data

MDPI

Article

Perspectives on Research and Personalized Healthcare in the Context of Federated FAIR Data Based on an Exploratory Study by Medical Researchers

Elena Poenaru [1], Monica Dugăeşescu [1,2,*], Călin Poenaru [1], Iulia Andrei-Bitere [1,3], Livia-Cristiana Băicoianu-Nițescu [1,4,*], Traian-Vasile Constantin [1,5], Aurelian Zugravu [1], Brandusa Bitel [1], Maria Magdalena Constantin [1,6] and Smaranda Stoleru [1]

[1] Faculty of Medicine, "Carol Davila" University of Medicine and Pharmacy, 050474 Bucharest, Romania; elena.poenaru@umfcd.ro (E.P.); calin.poenaru@gmail.com (C.P.); traian.constantin@umfcd.ro (T.-V.C.); aurelian.zugravu@umfcd.ro (A.Z.); brandusa.bitel@umfcd.ro (B.B.); maria.constantin@umfcd.ro (M.M.C.); smaranda.stoleru@umfcd.ro (S.S.)
[2] Emergency University Hospital, 050098 Bucharest, Romania
[3] Fundeni Clinical Institute, 022328 Bucharest, Romania
[4] "Elias" University Emergency Hospital, 011461 Bucharest, Romania
[5] Prof. Dr. Theodor Burghele Clinical Hospital, 061344 Bucharest, Romania
[6] Colentina Clinical Hospital, 021156 Bucharest, Romania
* Correspondence: monica.dugaesescu@drd.umfcd.ro (M.D.); livia.nitescu@rez.umfcd.ro (L.-C.B.-N.)

Abstract

Background: Research in personalized medicine, with applications in oncology, dermatology, cardiology, urology, and general healthcare, requires facile and safe access to accurate data. Due to its particularly sensitive character, obtaining health-related data, storing it in repositories, and federating it are challenging, especially in the context of open science and FAIR data. **Methods**: An online survey was conducted among medical researchers to gain insights into their knowledge and experience regarding the following topics: health data repositories and data federation, as well as their opinions regarding data sharing and their willingness to participate in sharing data. **Results**: The survey was completed by 189 respondents, the majority of whom were attending physicians and PhD candidates. Most of them acknowledged the complex, beneficial implications of data federation in the medical field but had concerns about data protection, with 75% declaring that they would agree to share data. A general lack of awareness (80%) about the importance of interoperability for federated data repositories was observed. **Conclusions**: Implementing federated data repositories in the health field requires thorough understanding, knowledge, and collaboration, enabling translational medicine to reach its full potential. Understanding the needs of all involved parties can shape the success of medical data federation initiatives, with this study serving as a foundation for further research.

Keywords: FAIR data; data federation; personalized medicine; health data repository; readiness score; medical research data

check for updates

Academic Editors: Johannes Schobel and Jorge dos Santos Oliveira

Received: 21 August 2025
Revised: 12 October 2025
Accepted: 23 October 2025
Published: 11 November 2025

Citation: Poenaru, E.; Dugăeşescu, M.; Poenaru, C.; Andrei-Bitere, I.; Băicoianu-Nițescu, L.-C.; Constantin, T.-V.; Zugravu, A.; Bitel, B.; Constantin, M.M.; Stoleru, S. Perspectives on Research and Personalized Healthcare in the Context of Federated FAIR Data Based on an Exploratory Study by Medical Researchers. *Data* **2025**, *10*, 187. https://doi.org/10.3390/data10110187

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The traditional medical approach, "one-size-fits-all," involves health interventions targeted at an "average" person, making medical decisions efficacious for a relatively small proportion of individuals who resemble this average. Personalized approaches have been proven to be cost-effective and to lead to the best outcomes for patients. Personalized

medicine, also known as precision medicine, involves customizing healthcare based on the characteristics of each patient, considering their genetic, environmental, lifestyle-related, and other factors [1,2]. Understanding the influence of these various factors on patient management and establishing impactful personalized clinical decisions require access to diverse and large-scale medical data, but the accessibility and the possibility to integrate these data are limited in the current context by fragmented data ownership, privacy concerns, and technical barriers [3].

The emerging high-throughput sequencing technologies generate a large amount of data, in addition to data collected in clinical practice and various research studies, leading to a data overload in the biomedical field and requiring a shift in omics data collection, integration, and analysis [4,5]. Facing the current data overload has two sides: It can be seen as both a challenge and an opportunity. Organizing this valuable data in repositories with broad accessibility and in compliance with FAIR (Findable, Accessible, Interoperable, Reusable) principles can advance both biomedical research and patient care [5,6].

In the current context, when the demand for open science and data is high, and transparency and reproducibility in research are required, data repositories, which allow research data management, preservation, and sharing, are becoming increasingly indispensable. Through data repositories, researchers can store the data they generate in their studies while simultaneously making data available for the scientific community, allowing different researchers to find, access, and use data, thereby promoting collaboration in this field [7].

FAIR principles offer an adequate framework for the sharing of data, including common phenotypes, as well as multi-omics data, including genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics [4,8,9]. According to FAIR principles, research data and linked metadata must be "Findable", "Accessible", "Interoperable", and "Reusable", implying a series of key factors that would enable personalized medicine, including facilitating efficient data retrieval and access for those researchers who can use it for the progress of medicine, data harmonization and integration through standardized formats, terminologies, protocols, and data reuse for new purposes [10]. A very important aspect is that FAIR principles apply not only to the data itself but also to the tools that helped generate the data and are relevant for both human- and machine-driven activities [11]. Using data from multiple sources allows larger sample sizes, increases statistical power, and leads to more accurate generalizability of research conclusions, as well as to valuable discoveries in the study of rare diseases, low-incidence adverse effects, and other investigations that occur rarely and require extensive samples [12]. The concept of data federation implies a collection of potentially heterogeneous, autonomous databases, with a database management system that allows information integration [13,14]. Therefore, through a federation mechanism, data from various locations belonging to different institutions are put together, allowing a unified analysis [15]. Performing data federation in the field of personalized medicine can play a pivotal role in advancing this field [16]. Making data available through adequately designed and implemented data federation systems provides the means to address research questions that cannot be answered otherwise [12].

However, in the field of medical research, data can be highly sensitive due to its personal nature, involving health status, lifestyle, environmental factors, family history, clinical examinations, imagistic investigations, laboratory tests, omics, and other health-related and biological data from living humans. This particular type of data involves ethical, legal, and societal challenges when used for research [10].

In the context of FAIRification and data federation, medical data pose particular challenges due to their sensitivity and the involvement of multiple stakeholders, including patients as data owners, clinicians and medical researchers as data collectors and users, and technical staff as developers and managers of repositories. Clinicians and medical re-

searchers play a central role in bridging these stakeholders, and their opinions, knowledge, and perceptions can significantly influence the success of such initiatives. However, this perspective has not yet been explored in detail.

This article aims to initiate a discussion on data federation and data repositories in healthcare, based on an exploratory study conducted at a Romanian medical university. Through a simple, question-based instrument, a quick snapshot of the knowledge, opinions, and perceptions regarding federated data repositories that respect FAIR principles was obtained from a group of researchers who are involved in both research and patient care, mainly as physicians, have access to patients and real-world data, and might possess a complex perspective on the topic. For this early phase of research, an open and flexible question-based instrument, rather than a standard questionnaire-based methodology, was preferred, as it can serve exploratory purposes more adequately.

While this research does not involve a validated quantitative methodology that enables the generalizability of conclusions, it is essential for exploring a novel field. To our knowledge, there are no studies specifically investigating the perceptions and knowledge of clinicians and medical researchers on federated medical data under FAIR principles. Therefore, this study aimed to identify potential trends and insights for further research.

2. Materials and Methods

2.1. Data Collection

A short, simple, question-based instrument (Supplementary File) was used for this exploratory study to capture insights on the topic of federated data repositories from medical researchers at Carol Davila University of Medicine and Pharmacy, using Google Forms. The documentation was conducted under the FAIR–Impact Project, Support Program for Repositories and Data Services. The questions were distributed through Google Forms.

2.2. Data Description

Google Spreadsheets was used for the descriptive statistical analysis of the data collected through the above-described question-based instrument, with frequencies calculated for the answers provided to each question by the respondents.

2.3. Data Analysis

While the main focus was on descriptive analysis, a few basic associations and correlations were studied using the software IBM SPSS Statistics (Version 20) to highlight potential trends within the group of respondents. For further analysis, we defined a set of variables as follows:

- "The awareness of the existence of data repositories within the institution"—either present or absent.
- "The level of interaction with health-related data"—a score based on the responses to the related question, as follows: none (score 0), data collection (+1), access (+1), analysis (+1), and storage (+1).
- "The experience with a diversity of types of data from repositories"—a score based on the responses to the related question, as follows: none (score 0), health status (+1), lifestyle (+1), environment (+1), family history (+1), clinical examination (+1), imagistic investigations (+1), laboratory tests (+1), omics (+1), and other data (+1).
- "The number of categories of data considered useful if stored in repositories"—a score based on the responses to the related question, as follows: none (score 0), health status (+1), lifestyle (+1), environment (+1), family history (+1), clinical examination (+1), imagistic investigations (+1), laboratory tests (+1), omics (+1), and other data (+1).

- "The agreement to share the data collected during professional activity in a data repository"—either present or absent.

The relevant questions for the above-defined variables are as follows:

- Data repositories within respondents' institution—respondents could choose to answer with "Yes", "No", or "I do not know".
- Involvement with health-related data—the respondents were asked to respond if they collect, analyze, store, or access health-related data in their professional activity (multiple choice question).
- Use of data from repositories—the respondents were asked if they had previously used data from repositories. If the answer was yes, they could choose from the following categories (multiple choice question): health status, lifestyle, environment, family history, clinical examination, medical imaging, omics data, laboratory investigations, or other data.
- Storing health-related data in repositories—the respondents were asked to choose which types of health-related data should be stored in a data repository (multiple choice question): health status, lifestyle, environment, family history, clinical examination, imagistic investigations, laboratory tests, omics, or other data.
- Agreement to store data in a repository—the respondents were asked to mention if they would agree to store the data collected in their professional activity in a data repository by choosing one of the following answers: "Yes", "I do not know", or "No".

The variables defined above will be referenced in the results section using the expressions established in quotation marks. These variables were used to analyze if the awareness of the existence of data repositories within the institution, the level of interaction with health-related data, or the experience with a diversity of types of data from repositories might influence the respondents' opinions related to the data federation's wider implementation, such as the number of categories of data considered to be useful if stored in repositories or the agreement to share the data collected during professional activity in a data repository. The following tests were applied: Mann–Whitney U Test, Chi-Square Test of Independence, and Spearman's Rank Correlation.

2.4. Organisational Readiness Score (ORS)

The questions have been grouped into two categories reflecting the existing experience in working with medical data (federated or not)—the first concerning experience, and the second concerning awareness about working with data in repositories. The grouping of questions into the "Experience" and "Awareness" categories was based on conceptual considerations, but it does not have validation at this stage. It was designed as a pilot indicator for internal monitoring within the institution.

For each question, we evaluated the number of participants who were familiar with at least half of the attributes enumerated and considered this percentage as a partial score of completeness for the capability the question addresses (in an ideal organizational case, all the participants should be familiar with at least half of them). We further used these scores to calculate a score for the entire category (experience or awareness) as the mean of the question scores.

$$\boldsymbol{Score_{EX}|Score_{AW}} = \sum_{\boldsymbol{i}} \frac{\boldsymbol{ScoreQ_i}}{\boldsymbol{i}}$$

where $ScoreQ_i$ is the sum of the percentages of respondents who checked at least half of the attributes under question.

As this study is an exploratory one, we did not add any weights to the question scores when we calculated the mean, but such an approach may prove to better match the reality.

Further, we can assemble ORS (Organisational Readiness Score) as a tuple, including the scores for the 2 categories

$$\text{ORS} = \{\text{Score}_{\text{EX}}, \text{Score}_{\text{AW}}\}$$

to be used as a metric showing future progress.

The ORS is an exploratory, pilot concept, and not a validated instrument. For the Organisational Readiness Score, we defined the components below for the partial scores.

Components of "Experience":

- "Purpose of the activities" as the reasons behind the use of data by the respondents.
- "Reasons for using federated data" as the reasons behind the potential use of federated data by the respondents.
- "Already used types of data from repositories" as the categories of data used by the respondents (health status, imagistic investigations, laboratory tests, omics, etc.).
- "Reasons for the lack of use of data stored in repositories" was the justification provided by the respondents when the data was not used.

Components of "Awareness":

- "Useful types of data when stored in a repository" as the categories of data considered beneficial by the respondents when stored in repositories (health status, imagistic investigations, laboratory tests, omics, etc.).
- "Knowledge of the concept of data federation" as the statements regarding data federation considered true by the respondents.
- "Data federation challenges" as the potential challenges for data federation reported by the respondents.
- "Data federation usefulness" as the benefits of data federation from respondents' perspectives.
- "Interest in related and additional concepts" as the topics mentioned by the respondents as being of interest.
- "Potential limitations to data federation implementation" as the factors that influence the decision of the respondents to share data.

To assess the research community's readiness for adopting a university-wide data policy, this study investigates two complementary dimensions: retrospective data handling experience and prospective data awareness. The former dimension evaluates the practical skills and competencies acquired by researchers throughout their careers, while the latter identifies their anticipated needs and future data utilization challenges. This dual analysis provides a comprehensive understanding of the community's current capabilities and future trajectory, informing the development of a data policy that is responsive to both present realities and foreseeable demands.

As a pilot study, this investigation represents the initial phase of a broader research program aimed at developing data policies tailored to the needs of researchers. For this preliminary analysis, an initial metric was derived by dividing the dataset, categorizing respondents based on affirmative answers to at least 50% of the options for each question. We acknowledge the provisional nature of this classification criterion, which will be refined in subsequent phases through a rigorous psychometric validation process.

3. Results

The survey form was filled in by 189 respondents, covering a variety of specializations and levels of training: physicians (residents and specialists), pharmacists, PhD candidates, and researchers. In total, 86% of the respondents self-reported their level of study as being

currently enrolled in PhD studies, and 72% of the respondents reported their profession as physicians with a clinical specialty.

3.1. Self-Reported Involvement with Data in Respondents' Professional Activity

To reflect the overall involvement of respondents with data in their professional activity, the percentages of responders who selected each of the following options—no contact with data, data collection, data access, data analysis, and data storage—are presented in Figure 1. Regarding the purpose of the use of data, 51% of the respondents use health-related data for research, 53% for clinical purposes, and 43% for trials.

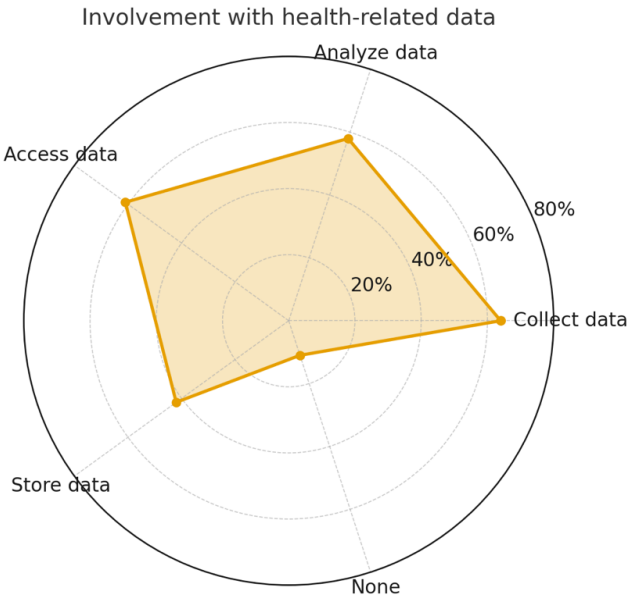


Figure 1. Respondents' experience with health-related data.

The respondents who accessed data repositories used the following categories of data: health status data (48%), lifestyle data (22%), environmental data (13%), family history data (28%), omics data (19%), laboratory tests (54%), clinical examination observations (54%), and imagistic explorations (49%). The lack of availability of certain categories of data was the main reason why the respondents did not access all the categories of data (45%), while 27% did not use other categories of data because they was not needed, and 21% did not access all data categories because of both reasons.

In total, 42% of the respondents do not use or do not intend to use federated data in their activity, while 37% of them declared that they use federated data for research purposes, 29% for clinical purposes, and 25% for trials. In total, 41% of the respondents reported the existence of data repositories in their institution, 23% reported the absence of data repositories in their institution, and 36% of them declared that they did not know this information.

3.2. Opinions and Perceptions on Shared Data

The respondents considered it useful to have the following categories of data stored in a medical data repository: health status data (81%), lifestyle data (65%), environmental data (46%), family history data (77%), omics data (54%), laboratory tests (81%), clinical examination (87%), and imagistic explorations (84%). Sixty percent of the respondents considered that having all the data stored in one location is useful for a data repository in their institution. Sixty-five percent of them agreed that clear policies for access, use, and storage of data are useful for a data repository from their institution. Fifty-six percent considered data security assurance useful for a data repository established within their institution.

In total, 75% of the respondents declared that they would agree to share the data collected in their professional activity in a data repository, 23% are not sure, and almost 3% do not wish to store this data in a repository. Regarding the decision to share or not to share the data collected in their activity in a data repository, the respondents' choices regarding the factors that are important are presented in Figure 2.

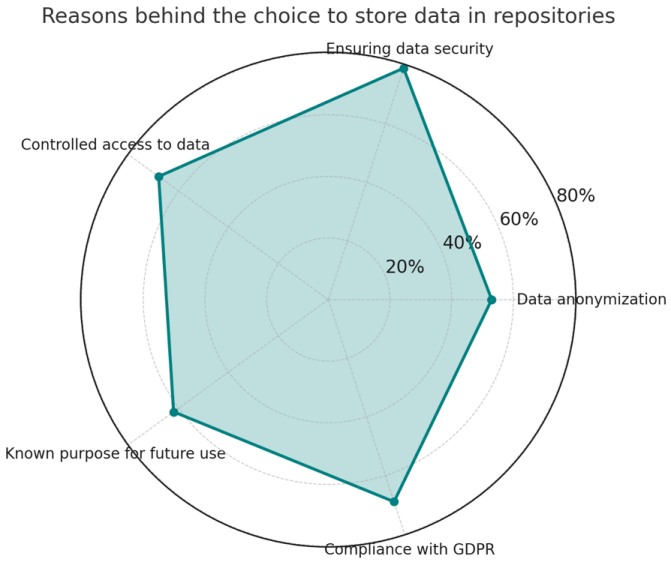


Figure 2. The factors that contributed to the willingness to store data in repositories.

The understanding of the concept of data federation varied among the respondents, with their agreement with several statements being presented below in Figure 3.

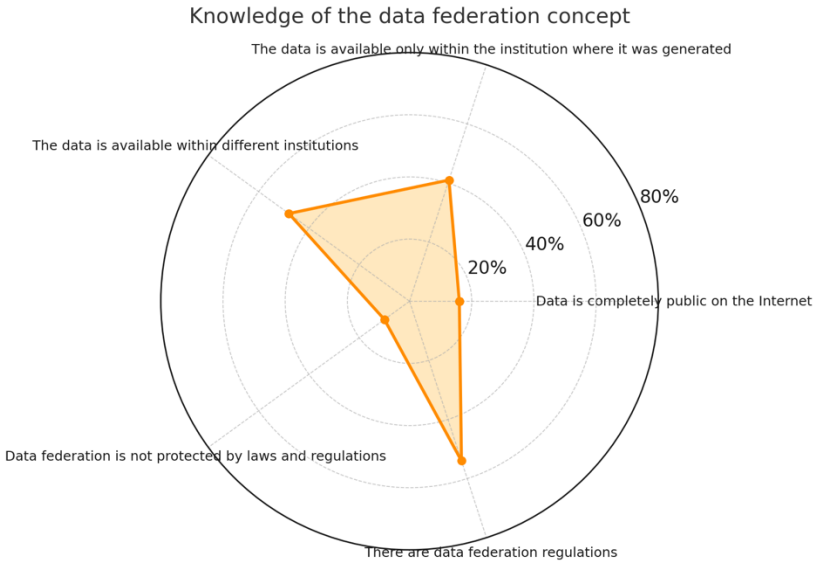


Figure 3. Respondents' understanding of the data federation concept, reflected through the agreement with several statements.

In Table 1, we present the respondents' choices as the main fields that can benefit through wider implementation of data federation in healthcare (Table 1).

Table 1. Respondents' opinion about the relevance of federated data for medicine, presented as percentages of respondents who agreed with each statement for potential uses of federated data.

The Relevance of Federated Data in Medicine	% of Respondents
Accelerating discoveries in medicine	88%
Identifying personalized therapies	62%
Developing new therapeutic molecules	43%
Understanding disease mechanisms	58%
Assessing environmental and lifestyle impacts on health	58%
Establishing the clinical significance of rare genetic variants	49%
Understanding the role of genetic factors in pathology	47%
Determining causes of rare diseases	19%

In Table 2, we present the respondents' opinions regarding the main challenges for the implementation of data federation in medicine, from data collection, which was the most selected option, to interoperability, which was the least chosen answer (Table 2).

Table 2. Respondents' opinions about the challenges related to data federation in medicine (percentages of respondents who agreed with each statement for potential challenges of federated data).

Challenges Related to Data Federation in Medicine	% of Respondents
Data collection—obtaining consent from participants	44%
Required infrastructure (communication, storage, access speed, etc.)	73%
Data storage security	72%
Guaranteeing data confidentiality	65%
Facilitating data access	50%
Lack of digitalized data	53%
Standardizing data formats	48%
Quality and accuracy of data	44%
The need for collaboration between professionals and institutions	49%
Achieving interoperability	20%

The respondents declared that they are interested in learning more about the following topics: FAIR principles (52%), open science (70%), data federation (60%), repository and analysis tools (63%), data quality (46%), and ethics related to data federation (44%).

3.3. Potential Trends Within the Group of Respondents

A low positive association ($p < 0.0001$, R-0.358, Spearman's Rank Correlation) was observed between the level of interaction with health-related data and the number of categories of data considered useful if stored in a repository. The level of interaction with health-related data was not associated with the agreement to share the data collected during professional activity in a data repository (p-0.842, Mann-Whitney U Test).

The awareness of the existence of data repositories within the institution was statistically significantly associated with the number of categories of data considered useful if stored in a repository (p-0.035, Mann–Whitney U Test). There was no association observed between the awareness of the existence of data repositories within the institution and the agreement to share the data collected during professional activity in a data repository (p-0.867, Pearson Chi-Square-0.036).

The experience with a diversity of types of data from repositories was weakly correlated with the number of categories of data considered useful if stored in a repository ($p < 0.0001$, R-0.264, Spearman's Rank Correlation). The experience with a diversity of

types of data from repositories was not associated with the agreement to share the data collected during professional activity in a data repository (p-0.74, Mann–Whitney U Test).

A statistically significant association was observed between the number of categories of data considered useful if stored in a repository and the agreement to share the data collected during professional activity in a data repository (p-0.020, Mann–Whitney U Test).

Regarding the calculation of the above-defined Organisational Readiness Score, for the "Experience" category, we obtained the partial scores presented in Table 3, which led to a Score_{EX} of 31.6%.

Table 3. Partial scores for the components of "Experience" (percentage of respondents).

Experience	% of Respondents
Performed data-related activities	47%
Defined purpose of the activities	42%
Reported reasons for using federated data	24%
Used types of data from repositories	39%
Reported reasons for the lack of use of data stored in repositories	7%

For the "Awareness" category, we obtained the partial scores presented in Table 4, which led to a Score_{AW} of 63.3%.

Table 4. Partial scores for the components of "Awareness" (percentage of respondents).

Awareness	% of Respondents
Useful types of data when stored in a repository	84%
Knowledge of the concept of data federation	23%
Data federation challenges	68%
Data federation usefulness	59%
Interest in related and additional concepts	67%
Identified potential limitations to implementing data federation	79%

With these values, we can calculate an actual metric as follows:

$$\text{ORS}_{2025} = \{32, 63\}$$

4. Discussion

A survey was conducted among the medical researchers from Carol Davila University of Medicine and Pharmacy, Bucharest, Romania, which involved 189 respondents. The main findings include a lack of awareness about repositories and data federation, involving around one-third of the respondents. The majority of respondents acknowledge the complex beneficial implications of data federation in the medical field but highlighted the importance of sensitive data protection. More than half of the respondents consider data collection and obtaining consent from participants, data storage security, data confidentiality, data access, and lack of digitized data as challenges for data federation. The ethical aspects and the application of appropriate procedures when collecting and analyzing the data from repositories are important for the respondents.

However, the least important challenge for data federation, in the opinion of the majority of the respondents, was interoperability, highlighting a potential fault in understanding the concept and a need for specific training to raise awareness about it.

The results of this study suggest that there might be valuable data available in universities and healthcare institutions in Romania, which can be stored in local repositories

and then federated together and used further to enhance medical research and precision medicine. More than one-third of respondents (41%) declared that their institution has established data repositories, which suggests the foundation of a distributed data system (based on multiple individual local data sets) is present.

In 2024, the concept of storing medical data in repositories is not new, and large medical and academic centers have already established complex initiatives in the direction of sharing data collected within their institution. For example, the University of Virginia and Virginia Commonwealth University Health System are developing Clinical Data Repositories as a large and linked database that stores data from primary electronic resources [17]. Carol Davila University of Medicine and Pharmacy is in the process of aligning with these worldwide tendencies observed within healthcare and research institutions, currently developing its first genomic data repository and the associated policy regarding data governance and access.

One striking aspect of the opinions of the respondents in this survey on the usefulness of diverse categories of health-related data concerns omics data. This type of data was the second least chosen category by the respondents when asked about their opinion on the types of data useful for a medical data repository. Combined with the low level of use of omics data, as reported in the survey, in comparison to other categories of health-related data, these results suggest a low level of knowledge and awareness of the importance of omics data availability for research and the enhancement of precision medicine [18] among the respondents.

Since its first mention in 2002, the quantity of scientific literature has doubled in 2022–2023, in connection with technological advances. Multi-omics data integration leads to a complex understanding of biological systems, including human physiology and pathology. Being the foundation for selecting the best treatment for a patient from an available spectrum of therapeutic choices, leading to improved outcomes and minimized side effects, deciphering molecular mechanisms that lead to disease development and progression, enabling new biomarker discovery for diagnosis, and facilitating new target identification for drug development, omics data has a central role in medical advances and innovation [9,17]. While the impact of the availability of omics data nowadays is high, awareness of its importance is low among healthcare professionals and scientists who do not have contact with this specific field, as shown by the results of this survey.

More than two-thirds of the respondents highlighted the importance of sensitive data protection when health and biological data from living humans are stored in repositories. Data protection assurance is a complex concept. Access to medical databases is subject to strict legal and regulatory requirements due to the sensitive nature of health data. Frameworks such as the EU GDPR regulate data processing, requiring informed consent, secure handling, and formal data access agreements. Additional considerations include data ownership, cross-border sharing restrictions, and accountability for breaches or misuse. These legal constraints play a critical role in shaping how medical data can be shared, accessed, and integrated, particularly in federated and FAIR data systems [10].

Within the European Union, GDPR has the purpose of preventing breaches of data privacy, but even at the level of the European Union, there are multiple discrepancies in achieving this goal, due to various national regulations and different understandings of concepts such as data anonymization [10]. While the ethical aspects of storing data in federated repositories are a major concern for the respondents, open science aims to improve the ethics of research by increasing the transparency of the way new medical knowledge is generated and building trust in research and its results' impact on saving patients' lives [19].

Open science in the context of medical data, which is particularly sensitive, might seem contradiction, but the goal of assuring data protection under open science is achievable. Implementing the FAIR principles aims to maximize the use of data for research, but only by respecting ethical and legal aspects. The FAIR principles emphasize machine-actionability (as opposed to human intervention) to find, access, interoperate, and reuse the data. This requires the unification of metadata and data models to allow the fast identification and usage of the required data for a specific study (findability), while the data itself is structured to be usable on different systems (interoperability) [20].

Data sharing, as supported by the FAIR principles, in the medical field can involve sensitive data [21]. While in the era of Big Data, privacy protection is more difficult to obtain, computational solutions for data de-identification have been made, which minimize risks [22]. The main difficulties in integrating the data start with the inability to identify it, as very often there is a lack of a common identifier to link the data sets of interest. Another problem to solve is the usage of a suitable anonymization technique, like differential privacy [23], which should prevent the identification of a specific patient as the available data grows. This is particularly important because not all data sets suffering from weak anonymization are suitable for research, contradicting the purpose of these efforts. The adoption of a suitable anonymization technique that allows adding newly available data regarding an individual but prevents identification is of extreme importance for any research repository [24].

Furthermore, policies have been established at the international level. For example, the European Health Data Space has been developed as a framework that covers health data protection under the technical solutions available [25]. Transparency and effective communication regarding how data privacy is handled can increase trust in federated data repositories.

An important observation was that the least relevant challenge for data federation, in the opinion of the respondents, was interoperability (20%). Medical researchers may underestimate the importance of interoperability because their daily work often focuses on clinical practice or individual research projects rather than on the technical complexities of data integration across institutions. Additionally, many researchers may be unaware of the diversity of medical coding systems and data standards that are necessary for successful data exchange, since these topics are not covered by traditional medical research training.

Previous scientific literature is consistent with this finding, highlighting the lack of awareness among healthcare professionals regarding data interoperability, which is an essential aspect when it comes to achieving the full potential of digital data. Interoperability assurance, which implies that data can be exchanged and used between systems, relies on the involved organizations, policies, and regulations [26].

The observed awareness gap regarding interoperability may also impact the successful implementation of national and European digital health initiatives, such as the European Health Data Space, as it can be more than a knowledge issue; it can represent a critical barrier to achieving their specific goals. Without a foundational understanding and adoption of interoperability standards by the researchers and clinicians who collect and manage health data, the vision of data sharing in medical research remains practically unachievable. Interoperability should not be solely considered a technical problem but also as an organizational one, requiring shared vocabularies, coordinated governance, and alignment across institutions. A lack of awareness at the point of data collection leads to the collection and generation of unstructured, non-codified data, which makes subsequent "FAIR-ification" and integration efforts either impossible or extensive. Highlighting this didactic dimension is essential for ensuring that those involved in data collection recognize their role in shaping interoperable health ecosystems to enhance research in medicine and healthcare.

Policymakers and institutions should consider targeted educational programs and awareness campaigns to ensure that the stakeholders responsible for data collection and use understand the requirements and challenges of interoperability. Such initiatives should not only aim to improve technical literacy but also promote a culture of semantic and organizational alignment across the healthcare and medical research institutions. Without such efforts, the full potential of federated and FAIRified medical data may not be realized. Increasing awareness and competence in these areas could empower future clinicians and researchers to actively contribute to the design and governance of interoperable health data systems.

The federated approach may provide an easier way to make the available data visible and usable both for research projects and for clinical usage. Studies have shown over the years that HL7 FHIR [27], SNOMED CT [28], and LOINC [29] may provide the basis for medical data interoperability. Ensuring that these standards are understood and applied from the stage of data collection and generation is essential for reducing fragmentation and supporting scalable data integration within the European Health Data Space and similar frameworks. Following the existing standards, the federated approach can combine omics data, clinical data collected from family medicine practice, electronic health records from hospitals, laboratories, specialized clinics, and even data collected by wearable sensors outside of medical facilities for integrated analysis [30,31].

Interoperable real-world data will accelerate medical research. The federated approach needs higher levels of transparency and trust between the participating partners to provide the necessary level of overall security and legal compliance for the entire system. The existing segmentation in data silos has a significant impact on the analysis of large sets of data that are necessary for the progress of personalized medicine, requiring significant effort and resources (time included) for data identification and data curation to make it suitable for analysis. Structuring medical data in local repositories in compliance with international standards facilitates analysis by eliminating the need for local data cleaning and preprocessing prior to actual analysis, as these steps have already been performed by the data owner. Therefore, the analysis can be performed similarly and coherently across multiple data sources from various institutions in the context of interoperable data [26].

Where are we now, and where to from here?

Data access is essential for medical research, but it requires data integration from different sources, either centralized or decentralized, to be used to their full potential. Valuable data exist, being collected in various separate, possibly smaller studies, and after its primary use, it is often lost. This data is commonly collected by individual research teams according to their specific needs, without a common framework or unitary structure that facilitates further use.

One set of difficulties in dealing with data for precision medicine is raised by the regular way the medical and research institutions are organized, which favors data silos. Technology silos, with differences starting from data structures, dictionaries, and taxonomies, and ending with storage and processing pipelines, are quite common even within the same organization, especially when dealing with different types of data (e.g., omics vs. images) or with different origins (e.g., research vs. clinical). Even if the associated metadata plays its central role in the effort of identification, different sets of metadata collected at different levels of the study (e.g., study-level metadata vs. sequencing-level metadata) have to be connected together in order to be useful. These issues are rapidly growing when we add non-technical dimensions to the equation such as institution status, which plays a major role in large studies conducted by multiple organizations with different types of data governance resulting from their affiliation (e.g., public vs. private), or budget sources (e.g., private versus centralized) [32].

The shift from single-use and limited data to expanding data roles in medical discoveries is desired at present, as data that have already been collected in single-specialty studies can be used in various investigations, leading to further benefits for precision medicine. Establishing a framework for data collection and data formats, and storing data in data repositories in compliance with FAIR principles, supports the process of data federation for the scientific community [33,34].

At present, there is an overload of available multimodal health-related data, which must be used to its full potential. Repositories in this field aim to integrate omics (genomics, transcriptomics, proteomics, metabolomics, and others), clinical, environmental, lifestyle, laboratory, and imaging data. All of these categories of data can be used in personalized prevention, diagnosis, and treatment [35]. Due to technological advances, medical data acquisition and analysis are feasible, but their use at full potential for personalized medicine is limited by ethical aspects, including access to the existing data, privacy control, and data governance requirements, as well as the integration of distinct formats and the distribution of diverse data sets all over the world, which are subject to various restrictions regarding storage, access, and use [36,37].

Standardized data-sharing systems and comprehensible data-related policies are pivotal in modern medical research, especially as the amount of sensitive data continues to increase. The need for international collaboration and innovation has driven the development of systems that can provide both data accessibility and privacy. Federated data systems offer a decentralized design where sensitive data can be analyzed without being transferred, addressing concerns about data security [38]. This approach aligns with initiatives such as the Global Alliance for Genomics and Health (GA4GH), which promotes responsible and ethical sharing of genomic data while ensuring that the rights to benefit from scientific advancements are upheld [39,40].

Early data-sharing initiatives, such as the Human Genome Project, emphasized open access to data, fostering scientific collaboration. However, concerns about reidentification and privacy risks in genomic data have led to the adoption of controlled-access models [41]. These models regulate how and when data can be accessed, often requiring approval for data use [38].

Implementing a federated data-sharing model requires high data quality, data source synchronization, and adequate connectivity between repositories. Continuous infrastructure maintenance and synchronization of roles, responsibilities, and policies on all sites are essential. Federated data can increase available data visibility, which can be used for both research and clinical purposes [42,43].

Even without being a specific characteristic, the decentralized architecture of federated systems is enabled by application programming interfaces (APIs), which allow secure programmatic access to different data systems, either when data is temporarily transferred for local processing or without transferring the actual data and performing the processing remotely [38,40]. The ongoing efforts by organizations like GA4GH to develop international standards for data sharing reflect the growing importance of these systems in both research and healthcare [40].

Adherence to FAIR principles contributes to standardization, a unified approach, and fragmentation avoidance. It emphasizes metadata unification and models for fast identification and usage of the required data, with a data structure that allows its use on different systems. This approach can provide the most efficient framework for data analysis, which generates the highest quantity and quality of new information, essential for the progress of personalized medicine [34,44].

The current study, even limited in its coverage and including a modest number of respondents, provides a snapshot of the existing understanding of the needs and challenges

of a medical research data repository and the expectations the scientific community regarding it. The study was conducted at an early stage of the development of the University's data repository, and a data access policy is currently being elaborated.

Despite the subjectivity of the respondents on the topic regarding their previous knowledge and understanding of concepts such as open science, data repositories, data federation, and FAIR principles, this study highlights a series of important factors in developing and implementing federated medical data systems, including the priorities of potential users and data providers, such as data protection, the categories of data with the highest potential for use, and the need for education and transparency on these topics for those who work in healthcare and medical research to improve trust.

Carol Davila University of Medicine and Pharmacy is currently in the first cycle of developing its own organization-wide scientific data repository. Currently, the University tests the architecture released by EGA (European Genome–Phenome Archive) for its repository. The Federated EGA architecture has been selected as it imposes strict protocols that govern how information is managed, stored, and distributed, combined with appropriate security measures to control access and maintain individual confidentiality, while providing access to researchers and clinicians who have obtained authorization. Data from each node in the Federated EGA Network are stored and managed locally, and metadata about the data it holds is shared publicly with the Central EGA, enabling researchers around the world to use a single-entry point to search for and discover data relevant to their research [10,45].

Being established within a community with a background mainly oriented towards clinical research, the potential data repository users from Carol Davila University of Medicine and Pharmacy might not be familiar with advanced methods for data analysis and processing. This implies that the repository should be user-friendly and facilitate easier access and use of data. Federated EGA architecture is adequate for this specific need. At this starting point, the current study shows the importance of assessing how the community understands the necessity and benefits of such an initiative in order to prioritize actions that provide the most expected benefits and identify the gaps that need to be filled to make the initiative successful.

Further use of data collected during various studies held at Carol Davila University of Medicine and Pharmacy is not only beneficial for medical discoveries but also for training students and researchers in bioinformatics and data analysis. For example, data collected during the Genetic Epidemiology of Cancer in Romania (ROMCAN) Project were primarily used for research [46] and then successfully used to train more than 100 students (undergraduate, graduate, and PhD) to perform genome-wide association studies [47]. Some of the trainees who completed the practical training made interesting novel discoveries based on this data, leading to the publication of scientific papers [48]. Therefore, establishing data repositories would also facilitate access to high-quality training for researchers and even lead to new discoveries in various fields, such as oncology, dermatology, and many others [46,48–51].

The secondary use of electronic health data offers major opportunities to enhance healthcare quality, strengthen public health surveillance, and drive research and innovation. The European Health Data Space provides an important policy framework to guide future developments, aiming to balance the significant public health gains with the protection of individual rights [52]. Similarly to our results, a systematic review of qualitative studies found that the public is generally supportive of health data sharing for research, but only under certain conditions. Key factors shaping acceptance include strong safeguards for confidentiality and security, transparency about use, clear public benefit, and trustworthy governance. Awareness of current practices is often low, and preferences are not uniform, highlighting the need for greater public engagement and ongoing evaluation [53].

The calculated readiness factor (ORS) shows us that we are almost halfway to an effective organization ready to work with data in local and remote repositories. It also shows that our community has a reasonable understanding but still lacks experience in real-life data analysis. The ORS will be used as a progress indicator showing the efficacy of the actions and policies established at the organizational level.

The results show that, on average, the scientific community from "Carol Davila" University is quite aware of the organization and benefits of the data repositories, which create good premises for acceptability, indicating a solid level of education on this matter. One notable aspect is the knowledge of the concept of federated data repositories, where the score is lower; this suggests that more knowledge transfer is needed.

On the experience side, the perspective is different, the readiness score showing that more activities must be conducted at an organizational level to provide greater exposure and experience for the staff involved in the practical aspects of data collection and processing.

The current study was limited by a series of factors. Firstly, the use of inferential statistics and ensuring generalizability were not feasible because of the study design, which was based on open questions and not a validated, questionnaire-based methodology. Furthermore, this study was conducted in a single center reflecting a specific geographic area—Bucharest, Romania—and involved respondents who predominantly represent specific categories of medical researchers, namely physicians with clinical specialties and PhD candidates. As mentioned in the study rationale description, this design was preferred due to the exploratory purpose of the study.

While these characteristics could offer valuable insights into the limitations of data federation under the FAIR principles in medicine, the lack of data on respondents' specialties prevented any subgroup analysis. Furthermore, the potential trends identified were associated with low correlation coefficients, warranting caution in data interpretation. As a result, the findings of this exploratory study are more reliable for descriptive results.

The use of ORS has several limitations. First, it relies on a subjective survey with predefined answers, which restricts the depth of insights and allows only limited quantitative evaluation. The Organisational Readiness Score (ORS) we developed is an empirical, exploratory construct rather than a validated instrument. The grouping of questions into the "Experience" and "Awareness" categories was based on conceptual reasoning and has not yet been empirically validated. In addition, when calculating the mean score, we did not apply weights to individual questions, although such an approach could potentially provide a closer reflection of reality. As such, the ORS should be seen as a pilot indicator designed primarily for internal monitoring within the institution, rather than a fully validated measurement tool. Nevertheless, despite these limitations, it offers a useful first step toward developing an internal readiness indicator.

However, evaluating organizational readiness is challenging. A systematic review examined instruments used to assess organizational readiness for knowledge translation in healthcare, and concluded that, from 39 publications, 26 instruments were identified, but only 18 demonstrated both validity and reliability [54].

Further research needs to be conducted to more accurately characterize the opinions and perceptions of medical researchers about federated data repositories and to help devise specific implementation policies that can aid the adoption of these tools and resources on a larger scale within the scientific medical community. Secondly, interpreting the data presented challenges due to the subjective nature of responses and the inability to apply rigorous statistical analysis.

5. Conclusions

The results of the current study highlight that medical researchers, particularly physicians, are generally open to the concept of sharing data for medical research and recognize its potential benefits. However, their readiness to adopt federated data repositories in daily practice may be limited by privacy concerns, subjective understanding, and a lack of awareness about the advantages of federated data sharing. Transparency and targeted education—especially on less-understood aspects such as interoperability—are key to increasing trust and enabling the full potential of data sharing in healthcare, particularly in fields like oncology, dermatology, cardiology, and urology.

To further support effective data sharing, targeted training programs might be beneficial to enhance knowledge of FAIR principles, interoperability, common data standards, and the implementation of federated infrastructures that preserve privacy and data sovereignty. By addressing these areas, future initiatives can facilitate secure, collaborative, and impactful use of medical data, ultimately advancing research and improving healthcare outcomes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data10110187/s1, The data collection instrument is included in the Supplementary File associated with this manuscript.

Author Contributions: Conceptualization, E.P., M.D. and C.P.; Methodology, E.P., M.D. and C.P.; Investigation, E.P., M.D., C.P., I.A.-B., L.-C.B.-N., T.-V.C., A.Z., M.M.C., B.B. and S.S.; Data Curation, E.P., M.D. and C.P.; Writing—original draft preparation, E.P., M.D., C.P., S.S., I.A.-B., L.-C.B.-N., T.-V.C., A.Z., M.M.C. and B.B.; Writing—review and editing, E.P., M.D., C.P., I.A.-B., L.-C.B.-N., T.-V.C., A.Z., M.M.C., B.B. and S.S.; Supervision, E.P., M.D., C.P., T.-V.C., A.Z., M.M.C. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Approved by the Research Ethics of "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania, No. 1846/31.01.2025.

Informed Consent Statement: Informed consent was obtained through the first section of the online form, where participants could proceed only if they agreed to participate. Data were collected anonymously via an online form, with no personal information recorded. All participants voluntarily completed the survey.

Data Availability Statement: The data are available from the corresponding author upon reasonable request.

Acknowledgments: Publication of this paper was supported by the Carol Davila University of Medicine and Pharmacy through the institutional program Publish not Perish. Data were collected with the aid of the FAIR–Impact Project, Support Programme for Repositories and Data Services.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application programming interfaces
EGA	European Genome-Phenome Archive
FAIR	Findable, Accessible, Interoperable, Reusable
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
ORS	Organisational Readiness Score

References

1. Mathur, S.; Sutton, J. Personalized medicine could transform healthcare. *Biomed. Rep.* **2017**, *7*, 3–5. [CrossRef]
2. Jakka, S.; Rossbach, M. An economic perspective on personalized medicine. *HUGO J.* **2013**, *7*, 1. [CrossRef]
3. Badr, Y.; Abdul Kader, L.; Shamayleh, A. The Use of Big Data in Personalized Healthcare to Reduce Inventory Waste and Optimize Patient Treatment. *J. Pers. Med.* **2024**, *14*, 383. [CrossRef]
4. Athieniti, E.; Spyrou, G.M. A guide to multi-omics data collection and integration for translational medicine. *Comput. Struct. Biotechnol. J.* **2022**, *21*, 134–149. [CrossRef]
5. Wade, T.D. Traits and types of health data repositories. *Health Inf. Sci. Syst.* **2014**, *2*, 4. [CrossRef] [PubMed]
6. Holub, P.; Kohlmayer, F.; Prasser, F.; Mayrhofer, M.T.; Schlünder, I.; Martin, G.M.; Casati, S.; Koumakis, L.; Wutte, A.; Kozera, Ł.; et al. Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. *Biopreserv. Biobank.* **2018**, *16*, 97–105. [CrossRef]
7. Lin, D.; McAuliffe, M.; Pruitt, K.D.; Gururaj, A.; Melchior, C.; Schmitt, C.; Wright, S.N. Biomedical Data Repository Concepts and Management Principles. *Sci. Data* **2024**, *11*, 622. [CrossRef]
8. Niehues, A.; de Visser, C.; Hagenbeek, F.A.; Kulkarni, P.; Pool, R.; Karu, N.; Kindt, A.S.D.; Singh, G.; Vermeiren, R.R.J.M.; Boomsma, D.I.; et al. A multi-omics data analysis workflow packaged as a FAIR Digital Object. *Gigascience* **2024**, *13*, giad115. [CrossRef] [PubMed]
9. Mohr, A.E.; Ortega-Santos, C.P.; Whisner, C.M.; Klein-Seetharaman, J.; Jasbi, P. Navigating Challenges and Opportunities in Multi-Omics Integration for Personalized Healthcare. *Biomedicines* **2024**, *12*, 1496. [CrossRef] [PubMed]
10. Rujano, M.A.; Boiten, J.W.; Ohmann, C.; Canham, S.; Contrino, S.; David, R.; Ewbank, J.; Filippone, C.; Connellan, C.; Custers, I.; et al. Sharing sensitive data in life sciences: An overview of centralized and federated approaches. *Brief. Bioinform.* **2024**, *25*, bbae262. [CrossRef]
11. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
12. Hunger, M.; Bardenheuer, K.; Passey, A.; Schade, R.; Sharma, R.; Hague, C. The Value of Federated Data Networks in Oncology: What Research Questions Do They Answer? Outcomes From a Systematic Literature Review. *Value Health* **2022**, *25*, 855–868. [CrossRef]
13. Sheth, A.P.; Larson, A.J. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.* **1990**, *22*, 183–236. [CrossRef]
14. Haas, L.M.; Lin, E.T.; Roth, M.A. Data integration through database federation. *IBM Syst. J.* **2002**, *41*, 578–596. [CrossRef]
15. Gu, Z.; Corcoglioniti, F.; Davide, L.; Mosca, A.; Xiao, G.; Xiong, J.; Calvanese, D. A systematic overview of data federation systems. *Semant. Web* **2024**, *15*, 107–165. [CrossRef]
16. Thorogood, A.; Rehm, H.L.; Goodhand, P.; Page, A.J.H.; Joly, Y.; Baudis, M.; Rambla, J.; Navarro, A.; Nyronen, T.H.; Linden, M.; et al. International federation of genomic medicine databases using GA4GH standards. *Cell Genom.* **2021**, *1*, 100032. [CrossRef]
17. Mullins, I.M.; Siadaty, M.S.; Lyman, J.; Scully, K.; Garrett, C.T.; Miller, W.G.; Muller, R.; Robson, B.; Apte, C.; Weiss, S.; et al. Data mining and clinical data repositories: Insights from a 667,000 patient data set. *Comput. Biol. Med.* **2006**, *36*, 1351–1377. [CrossRef]
18. De Maria Marchiano, R.; Di Sante, G.; Piro, G.; Carbone, C.; Tortora, G.; Boldrini, L.; Pietragalla, A.; Daniele, G.; Tredicine, M.; Cesario, A.; et al. Translational Research in the Era of Precision Medicine: Where We Are and Where We Will Go. *J. Pers. Med.* **2021**, *11*, 216. [CrossRef]
19. Winker, M.A.; Bloom, T.; Onie, S.; Tumwine, J. Equity, transparency, and accountability: Open science for the 21st century. *Lancet* **2023**, *402*, 1206–1209. [CrossRef]
20. The Lancet. Making sense of our digital medicine Babel. *Lancet* **2018**, *392*, 1487. [CrossRef]
21. Vesteghem, C.; Brøndum, R.F.; Sønderkær, M.; Sommer, M.; Schmitz, A.; Bødker, J.S.; Dybkær, K.; El-Galaly, T.C.; Bøgsted, M. Implementing the FAIR Data Principles in precision oncology: Review of supporting initiatives. *Brief. Bioinform.* **2020**, *21*, 936–945. [CrossRef] [PubMed]
22. Kayaalp, M. Patient Privacy in the Era of Big Data. *Balkan Med. J.* **2018**, *35*, 8–17. [CrossRef] [PubMed]
23. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
24. Murthy, S.; Abu Bakar, A.; Rahim, F.A.; Ramli, R. A Comparative Study of Data Anonymization Techniques. In Proceedings of the IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019. [CrossRef]
25. Attieh, H.A.; Haber, A.; Wirth, F.N.; Buchner, B.; Prasser, F. Enabling Open Science in Medicine Through Data Sharing: An Overview and Assessment of Common Approaches from the European Perspective. *J. Open Access Law* **2023**, *11*. [CrossRef]

26. Lehne, M.; Sass, J.; Essenwanger, A.; Schepers, J.; Thun, S. Why digital medicine depends on interoperability. *npj Digit. Med.* **2019**, *2*, 79. [CrossRef] [PubMed]
27. Mandel, J.C.; Kreda, D.A.; Mandl, K.D.; Kohane, I.S.; Ramoni, R.B. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 899–908. [CrossRef]
28. Benson, T.; Grieve, G. *Principles of Health Interoperability: SNOMED CT, HL7 and FHIR*, 3rd ed.; Springer: London, UK, 2016.
29. McDonald, C.J.; Huff, S.M.; Suico, J.G.; Hill, G.; Leavelle, D.; Aller, R.; Forrey, A.; Mercer, K.; DeMoor, G.; Hook, J.; et al. LOINC, a universal standard for identifying laboratory observations: A 5-year update. *Clin. Chem.* **2003**, *49*, 624–633. [CrossRef]
30. Glicksberg, B.S.; Johnson, K.W.; Dudley, J.T. The next generation of precision medicine: Observational studies, electronic health records, biobanks and continuous monitoring. *Hum. Mol. Genet.* **2018**, *27*, R56–R62. [CrossRef]
31. Raab, R.; Kuderle, A.; Zakreuskaya, A.; Stern, A.D.; Klucken, J.; Kaissis, G.; Rueckert, D.; Boll, S.; Eils, R.; Wagener, H.; et al. Federated electronic health records for the European Health Data Space. *Lancet Digit. Health* **2023**, *5*, E840–E847. [CrossRef]
32. Bernier, A.; Molnár-Gábor, F.; Knoppers, B.M. The international data governance landscape. *J. Law Biosci.* **2022**, *9*, lsac005. [CrossRef]
33. Hoffmann, K.; Pelz, A.; Karg, E.; Gottschalk, A.; Zerjatke, T.; Schuster, S.; Böhme, H.; Glauche, I.; Roeder, I. Data integration between clinical research and patient care: A framework for context-depending data sharing and in silico predictions. *PLoS Digit. Health* **2023**, *2*, e0000140. [CrossRef] [PubMed]
34. Inau, E.T.; Sack, J.; Waltemath, D.; Zeleke, A.A. Initiatives, Concepts, and Implementation Practices of the Findable, Accessible, Interoperable, and Reusable Data Principles in Health Data Stewardship: Scoping Review. *J. Med. Internet Res.* **2023**, *25*, e45013. [CrossRef]
35. Molla, G.; Bitew, M. Revolutionizing Personalized Medicine: Synergy with Multi-Omics Data Generation, Main Hurdles, and Future Perspectives. *Biomedicines* **2024**, *12*, 2750. [CrossRef] [PubMed]
36. Blobel, B.; Lopez, D.M.; Gonzalez, C. Patient privacy and security concerns on big data for personalized medicine. *Health Technol.* **2016**, *6*, 75–81. [CrossRef]
37. Williamson, S.M.; Prybutok, V. Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare. *Appl. Sci.* **2024**, *14*, 675. [CrossRef]
38. World Economic Forum. Available online: https://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf (accessed on 29 November 2024).
39. Bahcall, O.G. Partnerships to enable the responsible sharing of biomedical data. *Cell Genom.* **2021**, *1*, 100037. [CrossRef]
40. Rehm, H.L.; Page, A.J.H.; Smith, L.; Adams, J.B.; Alterovitz, G.; Babb, L.J.; Barkley, M.P.; Baudis, M.; Beauvais, M.J.S.; Beck, T.; et al. International policies and standards for data sharing across genomic research and healthcare. *Cell Genom.* **2021**, *1*, 100029. [CrossRef] [PubMed]
41. Dyke, S.O.M. Genomic data access policy models. In *Responsible Genomic Data Sharing—Challenges and Approaches*; Jiang, X., Tang, H., Eds.; Academic Press: Amsterdam, The Netherlands, 2020; pp. 19–32. [CrossRef]
42. Scheibner, J.; Raisaro, J.L.; Troncoso-Pastoriza, J.R.; Ienca, M.; Fellay, J.; Vayena, E.; Hubaux, J.P. Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis. *J. Med. Internet Res.* **2021**, *23*, e25120. [CrossRef]
43. González-García, J.; Estupiñán-Romero, F.; Tellería-Orriols, C.; González-Galindo, J.; Palmieri, L.; Faragalli, A.; Pristās, I.; Vuković, J.; Misinš, J.; Zile, I.; et al. InfAct Joint Action consortium. Coping with interoperability in the development of a federated research infrastructure: Achievements, challenges and recommendations from the JA-InfAct. *Arch. Public Health* **2021**, *79*, 221, Erratum in *Arch. Public Health* **2022**, *80*, 116. https://doi.org/10.1186/s13690-022-00877-4. [CrossRef]
44. Jacobsen, A.; de Miranda Azevedo, R.; Juty, N.; Batista, D.; Coles, S.; Cornet, R.; Courtot, M.; Crosas, M.; Dumontier, M.; Evelo, C.T.; et al. FAIR Principles: Interpretations and Implementation Considerations. *Data Intell.* **2020**, *2*, 10–29. [CrossRef]
45. European Genome-Phenome Archive. Available online: https://ega-archive.org/about/projects-and-funders/federated-ega/ (accessed on 12 November 2024).
46. Sima, C.; Iordache, P.; Poenaru, E.; Manolescu, A.; Poenaru, C.; Jinga, V. Genome-wide association study of nephrolithiasis in an Eastern European population. *Int. Urol. Nephrol.* **2021**, *53*, 309–313. [CrossRef] [PubMed]
47. Poenaru, E.; Chirică, V.I.; Poenaru, C.; Curcă, G.C.; Poenaru, A.; Jinga, V.; Vinereanu, D.; Sima, C.; Iacob, D.; Iordache, P.; et al. Considerations on the results evaluation of the genetic epidemiology training of AppGenEdu project. *Rom. J. Leg. Med.* **2021**, *29*, 147–156. [CrossRef]
48. Balan, A.; Dugaesescu, M.; Bejan, I.; Brinduse, O.; Bonciu, S.; Ferdoschi, C.; Iacob, D.; Poenaru, E. Lung Cancer Variants in Romanian Population—A Genome-Wide Association Study. In Proceedings of the 2021 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 18–19 November 2021. [CrossRef]
49. Mateescu, L.A.; Savu, A.P.; Mutu, C.C.; Vaida, C.D.; Șerban, E.D.; Bucur, Ș.; Poenaru, E.; Nicolescu, A.C.; Constantin, M.M. The Intersection of Psoriasis and Neoplasia: Risk Factors, Therapeutic Approaches, and Management Strategies. *Cancers* **2024**, *16*, 4224. [CrossRef] [PubMed]

50. Sandru, F.; Poenaru, E.; Stoleru, S.; Radu, A.-M.; Roman, A.-M.; Ionescu, C.; Zugravu, A.; Nader, J.M.; Baicoianu-Nitescu, L.-C. Microbial Colonization and Antibiotic Resistance Profiles in Chronic Wounds: A Comparative Study of Hidradenitis Suppurativa and Venous Ulcers. *Antibiotics* **2025**, *14*, 53. [CrossRef] [PubMed]
51. Matei, A.; Poenaru, E.; Dimitriu, M.C.T.; Zaharia, C.; Ionescu, C.A.; Navolan, D.; Furau, C.G. Obstetrical Soft Tissue Trauma during Spontaneous Vaginal Birth in the Romanian Adolescent Population—Multicentric Comparative Study with Adult Population. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11491. [CrossRef]
52. Cascini, F. *Secondary Use of Electronic Health Data: Public Health Perspectives, Use Cases and Challenges*; Springer Nature: Cham, Switzerland, 2025; ISBN 978-3-031-88496-2. [CrossRef]
53. Aitken, M.; de St. Jorre, J.; Pagliari, C.; Jepson, R.; Cunningham-Burley, S. Public Responses to the Sharing and Linkage of Health Data for Research Purposes: A Systematic Review and Thematic Synthesis of Qualitative Studies. *BMC Med. Ethics* **2016**, *17*, 73. [CrossRef]
54. Gagnon, M.P.; Attieh, R.; Ghandour, E.K.; Légaré, F.; Ouimet, M.; Estabrooks, C.A.; Grimshaw, J. A Systematic Review of Instruments to Assess Organizational Readiness for Knowledge Translation in Health Care. *PLoS ONE* **2014**, *9*, e114338. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.