "health data access bod" or HDAB or "data protection officer" or DPO or "ethics committee" or "institutional review board" or IRB or "research ethics board" or REB or "data sharing agreement" or "data use agreement" or "data license" or	gdpr:ti,ab OR 'general data protection regulation':ti,ab OR hipaa:ti,ab OR 'health insurance portability':ti,ab OR 'data protection law':ti,ab OR 'data protection regulation':ti,ab OR 'privacy law':ti,ab OR 'privacy regulation':ti,ab OR 'health data regulation':ti,ab OR 'legal framework':ti,ab OR 'regulatory framework':ti,ab OR 'ethical	#5 OR #6 OR #7 OR #8

Table 1. *Continued*

Ovid (Medline)	EMBASE	Web of Science
"governance framework" *or"governance model*" or "governance structure*").ti,ab,kf.	*framework*':ti,ab OR 'compliance': ti,ab OR 'interoperability':ti,ab OR 'fair data':ti,ab OR 'fair principles': ti,ab OR 'open data':ti,ab OR 'responsible data':ti,ab OR 'european health data space':ti,ab OR ehds:ti,ab OR'health data space*':ti,ab	
("informed consent" or "broad consent" or "dynamic consent" or "tiered consent" or "community consent" or "opt-in" or "opt-out" or "data sovereignty" or "benefit sharing" or "benefit-sharing" or "secondary use" or "secondary data use" or "data reuse" or "data re-use" or "social license" or "social licence" or "public trust" or "patient rights" or "participant rights" or "data subject*" or "right to withdraw" or "withdrawal of consent").ti,ab,kf.	#7 OR #8 OR #9	#4 AND #9 Timespan: 2000-01-01 to 2026-03-17
(GDPR or "General Data Protection Regulation" or HIPAA or "Health Insurance Portability" or "data protection law" *or" data protection regulation*" or "privacy law" *or" privacy regulation*" or "health data regulation" *or" legal framework*" or "regulatory framework" *or"ethical framework*" or "compliance" or "interoperability" or "FAIR data" or "FAIR principles" or "open data" or "responsible data" or "European Health Data Space" or EHDS or "health data space*").ti,ab,kf.	#6 AND #10	#9 AND TS = ("data sharing" OR "distributed research network" *OR"data network*")
7 or 8 or 9 or 10 or 11	#11 AND [2000–2026]/py	#10 OR #11
6 AND 12		
exp animals/NOT exp humans/		
13 NOT 14		
limit 15 to yr = "2000-Current"		

databases. We will not limit the search by geography or publication language and will search all databases from January 1, 2000, through the date of the search.

Stage III: Study selection and eligibility criteria

Title and abstract screening will be conducted by one reviewer, with ASReview, an open-source active learning tool for systematic review screening, serving as the second reviewer.[12] Full-text screening will be performed independently by two reviewers. Disagreements between reviewers at any stage of the process will be resolved through consensus. Consistent with the iterative nature of a scoping review, the search strategy may be refined to address emerging questions identified during screening. The screening process will be conducted using the COVIDENCE systematic review software.

Inclusion criteria

Included publications will comprise studies, reviews, reports, or recommendations that address data governance for the federated reuse of health or biomedical data. Eligible study designs include qualitative, quantitative, and mixed-methods

studies; systematic reviews; scoping reviews; case studies; reports; and white papers. Eligible sources include peer-reviewed articles, grey literature, policy documents, and official reports from recognised health organisations.

Exclusion criteria

Studies or reports that do not address data governance or that focus exclusively on the technical implementation of federated data reuse without discussing data governance will be excluded. Studies that focus solely on centralised data-sharing systems or on individual data sharing without a federated context are ineligible. Studies, reports, or recommendations on federated data governance outside the health or biomedical domains will be excluded. Editorials, letters, and opinion pieces without substantive recommendations will not be included. Studies that report only technical metrics for federated data reuse, such as algorithm performance, without linking to governance structures or objectives, 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 before commencing charting. Data extraction will be conducted in COVIDENCE, with one investigator extracting the data and a second investigator reviewing the data extraction. Key domains and the items to be extracted from the included studies or reports are presented in Table 2. Specifically, we will chart data on the types of governance models used, how they address participant rights and benefit sharing, and their alignment with ethical and legal standards. We will also extract data on metrics used to evaluate governance effectiveness and how governance adapts to different federated technical approaches (e.g., swarm learning, federated analysis via APIs, multiparty computation). Emphasis will be placed on whether the governance models enable privacy-preserving reuse of participant-level data, facilitate interoperability and inclusivity, and comply with regional regulations such as the GDPR. We will also capture how data custodians and other stakeholders

Table 2. Draft items for data charting.

Domain	Items
Study information	Study title
	Authors
	Publication year
	Study design (e.g., review, case study, qualitative research)
	Country/region
Data governance type	Governance structure
	Actors involved in governance (e.g., data providers, data protection officer, ethical review committee, national data access board)
	Objectives of governance
	Governance-related metrics (e.g., time-to-data reuse)
	Source of metrics (e.g., researcher reports, institutional records)
	Legal and ethical concerns, including the legislative context
	GDPR roles of different actors
Federated data reuse approach and context	Mechanism for reuse (e.g., swarm learning, access through code, software-based access approach)
	Data protection strategies (e.g., differential privacy)
	Technical approach to privacy/utility trade-off
	Data types
	Disease areas
	Population
Participant rights and benefit sharing	Participant or subject rights in data
	Benefit sharing with participants or data providers.
Comparative insights	Variation by data type, geography, population, technical approach, approach to privacy, context (epidemic or non-epidemic)

Table 3. GANTT Chart.

	2025	2026				2027
	Q4	Q1	Q2	Q3	Q4	Q1
Develop a scoping review protocol.						
Pilot test MeSH and text term strategy						
Implement a systematic search.						
Convene a stakeholder group to refine domains and items for data extraction.						
Screen title-abstract and full texts of publications identified through a systematic search of electronic databases						
Complete initial data charting						
Draft and submit findings for publication						
Disseminate findings to patient advocate groups, policymakers, and funders.						

are involved in governance processes and whether mechanisms such as dynamic consent or community-based governance are used.

Critical appraisal of sources of evidence

Consistent with established scoping review methodology, we will not undertake a formal critical appraisal or risk-of-bias assessment of the individual sources included. The aim is to map the range of recommendations and practices for data governance rather than to assess the methodological quality of individual studies or reports.

Stage V: collating, summarising and reporting results

The search and inclusion results will be presented in a PRISMA flow diagram. The results of the data charting process will be presented in narrative and tabular form.

Stage VI: stakeholder consultation

Stakeholder consultation is essential to ensure that the findings of this scoping review are grounded in the needs and practices of those who develop, implement, and are affected by federated data governance models. We will engage stakeholders through structured consultations at two key stages: after the preliminary mapping of governance models and again during the interpretation and synthesis of findings.

Stakeholders will include representatives from Health Data Commons and experts in data governance identified through the Health Data Commons Research Data Alliance (RDA) working group and the ESCMID-TGHN Online Biomedical Data and Sample Governance Training. These groups represent the diverse actors who contribute to and are affected by the governance of federated biomedical data reuse.

Consultations will focus on three main areas: (1) evaluating the relevance and completeness of the governance models identified, (2) identifying practical challenges in implementing these models, and (3) co-developing a roadmap for governance alignment across federated health data reuse infrastructures. Particular attention will be paid to how governance structures support participants' rights to data, benefit sharing, equity, and inclusivity.

The final output of the stakeholder consultation will include a synthesis report summarising key insights, recommendations for good governance in federated data reuse, and stakeholder-validated indicators for assessing the effectiveness of governance structures. These results will be integrated into the final review and disseminated alongside the main publication and policy briefs.

Ethics and dissemination

This scoping review will summarise evidence from published studies and does not require ethical approval. We will publish the scoping review results in an open-access journal. Findings will be shared with stakeholders via policy briefs, technical summaries, and webinars. The outputs will also contribute to a CoMeCT deliverable on the evidence for best

practices in data governance for federated reuse of biomedical data, with a focus on aligning governance models across European health data infrastructures.

Conclusion

This scoping review will provide the first structured synthesis of data governance models for federated reuse of biomedical data. It will map governance mechanisms, assess their relevance across different federated technical approaches, and explore how these models address ethical, legal, and social concerns, including participant rights, benefit sharing, and inclusivity. By highlighting key patterns and gaps, the review will inform national and cross-national efforts to establish transparent, equitable, and scalable governance frameworks in health data reuse.

The findings will serve as a critical resource for ongoing European and international initiatives seeking to implement federated data analysis infrastructures that align with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles and comply with regulatory requirements such as the GDPR. This work will also support the development of policy and technical guidance to ensure that governance strategies foster public trust and enable responsible, privacy-preserving reuse of biomedical data for research and public health purposes.

Study status

The current status and anticipated timeline of the study are presented in the GANTT chart in Table 3.

Ethics and consent

Ethical approval and consent were not required.

Data availability

No data are available for this protocol in the context of a scoping review.

Acknowledgements

The authors would like to thank the information scientist at the Norwegian Institute of Public Health for peer-reviewing the search strategy in accordance with the PRESS guidelines.

References

1. Rieke N, Hancox J, Li W, *et al.*: **The future of digital health with federated learning.** *npj Digit Med.* 2020; **3**: 119.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Warnat-Herresthal S, Schultze H, Shastry KL, *et al.*: **Swarm Learning for decentralised and confidential clinical machine learning.** *Nature.* 2021; **594**(7862): 265–270.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Bernier A, Molnár-Gábor F, Knoppers BM: **The international data governance landscape.** *J Law Biosci.* 2022; **9**.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Dayan I, Roth HR, Zhong A, *et al.*: **Federated learning for predicting clinical outcomes in patients with COVID-19.** *Nat Med.* 2021; **27**(10): 1735–1743.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Gaye A, Marcon Y, Isaeva J, *et al.*: **DataSHIELD: taking the analysis to the data, not the data to the analysis.** *Int J Epidemiol.* 2014; **43**(6): 1929–1944.
PubMed Abstract | Publisher Full Text | Free Full Text

6. 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.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Kalkman S, van Delden J, Banerjee A, *et al.*: **Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence.** *J Med Ethics.* 2022; **48**(1): 3–13.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Cascini F, Pantovic A, Al-Ajlouni YA, *et al.*: **Health data sharing attitudes towards primary and secondary use of data: a systematic review.** *eClinicalMedicine.* 2024; **71**: 102551.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Kaye J, Whitley EA, Lund D, *et al.*: **Dynamic consent: a patient interface for twenty-first century research networks.** *Eur J Hum Genet.* 2015; **23**(2): 141–146.
PubMed Abstract | Publisher Full Text | Free Full Text

10. Tricco AC, Lillie E, Zarin W, *et al.*: **PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation.** *Ann Intern Med.* 2018; **169**(7): 467–473.
Publisher Full Text

11. Arksey H, O'Malley L: **Scoping studies: towards a methodological framework.** *Int J Soc Res Methodol.* 2005; **8**(1): 19–32.
Publisher Full Text

12. van de Schoot R, de Bruin J, Schram R, *et al.*: **An open source machine learning framework for efficient and transparent systematic reviews.** *Nat Mach Intell.* 2021; **3**(2): 125–133.
Publisher Full Text

13. McGowan J, Sampson M, Salzwedel DM, *et al.*: **PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement.** *J Clin Epidemiol.* 2016; **75**: 40–46.
PubMed Abstract | Publisher Full Text