EP PerMed
European Partnership
for **Personalised Medicine**

# Guidelines for Data Reusability

## For personalised medicine researchers and other stakeholders

March 2025

The French National Research Agency (ANR)

Co-funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101137129.

# Imprint

**EU grant**
The European Partnership for Personalised Medicine, EP PerMed, received funds from the European Union under the Horizon Europe Framework Programme Grant Agreement N°: 101137129.

**Authors**
Mylene Vaillancourt and Dr Monika Frenzel on behalf of the French National Research Agency (ANR) and the European Partnership for Personalised Medicine, with valuable contribution of Dr Ilaria Colussi (BBMRI-ERIC), Dr Eva García Álvarez (BBMRI-ERIC), Dr Kaya Akyüz (BBMRI-ERIC), Dr Emanuela Oldoni (EATRIS), Dr Nektarios Liaskos (EATRIS) and Dr Jacques Demotes (ECRIN).

**Acknowledgement**
This document was developed with the help and approval of the EP PerMed Coordination Unit and EP PerMed Work Package 2 partners.

**Contact**
The French National Research Agency (ANR)
Mylene Vaillancourt and Dr Monika Frenzel
E-mail: **EPPerMed@agencerecherche.fr**

**Publisher**
Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR) / DLR Project Management Agency, Department Health Linder Höhe, 51147 Köln, Germany

**Date**
March 2025

**Links to external websites**
This EP PerMed Guidelines for Data Reusability 2025 contains links to external third-party websites. These links to third-party sites do not imply approval of their contents. The DLR Project Management Agency has no influence on the current or future contents of these sites. We therefore accept no liability for the accessibility or contents of such websites and no liability for damages that may arise as a result of the use of such content.

**Using the content and citation**
If you wish to use some of the written content, please refer to: The EP PerMed – Guidelines for Data Reusability (2025).

# Table of contents

# 1 Introduction

Personalised medicine marks a paradigm shift from a "one-size-fits-all" approach to a tailored strategy for disease prevention, diagnosis, and treatment. It leverages an individual's unique biological characteristics (e.g., phenotype, endotype, genotype) alongside lifestyle and environmental factors to optimise healthcare outcomes. With rapid advancements in the field, the scope of technologies, methodologies, and data sources has expanded significantly, enabling more precise diagnostics, targeted treatments, and personalised strategies for prevention, rehabilitation, and overall healthcare improvement.

In alignment with the core values of the European Union (EU), research data generated under the EP PerMed programme is expected to remain accessible and reusable, even beyond the project's funding period. However, ensuring long-term accessibility and reusability presents significant challenges.

This document aims to guide personalised medicine researchers and other stakeholders by outlining key principles of data reusability and its associated challenges.

Specifically, it explores the following aspects:

1. Data life cycle overview and introduction to the FAIR principles[1] (Findability, Accessibility, Interoperability and Reusability). Integrating FAIR principles is imperative for improving data reusability and ensuring sustainable data management.
2. Data management and stewardship plans, essential to ensure high quality digital publications that facilitate research data discoverability and reusability.
3. Ethical and legal considerations and challenges in term of data interoperability and harmonisation when generating or reusing healthcare data.
4. Challenges of data reusability in personalised medicine and initiatives to address these challenges.

**Please note:**
**As the fields of big data and personalised medicine continue to evolve, this document remains a living resource and will be adapted to ongoing advancements and emerging challenges.**

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

# 2 Data life cycle

The data life cycle refers to the series of stages that data goes through, from its initial creation to its final disposal. In healthcare, research, and AI-driven applications, managing data effectively is essential for ensuring accuracy, security, compliance, and reusability. The following figure provides an overview of the key stages in the data life cycle.

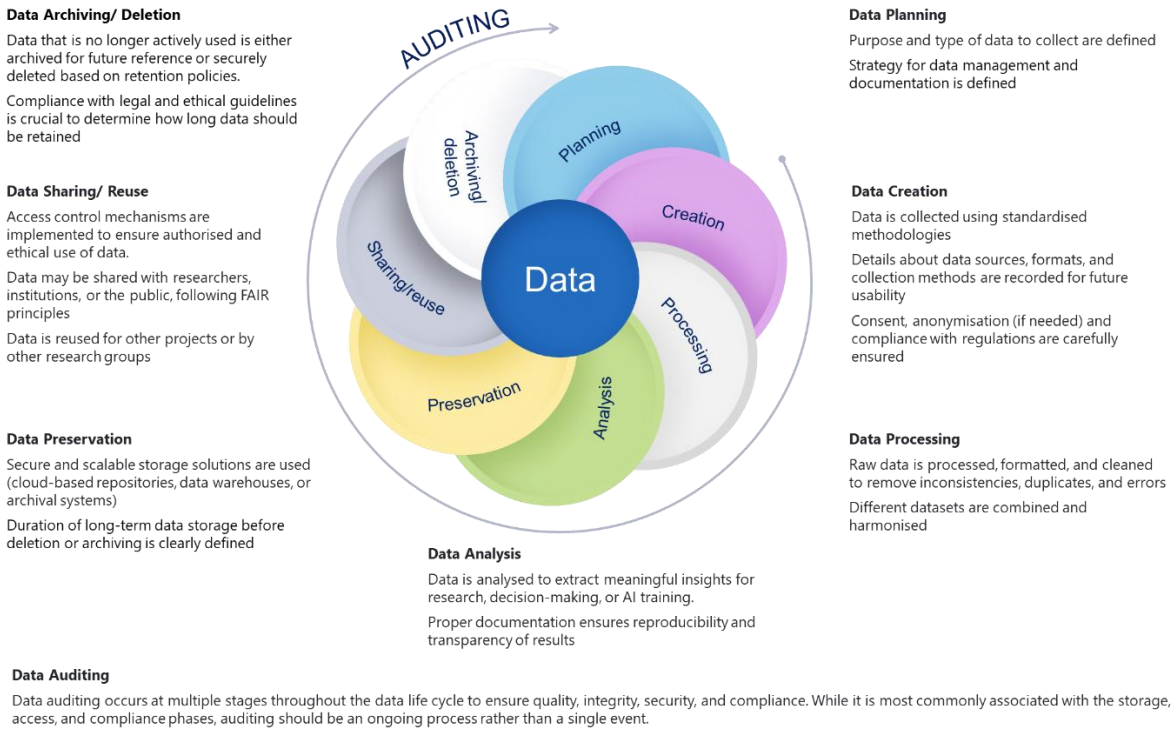


**Figure legend:** The data life cycle in 7 steps: data planning, creation, , processing, analysis, preservation, sharing/reuse and archiving/deletion. To ensure data quality, integrity, security, and compliance, auditing should occur at multiple stages throughout the data life cycle.

# 3 FAIR principles

The insufficient utilisation of data and biological material for medical research is a subject of ongoing debate. FAIR principles[1] provide guidance for improving Findability, Accessibility, Interoperability and Reusability of digital resources. FAIRness should be applied to both human- and machine-driven activities. In this section, FAIR principles[1] are listed and summarised.

## 3.1 FAIR

**F**indability

1. (Meta)data assigned a globally unique and persistent identifier (e.g. DOI)
2. Data are described with rich metadata (defined in Reusability below)

**A**ccessibility

1. (Meta)data are retrievable using an open, standardised and secure protocol
2. Access is controlled and compliant with regulations to ensure privacy and security
3. Metadata are accessible, even when the data are no longer available

**I**nteroperability

1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
2. (Meta)data use standardised vocabularies
3. (Meta)data include qualified references to other (meta)data

**R**eusability

1. (Meta)data are released with a clear and accessible data usage license
2. (Meta)data are associated with detailed provenance
3. (Meta)data meet domain-relevant community standards

### 3.1.1 FAIR-Health

The FAIR-Health principles apply the FAIR principles to health data. These principles aim to improve the usability and sharing of health-related data for research, personalised medicine, and healthcare innovation while maintaining privacy and security.

FAIR-Health[2] extension includes:

Quality and traceability

1. The reproducibility and significance of data should be defined by evaluating the suitability of biological material, based on its provenance information, for the specific research purpose and analytical method.
2. The data provenance, as well as collection and processes methodologies should be documented as completely as possible.

---

[2] Holub P et al. Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. Biopreserv Biobank. 2018 Apr;16(2):97-105. doi: 10.1089/bio.2017.0110. Epub 2018 Jan 23. PMID: 29359962; PMCID: PMC5906729.

Incentive schemes

1. Positive and effective incentive schemes should be developed and adopted in wide research communities, mainly providers and accessors.

Privacy regulation compliance

1. There must be clear identification of the responsible data controller for any given biological material and data set, who can be contacted by data subjects or authorities
2. Compliance of research projects with informed consent and the ethics approval of the research project should be evaluated before providing access to sensitive biological material and data
3. Privacy-enhancing technologies should be applied to personal data before the data can be used for research purposes, in compliance with the data minimisation principle
4. Data provenance should be implemented in a way that allows for identification of relevant data sets in case of informed consent withdrawal

More resources on FAIR considerations and FAIRification workflows are available in the section 7. Additional resources.

## 3.2 Open data access

Open data access makes research results more accessible to users and contributes to better and more efficient science, and to innovation in the public and private sectors.

Open access to scientific data is related to two main categories:

- peer-reviewed scientific publications (primarily research articles published in academic journals)
- scientific research data: data underlying publications and/or other data (such as curated but unpublished datasets or raw data)

Publication of the scientific outcomes of the research projects funded through EP PerMed should be as open as possible and as closed as necessary. They are eligible to publish on **Open Research Europe**[3] (ORE), an open access publishing platform of the EC.

[3] https://open-research-europe.ec.europa.eu/

RECOMMENDATIONS:

1. Researchers supported by European Commission (EC) or EU- cofunded programmes, such as EP PerMed, commit to keep their data as open and accessible as possible without compromising the protection of sensitive information, commercialisation and intellectual property rights, privacy concerns, and security. Researchers can refer to the **OECD Principles and Guidelines for Access to Research Data from Public Funding**[4] policy for recommendations on access to research data from public funding.
2. Researchers can adopt an open license or dedicate their work to the public domain. EP PerMed funded researchers are encouraged to apply for the least restrictive license possible (e.g. CC Zero or CC-BY) or acquire the license by depositing their data in a trustworthy repository. For instance, well-known data repositories like Zenodo, Figshare, and Dryad also use CC Zero and CC-BY licenses.

More information and resources are available in the **ICPerMed - Addressing Open Access in Personalised Medicine**[5] document.

More resources on how to choose a license, data repositories and data infrastructures are available in section 7. Additional resources.

## 3.3 Data provenance

Accurate and detailed data provenance is crucial for data reusability. Provenance information is essential to determine whether a data resource meets one's criteria for the intended reuse, and what data manipulation procedures may be necessary in order to reuse it appropriately.

Data provenance refers to the 1) data source, i.e. where the data was generated or collected, 2) data transformation such as cleaning, aggregation, enrichment, etc., and 3) data destination, i.e. how is data stored or delivered after processing.

RECOMMENDATIONS:

1. To foster data sharing, researchers funded under EP PerMed programme must clearly describe all tools, technologies, methodological approaches and digital supports used in their publications.

---

[4] OECD (2007), OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD Publishing, Paris, https://doi.org/10.1787/9789264034020-en-fr

[5] https://www.icpermed.eu/wp-content/uploads/ICPerMed_Open_Science_Personalised_Medicine.pdf

2. In addition, they should include descriptions of how data from different sources (such as different institutions) were combined and how different data streams were merged.
3. Data provenance records should be audited routinely to identify and solve discrepancies.

# 4 Data management plan

The data management plan (DMP) represents an essential document for the implementation of the research, as it helps to define the responsibilities of research data management ahead of the start of the project.

RECOMMENDATIONS:

1. In the DMP, researchers should explain how the data, tools, code or algorithms gathered, developed or used through the project will be maintained after the project end and would be available (findable, accessible, interoperable and reusable) or communicated to the wider research community, during and after the end of the project period. In brief, the DMP explains how data will be FAIR and shared.

   The DMP should include the following core requirements[6]:

   - Data description and collection or reuse of existing data
     - methodologies used if new data are collected or produced
     - description of how data provenance will be documented
   - Documentation and data quality
     - metadata and documentation that will accompany the data
     - data quality control measures that will be used
   - Storage and backup during the research process
     - data storage and back up
     - technical measures in place to ensure data security
   - Legal and ethical requirements
     - Material and Data Transfer Agreements (MDTA) developed between the different stakeholders
     - technical method (e.g., pseudonymisation, anonymisation) used to protect participants' privacy
     - description of any remaining re-identification risks, if applicable
   - Data sharing and long-term preservation

[6] **Science Europe - Practical guide to the international alignment of research data management**

  - description of how data will be shared
  - long-term preservation plan
- Data management responsibilities and resources
  - description of the governance structures overseeing the entire lifecycle of biological material and data to ensure accountability

2. When dealing with personal data, which is at the centre of personalised medicine, DMPs should also describe how researchers will comply to data protection laws, such as the **General Data Protection Regulation (GDPR)**[7], and in accordance with Ethical principles for data management[8] (see section 5. Ethical and legal considerations).
3. DMPs are typically updated multiple times throughout a project. Any changes in data management should be documented through revised versions of the plan to ensure consistency and transparency.

Practical guides and DMP templates are available in the section 7. Additional resources.

## 4.1 Data sharing

Data sharing refers to the process of making data resources available to multiple applications or users. There are several benefits to sharing data: 1) it increases the possibilities to new discoveries and decreases time needed for drug research and development lifecycle, 2) organisations have access to data resources that would otherwise be costly or laborious (or sometimes impossible) to collect or create, 3) it favours the development of collective and sustainable standardisation procedures related to the collection and use of data, and 4) it benefits patient communities by preventing data silos.

Despite significant work on the topic in the past decade, the sharing of data is still challenged in different ways, e.g. lack of standardisation in data collection, storage, transformation and harmonisation, legal and intellectual property issues, internal processes, data protection regulations, etc.

RECOMMENDATION:

To overcome these challenges and foster healthy partnerships, it is essential for researchers to have a thorough reflection about the different aspects related to data sharing as early as possible in the project and define a clear plan for data to be shared. This reflection

[7] **General Data Protection Regulation (GDPR): https://gdpr-info.eu/**

[8] **https://commission.europa.eu/system/files/2020-06/5._h2020_ethics_and_data_protection_0.pdf**

can be summarised in key decisions[9] that must be taken regarding data sharing during the project life cycle:

- Establish the type of data needed to answer the research question/s
  - Identify the internal stakeholders and engage them early in the project (academic and therapeutic area leads, senior managers, statisticians, lawyers, data protection officers)
  - Check IP sensitiveness
  - Explore ownership or sufficient rights to use the data
- Decide on data sharing model and develop data flow/s
  - Determine the appropriate data anonymisation strategy
  - Define who will be data controller (or joint controllers) and data processor
- Check if data needs to be transferred outside the EU
  - Negotiate data related agreements
  - Define standards and security requirements

### 4.1.1 Data sharing models

In personalised medicine, the integration and analysis of diverse clinical datasets are crucial for advancing patient-specific treatments. Various data management models (centralised, federated, and hybrid) have been explored to facilitate this integration while addressing challenges related to data privacy, security, and heterogeneity.

Below is a brief overview of the different data sharing models[10] with their main pros (✅) and cons (❌):

**Centralised data management**: All data is stored in a single location, such as a cloud or local server.

- ✅ Single source of truth.
- ✅ Easier data access and analysis
- ✅ High level of control and security
- ✅ Easier governance, consistency, and compliance
- ❌ Can create bottlenecks and scalability challenges
- ❌ Higher risk of data breaches

---

[9] Magda Chlebus and Hugh Laverty. 2024. IMI_IHI Data Sharing Play Book – How to unlock the potential of data sharing in collaborative projects. https://www.ihi.europa.eu/sites/default/files/uploads/Documents/ProjectResources/IMI_IHI_DataSharingPlayBook_2024.pdf

[10] Rujano MA et al. Sharing sensitive data in life sciences: an overview of centralized and federated approaches. Brief Bioinform. 2024 May 23;25(4):bbae262. doi: 10.1093/bib/bbae262. PMID: 38836701; PMCID: PMC11151787.

- Can be slow for large-scale data sharing

**Federated data management**: Data remains in multiple locations, but models/algorithms are sent to the data for processing.

- No raw data transfer
- Enhances privacy and security
- Complies with regulations (GDPR, Health Insurance Portability and Accountability Act (HIPAA), other national regulations)
- Allows collaboration without exposing raw data
- More complex to implement
- Harder to ensure data consistency
- Requires strong computing infrastructure

**Hybrid data management**: Combines aspects of centralised and federated models, where some data is centralised while other data remains local.

- Flexible data access
- Selective centralisation of non-sensitive data
- Balances privacy and accessibility
- Optimises performance for different needs
- Allows institutions to control what is shared
- Still requires strong security
- Implementation complexity
- Risk of data fragmentation

RECOMMENDATION:

Researchers should assess their needs, anticipate potential barriers, and determine the most suitable data-sharing model for their project.

Centralised data management is ideal for organisations requiring strong governance, security, and a single source of truth. It is the preferred choice when artificial intelligence (AI)training and big data analytics are top priorities, and privacy concerns are minimal.

Federated data management is ideal for organisations that require decentralised control while ensuring data accessibility. It is the preferred choice when privacy and compliance are paramount, and data must remain local.

Hybrid models are ideal for organisations requiring a blend of on-premise and cloud storage with seamless integration. They offer a balanced approach to scalability and security, ensuring privacy while enabling data-driven innovation

Additional resources on data sharing and tools are available in the section 7. Additional resources.

# 5 Ethical and Legal considerations

Research for personalised medicine may involve sensitive personal data, which can raise significant ethical and legal concerns.

Ethical and legal issues concern not only data collection, processing, storage, and transfer for a primary usage, but also the reuse of those data.

RECOMMENDATION:

Researchers must identify the potential ethical and legal issues of their research and address those issues in accordance to the GDPR and other appropriate regulations.

## 5.1 Data Protection, ethical issues and different types of research

Ensuring data protection and addressing ethical considerations across various types of research are essential for safeguarding privacy, maintaining integrity, and ensuring compliance in scientific and medical studies.

RECOMMENDATIONS:

1. Under the **GDPR**[7], **all research involving the processing of personal data** (including pseudonymised or encrypted data, which are personal data) must provide information about the data protection in their proposal. **This also applies for the reuse of data.**

   Providing information on data protection means explaining the legal basis for processing data, defining the roles of controller or processor, identify the technical and organisational measures adopted for protecting data, ensuring the rights of data subjects, etc.

   As regards technical measure, one of the main topics related to personal data concerns anonymisation or pseudonymisation. In the first case, data are no longer related to identifiable persons, and only complete anonymised data are no longer considered personal data. Pseudonymised data refers to personal data where direct identifiers have been replaced or removed, but re-identification remains possible using additional information. Pseudonymised data is often far more valuable for research than fully anonymised data. When complete anonymisation is not feasible, pseudonymisation or data encryption should be used. However, both pseudonymised and encrypted data are still classified as personal data. The highest level of protection should always be the one privileged.

2. If the research programme or project involves **higher ethic risks,** researchers should provide a detailed analysis of the ethics issues raised by the project and should include:

    - an overview of all planned data collection and processing operations
    - identification and analysis of the ethics issues that these raise
    - an explanation of how these issues will be mitigated in practice.

    A project might raise higher ethics risks if it involves[7]:

    - Processing of 'special categories' of personal data (racial or ethnic origin, genetic, biometric or health data, etc.)
    - processing of personal data concerning children, vulnerable people or people who have not given their consent to participate in the research
    - complex processing operations and/or the processing of personal data on a large scale and/or systematic monitoring of a publicly accessible area on a large scale
    - data processing techniques that are invasive and deemed to pose a risk to the rights and freedoms of research participants, or techniques that are vulnerable to misuse
    - collecting data outside the EU or transferring personal data collected in the EU to entities in non-EU countries.

    In certain cases, a **data protection impact assessment** (DPIA) may be mandated[11]. A DPIA is a process designed to help assess and minimise the data protection risks of a project. It should include a description and purpose of the of the data processing, an assessment of the necessity and proportionality in relation to the purpose, identification of the risks for the data subjects and the implementation of measures to mitigate the risks[7].

    Here are examples of situations when data processing may result in high ethic risks:

    - A hospital processing its patients' genetic and health data (hospital information system)
    - Storage for archiving purpose of pseudonymised personal sensitive data concerning vulnerable data subjects of research projects or clinical trials.

---

[11] *"In order to enhance compliance with this Regulation where processing operations are likely to result in a high risk to the rights and freedoms of natural persons, the controller should be responsible for the carrying-out of a data protection impact assessment to evaluate, in particular, the origin, nature, particularity and severity of that risk. The outcome of the assessment should be taken into account when determining the appropriate measures to be taken in order to demonstrate that the processing of personal data complies with this Regulation."* General Data Protection Regulations: **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679**

3. When using personal data **from a previous research project**, details regarding the initial data collection, methodology and informed consent procedure must be described. The permission to use the datasets must also be provided. How consent has been collected and what information has been provided is crucial, and it is a sensitive ethical issue.

   Consent documents should include the following aspects[12]:

   - The potential transfer of biological material and data to non-EU entities
   - The right of participants to withdraw consent
   - The type of results that may be returned to participants (general research outcomes, individual results, incidental findings)
   - Measures to protect participants' privacy (e.g., coding or anonymisation)
   - Commercial use and benefit-sharing provisions

   Furthermore, researchers should make sure to establish a process for consent withdrawal that aligns with the local consent documents.

4. When using **data that are publicly available**, details of the source(s) and the confirmation that the data are openly and publicly accessible and may be used for research purposes must be provided.

Resources on GDPR guidelines and check list, ethic self-assessment and domain-specific guidelines are available in the section 7. Additional resources.

## 5.2 International personal data transfer

Data transfer is a major aspect of data reusability. In terms of data protection, researchers and institutions based outside the EU may be subject to different ethical and legal rules that do not fully comply with the EU standards.

These regulations apply not only for the "physical transfer" of personal data, but also **in any situation where personal data would be accessible** to a partner or institution located outside the EU. Data transfer regulations also apply to all personal data, regardless of the sensitivity of the data.

[12] **World Medical Association – Taipei Declaration: https://www.wma.net/what-we-do/medical-ethics/declaration-of-taipei/**

RECOMMENDATIONS:

1. Researchers who plan to either import or export personal data internationally need to be aware of the foreign national regulations (i.e. local regulations in the receiving country), in addition to the GDPR.
2. When collaborating with partners from non-EU countries, researchers should ensure that each Research Ethics Committee approval is verified for the different countries, particularly concerning the reuse of biological material and data.
3. Researchers should identify a proper legal basis for data transfer. According to GDPR, personal data transfers to non-EU countries must include one of the following[8]:
    - the explicit consent of the data subject (which requires them to be informed in advance of any such transfers)
    - a data-transfer agreement containing EC standard contractual clauses giving effect to EU data protection law
    - binding corporate rules covering both sender and recipient and approved by a national supervisory authority
    - an 'adequacy determination' by the EC in respect of the country in question.

    Non-EU countries that have received an "adequacy decision" were recognised by the EC as having an adequate level of data protection, equivalent to the GDPR. To this date, few countries have received an "adequacy decision"[13]. An "adequacy decision" for a country means that personal data can be transferred from the EU (and Norway, Liechtenstein and Iceland) to this third country without any further safeguard being necessary.
4. Researcher should also implement safeguards by encrypting / pseudonymising or anonymising personal data where possible, inform data subjects about where their data is going and how it will be used and obtain explicit consent when required for sensitive data transfers. Application of rules to EU researchers

### 5.2.1 Application of rules to EU researchers

The GDPR applies to any data processing activity conducted by an entity established in the EU, regardless of where the data is collected or processed (extraterritorial scope, art. 3 GDPR[7]).

Even if research is conducted outside the EU, an EU-based researcher or institution must still comply with EU data protection rules. Horizon Europe and other EU-funded research

[13] **Data protection adequacy for non-EU countries: https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en**

programmes require compliance with EU ethical and legal standards, even when conducted abroad. Researchers working outside the EU must also comply with the national laws of the host country, including local data protection laws.

RECOMMENDATIONS:

1. Even if personal data is collected and processed outside the EU, researchers based in Europe must ensure to comply with EU laws (as provided by art. 3 GDPR[7]).
2. Researchers have to comply with the laws of the country in which they conduct their research, including any national data-protection laws. It is the researcher's responsibility to determine which legal obligations apply to their research and request further authorisations when required.
3. Researchers must determine applicable laws and obtain necessary approvals (e.g., ethics committee clearance, data transfer authorisations).
4. Failure to comply could lead to legal liabilities, research delays, or funding issues.

## 5.3 AI/ML considerations

European regulations surrounding artificial intelligence (AI) and machine learning (ML) in personalised medicine are evolving rapidly, aiming to ensure patient safety, ethical compliance, and technological transparency while fostering innovation. Personalised medicine uses AI/ML to tailor medical treatments to individual patients based on their unique genetic makeup, environment, and lifestyle.

Given the high stakes in healthcare, AI/ML applications in this field are subject to rigorous legislative frameworks. Below is an overview of the different regulatory frameworks and their relevance to AI/ML-driven personalised medicine:

**European Union AI Act[14]**

- Risk-based categorisation: AI models in personalised medicine may be classified as "high-risk", requiring strict validation and compliance.
- Transparency and explainability: Developers must ensure AI decisions are interpretable, particularly in healthcare applications.
- Human oversight: Medical AI cannot function autonomously in critical decisions; clinicians must remain in control.

[14] **https://artificialintelligenceact.eu/**

- Data quality and bias prevention: AI in personalised medicine must mitigate biases in training data that could impact patient outcomes.

**General Data Protection Regulation**[7]

- Patient data protection: AI models processing genomic, biomarker, or health data must comply with data minimisation, consent, and security requirements.
- Right to explanation: Patients have the right to understand how AI-driven decisions affect their treatment.
- Cross-border data sharing: Special rules apply for training AI models using multi-country datasets, requiring strict compliance.

**Medical Devices & In Vitro Diagnostics Regulations**[15]

- AI as a medical device: AI used for diagnosis, prognosis, or treatment recommendations is considered a medical device and must undergo *Conformité Européenne* (CE) marking (European Conformity Certificate).
- Clinical validation: AI models require rigorous testing similar to pharmaceuticals, including clinical trials for high-risk applications.
- Real-world performance monitoring: Continuous assessment post-market is mandatory.

**Health Technology Assessment (HTA) Regulation**[16]

- AI-driven healthcare innovations must undergo HTA evaluations to assess their cost-effectiveness and clinical utility before broad adoption.

RECOMMENDATIONS:

1. Researchers using AI/ML for personalised medicine should prioritise high-quality, diverse datasets to prevent biases related to age, gender, ethnicity, and socioeconomic status.
2. When handling patient data, researchers should address data privacy, security, and compliance by following the GDPR, HIPAA, and national regulations and by using privacy-preserving AI techniques.
3. Researchers must uphold ethical AI and patient trust by ensuring AI decisions are fair, unbiased, and explainable while providing patients with control over their data through transparent consent mechanisms.

---

[15] **https://health.ec.europa.eu/medical-devices-vitro-diagnostics_en**

[16] **https://health.ec.europa.eu/health-technology-assessment_en**

4. Researchers should engage with regulatory bodies early and understand their requirements throughout AI model development to ensure compliance.
5. Researchers should perform real-world validation studies before deploying AI models in clinical settings.

# 6 Challenges of data reusability in personalised medicine
(and initiatives to address these challenges)

Data reusability is essential for advancing personalised medicine, yet challenges such as data interoperability and harmonisation, cross-border data sharing, and limitations in bioinformatics and computing resources hinder effective data integration and collaboration. To address the various challenges associated with using data in personalised medicine applications, guidelines for designing and implementing personalised medicine research have been established[17].

Furthermore, various initiatives and tools have been developed and continue to evolve to address these challenges.

## 6.1 Data interoperability and harmonisation

Personalised medicine leverages diverse data sources, including health records, clinical data, -omics, and imaging data, to enhance treatments and improve individual health outcomes. However, a lack of standardisation persists in both data formats and the structuring of data for personalised medicine applications. This presents major challenges for data interoperability and harmonisation across research and healthcare systems.

To address this, standardised metadata and ontologies are essential for integrating the vast array of data generated in these fields. While efforts are underway to develop and refine these systems, further advancements are needed to ensure seamless data exchange and optimal support for personalised medicine.

The **European Health Data Space (EHDS)**[18] will play a crucial role in mitigating harmonisation challenges in personalised medicine by establishing standardised, interoperable, and secure frameworks for health data exchange across the EU.

EHDS addresses these challenges by:

- Promoting the adoption of (meta)data standards and ontologies

---

[17] Garcia et al. The PERMIT guidelines for designing and implementing all stages of personalised medicine research. Scientific Reports (2024) 14:27894. DOI: 10.1038/s41598-024-79161-0

[18] **https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en**

- Establishing interoperable digital infrastructures built on existing European and international standards
- Facilitating secure and ethical data access through Health Data Access Bodies and strict privacy safeguards
- Supporting AI-driven analytics to advance personalised treatment approaches
- Driving collaborative research and innovation by fostering pan-European ecosystems where health data can be shared between clinicians, researchers, and policymakers

By bridging the gap in the data harmonisation, implementing common standards, enhancing interoperability, ensuring secure access, and supporting AI-driven innovation, the EHDS initiative is set to have far-reaching implications for personalised medicine, advancing tailored treatments and enhancing patient outcomes across Europe.

## 6.2 Data sharing across borders

Delays in research and innovation projects due to data sharing ethics regulations are a common challenge, particularly in fields like healthcare, AI/ML and genomics, where sensitive data is involved. These delays are due to strict data protection laws, lengthy ethical review processes, data anonymisation and de-identification, interoperability and compliance issues, legal and institutional barriers, etc.

Practical solutions to address these challenges could include:

- Choosing federated or hybrid data sharing models: Instead of transferring data, researchers can access and analyse data remotely while keeping it secure.
- Using AI-generated synthetic data: Artificially generated data mimics real-world datasets while preserving statistical properties, patterns, and relationships. Using synthetic datasets could mitigate the delays related to the sharing of personal data.
- Automated compliance and consent management: AI-powered smart contracts and blockchain could streamline data consent tracking.

## 6.3 Bioinformatics and computing

Developing efficient and widely applicable personalised medicine approaches requires the generation and integration of vast amounts of data. The analysis of these large, diverse, and complex datasets demands advanced bioinformatics tools, including machine learning algorithms. However, many of these technologies are still evolving, requiring further refinement to fully unlock their potential in personalised medicine.

Europe is at the forefront of computing and AI-driven bioinformatics, with several initiatives aimed at enhancing biomedical data analysis and computational infrastructure. These initiatives support genomics, drug discovery, imaging analysis and personalised medicine by leveraging high-performance computing and AI/ML.

**EuroHigh Performance Computing Joint Undertaking (EuroHPC JU)**[19]: Develops supercomputing infrastructure for large-scale biomedical and genomics research and supports AI-driven bioinformatics workflows in personalised medicine.

**EOSC (European Open Science Cloud)**[20]: Offers cloud-based storage and computing resources for genomic and clinical data analysis.

**GAIA-X (European Cloud Infrastructure Initiative)**[21]: Secure, federated cloud for data-driven AI applications, including bioinformatics.

**BioExcel (Excellence in Computational Biomolecular Research)**[22]: Uses AI-enhanced molecular simulations for drug design and personalised medicine and provides bioinformatics software for protein structure modeling and molecular dynamics.

**Instruct-ERIC (European Research Infrastructure Consortium)**[23]: Provides open access to facilities and resources, including sample preparation, biomolecular and 3D structural analysis, computational services, and training courses.

**ELIXIR (European Life-Science Infrastructure for Biological Information)**[24]: Provides bioinformatics tools and training for computational biology research

---

[19] https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/european-high-performance-computing-joint-undertaking-eurohpc-ju_en#:~:text=The%20EuroHPC%20JU%20allows%20the%20European%20Union%2C%20its,to%20make%20Europe%20a%20world%20leader%20in%20supercomputing.

[20] https://open-science-cloud.ec.europa.eu/

[21] https://gaia-x.eu/

[22] https://eurohpc-ju.europa.eu/research-innovation/our-projects/bioexcel-centre-excellence-computational-biomolecular-research_en#:~:text=BioExcel%20is%20a%20major%20innovation%20hub%20for%20scientific,networking%20opportunities%20to%20the%20Life%20Science%20research%20communities

[23] https://instruct-eric.org/about-us#:~:text=Instruct-ERIC%20is%20a%20pan-European%20distributed%20research%20infrastructure%20making,basis%20within%20the%20scope%20of%20the%20ERIC%20Regulation.

[24] https://elixir-europe.org/

# 7 Additional resources

This section contains additional helpful resources for researchers on the different aspects of data reusability addressed in this document.

## FAIR guidelines and tools

Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). **https://doi.org/10.1038/sdata.2016.18**

Jacobsen et al. FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence 2020; 2 (1-2): 10–29. doi: **https://doi.org/10.1162/dint_r_00024**

Jacobsen et al. Generic Workflow for the Data FAIRification Process. Data Intelligence 2020; 2 (1-2): 56–65. doi: **https://doi.org/10.1162/dint_a_00028**

Special Issue: Emerging FAIR Practices. Issue Editors: Barend Mons, Erik Schultes & Annika Jacobsen. Data Intelligence (2020) 2 (1-2): iii–vi.

ELIXIR FAIR Cookbook: **https://faircookbook.elixir-europe.org/content/home.html**

GO FAIR Initative: **https://www.go-fair.org/**

FAIRsharing working group: **https://fairsharing.org/**

Data Stewardship Wizard: **https://ds-wizard.org/**

## How to choose a license

ELIXIR Data Management for information and tool for data licensing:
**Your tasks: Licensing | RDMkit**

Common Creative license chooser: **https://chooser-beta.creativecommons.org/**

GitHub tools for license selection: **Choose an open source license | Choose a license** and **license Selector**

How to License Research Data: **https://www.dcc.ac.uk/guidance/how-guides/License-research-data**

Video tutorial from Kingsborough E-Learning: **How to add a Creative Commons License to your work**

## Data repositories

Zenodo: **https://zenodo.org/**

Figshare: **https://figshare.com/**

Dryad: **https://datadryad.org/stash**

## Life science infrastructures

Biobanking and Biomolecular Resources Research Infrastructure (BBMRI): **https://www.bbmri-eric.eu/**

The European Life Sciences Infrastructure for Biological Information (ELIXIR): **https://www.elixir-europe.org/personalised-medicine**

European Infrastructure for translational medicine (EATRIS): **http://eatris.eu/**

European Clinical Research Infrastructure Network (ECRIN): **http://www.ecrin.org/**

European High Capacity Screening Network (EU-Openscreen): **http://www.euopenscreen.eu/**

European Infrastructure for Phenotyping, Archiving and Distribution of Mouse Models (INFRAFRONTIER): **https://www.infrafrontier.eu/**

Integrated Structural Biology Infrastructure for Europe (INSTRUCT): **http://www.structuralbiology.eu/**

European Strategy Forum on Research Infrastructures (ESFRI): **https://www.esfri.eu/**

The European Intergovernmental Research Organisation forum (EIROforum): **https://www.eiroforum.org/about-eiroforum/**

Coordinated Research Infrastructures Building Enduring Life-science Services (CORBEL): **http://www.corbel-project.eu/services.html**

## Data provenance and sharing

Identity: What Is Data Provenance?: **https://www.identity.com/what-is-data-provenance/**

Lightweight Distributed Provenance Model for Complex Real-world Environments: **https://pubmed.ncbi.nlm.nih.gov/35977957/**

Acceldata: A Comprehensive Definition of Data Provenance: **https://www.acceldata.io/blog/data-provenance**

Data Sharing Play Book – How to unlock the potential of data sharing in collaborative projects: **https://www.ihi.europa.eu/sites/default/files/uploads/Documents/ProjectResources/IMI_IHI_DataSharingPlayBook_2024.pdf**

Enhancing the QUAlity and Transparency Of health Research: **https://www.equator-network.org/**

Open Science Framework: **https://osf.io/**

## Data management plan tools

Science Europe – Practical guide to the international alignment of research data management: **https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf**

Horizon Europe - Data Management Plan Template: **Horizon-Europe-Data-Management-Plan-Template.pdf**

ELIXIR RMD kit: **https://rdmkit.elixir-europe.org/**

ELIXIR Data Stewardship Wizard: **https://ds-wizard.org/**

## Ethics and regulations

European Commission - Ethics and data protection: **https://commission.europa.eu/system/files/2020-06/5. h2020 ethics and data protection 0.pdf**

European Research Council - Ethics guidance: **https://erc.europa.eu/manage-your-project/ethics-guidance**

Horizon 2020 Programme: How to complete your ethics self-assessment: **https://eneri.eu/wp-content/uploads/2018/10/H2020-Guidance-How-to-complete-your-ethics-self-assessment.pdf**

General Data Protection Regulations: **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679**

GDPR checklist for data controllers: **https://gdpr.eu/checklist/**

Swiss Personalised Health Network – Ethics, Legal & Governance: **https://sphn.ch/services/documents/ethics-legal-governance/**