REVIEW PAPER

FAIR sharing of health data: a systematic review of applicable solutions

Paul Guillot[1,2,3] · **Martin Bøgsted**[1,2] · **Charles Vesteghem**[1,2]

Abstract

Purpose Data sharing is essential in health science research. This has also been acknowledged by governments and institutions who have set-up a number of regulations, laws, and initiatives to facilitate it. A large number of initiatives has been trying to address data sharing issues. With the development of the FAIR principles, a set of detailed criteria for evaluating the relevance of such solutions is now available. This article intends to help researchers to choose a suitable solution for sharing their health data in a FAIR way.

Methods We conducted a systematic literature review of data sharing platforms adapted to health science research. We selected these platforms through a query on Scopus, PubMed, and Web of Science and filtered them based on specific exclusion criteria. We assessed their relevance by evaluating their: implementation of the FAIR principles, ease of use by researchers, ease of implementation by institutions, and suitability for handling Individual Participant Data (IPD).

Results We categorized the 35 identified solutions as being either online or on-premises software platforms. Interoperability was the main obstacle for the solutions regarding the fulfilment of the FAIR principles. Additionally, we identified which solutions address sharing of IPD and anonymization issues. Vivli and Dataverse were identified as the two most all-round solutions for sharing health science data in a FAIR way.

Conclusions Although no solution is perfectly adapted to share all type of health data, there are work-arounds and interesting solutions to make health research data FAIR.

Keywords Databases, Factual · Information Dissemination · Metadata · Medical Informatics · Information Storage and Retrieval

1 Introduction

Research heavily relies on data-driven methods to provide knowledge about complex systems. In particular, health science research requires a large variety and amount of individual participant data (IPD), e.g., genomics, imaging, and clinical data, too large for a single health care institution to possibly produce on its own. Thus, there is a need for sharing data across institutions. For the past decades, governments and institutions have been trying to create the right infrastructures for sharing data, through laws, regulations, and initiatives. For instance, the Research Data Alliance (RDA) was created in 2013 as a collaboration between the EU, the USA, and Australia, to build the infrastructures for open sharing and re-use of research data [1]. It delivers recommendations regarding good practices and standards. In terms of legislation, the recent Data Governance Act [2] and the new Data Act [3] aim at improving data access and interoperability in Europe. More concretely, some initiatives have implemented data sharing between institutions. In the USA, the National Institute of Health (NIH) launched the Generalist Repository Ecosystem Initiative (GREI) [4]; in the EU, the Horizon 2020 program has funded numerous research data sharing projects [5], such as BBMRI-ERIC [6], EUDAT [7], and ELIXIR [8]. More recently, the European Commission even announced the creation of a European Health Data Space [9].

However, open sharing does not guarantee that research data will be understood and reused in the right way. When

✉ Paul Guillot
paul.guillot@etu.minesparis.psl.eu

Martin Bøgsted
martin.boegsted@rn.dk

Charles Vesteghem
charles.vesteghem@rn.dk

1 Center for Clinical Data Science, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

2 Research, Education, and Innovation, Aalborg University Hospital, Aalborg, Denmark

3 Mines Paris - PSL, Paris, France

looking for relevant shared data, researchers often stumble upon poorly-described, non-standard, and therefore not reusable data [10]. Therefore, developing systematic data sharing is crucial, but it must also be done according to community standards in order to be efficient.

These considerations led to the development of the FAIR principles in 2016 [11]. FAIR stands for "Findable", "Accessible", "Interoperable", and "Reusable". Each term covers a precise set of criteria that must be respected to design a FAIR data sharing environment. It has become a reference in research data sharing, so much so that the European Commission elects the FAIR principles as the gold standard for data sharing and adopted them in the EU directive 1024 of 2019 [12]. The RDA also fosters their development by publishing them in their recommendations, and it has been used by a large number of recent initiatives [13–15].

Health science researchers face additional specific challenges. Firstly, ethical and legal issues are barriers regarding the sharing of IPD. Legislation, like the General Data Protection Regulation (GDPR [16]) in Europe or the Health Insurance Portability and Accountability Act (HIPAA [17]) in the USA, prevents research data from being openly shared. IPD can only be shared publicly after the removal of all information allowing the identification of the individual participants, unless explicit consent has been obtained from the individual participants. Furthermore, the legislation has been growing stronger over the years. State laws have emerged in the USA, like the CCPA in California [18], as well as European legislation such as the Convention 108 [19] or the proposal for a reform of ePrivacy legislation [20].

Secondly, health data are diverse and heterogeneous and can be of very different types and formats, depending on the field they belong to, e.g., imaging, genomics, and mass spectrometry. Handling these data requires specific expertise and tools which can usually only be found in the specialized, dedicated communities.

The objective of this paper is to identify and evaluate technical solutions to implement systematic data sharing in an academic context, in order to help researchers making their data FAIR. We will evaluate various software programs and online platforms used in academic projects to manage and store data through a systematic literature review focusing on the implementation of the FAIR principles and the ability to support sharing of Individual Participant Data (IPD).

2 Materials and Methods

2.1 Selection process

We conducted a systematic literature review of articles detailing, reviewing, or illustrating the use of solutions for data management and sharing. A solution is defined as any infrastructure that can be used to share health research data and meeting the following criteria.

Firstly, the solution must be an online platform or a software program. For instance, a solution can be a public online repository where data can be downloaded directly or an open-source software program for data management and sharing that can be installed on the individual institution's IT infrastructure and made available to researchers.

Secondly, the solution must implement the FAIR principles. Since the FAIR principles were only introduced in 2016, we extended our research not only to solutions explicitly mentioning the FAIR data principles but also solutions facilitating FAIR alike data sharing.

Thirdly, the solution must be adapted to sharing of IPD. Concretely, we looked for infrastructures that contained data relating to patients and health sciences.

Finally, the solution must be usable by any researcher or institution. For an online platform, it means that the protocol to share and access data is fully open, i.e., there is no restriction a priori regarding the persons that can make a data deposit request or a data access request. For a software, it means that any institution or researchers have the right to implement it. For instance, an institutional repository would not meet this criterion, even if it is publicly open, because researchers outside the institution cannot upload data to it. This criterion does not mean that data are open, or that the solution is free to use.

Technical solutions matching these criteria will be referred to as just "solutions" in the following.

To select the solutions of interest for review, we implemented a three-step process. First, we defined the search query (see Section 2.1.1). Second, we selected the articles following a systematic approach (see Section 2.1.2). Then, we extracted and filtered the solutions described in these articles (see Section 2.1.3).

2.1.1 Search query

The search for articles was conducted in PubMed, Scopus, and Web of Science, following an approach based on the PRISMA guidelines [21]. We searched for articles looking at titles and keywords and comparing those to a combination of relevant vocabulary, using the inclusion criteria detailed above. Articles reviewing multiple solutions of interest were also included, and referred as "reviews" in the following. The detailed search queries can be seen in Supplementary Table S1. These queries are somewhat different from one database to another, because they have been adapted to each platform. We obtained 1072 solutions from Scopus, 479 from PubMed and 93 from Web of Science.

2.1.2 Filtering of articles

We found 1644 references, among which 214 were duplicates and three were retracted.

We defined the following exclusion criteria, based on the selection requirements described above.

Exclusion criteria 'Not data sharing': the article is not tackling research data sharing issues. 'Not sharing-focused': the solution described is implementing sharing as a secondary or optional feature. 'Context and guidelines': the article is describing guidelines or state of the art instead of concrete solutions.'Not research': the solution described is not applicable for research projects.'Not platform': the solution described is not a software, a platform, or a repository'.'Not patients': the data that can be hosted are not related to health sciences or IPD. 'Private': the solution described is not usable by every researcher or institution, its use is reserved to a specific group of persons.'Not available': the article was not available. 'Legal': the article tackled legal issues and concerns rather than concrete ways of sharing data.

Articles filtered We filtered the 1427 references based on their title (see Fig. 1), using the exclusion criteria. From the remaining 294 articles, we filtered out 180 based on the abstract (see Fig. 1), using the exclusion criteria. Among the remaining 114 articles, 107 described a single solution and seven were reviews describing multiple solutions (see Fig. 1).

From articles to solutions 112 and 49 solutions were extracted from the single solution articles and the reviews, respectively. 9 solutions were described by both. Additionally, some of the solutions presented multiple use cases. We decided to consider these use cases as individual solutions. After removing duplicates and separating solutions based on individual use case, 152 solutions were extracted from the 114 articles (see Fig. 1).

2.1.3 Filtering of solutions

Exclusion criteria A partial review was conducted on the solutions extracted from the articles selected above. The objective of this partial review was to obtain enough information on the potential solutions of interest to assess whether they would fit with the criteria of this review. Concretely, we went through each potential solution by reading the associated article(s) and consulting the associated website, when available, to select only those that were not private, inactive, or too customizable (see Fig. 1).

'Private', "Not patients" and 'Not research' refer to the criteria already mentioned and used in the filtering of articles (see Section 2.1.2). 'Too specific' means that the initiative tackles only a very specific field of health science, such as a specific disease. The remaining initiatives could then be grouped as belonging to four fields: general-purpose; clinical; imaging; and "omics", e.g., genomics, metabolomics, and proteomics. 'Inactive' means that a solution cannot be used anymore. For instance, the associated website or links are expired, the solution was only temporary, or it was a service that has been officially closed. 'Too customizable' refers to solutions that are not data sharing infrastructures by themselves, but rather software or platforms used to build very customizable data sharing solutions through an extensive amount of development. Although they can be interesting to use for an institution, they offer so many possibilities that it is not possible to define a specific use case, specific capacities, or FAIR implementation. Therefore, we considered their inclusion irrelevant, as they could not be compared to the other solutions.

Using the whole article and sometimes the corresponding website, we were therefore able to decrease significantly the number of solutions to consider, which allowed us to spend more time on each of these solutions.

Solutions filtered From the 152 solutions, we filtered out 117 solutions through the partial review. Sometimes, a solution would not meet the requirements described above, but the software with which it was built would meet them. In this case, we included the software itself, rather than the actual solution described by the article.

50 were considered too specific, 35 were private, 15 were inactive, 8 were deemed too customizable, 4 were not interested in patients, and 4 were not applicable for research projects (see Fig. 1). Finally, we ended up conducting a full analysis of 35 solutions.

Selection process based on the PRISMA guidelines. The difference is that, instead of looking for studies, we are looking for solutions described in articles. The filtering of solutions was performed using all available information in the original article and associated website (if available). Such a process was particularly necessary to obtain up-to-date and practical information on the level of activity of the solution, its accessibility, and the specificities of the hosted data. The exclusion criteria are described in detail in Sections 2.1.2 and 2.1.3.

2.2 Analysis

Our first objective was to obtain an overview of each solution. We looked into the type of sharing (controlled or open), the type of data network (how the data are maintained by the institutions and if the solution is deployed as one instance or as multiple instances), the presence of project management features, the ability to store data, and the presence of analysis features (see Table 1).
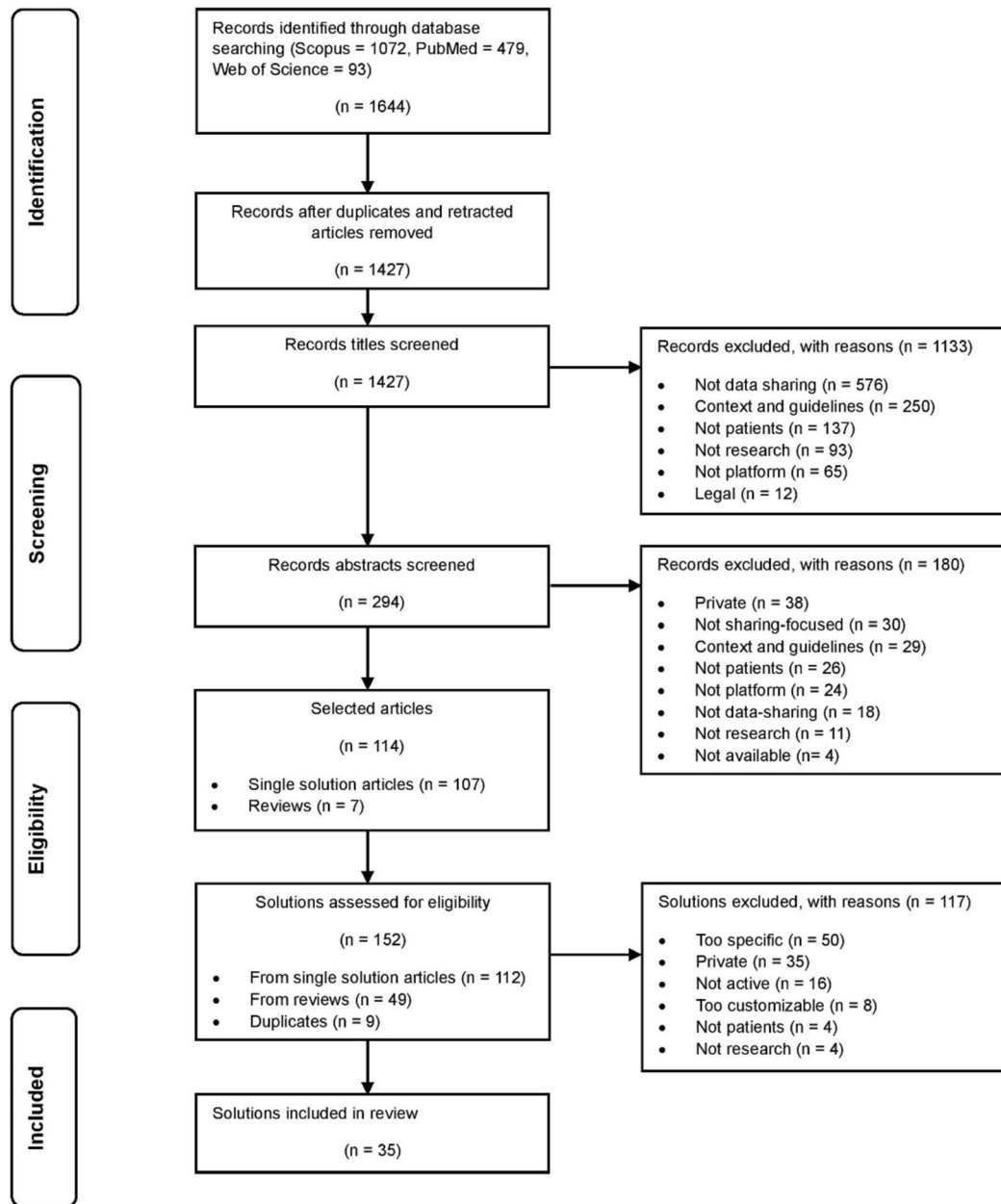
Fig. 1 Selection process of the reviewed solutions

We then assessed each solution regarding three features: search, retrieval, and upload features (see Supplementary Table S2). Finally, we gathered extensive information on the type of supported data, provided access controls, policy regarding IPD, hosting location, funding and affiliation, cost, current implementation/uptake, customizability, code availability, support, and maintenance (see Supplementary Tables S3 and S4). The final step consisted in evaluating each solution according to the

three following criteria, following an approach similar to Vesteghem et al.[22] (see Table 2).

2.2.1 FAIRness

Evaluating the FAIRness of a solution consisted in assessing if it is implementing the FAIR principles. It can be either an explicit or implicit implementation. Indeed, some solutions described how they are respecting the FAIR criteria,

Table 1 Reviewed solutions and features

Name	Data network Federated	Independent	Centralized	Sharing Open	Controlled	P	S	A
B2SHARE - for institutions [23, 24]		x			x	x		
B2SHARE - repository [23, 25]			x		x	x	x	
BBMRI-ERIC [6, 26]	x				x[a]			
BRICS [27, 28]		x			x[b]	x	x	
COMPARE Data Hub [29]		x			x	x	x	x
CyVerse Data Store [30, 31]			x		x	x	x	x
CyVerse instance [30, 32]		x			x	x		
DataFed [33, 34]	x				x	x		
Dataverse [35–37]		x			x	x		
Dataverse Harvard [35–38]			x		x	x	x	
dbGaP [39, 40]			x		x[c]		x	
DECIPHER [41, 42]			x		x	x	x	
Digital Commons [43–45]		x			x		x	
Dryad [46, 47]			x	x			x	
e!DAL [48–50]	x			x				
EGA [40, 51]			x		x		x	
FAIRDOM SEEK [52, 53]		x			x	x		
FAIRDOMHub [53, 54]			x		x	x	x	
figshare - for institutions [55, 56]		x			x	x	x	
figshare - repository [55, 57]			x		x		x	
GDC [58, 59]			x		x		x	x
GNPS [60, 61]			x		x		x	x
IDA [62, 63]			x		x	x	x	
Mendeley Data Repository [64, 65]			x		x		x	
Menoci [66, 67]		x			x	x	x	x
MetaboLights [68, 69]			x	x			x	
NIAID ImmPort [40, 70]			x		x		x	x
NMDR [71, 72]			x	x[d]			x	
Open Science Framework [47, 73]			x		x	x	x	x
OpenNeuro [74, 75]			x	x[e]			x	
Project Datasphere [47, 76]			x	x			x	x
Vivli [77, 78]			x		x		x	x
XNAT Central [79, 80]			x		x		x	
XNAT for data sharing [81–83]		x			x	x		x
Zenodo [44, 45, 47, 84]			x		x		x	

Sharing: 'Open': data are in open access. 'Controlled': data owners or a DAC retain control on who can access their data. *Data Network:* 'Federated': all contributing institutions are nodes of a network that can be queried to access any data that are part of it. The data are maintained and protected locally by each institution. 'Independent': the institution owning the data has its own self-maintained data network. An instance of the solution is deployed for each project, institution, or community that is using it. 'Centralized': all data submitted to the solution are gathered in one place and can be queried through the same interface. 'P': project management tools are provided. 'S': storage capacity is provided. 'A': analysis tools are provided

[a] Control is retained by each institution that is part of the federated network through a Data Access Committee (DAC)

[b] Data owners control the access to their data for 1 year. Then a DAC is set up by the BRICS' instance

[c] Data access requests are reviewed by an appropriate DAC

[d] Data owners can keep their data private for 1 year at maximum

[e] Data owners can keep their data private for 3 years at maximum

Table 2 Evaluation of the reviewed solutions

Name	F	A	I	R	Use / Implementation	
					Researcher	Institution
B2SHARE - for institutions	green	green	yellow	green	green	orange
B2SHARE - repository	green	green	green	green	yellow	green
BBMRI-ERIC	green	yellow	white	green	green	yellow
BRICS	green	yellow	green	green	green	green
COMPARE Data Hub	yellow	yellow	yellow	yellow	green	orange
Cyverse Data Store	green	green	green	green	yellow	green
Cyverse instance	green	green	green	green	green	orange
DataFed	green	yellow	yellow	green	green	green
Dataverse	green	green	yellow	green	green	orange
Dataverse Harvard	green	green	green	green	yellow	green
dbGaP	green	green	green	yellow	red	green
DECIPHER	yellow	green	green	green	orange	green
Digital Commons	yellow	orange	yellow	green	green	yellow
Dryad	green	green	green	green	yellow	green
e!DAL	green	green	green	green	green	orange
EGA	yellow	green	green	yellow	orange	green
FAIRDOM SEEK	green	green	yellow	green	green	orange
FAIRDOMHub	green	green	yellow	green	green	green
Figshare - for institutions	green	green	green	green	green	orange
Figshare - repository	green	green	orange	yellow	green	green
GDC	green	green	green	green	red	green
GNPS	green	green	white	green	green	green
IDA	orange	yellow	green	green	yellow	green
Mendeley Data Repository	green	green	yellow	green	yellow	green
Menoci	green	yellow	green	green	green	orange
MetaboLights	yellow	green	green	yellow	yellow	green
NIAID ImmPort	green	green	green	green	yellow	green
NMDR	green	green	green	green	yellow	green
Open Science Framework	green	green	yellow	yellow	yellow	green
OpenNeuro	green	green	green	green	yellow	green
Project Datasphere	green	green	green	yellow	yellow	green
Vivli	green	green	green	green	orange	green
XNAT Central	yellow	green	orange	orange	yellow	yellow
XNAT for data sharing	white	white	white	white	green	orange
Zenodo	green	green	yellow	yellow	green	green

The cell color indicates the level of fulfillment of the associated criterion. 1. FAIR criteria: Red: none of the subcriteria where fulfilled. Orange: less than the majority of subcriteria and subsubcriteria were fulfilled. Yellow: more than the majority of subcriteria and subsubcriteria were fulfilled. Green: all subcriteria were fulfilled. The choice between orange and yellow was discussed between all authors when necessary. 2. Ease of use and implementation: from red to green, a solution is less time consuming and more straightforward

whereas some did not cite the FAIR principles, but implemented features that, in practice, allow to respect these principles. The FAIR principles contain four main criteria, which can all be subdivided further. To facilitate reading, we gave one grade for each main criteria, based on the number of sub criteria fulfilled.

2.2.2 Ease of use for the researcher

This criterion evaluates the protocols the researcher must go through to use the solution, i.e., registration, data access, and data submission. For instance, an online platform with easy registration, open data download, and open data upload, would be considered very straightforward to use, as opposed to a solution requiring complex procedures for registration, data download and upload. However, this criterion does not say anything about the compliance of a solution to GDPR rules. For example, a repository that is completely open and contains only anonymized data would respect GDPR while also being very straightforward to use. The contrary is also possible. A repository that implements access controls would be considered less straightforward to use than an open access repository, but it can still suffer a leak of confidential information. Its safety is therefore never entirely guaranteed.

2.2.3 Ease of implementation for the institution

This criterion evaluates the implementation process of the solution. For example, a software as a service with provided support and storage would be considered straightforward to roll out. Conversely, a software that has to be installed and maintained locally would be more difficult to set-up and maintain.

3 Results

3.1 Categorization of the results

The 35 reviewed solutions are heterogeneous regarding their features, but they can be broadly divided into two groups.

Firstly, the solutions that are implemented as a single global instance (referred to as "Centralized" in Table 1, n = 22) are online platforms that can be used after a simple registration or sometimes an approval process, like EGA and GDC. They aggregate data in a single place, through a centralized network, enabling researchers to easily work on and publish their data. These platforms are usually not customizable but are straightforward to use. Some are suited for project management (n = 7/22) and even provide analysis tools (n = 7/22), but most are only meant to be used as publishing and archiving platforms.

Secondly, the solutions that can be distributed across institutions (referred to as "Independent" in Table 1, n = 10)

are instances of a software, managed by the researcher's organizations, with a few customizable options like metadata standards, access controls, or members permissions. Data are maintained on an independent network (n = 8/10) but not necessarily on-premises since storage can sometimes be provided as a service (n = 5/10). They implement tools for data and project management (n = 9/10) but can also be used as publishing platforms. Lastly, three solutions are referred to as "Federated" in Table 1. Their specificity is that the data are managed by the owning institutions (which is similar to an independent network), but they are made available for query through one global interface, and not through separated instances (which is similar to a centralized network).

3.2 Evaluation

3.2.1 Findability

Findability was almost always satisfied (see Table 2). Digital Object Identifier (DOI) implementation is becoming increasingly popular, and most solutions have a search tool allowing to easily browse the data and metadata. A DOI is a very relevant item for Findability, because it references a dataset in a unique manner, while also making it searchable through any web browser. Moreover, all the solutions presented here were selected after verifying that public data were in practice findable by anyone on any web browser, since we excluded 'private' solutions during the selection process (see Fig. 1). Some solutions, such as CyVerse Data Store [31], DataFed [33, 34], IDA [62, 63], and XNAT for data sharing [81], need an account to be explored but they all have a free and open process. Data on IDA are especially hard to find even with an account because they are organized in projects with little to no description of the data, even though external links to the projects' websites are provided.

3.2.2 Accessibility

Accessibility was satisfied in almost all cases (see Table 2). Indeed, the data are usually fully open, or controlled with the possibility to submit an access request, or equipped with enough information to contact the owner by other means (see Supplementary Table S3). The access requests can be very different from one solution to another (see Supplementary Table S3). The simplest ones are the equivalent of sending a message to the owner through the platform. The more complicated require researchers to upload a study proposal which is then reviewed by a committee. In all cases, these protocols are "open, free, and universally implementable" (criterion A.1.2 of the FAIR principles [11]). It is worth noting that apart from OpenNeuro no solution explicitly mentions that metadata would stay accessible even if the data were to be no longer available (criterion A.2 of the FAIR

principles [11]). However, many solutions suggest that they can be used to display metadata while the actual data are stored elsewhere and linked by a persistent and unique identifier. Digital Commons is the most limited of all solutions, because it in practice hosts articles rather than datasets, even though it can theoretically do both. Therefore, we deemed the data underlying these articles to be poorly Accessible.

3.2.3 Interoperability

Interoperability is the main limiting factor for FAIR-compliance (see Table 2). On the one hand, general-purpose platforms will be able to store any type of data but do not implement all the appropriate vocabularies and formats necessary to curate, standardize, and visualize specialized data. For instance, regarding figshare repository, we could not find information indicating that community standards or vocabularies are suggested to the uploaders. On the other hand, specialized platforms will provide all the tools necessary to share any data associated with a specialized research community, e.g., genomics or neuroimaging, and might have well-curated metadata schemas and standards specifically designed to handle that type of data. They inherently lack the ability to support any data that are outside the scope of their field. XNAT Central is not scoring well in terms of interoperability. It is a neuroimaging repository whose data are not checked or curated upon upload. In practice, it contains a lot of empty or poorly described datasets.

3.2.4 Reusability

Reusability was satisfied most of the time (see Table 2). Many solutions implement popular metadata standards such as Dublin Core or DataCite, enabling metadata to be both rich and adapted. However, the main drawback of these general metadata standards is that most solutions only require a few fields to be filled, leading to datasets described with minimal information: title, author, contact, and a small description. XNAT Central is a good illustration of this situation, since it is an open, non-curated repository. It contains non-standardized and poorly described datasets, which prevents data from being Reusable. Some solutions like EGA, GDC, and Dryad, however, implement a review process upon data submission, to ensure compliance with standards.

3.2.5 Ease of use and implementation

This evaluation is a result of the various specificities of each solution, such as access protocols, conditions for data deposit, and more generally all the characteristics described in Table 1, Supplementary Tables S3 and S4. Online platforms are mostly straightforward to use (n = 17/22 "researcher" cells colored yellow or green in Table 2). A

simple registration is usually enough to use all the functionalities of the platforms. Only dbGaP and GDC are time-consuming. This is due to the fact that they host sensitive data that have not been anonymized, and thus implement a long protocol of submissions and reviews. EGA implements a lighter version of such a protocol.

Software instances, on the other hand, usually require institutions to spend time on the installation and customization (n = 8/10 "institution" cells colored orange in Table 2). Therefore, data submitters have the comfort of working with locally managed data.

3.3 IPD policy

We found that only a bit more than half of the solutions is mentioning anonymization considerations (n = 19/35) (see Supplementary Table S3). Almost all of them require data to be anonymized before the upload even when access controls are provided. Vivli and IDA are the only platforms offering help with the anonymization. Some solutions like EGA, dbGap, and GDC, accept sensitive data, but they also have much more advanced access protocols, ensuring compliance with the data sharing legislation. In fact, these platforms are directly in relation with the instances responsible for the legislation: dbGaP and GDC are funded and administered by the NIH and EGA is part of the ELIXIR consortium which is partly funded by the European Commission.

3.4 Examples

In this section, we present in more details two solutions that illustrate the previously identified categories: Vivli for the online platforms, which are implemented as a single instance and Dataverse for the software instances, which are distributed across institutions (see Section 3.1). These solutions were chosen because they present enough differences to be representative of the landscape of available solutions. They also provide a large amount of relevant up-to-date documentation and are well-implemented in their respective communities.

3.4.1 Dataverse

Dataverse is an open-source web application to share, preserve, cite, and explore research data [36]. The underlying software must be installed and configured by the institution. It then constitutes a Dataverse repository which can host multiple virtual archives called Dataverse collections, which contain datasets, which consists of files and metadata. figshare [56] and B2SHARE for institutions [23, 24] as well as CyVerse [30, 32], Digital Commons [43], and XNAT for data sharing [81] also provide similar institutional repositories, with variability regarding maintenance, storage, and cost. Researchers of the institution can create

Dataverse collections to deposit data featured in a project or in a published article.

The FAIR principles are explicitly mentioned as the first feature of the software [85], and Dataverse is cited as a viable tool for data sharing in the article that introduces the FAIR principles [11]. The following information were extracted from the articles screened above and the corresponding website [35–37, 86].

Findability The structure of the Dataverse instances systematically guarantees Findability. Dataverse uses DOIs as well as Universal Numeric Fingerprint (UNF) which are globally unique and persistent identifiers (F1: first criterion of the Findability criteria). These identifiers are registered in the metadata, which cannot be separated from the data themselves, as they are bundled together in a single entity called dataset (F3). These datasets are contained in collections and can be searched and accessed through the Dataverse instance (F4). The search tool itself can be easily integrated into an institutional website. It is, however, the choice of the said institution to make the Dataverse instance searchable for everyone or to remain private.

Accessibility Like for the search tool, the rest of the Dataverse infrastructure can be integrated in a website. This allows all interactions (e.g., data access, submission, requests) to take place in a single interface, and guarantees that the data owner can be contacted.

The retrieval of (meta)data from a Dataverse collection or dataset depends on the level of protection the data owner has chosen. If the data are in open access, one can simply download the data in a few clicks. If the access to data is controlled, it is necessary to authenticate oneself (A1.2) and submit an access request, through the Dataverse interface. The data owner can then choose to allow access, or not, to its data. In every case, this is a free, open and universally implementable protocol (A1.1). Once again, it is possible for the owner to make the data totally private, or hidden, to all users. It is not mentioned that metadata remain accessible "even when the data are no longer available" (A2). However, it is possible to create empty datasets, which means that metadata can be hosted on Dataverse, without any uploaded data. This helps respecting the criterion A2, although not explicitly.

Interoperability The evaluation of the interoperability was more delicate. As Dataverse is not a specialized software, it does not provide community guidelines, nor curation of data, even though it advises to follow vocabularies and good practices. It is the responsibility of the institution to make sure the uploaded data are interoperable. Nonetheless, Dataverse provides tools that help facing these challenges, such as customizable metadata schemas, and some community schemas and ontologies: Data Documentation Initiative

(DDI) for social and health sciences, DATS for life sciences and The Gene Ontology for molecular biology and genetics. Moreover, metadata can contain references to other data (I3), such as a scientific publication, or any website.

Reusability Dataverse implements rich metadata schemas, such as Schema.org, DataCite, and DublinCore. License and terms of use are available in the metadata (R1.1), and detailed provenance can be provided (R1.2). Once again, some metadata fields are mandatory, but it is the responsibility of the researcher to fill the additional fields necessary to make its data Reusable. Dataverse provides the infrastructure for it but cannot guarantee it.

NB: for software instances, it is always the responsibility of the institution to decide the level of findability and accessibility. This evaluation of the FAIR criteria is solely based on the possibilities offered by the software, not on the practical choices made by the users.

However, Dataverse does not explicitly mention the sharing of IPD. Nonetheless, all data are stored on-premises by the managing institution and Dataverse collections can contain metadata alone when the data files are too sensitive to be shared. This is a way to ensure the Findability of the data while respecting data sharing legislation. Additionally, it is not a specialized repository which means that every file format is accepted. However, only a few can be previewed, e.g., images, PDF, text, video, tabular data, and other basic formats. To better understand the concrete implementation of Dataverse, one can browse one of the 93 installations [86] at time of writing (12 December 2022) or try to use Dataverse Harvard [38] for free which is a public instance of the software.

3.4.2 Vivli

Vivli [77, 78] is an online repository hosting anonymized clinical data. Anyone can search for clinical studies on Vivli. However, access and upload of data are controlled by various protocols and agreements. These protocols are similar to the ones implemented by EGA [51], dbGaP [39], and GDC [58, 59] for individually identifiable data, although to some extent only, since the data on Vivli are systematically anonymized. Moreover, Vivli offers help for the anonymization of datasets before submission, ensuring the sharing of IPD in a secure manner. At time of writing, 12 December 2022, Vivli hosts 6907 studies and 621 data requests have been submitted.

Findability Vivli implements automatic DOI attribution (FAIR criterion F1) and open visualization of the studies through a search tool (F4). A study contains data files, metadata, and a description of the aim of the study. It is registered on ClinicalTrials.gov. The DOI is linked to the study,

which cannot be separated from the metadata and the actual datafiles (F3).

Accessibility The access protocol, although time consuming, is very clear and open (A1.1). Anyone can submit an access request, after creation of a free account, to allow for authentication (A1.2). The request is approved or refused by Vivli within 3 business days (A1.2). Some data can also be public, depending on the choices made by the data contributor.

Additionally, metadata are always public and available, even if there are no data files uploaded, although it is not explicitly mentioned what would happen if datafiles were deleted (A2).

Interoperability and reusability Vivli encourages researchers to use rich metadata, dictionaries, and ontologies (e.g., the Cochrane ontology) (I1, I2, R1.3). It also reviews all uploaded data, which improves Interoperability and Reusability. When requesting data from a study, it is necessary to sign a Data Use Agreement (DUA), with clear terms of use and license (R1.1). All studies are richly described, give provenance information (R1.2), and additional information are available on ClinicalTrials.gov.

4 Discussion

4.1 Summary of the findings

We analysed the 35 reviewed solutions, first categorizing their implementation (see Section 3.1) then focusing on their compliance with the FAIR principles (see Section 3.2).

Findability and Accessibility were satisfied most of the time. Interoperability, however, was shown to be the main obstacle regarding the fulfilment of the FAIR principles, since both general-purpose and specialized platforms have inherent interoperability limitations.

Additionally, we identified which solutions are mentioning the sharing of IPD and anonymization issues. Unsurprisingly, most of them are dedicated to the field of health sciences. Across all the reviewed solutions, it appears to us that there is no platform capable of hosting all IPD types in a FAIR way. The most adapted for sharing diverse IPD might be Vivli because it provides a service for anonymization, focuses on clinical data, is highly FAIR compliant, and is straightforward to use (see Section 3.4.2). Dataverse is also an interesting alternative for institutions that would like to retain complete control over their data and have enough customizability and flexibility to adapt to different data types (see Section 3.4.1).

Institutions and researchers face sharing challenges that call for many important considerations, especially regarding sensitive data, e.g., the hosting location, the choice of sharing anonymized or raw data, the cost of the solution, and its adaptability. In all likelihood, the best solution will not be perfect in all regards, but it will be the best compromise for all these issues.

4.2 Strengths and limitations

Through a systematic process, we included solutions dating from both before and after the creation of the FAIR principles, as well as relevant reviews. We aimed at defining the inclusion and exclusion criteria as clearly as possible, to make the selection process as transparent as possible. However, because the review was conducted in an academic context, it is possible that some relevant results were missed. The inclusion of reviews with different prospecting methods (e.g., using the Google search tool directly [47]) helped us gather additional solutions.

Moreover, most of the articles screened by the review were published in the last four years, which means that this article is only a snapshot of what exists at time of writing. Therefore, we do not claim that the list of solutions provided here is complete, but we believe it is representative of the current health data sharing landscape. The recent evolutions of legislation in Europe and in the USA might create changes in this landscape. Notably, the European Health Data Space [9] should be followed closely over the coming years due its potential large-scale impact on data sharing in Europe.

Solutions were described and evaluated through a clear grid, using information from both academic literature and corresponding websites. Not all the information needed was always findable or open, leading to some difficulties in the evaluation and the description. This was particularly true about sustainability/maintenance, funding and hosting location. Regarding the latter, we tried to identify at least the country or area of storage: for instance, the USA or the European Union. Also, a more detailed look at the solutions remains necessary to fully understand how they work. In the case of a software, it is often possible to ask for a demo. For online platforms, it is often possible to create a free account.

Additionally, the evaluation of the FAIR principles was a very heterogeneous process. Some solutions had explicit and clear justifications of their FAIR-compliance, while others—especially those created before the FAIR principles—had these information not directly available, notably for Interoperability and Reusability. In these cases, we had to base our evaluation on the data available on the platform. The evaluation was primarily done by one person, but difficulties arising from the lack of information were discussed by all co-authors after evaluation.

Even though the FAIR evaluation is the main contribution of this review, it must be read and understood in the

context of the descriptions provided in Table 1, Supplementary Tables S2, S3, and S4. Indeed, the choice of a relevant solution cannot rely solely on the FAIR principles' implementation. For any institution, a broader understanding of how each solution works and interacts with its users is necessary in order to make an informed decision, which should be discussed thoroughly between researchers, administrators and data owners. In the end, great communication and coordination are as important as complying to the FAIR principles.

4.3 Alternative strategies

An interesting strategy would be to use a combination of specialized repositories with more general and well implemented platforms. For instance, research data could be uploaded to several specialized repositories (e.g., Open-Neuro, MetaboLights, and IDA), while metadata are being displayed on a general-purpose platform (e.g., Dataverse, or figshare) with persistent identifiers linking back to the specialized repositories. The former ensures Interoperability and Reusability, while the latter implements Findability and Accessibility.

Some potentially relevant solutions that were excluded from the analysis due to their lack of data hosting functionalities, could also be used to build a FAIR data infrastructure. Such a solution is ClinicalTrials.gov [87]. It is a central resource in health science (with more than 400,000 registered research studies), provided by the U.S. National Library of Medicine, and is relevant for ensuring Findability of clinical trials data, because it hosts rich descriptions of the studies. It can very well contain metadata with an identifier pointing towards the actual hosting location of the data.

A last possibility could be to build an institutional portal either from scratch or using a framework solution, for example among the ones excluded during the selection process [88–99] because they were 'too customizable' (See Section 2.1.3). Such an approach would allow the institution to build precisely according to its needs but requires a lot more time and resources, notably in terms of specifications to ensure compliance with the FAIR data principles.

4.4 Comparison with other studies

To the best of our knowledge, Banzi et al. [47] is the closest related review. It assesses the suitability of 25 repositories for hosting clinical trials data and implicitly evaluates their FAIR-compliance.

Comparison of the exclusion criteria shows that Banzi et al. focus on a slightly different type of solutions. Firstly, they chose to include disease-specific repositories and national/institutional repositories. On the contrary, we

excluded these solutions, because they respectively focus on a specific field and a specific group of researchers. The idea was to provide a list of solutions that could be interesting for most health science researchers, instead of very specialized solutions that would only interest a few and might be already well-known within their dedicated communities. Secondly, Banzi et al. focused on clinical trials data alone whereas we included health science data in general. The rest of the criteria, however, are quite similar, such as focusing on research data sharing, including general-purpose repositories.

Regarding the evaluation of the solutions, different choices were made. We evaluated, e.g., the FAIR principles explicitly, whereas Banzi et al. partially evaluated some of the core criteria of the FAIR principles. We did not tackle long term preservation capacity. The main reason for this is that this information was usually not findable which is also highlighted by Banzi et al.

Banzi et al. also identified Vivli as the best solution for sharing clinical IPD (see Section 4.1).

5 Conclusion

In the vast and complex landscape of health science data sharing, with a significant and moving regulatory requirements, it is important to understand and characterize stakeholders' requirements and needs before investing time and efforts in one or several solutions.

In this study, we compared 35 solutions regarding their implementation of the FAIR principles, their ease of use and implementation, and we richly described their functionality. Vivli and Dataverse were identified as the two most all-round solutions for sharing health science data in a FAIR way.

Even though the FAIR principles are being increasingly implemented, this article shows that a lot of work still needs to be done to reach standards and formats that would make health science research data FAIR. Fostering data sharing practices, for instance by showing to researchers the self-benefits of sharing, is also necessary in order to improve the uptake on FAIR solutions and develop reliable, well-implemented communities.

Declarations

References

1. Berman F, Wilkinson R, Wood J. Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance. D-Lib Magazine 2014; 20. http://www.dlib.org/dlib/january14/01guest_editorial.html. https://doi.org/10.1045/january2014-berman.

2. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance). 2022. http://data.europa.eu/eli/reg/2022/868/oj/eng. Accessed 15 Oct 2023.

3. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act). 2022. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN. Accessed 15 Oct 2023.

4. NIH GREI - NIH Office of Data Science Strategy Announces New Initiative to Improve Access to NIH-funded Data. https://datascience.nih.gov/news/nih-office-of-data-science-strategy-announces-new-initiative-to-improve-data-access. Accessed 7 Oct 2022.

5. Horizon 2020 - Details of the EU funding programme which ended in 2020 and links to further information. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en. Accessed 13 Oct 2022.

6. Mayrhofer MTh, Holub P, Wutte A, Litton J-E. BBMRI-ERIC: the novel gateway to biobanks. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2016;59:379–84. https://doi.org/10.1007/s00103-015-2301-8.

7. EUDAT - Research Data Services, Expertise & Technology Solutions. https://eudat.eu/. Accessed 13 Oct 2022.

8. ELIXIR - A distributed infrastructure for life-science information. https://elixir-europe.org/. Accessed 14 Oct 2022.

9. Proposal for a regulation of the European parliament and of the council on the European Health Data Space. 2022. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197. (Accessed 15 Oct 2023).

10. Tenopir C, Allard S, Douglass K, et al. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. 2011;6: e21101. https://doi.org/10.1371/journal.pone.0021101.

11. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3: 160018. https://doi.org/10.1038/sdata.2016.18.

12. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). 2019. http://data.europa.eu/eli/dir/2019/1024/oj/eng. Accessed 15 Oct 2023.

13. Arguillas F, Christian T-M, Gooch M, Honeyman T, Peer L, WG C-F. 10 Things for Curating Reproducible and FAIR Research. 2022; https://doi.org/10.15497/RDA00074.

14. Making Data Work for Cross-Domain Grand Challenges: the CODATA Decadal Programme. https://codata.org/initiatives/decadal-programme2/. Accessed 7 Oct 2022.

15. IMI2 Call 12. http://www.imi.europa.eu/apply-funding/closed-calls/imi2-call-12. Accessed 7 Oct 2022.

16. EUR-Lex - 32016R0679 - EN - EUR-Lex. https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed 15 Oct 2023.

17. Rights (OCR) O for C. Health Information Privacy. 2021. https://www.hhs.gov/hipaa/index.html. Accessed 15 Oct 2023.

18. California Consumer Privacy Act (CCPA). 2018. https://oag.ca.gov/privacy/ccpa. Accessed 15 Oct 2023.

19. Convention 108 et Protocoles - Protection des données - www.coe.int. https://www.coe.int/fr/web/data-protection/convention108-and-protocol. Accessed 15 Oct 2023.

20. Proposal for an ePrivacy Regulation | Shaping Europe's digital future. 2023. https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation. Accessed 15 Oct 2023.

21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. The BMJ. 2009;339:332–6. https://doi.org/10.1136/BMJ.B2535.

22. Vesteghem C, Brøndum RF, Sønderkær M, et al. Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives. Brief Bioinform. 2020;21:936–45. https://doi.org/10.1093/BIB/BBZ044.

23. Berenji S, Hakansson CJ, Laure E, et al. B2SHARE: An Open eScience Data Sharing Platform. In: 2015 IEEE 11th International Conference on E-Science. 2015: 448–453. https://doi.org/10.1109/eScience.2015.44.

24. B2SHARE - For institutions. https://eudat.eu/catalogue/b2share. Accessed 12 Oct 2022.

25. B2SHARE - Repository. https://b2share.eudat.eu/help. Accessed 12 Oct 2022.

26. BBMRI-ERIC: Making New Treatments Possible. https://www.bbmri-eric.eu/. Accessed 18 Oct 2022.

27. BRICS - Introducing BRICS. https://brics.cit.nih.gov/intro. Accessed 13 Oct 2022.

28. Navale V, Ji M, Vovk O, et al. Development of an informatics system for accelerating biomedical research. F1000Research 2019; 8:1430. https://doi.org/10.12688/f1000research.19161.2.

29. Amid C, Pakseresht N, Silvester N, et al. The COMPARE Data Hubs. Database : the journal of biological databases and curation. 2019. https://doi.org/10.1093/database/baz136.

30. Wieser F, Stryeck S, Lang K, et al. A local platform for user-friendly FAIR data management and reproducible analytics. J Biotechnol. 2021;341:43–50. https://doi.org/10.1016/j.jbiotec.2021.08.004.

31. CyVerse Data Store - Discovery Environment. https://de.cyverse.org/. Accessed 12 Oct 2022.

32. CyVerse. https://cyverse.org/. Accessed 18 Oct 2022.

33. Stansberry D, Somnath S, Breet J, Shutt G, Shankar M. DataFed: Towards reproducible research via federated data management. In: Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019. Institute of Electrical and Electronics Engineers Inc., 2019: 1312–1317. https://doi.org/10.1109/CSCI49370.2019.00245.

34. DataFed - A Scientific Data Federation. https://ornl.github.io/DataFed/. Accessed 13 Oct 2022.

35. Jamwal V, Kaur S. Global presence of open-source research data management platform for libraries: the Dataverse project. Library Hi Tech News. 2021;38:8–12. https://doi.org/10.1108/LHTN-10-2021-0066.

36. King G. An introduction to the dataverse network as an infrastructure for data sharing. Sociological Methods and Research. 2007;36:173–99. https://doi.org/10.1177/0049124107306660.

37. Anggawira D, Mayesti N. The Indonesian national scientific repository: A case study of research data sharing. Preservation, Digital Technology and Culture. 2020;49:14–25. https://doi.org/10.1515/pdtc-2019-0015.

38. The Harvard Dataverse. https://dataverse.harvard.edu/. Accessed 12 Oct 2022.

39. dbGaP - NCBI. https://www.ncbi.nlm.nih.gov/gap/. Accessed 10 Nov 2022.

40. Huser V, Shmueli-Blumberg D. Data sharing platforms for de-identified data from human clinical trials. Clin Trials. 2018;15:413–23. https://doi.org/10.1177/1740774518769655.

41. Chatzimichali EA, Brent S, Hutton B, et al. Facilitating Collaboration in Rare Genetic Disorders Through Effective Matchmaking in DECIPHER. Hum Mutat. 2015;36:941–9. https://doi.org/10.1002/humu.22842.

42. DECIPHER - Mapping the clinical genome. https://www.deciphergenomics.org/. Accessed 13 Oct 2022.

43. Digital Commons. https://bepress.com/products/digital-commons/. Accessed 10 Nov 2022.

44. Amorim RC, Castro JA, Rocha da Silva J, Ribeiro C. A comparison of research data management platforms: architecture, flexible metadata and interoperability. Univers Access Inf Soc 2017; 16:851–862. https://doi.org/10.1007/S10209-016-0475-Y/TABLES/3.

45. Amorim RC, Castro JA, da Silva JR, Ribeiro C. A comparative study of platforms for research data management: Interoperability, metadata capabilities and integration potential. Adv Intell Syst Comput. 2015;353:101–11. https://doi.org/10.1007/978-3-319-16486-1_10.

46. Dryad - Publish and Preserve your Data. https://datadryad.org/stash. Accessed 13 Oct 2022.

47. 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;20:1–10. https://doi.org/10.1186/S13063-019-3253-3/FIGURES/2.

48. Arend D, Koenig P, Junker A, Scholz U, Lange M. The on-premise data sharing infrastructure e!DAL: Foster FAIR data for faster data acquisition. Gigascience 2020; 9. https://doi.org/10.1093/gigascience/giaa107.

49. e!DAL - electronic Data Archive Library. https://edal.ipk-gatersleben.de/. Accessed 13 Oct 2022.

50. Arend D. e!DAL - a framework to store, share and publish research data. 2014; https://doi.org/10.1186/1471-2105-15-214.

51. EGA - European Genome-Phenome Archive. https://ega-archive.org/. Accessed 14 Oct 2022.

52. FAIRDOM-SEEK - An open source web-based cataloguing and commons platform. https://seek4science.org/. Accessed 12 Oct 2022.

53. Wolstencroft K, Krebs O, Snoep JL, et al. FAIRDOMHub: A repository and collaboration environment for sharing systems biology research. Nucleic Acids Res. 2017;45:D404–7. https://doi.org/10.1093/nar/gkw1032.

54. FAIRDOMHub - The better way to manage your data. https://fairdomhub.org/. Accessed 12 Oct 2022.

55. Thelwall M, Kousha K. Figshare: a universal repository for academic resource sharing? Online Inf Rev. 2016;40:333–46. https://doi.org/10.1108/OIR-06-2015-0190.

56. figshare - For institutions. https://knowledge.figshare.com/institutions. Accessed 12 Oct 2022.

57. figshare - Repository. https://figshare.com/browse. Accessed 12 Oct 2022.

58. Heath AP, Ferretti V, Agrawal S, et al. The NCI Genomic Data Commons. Nat Genet. 2021;53:257–62. https://doi.org/10.1038/s41588-021-00791-5.

59. GDC - Genomic Data Commons Data Portal. https://portal.gdc.cancer.gov/. Accessed 13 Oct 2022.

60. Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34:828–37. https://doi.org/10.1038/nbt.3597.

61. GNPS - Analyze, Connect, and Network with your Mass Spectrometry Data. https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp. Accessed 13 Oct 2022.

62. Crawford KL, Neu SC, Toga AW. The Image and Data Archive at the Laboratory of Neuro Imaging. Neuroimage. 2016;124:1080–3. https://doi.org/10.1016/j.neuroimage.2015.04.067.

63. IDA - Image Data Archive. https://ida.loni.usc.edu/login.jsp. Accessed 13 Oct 2022.

64. Mendeley Data. https://data.mendeley.com/. Accessed 18 Oct 2022.

65. Descoteaux D, Farinelli C, E Silva MS, de Waard A. Playing Well on the Data FAIRground: Initiatives and Infrastructure in Research Data Management. Data Intelligence 2019; 1:350–367. https://doi.org/10.1162/DINT_A_00020.

66. Suhr M, Lehmann C, Bauer CR, et al. Menoci: lightweight extensible web portal enhancing data management for biomedical research projects. BMC Bioinformatics. 2020;21:582. https://doi.org/10.1186/s12859-020-03928-1.

67. Menoci - Lightweight data management for biomedical research. https://menoci.io/. Accessed 13 Oct 2022.

68. Salek RM, Haug K, Steinbeck C. Dissemination of metabolomics results: role of MetaboLights and COSMOS. GigaScience. 2013;2:8. https://doi.org/10.1186/2047-217X-2-8.

69. MetaboLights - Metabolomics experiments and derived information. https://www.ebi.ac.uk/metabolights/. Accessed 13 Oct 2022.

70. ImmPort Portal. https://www.immport.org/home. Accessed 16 Nov 2022.

71. NMDR - Metabolomics Workbench. https://www.metabolomicsworkbench.org/. Accessed 13 Oct 2022.

72. Smelter A, Moseley HNB. A Python library for FAIRer access and deposition to the Metabolomics Workbench Data Repository. Metabolomics. 2018;14:64. https://doi.org/10.1007/s11306-018-1356-6.

73. OSF - Open Science Framework. https://osf.io/. Accessed 13 Oct 2022.

74. OpenNeuro. https://openneuro.org/. Accessed 13 Oct 2022.

75. Wiseman S. A FAIR platform for data-sharing. Nat Neurosci. 2021;24:1640. https://doi.org/10.1038/s41593-021-00976-5.

76. Project Data Sphere - Share, Integrate & Analyze Cancer Research Data. https://data.projectdatasphere.org/projectdatasphere/html/home. Accessed 16 Nov 2022.

77. Vivli - A global clinical research data sharing platform. https://vivli.org/. Accessed 13 Oct 2022.

78. Bierer BE, Li R, Barnes M, Sim I. A Global, Neutral Platform for Sharing Trial Data. N Engl J Med. 2016;374:2411–3. https://doi.org/10.1056/NEJMp1605348.

79. Herrick R, Horton W, Olsen T, McKay M, Archie KA, Marcus DS. XNAT Central: Open sourcing imaging research data. Neuroimage. 2016;124:1093–6. https://doi.org/10.1016/J.NEUROIMAGE.2015.06.076.

80. XNAT Central. https://central.xnat.org/app/template/Index.vm. Accessed 13 Oct 2022.

81. XNAT - For data sharing. https://www.xnat.org/. Accessed 12 Oct 2022.

82. Alpert K, Kogan A, Parrish T, Marcus D, Wang L. The Northwestern University Neuroimaging Data Archive (NUNDA). Neuroimage. 2016;124:1131–6. https://doi.org/10.1016/j.neuroimage.2015.05.060.

83. Kennedy DN, Haselgrove C, Riehl J, Preuss N, Buccigrossi R. The NITRC image repository. Neuroimage. 2016;124:1069–73. https://doi.org/10.1016/J.NEUROIMAGE.2015.05.074.

84. European Organization For Nuclear Research, OpenAIRE. Zenodo. 2013; https://www.zenodo.org/. https://doi.org/10.25495/7GXK-RD71.

85. Dataverse - Features. https://dataverse.org/software-features. Accessed 12 Oct 2022.

86. Dataverse - Installations Around the World. https://dataverse.org/installations. Accessed 12 Oct 2022.

87. ClinicalTrials.gov. https://clinicaltrials.gov/. Accessed 10 Nov 2022.

88. CKAN - The open source data management system. https://ckan.org/. Accessed 13 Oct 2022.

89. LabKey - Life Science Software. https://www.labkey.com/. Accessed 13 Oct 2022.

90. Nelson EK, Piehler B, Eckels J, et al. LabKey Server: an open source platform for scientific data integration, analysis and collaboration. BMC Bioinformatics. 2011;12:71. https://doi.org/10.1186/1471-2105-12-71.

91. Tripal v3. https://tripal.info/about/tripal_v3. Accessed 13 Oct 2022.

92. Spoor S, Cheng C-H, Sanderson L-A, et al. Tripal v3: an ontology-based toolkit for construction of FAIR biological community databases. Database: J Biol Databases Curation. 2019. https://doi.org/10.1093/database/baz077.

93. DSpace. https://dspace.lyrasis.org/. Accessed 13 Oct 2022.

94. Ohmann C, Tilki B, Schulenberg T, Canham S, Banzi R, Kuchinke W. Assessment of a demonstrator repository for individual clinical trial data built upon DSpace. F1000Research 2020; 9. https://doi.org/10.12688/f1000research.23468.1.

95. EPrints. https://www.eprints.org/uk/. Accessed 10 Nov 2022.

96. Fedora. https://getfedora.org/fr/. Accessed 10 Nov 2022.

97. HUBzero. https://hubzero.org/. Accessed 13 Oct 2022.

98. Dearborn CC, Barton AJ, Harmeyer NA. The Purdue University Research Repository: HUBzero customization for dataset publication and digital preservation. OCLC Syst Serv. 2014;30:15–27. https://doi.org/10.1108/OCLC-07-2013-0022.

99. Invenio. https://inveniosoftware.org/. (Accessed 10 Nov 2022).