Guidelines for data sharing of investigator-initiated clinical studies

Deliverable 16.5 WP16

Malik S, Contrino S, del Alamo M, Lémeret S, Demotes-Mainard J, Kubiak C, Matei M, Klammt S

ECRIN, February 2025

Co-funded by the European Union

This leaflet has been developed as part of the ERA4Health project, co-funded by the European Union under the Horizon Europe Framework Programme. Grant Agreement N° 101095426.

Table of Contents

These guidelines have been adapted from ERA4Health's original document D16.5 Guidelines for data sharing of investigator-initiated clinical studies (1).

Data Sharing Plan Preparation Guidelines

The guidance presented here is structured in correspondence to *Figure 1: Data sharing plan guidelines* (next page).

These guidelines are meant to promote and support data sharing and reuse among researchers, adequately inform trial participants and protect their rights and provide effective and efficient systems for preparing, storing and accessing data. By enforcing implementation of these guidelines, funders of clinical trials contribute to the principle that publicly funded research data is a public good, produced in the public interest and should be made openly available with as few restrictions as possible in a timely and responsible manner (2).

Next to each of the subsections you will find a summary of the topic to be covered in the data sharing plan along with the useful resources that will facilitate the preparation of a high quality data sharing plan.

Co-funded by the European Union

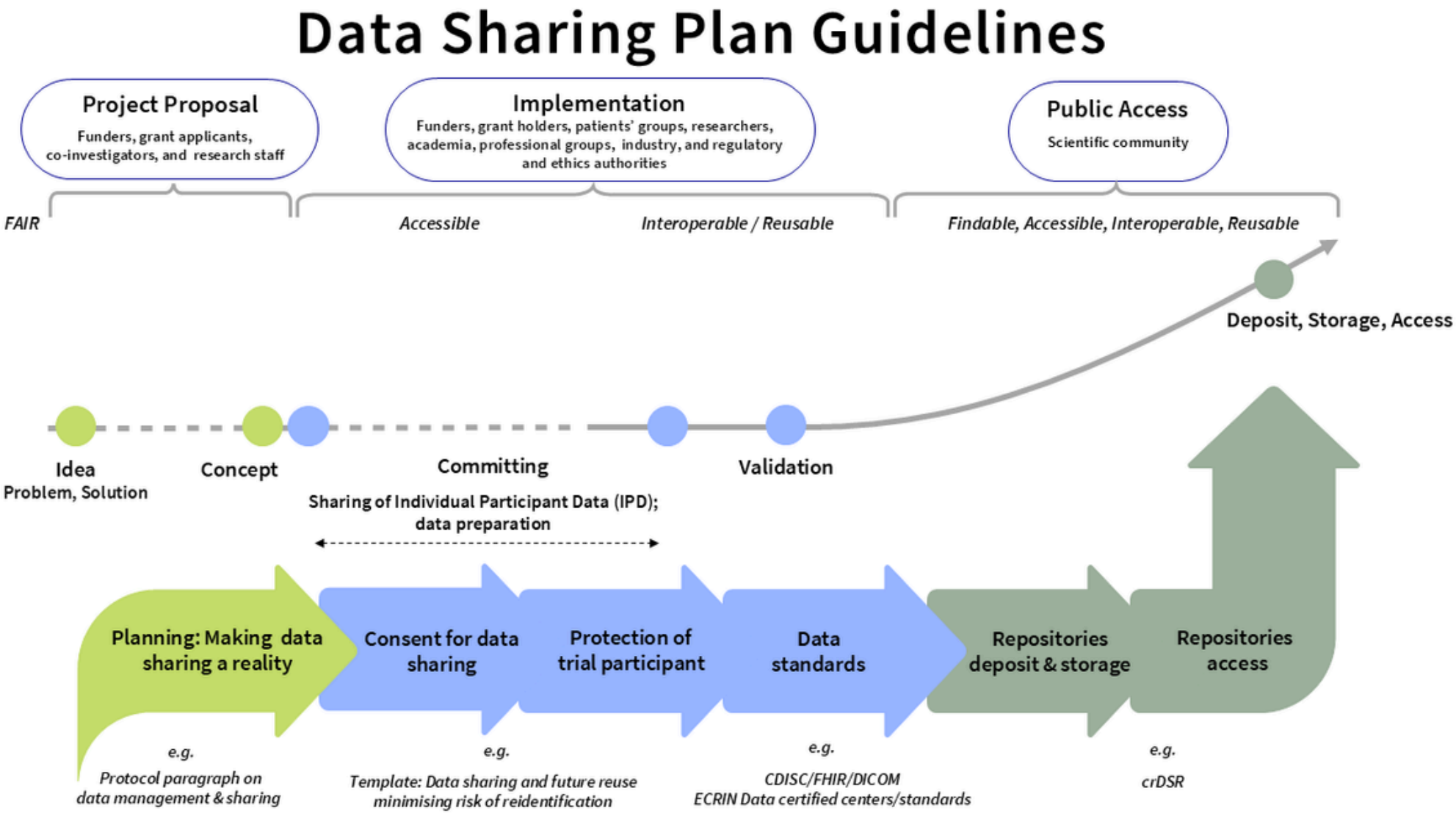


Figure 1: Data sharing plan guidelines

Points to consider	Information to be provided in the plan	Guidance
PLANNING (PROJECT PROPOSAL)		
Making data sharing a reality. Describe research data management and sharing (data management plan).	Researchers should ensure that data sharing is considered from the very beginning of study planning and should be included within the trial protocols, data management, sharing plan and trial registry entry as well as in patients/participants information and informed consent, if appropriate (see below). Funds for responsible data sharing should be requested by the original trial team from trial funders as part of initial trial grant applications, e.g. to fund dataset preparation and anonymisation/ pseudonymisation. Reasonable costs may be recovered from data requesters if appropriate (as mentioned in the European Health Data Space (EHDS) Regulation), but data sharing activities should not be profit-generating.	The protocol of the study should clearly indicate that the data, study documents and participant-level dataset (either as anonymised or pseudonymised), will be made available and summarised in the relevant section of the trial registration record. Recommended resources: • SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials, item 31a (3) • **Annex 1**: Example of data sharing specific paragraph to be included in the clinical study protocol • Data management and sharing plan templates for consideration https://dmponline.dcc.ac.uk/public_plans • International Trial registry for consideration https://clinicaltrials.gov/ • Primary registries in the WHO registry network for consideration https://www.who.int/clinical-trials-registry-platform/network/primary-registries

IMPLEMENTATION (DATA PREPARATION AND STANDARDS)		
Consent for data sharing	All the trial documents (e.g. participants' information leaflet and consent forms, ethical submission documents) should be written and translated considering the planned data sharing strategy.	Individual Participant Data (IPD) sharing should be based on explicit broad consent by study participants (or if applicable by their legal representatives) to the sharing and reuse of their data for scientific purposes. The process of informing study participants about possible sharing of their data, and then gaining their explicit consent to it, is of fundamental importance, and is normally a prerequisite for the sharing of pseudonymised data. The consent for secondary use of IPD should be as broad as possible. No consent is needed if the data are to be shared only after anonymisation, or according to the opt-out procedure after obtention of a data permit from a Heath Data Access Board (HDAB), or because of the 'public interest' underpinning the General Data Protection Regulation (GDPR) implementation in some countries. **Annex 2**: In the context of the VACCELERATE project (Task 9.3 Promotion of activities data sharing and future reuse), an informed consent including clauses for data sharing and reuse for future research purposes was generated.

IMPLEMENTATION (DATA PREPARATION AND STANDARDS)		
Protection of study participants **GDPR/Anonymisation/ Pseudonymisation**	Shared IPD from clinical trials used for further scientific research should be stored in a secure processing environment and, depending on the request for secondary use, will be made accessible either as pseudonymised or anonymised datasets. Prepare the dataset for sharing (pseudonymisation, minimisation, anonymisation if relevant), taking into account the original patient consent and method of data transfer. Dataset preparation should be done by individuals with an understanding of data management and basic statistics, with quality control provided by a further individual who is independent to the process.	Sharing of pseudonymous data should be considered as the default option – this means making data accessible to re-users, after obtention of a data permit, in a secure processing environment, without the possibility to download the data. Depending on the objective of the request for secondary use, the data should be minimised, (e.g. removal of data unnecessary for the planned secondary use). Depending on the request for secondary use, the sharing of anonymised data may be possible if the anonymisation process does not alter the informative value of the dataset. Anonymisation is based on the risk of re-identification, and anonymised data is not governed by the GDPR, therefore anonymised data can be downloaded and circulated. Various anonymisation techniques can be used by a data expert or through anonymisation tools like Amnesia https://amnesia.openaire.eu/amnesia/. An assessment of the residual risks for re-identification of participants in pseudonymised data sets should be performed. Under the GDPR in Europe there is an obligation on the data controller to carry out a data protection impact assessment (DPIA), “to evaluate the origin, nature, particularity, and severity of the risk to the rights and freedoms of natural persons” before processing personal data. The impact assessment “should include the measures, safeguards and mechanisms envisaged for mitigating” the identified risks.

IMPLEMENTATION (DATA PREPARATION AND STANDARDS)		
Data Preparation	Selection of appropriate data standards has an impact on the secondary use. For instance, meta-analyses require alignment of data standards (e.g. CDISC for clinical trial data). This is critical to the success of IPD sharing. Besides the IPD sets, other clinical trial data objects (metadata) should be made available for sharing (e.g. protocols, clinical study reports, statistical analysis plans, blank consent forms) to allow a full understanding of any data set.	**Annex 3**: Example of clinical trial DPIA content Services to support pseudonymisation of data sets, which could range from simple guidance, through consultancy, to performing and documenting the de-identification process, should be established. Data and coding standards should be built into any study's data design prospectively, from the beginning of the study. While many scientific fields have developed and adopted common data standards, others have not. In such cases, the data sharing plan may indicate that no consensus data standards exist for the scientific data and metadata to be generated, preserved, and shared. If no standards exist, then the data should be managed so that is convertible to Study Data Tabulation Model (SDTM) format. **Annex 4**: Health data standards. Consider sharing study documents (study protocol and reports, statistical analysis plan, blank consent form, and other related documents that could help the reanalysis or understanding of the data).

PUBLIC ACCESS (DEPOSIT, STORAGE AND ACCESS)		
Data storage (repositories)	Search for a suitable repository. Data repositories have the potential to play an important role in the effective and safe sharing of clinical study data. This is because they can provide a stable, long-term, secure environment for the data, improve the security and quality of archiving through active data curation, increase the discoverability of data through the application of metadata schemes.	Provide the name of the repository where scientific data and metadata arising from the project will be stored. Data sharing requirement may vary between funding bodies and Data Holding Organisations (DHO). Data deposition repositories for suitable datasets can be searched using an online data repository finder: https://edctpknowledgehub.tghn.org/data-sharing-toolkit/repository-finder/ An evaluation of repositories for sharing individual-participant data from clinical studies analysed, in 2019, the landscape of data repositories to create a detailed description of available repositories and assess their suitability for hosting data from clinical studies, from the perspective of the clinical researcher (4). However, the landscape has evolved as the EU Regulations (GDPR, EHDS) will require the development of European, GDPR- and EHDS-compliant repositories, both from the technical (secure processing environment) and from the procedural point of view (Health Data Access Body).

PUBLIC ACCESS (DEPOSIT, STORAGE AND ACCESS)		
Data storage (repositories) - continued		**Annex 5**. Examples of repositories that enable storage, sharing, discoverability, reuse of the IPD and associated documents from clinical studies. When selecting a repository to manage and share data, here are some desirable characteristics to look for: **Unique persistent identifiers**: Assigns datasets a citable, unique persistent identifier, such as a digital object identifier (DOI). **Long-term sustainability**: Has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; building on a stable technical infrastructure. **Metadata**: Ensures datasets are accompanied by metadata to enable discovery, reuse, and citation of datasets, using schema that are appropriate to, and ideally widely used across, the community(ies) the repository serves. **Quality assurance**: Provides, or has a mechanism for others to provide, expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata. **Easy access:** Provides broad, equitable, and maximise open access to datasets and their metadata in a timely manner after submission, consistent with legal and ethical limits required to maintain privacy and confidentiality, and protection of other sensitive data.

PUBLIC ACCESS (DEPOSIT, STORAGE AND ACCESS)		
Data storage (repositories) - continued		**Broad and measured reuse**: Makes datasets and their metadata available with broadest possible terms of reuse; and provides the ability to measure attribution, citation, and reuse of data (e.g. through assignment of adequate metadata and unique persistent identifiers). **Clear use guidance**: Provides accompanying documentation describing terms of dataset access and use (e.g. particular licenses, need for approval by a data use committee or ethics committee, fees). **Security and integrity**: Has documented measures in place to meet generally accepted criteria for preventing unauthorised access to, modification of, or release of data, with levels of security that are appropriate to the sensitivity of data. **Confidentiality**: Has documented capabilities for ensuring that administrative, technical, and physical safeguards are employed to comply with applicable confidentiality, risk management, and continuous monitoring requirements for sensitive data. **Common format**: Allows datasets and metadata downloaded, accessed, or exported from the repository to be in widely used, preferably non-proprietary, formats consistent with those used in the community(ies) the repository serves. **Provenance**: Has mechanisms in place to record the origin, chain of custody, and any modifications to submitted datasets and metadata. **Retention policy**: Provides documentation on policies for data retention within the repository.

PUBLIC ACCESS (DEPOSIT, STORAGE AND ACCESS)		
Discoverability (metadata)	A metadata dictionary should be developed to make the clinical study datasets findable. Evaluate if sufficient metadata will be available to enable data reuse.	Metadata is “data about the data” which describe the properties of the data set. Metadata should be structured so that it can be machine-read (e.g. by an Internet search engine) to maximise findability.
Data Access (rights, types and management of access)	IPD should be made available as soon as reasonably possible. Check and clearly mention under which terms and conditions the data will be shared. Make sure that there is a licence, and that the licence gives you permission to do what you intend to.	Host organisations (e.g. Institute of Higher Education) may be able to provide funds for routine data sharing activities, e.g. ongoing maintenance of a data sharing system. Responsibilities of staff for data sharing should be determined and funding should be sourced.

DATA SHARING PLAN CHECKLIST

This checklist provides a step-by-step guide to determining whether your data sharing plan aligns appropriately with the guidelines provided. It includes elements which you should consider thoroughly in order to confirm that your data sharing plan complies with the EU and international funders expectations (**Annex 7**).

CONCLUSIONS

Many funding bodies require the data sharing plan or statements about data sharing for secondary use by the grant applicants in order to make their data available. The implementation of the guidelines outlined in this document will foster compliance with funder policies and expectation on data sharing. Researchers using these guidelines are encouraged to share their experience to inform future updates of this guidance.

REFERENCES

1. Malik S, Contrino S, del Alamo M, Lémeret S, Demotes-Mainard J, Kubiak C, Matei M, Klammt S. D16.5 Guidelines for data sharing of investigator-initiated clinical studies. 2024 Feb. doi.org/10.5281/zenodo.14904510.
2. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open. 2017 Dec 14;7(12):e018647. doi: 10.1136/bmjopen-2017-018647
3. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. Ann Intern Med. 2013 Feb 5;158(3):200. doi: 10.7326/0003-4819-158-3-201302050-00583
4. Banzi R, Canham S, Kuchinke W, Krleza-Jeric K, Demotes-Mainard J, Ohmann C. Evaluation of repositories for sharing individual-participant data from clinical studies. Trials. 2019 Dec;20(1):169. doi: 10.1186/s13063-019-3253-3

ANNEX 1: EXAMPLE OF DESCRIPTION OF DATA SHARING WITHIN STUDY PROTOCOLS

"In line with the EU data sharing policy, individual subject-level data will be shared with the scientific community (either as anonymised or pseudonymised data sets), while maintaining the integrity and privacy of the trial participants and in compliance with the EU General Data Protection Regulation (GDPR) and national or local rules. Data and other trial documents should be made available through an appropriate data repository, helping to ensure that the data objects are properly prepared, are available in the longer term, are stored securely and subject to rigorous governance. The terms and conditions of data transfers to a repository and the data sharing process shall be subject to specific data processing agreements to be established between the concerned parties as well as to a specific data sharing plan, where the details are specified".

ANNEX 2: CONSIDERATIONS FOR CONSENT ON DATA SHARING FOR FURTHER RESEARCH PURPOSES

The below considerations must be taken into account when developing consent forms in a clinical study. They are written from the participant perspective.

I. Request for participation in future research.

II. What is future research?

III. What is “pseudonymised” (OPTION 1)/ “anonymised” data (OPTION 2).

IV. What are the benefits of sharing my data?

V. Which categories of data are involved? (clinical data, laboratory records, genetic data).

VI. How will my data be shared?
- Anonymised/pseudonymised
- Within the European Economic Area (EEA) or outside
- Location of the data repository
- Time of storage
- Conditions to access

VII. What are the risks and how will my privacy be protected?

VIII. What if I want to withdraw from future research?

IX. How will I be informed about the results?

X. What are my rights and how can I exercise my rights?

ANNEX 3: EXAMPLE OF CLINICAL TRIAL DPIA CONTENT

1. Basic trial information and setting
2. Roles of *data controller* and *data processors*
3. Biobanking specifics
4. Separate research project on “Further Research”
5. Data of staff at clinical study site and at Sponsor’s office
6. Collection of data
7. Use of data
8. Storage of data
9. Details of data deletion
10. Characterisation of participants’ personal data
11. Sharing of participants’ personal data
12. Scope of data processing
13. Purpose of data processing steps
14. Security measures in place
15. International Transfer Impact Assessment (ITIA)
16. Description of Electronic Data Capture (EDC) system
17. Legal basis
18. Data exit strategy
19. Consultation of external partners
20. Assessment of necessity and proportionality
21. Rights of participants
22. Identification and assessment of risks
23. Signatures of the *data protection officer* and the *principal coordinating investigator*

ANNEX 4: KEY HEALTH DATA STANDARDS

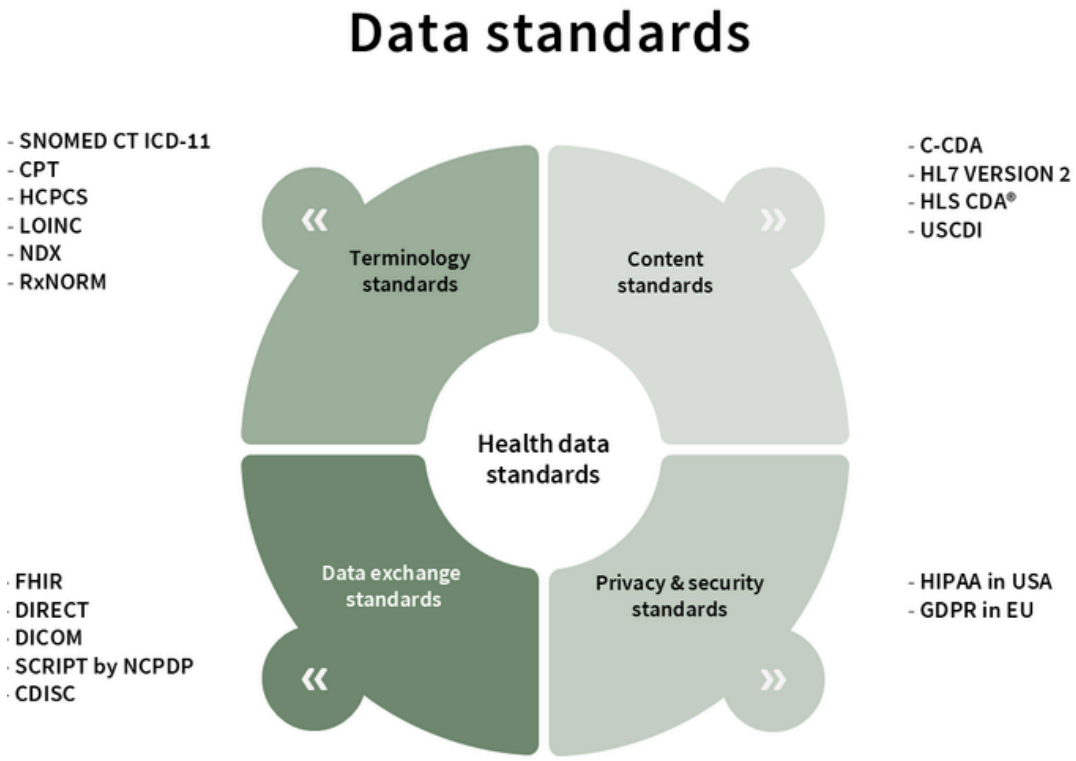
Key Health Data Standards

Terminology standards
Terminology standards utilise unique codes and classification systems to represent a wide range of health concepts and purposes.

Content standards
Content standards focus on defining the structure and data types of electronic medical documents. These standards ensure that medical data is properly organised and represented in a clear and easily understandable form.

Data exchange standards
Data exchange standards establish the information flow between health systems. These standards ensure interoperability by specifying formats, document architecture, methods, APIs, and other necessary components for effective data exchange.

Privacy and security standards
Privacy and security standards in healthcare are of utmost importance to protect sensitive health information from unauthorised access and ensure compliance with regulations such as HIPAA (Health Insurance Portability and Accountability Act) and EU-GDPR.



ANNEX 5 REPOSITORIES FOR DATA SHARING IN CLINICAL RESEARCH (some examples)

In 2019, Banzi et al., (4) made an assessment of the suitability of different repositories for hosting clinical study data.

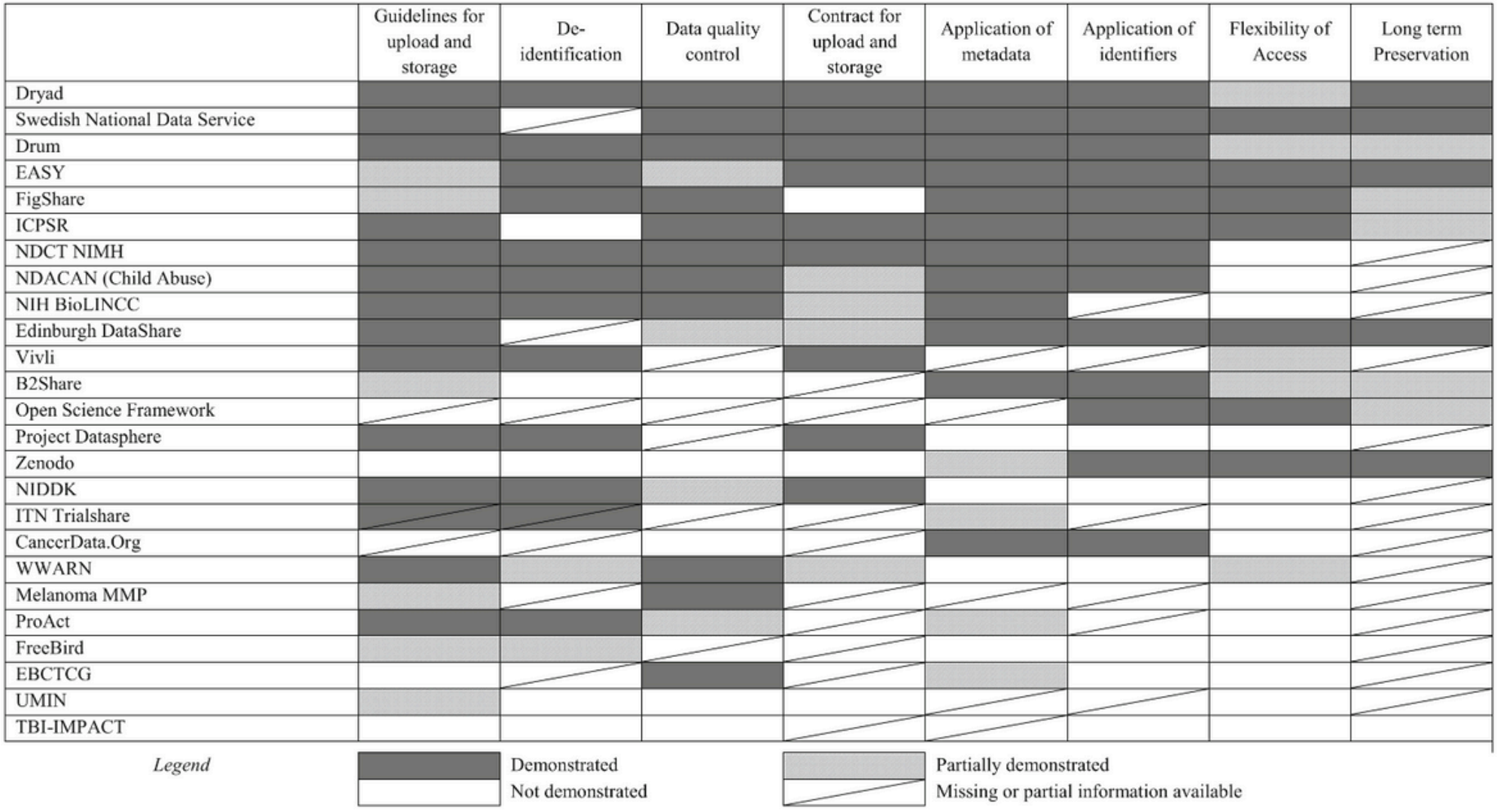
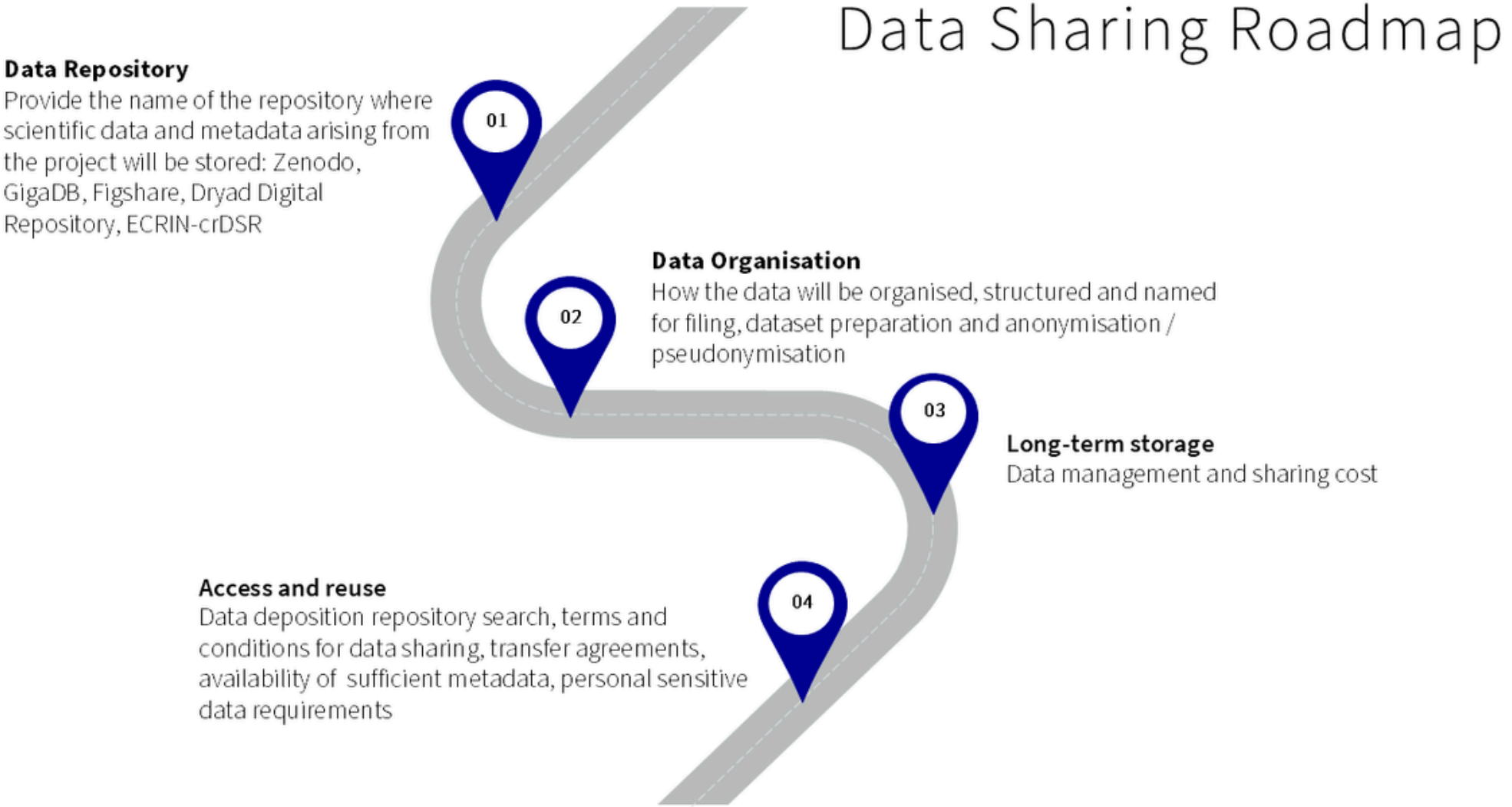


Fig. Suitability of the repository for hosting clinical study data

ANNEX 6: DATA SHARING ROADMAP



ANNEX 7: DATA SHARING PLAN CHECKLIST