Beyond One Million Genomes

D2.4

Report on data access and governance framework

Project Title (grant agreement No)	Beyond One Million Genomes (B1MG) Grant Agreement 951724		
Project Acronym	B1MG		
WP No & Title	WP2 - ELSI		
WP Leaders	Regina Becker (UNILU), Jasper Bovenberg (Legal Pathways)		
Deliverable Lead Beneficiary	20 - UNILU		
Deliverable	D2.4 - Report on data access and governance framework		
Contractual delivery date	31/05/2023	**Actual delivery date**	05/10/2023
Delayed	No		
Authors	Adrian Thorogood (UNILU), Regina Becker (UNILU), Olga Tzortzatou (BBMRI-GR), Mihaela Matei (ECRIN), Susanne Rebers (BBMRI-NL)		
Contributors	All members of B1MG WP2		
Acknowledgements (not grant participants)	1+MG WG2 (ELSI) members 1+MG Special Group B1MG stakeholders		
Deliverable type	Report		
Dissemination level	Public		

Document History

Date	Mvm	Who	Description

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

20/08/2023	0v1	Regina Becker (PNED)	Initial draft circulated to WP participants for feedback
05/09/2023	0v2	Regina Becker (PNED)	WP comments addressed. Version sent to Coordinator
27/09/2023	0v3	Nikki Coutts (ELIXIR Hub)	Version circulated to B1MG-OG, B1MG-GB & Stakeholders for feedback]
05/10/2023	1v0	Nikki Coutts (ELIXIR Hub)	Comments addressed. Version uploaded to the EC Portal]

Table of Contents

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. Executive Summary

Task 2.4 of Work Package (WP) 2 focuses on the development of a Data Access and Use Governance Toolkit Framework. Such a framework oversees the data linking and data management and checks the compliance with ethical and data protection requirements while it considers the responsibilities of different stakeholders. To achieve this, it is necessary to identify the critical elements for an efficient and transparent governance that allows to set up a digital infrastructure that will enable the cross-border linking of genomic and other health data for research in Europe.

The recommendations were developed specifically for use in a large-scale pan-European research infrastructure that aims to operate in a harmonised fashion where data, quality, IT infrastructure and data governance are harmonised in a way that creates a federated virtual infrastructure. Towards the user, the infrastructure is to appear homogeneous whereas the maximum possible freedom is given to the participating countries. We have targeted our recommendations towards the need of the 1+ Million Genome (1+MG) initiative where countries want to make data available based on a joint Declaration[1]. The 1+MG initiative has a wider scope of secondary use, including also healthcare reuse and policy development. While our focus in the project is on the research use, occasionally, we have also still taken the broader scope into account.

In task 2.4, the input from all other three tasks of WP2 was used to develop a set of specifications and guidelines needed to allow efficient cross-border access and use of genomes for research in compliance with legal and ethical requirements. The following documents form part of our Data Access and Use Governance Toolkit Framework:

1. A data governance, describing a legally and ethically responsible approach for data inclusion into the infrastructure, a data access governance for research and a data use governance.

2. A guidance for transparency and consent, covering both legal and ethical requirements. This document builds on a document compiled in the task 2.2 on minimal standards and best practice guidelines for consent forms.

3. A recommendation on a practical approach to the management of the generated intellectual property (IP) rights that emanate from cross-border access and use of personalised medicine data in a pan-European genome initiative.

4. A recommendation on the IT infrastructure that establishes a data protection by design and default approach to support the data governance and to provide sufficient security for the data.

5. A recommendation for an information management to accompany the data governance.

6. As an appendix, a recommendation on a 1+MG - EHDS alignment based on the comparison between the proposed data governance of 1+MG and the draft European Health Data Space (EHDS) Regulation proposal published in May 2022.

[1] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50964

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Below the scope of the different tools are described. The full recommendations and other relevant aspects, can be found further in this document.

Data governance for research

Following the 1+MG Declaration, one of the goals of the 1+MG Initiative is to establish a European research cohort of over 1 million genomes. The Initiative will be structured as a federated network that connects genomic data resources and supporting infrastructures within Member Countries. This federated approach ensures that authority, responsibility, and resources are primarily based within the Member Countries, and that Member Countries have a certain flexibility over how to implement their national networks. To ensure the 1+MG cross-border, federated network truly functions as a "virtual" European research cohort, a clear governance framework must be established with the following aims:

1) to ensure efficiency and feasibility of cross-border access processes and therefore procedures that scale;

2) to promote clarity over general data access and re-use rights, applicable data-specific access and use conditions, and access procedures; and

3) to ensure compliance with applicable laws and ethical principles, particularly those relating to transparency and the protection of data subjects.

Data access by researchers based outside the EEA is explicitly considered in the "Scope of the 1+MG" policy. However, this special case will be addressed in a separate policy document once the "standard" data governance is agreed and can subsequently be integrated into the overall 1+MG data governance.

Transparency and consent guidance

Recommendations are made that 1+MG adopt 1) minimum requirements (MUST); 2) best practices (SHOULD); and 3) points-to-consider (non-directive). If a minimum requirement is missing, this may mean that a Data provider cannot legally or ethically make data available through 1+MG, or can only do so subject to special data and access and use conditions. Best practices may also constitute national legal requirements in some countries.

The recommendations are informed by the legal requirements of the GDPR, the interpretive guidance of the European Data Protection Board (EDPB), research ethics principles[2] and guidelines, as well as legal data governance principles, such as those outlined in the Data Governance Act, and implemented in the 1+MG Data Governance Policy. Ethical requirements are in particular based on the International Ethical Guidelines for Health-related Research Involving Humans by CIOMS. Justifications and explanations are provided. Legal consent requirements depend on the legal basis selected under the GDPR Art. 6 and the legitimation under Art. 9. The guidance provided is applicable for all legal bases, but always points out where a consent legal basis under the GDPR may lead to additional requirements, a stricter regime with respect to information related to consent, scope of the consent, interpretation of what counts as

[2] Including respect for persons, beneficence, and justice with a focus on the main ethical concerns raised by the informed consent process in the context of genomics.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

"freely given" as well as in consequences of withdrawal.[3] Requirements for consent as a legal basis may also depend on national laws.[4] National advisory bodies (e.g., ethics committees) are expected to provide additional, nationally-tailored guidance. It is the ultimate responsibility of the organisations involved in collecting data to identify and comply with all norms applicable to their activities.

This guidance is agnostic to different collection and sequencing contexts across Europe, including: population databases, genomic research projects, precision medicine clinical trials, genomic medicine initiatives, as well as clinical care (such as predictive, diagnostic or confirmatory genome sequencing). The guidance is designed for any organisation who plans to make data collected in a primary context available through a repository for research projects, where the details of these projects cannot be fully identified at the time of the data collection (or even at the time of the transfer to the repository). Some practical implementation examples are provided to facilitate application of the guidelines in specific contexts.

As 1+MG has not yet determined all aspects of its organisation structure, data governance and legal framework, some key information elements have not yet been fully defined. These elements are relevant to provide transparency and to obtain a valid informed consent. 1+MG is working to clarify these elements so that concrete wording or even a 1+MG specific part of the information sheet, where applicable, can be provided as an appendix in future versions of these guidelines.

Practical approach towards the management of the generated IP rights

This recommendation aims to establish a practical approach to the management of the generated intellectual property (IP) rights that emanate from cross-border access and use of personalised medicine data in a pan-European genome initiative. It navigates the different IP rights that arise in the context of a pan-European genome initiative, including the copyright on data, patent on inventions and trade secret protection.

The report also critically assesses the Open Innovation scheme, presenting the pros and cons of adopting such an approach.

The recommendation includes a checklist with all the information on IP rights that should be included in data transfer agreements, facilitating researchers who are involved in cross-border research projects.

It is necessary to reconcile IP rights as a means to encourage research with the public interest which is served through advancing innovation. This could be achieved through the adoption of appropriate governance and contractual access arrangements.

Data protection by design and default (DPbDD) recommendations for the IT infrastructure

[3] E.g., depending on national law or authoritative interpretations, this may include greater specificity of purposes and recipients; more details about the scope of data subject rights (especially if data are accessed by downstream controllers); and potential power imbalances between controllers (public bodies) and data subjects precluding consent.
[4] E.g., National, regional or sectoral data protection law, medical research law, health law, bioethics law, biobanking law, health research regulations.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

DPbDD means that the compliance with these principles must be considered already when the processing is planned and not mapped afterwards ("by design"). The "by default" means that the default state of a system should be "closed" or "protected" and only those data necessary for the purpose should be processed. Disclosure should be an active step that has to be planned in compliance with the above principles.

The current document recommends a list of requirements that the 1+MG IT infrastructure should fulfil. "IT infrastructure" in the covers here all information and communication technology support of the operations of 1+MG. This goes beyond the management of data access and the provision of an analysis platform for data use and also includes information and workflow management of 1+MG.

We compiled recommendations that consider the specific situation of 1+MG and follow the journey of the data within the initiative. The recommendations also reflect the envisaged data governance, which must be supported by suitable IT tools to become feasible and efficient. The analysis of the different stages will be organised according to the data protection principles of the GDPR to allow an easier demonstration of compliance and subsequently performance of a DPIA and auditing of 1+MG infrastructure implementations.

Data protection by design and default recommendations for the information management

Data protection compliant data management includes the management of relevant information on how data can be used but also requires the information on the actors along the data life cycle with their responsibilities and contacts, information on the data use itself, information about organisational and technical safeguards, including the management of such safeguards such as for pseudonymisation and secondary pseudonymisation. Accountability means that the measures taken must be auditable, which again sets up certain requirements for the documentation around the entire life cycle of data in the 1+MG.

The considerations on information management build on the mission of 1+MG, the data governance implementing DPbDD workflows, requirements of the GDPR and practical considerations that link the various needs. The definition of relevant structured and (where applicable) machine readable information is an important input for the design of the IT infrastructure. The current version is a first draft that will further develop with e.g. the decisions on the data governance. It should also be considered to be complemented by information requirements relevant from the user's perspective. The 1+MG Working Groups for Standards (WG3), Interoperability and Secure IT environment (WG5) and ELSI (WG2) must work closely together to define how such information management can be implemented in 1+MG.

Recommendations on a 1+MG - EHDS alignment based on the comparison between the proposed data governance of 1+MG and the draft EHDS Regulation

This document has been added as an appendix as it is not strictly part of the toolbox for the implementation and was not part of task 2.4. However, the publication of the proposal for a

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

European Health Data Space (EHDS) as a Regulation by DG SANTE highlighted the need for additionally analysing the interactions, synergies and possibilities for integration were analysed. For this analysis, the document goes through the different elements of the 1+MG data governance and compares the different approaches. A summary of the conclusions is provided but also a detailed point by point listing of the comparison including references to the relevant articles.

2. Contribution towards project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives/key results:

[Select 'Yes' (at least one) if the deliverable contributed to the key result, otherwise select 'No'.]

	Key Result No and description	Contributed
Objective 1 Engage local, regional, national and European stakeholders to define the requirements for cross-border access to genomics and personalised medicine data	**1.** B1MG assembles key local, national, European and global actors in the field of Personalised Medicine within a B1MG Stakeholder Coordination Group (WP1) by M6.	No
	2. B1MG drives broad engagement around European access to personalised medicine data via the B1MG Stakeholder Coordination Portal (WP1) following the B1MG Communication Strategy (WP6) by M12.	No
	3. B1MG establishes awareness and dialogue with a broad set of societal actors via a continuously monitored and refined communications strategy (WP1, WP6) by M12, M18, M24 & M30.	No
	4. The open B1MG Summit (M18) engages and ensures that the views of all relevant stakeholders are captured in B1MG requirements and guidelines (WP1, WP6).	No
Objective 2 Translate requirements for data quality, standards, technical infrastructure, and ELSI into technical specifications and implementation guidelines that captures European best practice	**Legal & Ethical Key Results**	
	1. Establish relevant best practice in ethics of cross-border access to genome and phenotypic data (WP2) by M36	Yes
	2. Analysis of legal framework and development of common minimum standard (WP2) by M36.	Yes
	3. Cross-border Data Access and Use Governance Toolkit Framework (WP2) by M36.	Yes
	Technical Key Results	
	4. Quality metrics for sequencing (WP3) by M12.	No
	5. Best practices for Next Generation Sequencing (WP3) by M24.	No
	6. Phenotypic and clinical metadata framework (WP3) by M12, M24 & M36.	No

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	7. Best practices in sharing and linking phenotypic and genetic data (WP3) by M12 & M24.	No
	8. Data analysis challenge (WP3) by M36.	No
	Infrastructure Key Results	
	9. Secure cross-border data access roadmap (WP4) by M12 & M36.	No
	10. Secure cross-border data access demonstrator (WP4) by M24.	No
Objective 3 Drive adoption and support long-term operation by organisations at local, regional, national and European level by providing guidance on phased development (via the B1MG maturity level model), and a methodology for economic evaluation	**1.** The B1MG maturity level model (WP5) by M24.	No
	2. Roadmap and guidance tools for countries for effective implementation of Personalised Medicine (WP5) by M36.	No
	3. Economic evaluation models for Personalised Medicine and case studies (WP5) by M30.	No
	4. Guidance principles for national mirror groups and cross-border Personalised Medicine governance (WP6) by M30.	No
	5. Long-term sustainability design and funding routes for cross-border Personalised Medicine delivery (WP6) by M34.	No

3. Methods

All policy recommendations were conceived by combining existing knowledge and expert opinions, which were then used to set up the recommendations for a large-scale genome data sharing initiative. The 1+MG initiative and its goals as derived from the joint Declaration of signatory countries was taken as a baseline for which recommendations were made. This provided us with concrete use cases that could be analysed and ensures that our recommendations also find a practical implementation audience. The 1+MG ELSI Working Group served here both as input provider and as sounding board. In some recommendations, also the 1+MG Group with the representatives of the signatory countries were involved for feedback provision.

The following assumption on the data life cycle in 1+MG was made.

The data journey in 1+MG

The following stages are considered for the data journey.

1. Data preparation: Pre-processing to agreed standards, annotation with metadata etc.
2. Data inclusion: Physical transfer of data incl., legal transfer of data to 1+MG to enable visibility in data catalogue
3. Data storage and management: Including GDPR compliant processing environment, data versioning etc.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

4. Data discovery: Discovery of data using GDPR compliant application programming interfaces (APIs)
5. Data access: A mechanism(s) by which the party acting as the controller for data access/disclosure can authorise access to select dataset(s)
6. Data use: Data processing for the approved purposes to obtain a result
7. Data archiving for the approved purposes (where necessary for a respective purpose)

Data's life in 1+MG

Data included in 1+MG have already been generated in a different context (primary use) and are subsequently brought into 1+MG by the Data Provider. It is the responsibility of the Data Provider that the data are characterised with relevant metadata and curated into a data model accepted by 1+MG. 1+MG and potentially national / local structures will support Data Providers. Data are uploaded by a Data Provider, alongside with the associated metadata. The data are ingested into the system and relevant metadata made available in the data catalogue. A user may choose from the data catalogue relevant data collections or, after registration for research use, even individual records across different collections and countries. Based on the selection, the user launches an access request. The signing official of the user's institution must confirm this request unless the user is whitelisted in the system, which flags that the researcher has a permanently confirmed right in the home institutions to apply for access. The information on the request with all relevant information is sent to the 1+MG central office of the central 1+MG Data Access Committee (central DAC) and the relevant National Coordination Points. The central office communicates to the relevant National Coordination Points the justified access decision of the central DAC. In case of research access, the National Coordination Point (after consultation of relevant stakeholders, where applicable) may or may not communicate a veto in a certain time frame. Positive feedback can be sent as well to shorten the response time for the access request. In case of approval and a signed contract, the user receives access to the data set that was applied for. This may include subsets of records according to data types that can be chosen and therefore a definition of a "minimum data unit" that can be managed. The user may need an easy-to-use, intuitive interface for querying data and tracking their applications, data that they have been granted access and expiration date, how to extend them etc.. Other users need the possibility to run more complex analyses on the data; some may even bring in their own data and /or own algorithm. This may require a more complex compute environment. Data may change over time because of new data points added (additional data types, additional time points) or data subjects' requests (deletion, rectification). In addition, data use may lead to enriched or derived data also available through 1+MG. The users must be able to update datasets but also to request a stable version of the dataset that will be preserved in an unchanged form. For some use, archiving of the analysis and the possibility to re-run the analysis again up to 10 years later is required, in accordance with good scientific research and reproducibility practices. A data versioning (release) mechanism and tools are therefore needed. This should also include the possibility to identify when dataset versions can be deleted because there is no longer a need to retain them.

Detailed methodologies for the individual recommendations

As this deliverable addresses a heterogeneous scope of recommendations, a more precise description of the methods used for each recommendation is provided below.

Data governance for research

The data governance for research is a first proposition developed by B1MG WP2 and together with the 1+MG ELSI WG in the course of dedicated workshops, several ELSI WG meetings and written comments. We used guidelines by recognised organisations in ethics and built on GDPR requirements and guidelines by the European Data Protection Board (EDPB) to compile the data governance. It also builds on the policies for incidental findings, special (vulnerable) subjects and groups as well as the information provision on general research results. Often, there is only a

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

brief translation of outcome into this document; the more detailed information can be found in the respective policy.

Apart from the ethical and data protection requirements, practical feasibility considerations, the possibility to scale up operations and the experience made by other research infrastructures fed into the development. In particular, the considerations of a federated environment, the heterogeneity of participating countries who should be given flexibility as much as possible was a decisive element as well as the necessity that the resulting data infrastructure should function as a harmonised virtual cohort towards the user.

Input was also provided by other WPs of B1MG. The document was subsequently presented to the 1+MG Signatory Countries for information and to obtain feedback for further refining the data governance. The feedback was considered and led to a further refinement of the data governance recommendation.

Transparency and consent guidelines

For the guidelines we analysed ethics literature on practical and ethical challenges raised by the process of consent in the context of biobanks, genomic/genetic research, precision medicine, genomic medicine initiatives, and clinical care. In our search in the PubMed databases and other relevant specialised journals, we have prioritised two types of publications:

- findings and recommendations based on empirical studies, where conclusions of the study were drawn from empirical evidence (e.g. qualitative and quantitative studies, such as assessment of research participants' perceptions of research based on the information provided through the consent form/notice, etc.,);
- recommendations, reports, guidelines, models of consent developed by key leaders and initiatives in the field.

We used conclusions from project workshops organised in the course of the B1MG project as well as the 1+MG use cases both to compile and to subsequently revise the document. Members of the 1+MG ELSI WG tested a first version that was compiled as a policy. We subsequently restructured the document to have a more stringent guideline, which is easier for implementation.

The work was jointly pursued by Task 2.2 for ethical aspects and Task 2.3 for legal aspects as well as Task 2.4 for practical implementation. We also reviewed GDPR requirements and European Data Protection Board (EDPB) guidance. A sample of national data protection law implementations applicable in the research and healthcare sector were reviewed and are reported in the appendix.

Practical approach towards the management of the generated IP rights

The recommendation is based on a mixed method, consisting of findings of the paper "Ethical, Legal and Social Implications in Research Biobanking: A Checklist for Navigating Complexity" which is currently under publication at Developing World Bioethics (Annex 1) and literature review of relevant papers, reports, and international legal instruments. In particular, the outcomes of the paper emerged from research focusing on biobanking in Africa, which was funded by both the B1MG project and "B3Africa" Project. The findings of the research enabled the creation of a four-step checklist, which reflects the requirements that researchers should fulfil to ensure the Ethical, Legal and Social (ELSI) compliance of their research project. The research was mainly focused on the African continent. Nevertheless, the paper presents a comprehensive overview that transcends Africa and can be applied in various research settings. The findings of the paper have been complemented by subsequent literature review on IP rights in the context of cross-border access and use of personalised medicine data.

Data protection by design and default (DPbDD) recommendations for the IT infrastructure

The recommendations are grounded in Article 25 GDPR on data protection by design and default as well as the related guidelines by the EDPB. These legal requirements have found their way not just into recommendations for IT security but rather into the entire data governance of the planned 1+MG data infrastructure. IT tools can help to implement data governance

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

procedures more securely, which is why this recommendation builds on the data governance for research.

Several workshops were held with the 1+MG Working Group 5 on IT infrastructure and WP4 of B1MG. The outcomes were compiled as a recommendation to WP4 for the design of the infrastructure. A last joint workshop was scheduled to review and discuss the recommendations jointly. Also the 1+MG ELSI WG was invited to join the workshop as well as to provide written comments.

Data protection by design and default (DPbDD) recommendations for the information management

The work on the data governance and the DPbDD recommendations led to the recognition that beyond IT tools, there is also a need to pay attention to information management in a genomic resource as planned by 1+MG. Information refers here to ELSI related metadata but also other information that documents data inclusion, data access, measures implemented to safeguard the rights and freedoms of data subjects etc. to fulfil the accountability requirements under the GDPR.

Input for this deliverable was derived from all the other recommendations and deliverables in WP2 as well as from input of the 1+MG ELSI WG. The WP leaders of WP3 and WP4 were asked to review the content. A first review took place as part of the DPbDD workshops. A dedicated workshop on the recommendations with the relevant 1+MG Working Groups is scheduled for September.

Recommendations on a 1+MG - EHDS alignment based on the comparison between the proposed data governance of 1+MG and the draft EHDS Regulation

The comparison was performed as a desk exercise, comparing the data governance worked out for 1+MG within this task and the draft Regulation for the EHDS. The result was presented and discussed in the 1+MG ELSI Working Group and based on the feedback, an updated version was compiled. This version was also presented to the 1+MG Group.

4. Next steps

The recommendations are the basis of further work in a subsequent deployment project for the 1+MG initiative through the Genomic Data Infrastructure (GDI) project that builds on the B1MG work. The aim of the GDI project is to deploy the IT infrastructure for 1+MG and to provide a framework for sustainability, both legally and financially. The data governance and DPbDD recommendations will be used to ensure the responsible implementation of the data infrastructure. The legal implementation will build on the entirety of recommendations. An additional project that aims to create reference genomes in Europe will also build on the results to consider an integration of the collected genomes into a sustainable implementation of the 1+MG.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

5. Recommendations for a 1+MG Data Governance Framework for Research Use

Version 2.1 - 7 November 2022

Authors: Adrian Thorogood, Regina Becker

Main contributors: Mihaela Matei, Signe Mežinska, Susanne Rebers, Olga Tzortzatou

Additional contributors and/or reviewers: entire WP2 / 1+MG ELSI WG

Introduction

This document recommends standards from an ELSI point of view to ensure genomic and health-related data can be responsibly and effectively made available through 1+MG **for secondary use for research purposes** (healthcare reuse of 1+MG data and policy-making purposes to be addressed later).

The research referred to in this document is scientific research as envisaged in the GDPR. This is a fairly wide scope that also includes "applied research"[5]. Following the EU Framework for State aid for research and development and innovation[6], 'applied research' means industrial research, experimental development, or any combination of both. In addition, EDPB states in their Guidelines 05/2020[7]: 'scientific research' in this context means a research project set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice.

For good practice in methodological standards, we build on the guidelines for good scientific practice in research by the European Science Foundation[8]. For good ethical practice we build on the ethical guidelines by the Council for International Organizations of Medical Sciences (CIOMS)[9]

.

Scientific research on data governed by 1+MG must pursue the goals of the 1+MG Declaration, i.e. must aim for improving precision medicine and better health for the EU citizens in general, e.g. involving prevention.

Elements not covered and to be addressed in complementary documents:

- The data governance considers currently only access by users within the European Economic Area (EEA). For access by users outside the EEA, additional requirements will have to be applied, which are likely of a contractual nature but may also contain limitations of data processing operations. Regarding data governance for access in the

[5] GDPR Recital 159: "For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research."
[6] Communication from the Commission — Framework for State aid for research and development and innovation, 1.3 Definitions; https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:52014XC0627(01)
[7] EDPB, Guidelines 05/2020 on consent under Regulation 2016/679, para 153, https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf
[8] European Science Foundation: Good Scientific Practice in Research and Scholarship; https://wcrif.org/documents/293-2007-242-good-scientific-practice/file
[9] Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO), International Ethical Guidelines for Health-related Research Involving Humans; https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

context of rare diseases, it has not yet been decided where these fall between healthcare and research applications.
- This document does not address the technical standards for interoperability or data security. More detailed provisions on technical requirements are provided in a separate document. The data governance does also not cover aspects of the overall legal framework or the financial sustainability.

The document is a first proposition developed by B1MG WP2 and together with the 1+MG ELSI WG in the course of dedicated workshops, several ELSI WG meetings and written comments. It builds on the policies for incidental findings, special (vulnerable) subjects and groups as well as the information provision on general research results. Often, there is only a brief translation of outcome into this document; the more detailed information can be found in the respective policy.

Input was also provided by other WPs of B1MG. The document is presented to the 1+MG Signatory Countries for information and to obtain feedback for further refining the data governance. As such, it is to be seen as a dynamic document that will be finalised through adoption by the countries that represent 1+MG and ultimately by implementation. It will be complemented by other documents that are more operational and relevant for the implementation of this governance.

Following the 1+MG Declaration, one of the goals of the 1+MG Initiative is to establish a European research cohort of over 1 million genomes. The Initiative will be structured as a federated network that connects genomic data resources and supporting infrastructures within Member Countries. This federated approach ensures that authority, responsibility, and resources are primarily based within the Member Countries, and that Member Countries have a certain flexibility over how to implement their national networks. To ensure the 1+MG cross-border, federated network truly functions as a “virtual” European research cohort, a clear governance framework must be established with the following aims:

1) to ensure efficiency and feasibility of cross-border access processes and therefore procedures that scale;

2) to promote clarity over general data access and re-use rights, applicable data-specific access and use conditions, and access procedures; and

3) to ensure compliance with applicable laws and ethical principles, particularly those relating to transparency and the protection of data subjects.

Data access by researchers based outside the EEA is explicitly considered in the “Scope of the 1+MG” policy. However, this special case will be addressed in a separate policy document once the “standard” data governance is agreed and can subsequently be integrated into the overall 1+MG data governance.

The 1+MG considers secondary use of genomic data that were collected in a different context. To comply with minimum dataset requirements of 1+MG, however, some accompanying health, lifestyle or demographic data may be collected dedicatedly for 1+MG in addition. These subsets of data would formally fall under “primary use”. However, this is not emphasised in the following as the focus is on the genomic data.

There is potential overlap with and opportunities in the Data Governance Act (DGA), and the European Health Data Space (EHDS) legislation, governance, and infrastructure. This framework aims to ensure compatibility with DGA and EHDS principles, as they become clear. Such compatibility would potentially allow 1+MG to act as a pilot of the EHDS or to ensure parts or all of the 1+MG could be integrated into the EHDS as an authorised participant once the strategy becomes clear as to how these initiatives relate. This framework also encourages Member Countries to pursue opportunities to leverage access and re-use rights, competent bodies, and data altruism organisations where applicable. However, this policy was developed with the approach to allow as much flexibility as possible to Member Countries while providing a single coherent and consistent data resource / infrastructure to the user.

B1MG has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 951724

Stakeholders

1+MG Data providers	The institutions who collected relevant genomic and or health data in a primary context of their own; they may or may not have the right to grant or refuse access to certain genomic and health-related data.[10] A Data provider is a legal entity (i.e., controller), with a role distinct from the technical infrastructure. Where the Data Provider originally collected and generated the data (e.g., national genomic medicine initiative), it is solely responsible for meeting the obligations ascribed to Data Providers. Where the Data Provider receives data from an upstream controller such as a hospital, the Data Provider obligations must be coordinated with this other controller.
1+MG Data Requester/User	The individuals and/or institutions seeking access to data (Data Requester), or granted access to data for a specific research project (Data User).
1+MG (Federated) IT Infrastructure	Consisting of secure platforms in / for each Member Country where data are physically submitted, pre-processed, stored, and made discoverable and accessible. These platforms provide services on behalf of Data Providers to assist them with meeting 1+MG standards and obligations. Infrastructure also includes secure processing environments where data are remotely accessed and analysed by Data Users to carry out their research projects. The IT Infrastructure must comply with technical interoperability requirements that are defined in 1+MG based on the work of WG5 to allow operations across the entire 1+MG network, where required.
1+MG Working Groups	Develop recommendations for and maintain standards (e.g., data quality) and policy frameworks. Standards may be described as minimum requirements or best practices.
1+MG Central Coordination (CC)	The central coordination team established in the 1+MG legal entity.
The 1+MG Data Access	A "central" Data Access Committee, DAC (or several domain-specific committees) hosted by 1+MG that receives access requests, reviews access

[10] There are two different definitions of data holders in current EU legislation (DGA) and proposed EU legislation (EHDS). This definition aims to include both types of envisaged entities.
DGA: Data Holder "means a legal person, public body, international organisation, or a natural person who is not a data subject with respect to the specific data in question, which, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal data or non-personal data." See a proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) (Analysis of the final compromise text in view to agreement v 10 Dec 2021) Art 2(5).
EHDS: 'data holder' means any natural or legal person, which is an entity or a body in the health or care sector, or performing research in relation to these sectors, as well as Union institutions, bodies, offices and agencies who has the right or obligation, in accordance with this Regulation, applicable Union law or national legislation implementing Union law, or in the case of non-personal data, through control of the technical design of a product and related services, the ability to make available, including to register, provide, restrict access or exchange certain data; Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space; version 3.5.2022
To avoid confusions as none of the terms seems to meet "our" data holder, we have termed the entity that envisages to make data available through 1+MG a "Data Provider".

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Committee (DAC) [11]	requests to ensure they are relevant and meet published access and used conditions, and sends a provisional decision to approve access.
National Coordination Point	A national node within Member Countries responsible for providing information on national rules for data inclusion, coordinating publication of a data catalogue, as well as coordinating search/access requests, and translating/publishing access/use information for a national audience (may be a competent body, permit authority (i.e,. an entity competent by law to approve data access for secondary use), or "lead" Data Provider (i.e. one Data Provider acting as primus inter pares).
Member Countries	Countries that participate in the 1+MG legal entity and are subject to the 1+MG legal and/or governance frameworks. Ensure appropriate national legislation, governance, infrastructure and sustainability is in place and strive to harmonise these across countries.

Guidance

1. **General access and re-use rights for research purposes must be clear, and any data-specific access and use conditions must be transparent.**

1+MG is a federated network, meaning in part that Member Countries and Data Providers determine the scope of access and re-use rights (subject to any European legislation or agreements). While they must agree to certain general conditions of access and use, they retain some flexibility to additional *(in particular legally necessary)* access and use conditions (at the national, institutional, or data level). Some conditions necessarily differ across countries due to different national laws and interpretations, and across data collection contexts, due to different consents, inclusion / exclusion modes, approvals and opt-outs. 1+MG itself does not establish these rights or conditions. To mitigate the resulting fragmentation, 1+MG must ensure agreement over certain general conditions described below, as well as transparency over any special access and use conditions. Data must be tagged with rights metadata following a standard vocabulary, to enable discovery through the data catalogue and record-level search function. Data requesters must be able to determine if they have rights to access and re-use certain genomic and related health data for their particular research projects. See **Appendix A** for a list of permissible, special access and use conditions that may be applied in research contexts.

Actor-specific obligations relating to ensuring transparency and accuracy of access and use conditions:

Data Providers:

- Must clearly indicate in the 1+MG Data Catalogue[12] what genomic and related health datasets are available cross-border for precision-medicine research purposes (in particular distinguishing where data are (additionally) available for other purposes such as data reuse for healthcare or policy-making).
- Must provide scientific metadata and metadata related to ethical, legal and social aspects (ELSI metadata) in the format agreed upon by 1+MG.

[11] We assume a DAC is established by an entity in the context of the data access decision. Whereas the DAC gives the recommendation for the access decision, it is the entity establishing the DAC who is legally responsible for the data access.
[12] 1+MG will establish a data catalogue that characterises the data available through 1+MG based on agreed scientific and ELSI metadata models and ontologies. The data catalogue will offer functions for data selection, including a selection on the record level to allow data minimisation.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Must agree to 1+MG general standards applying to access and use for research purposes described below (e.g., relating to access timelines, conditions of procedural fairness (justified refusals/appeals), IP conditions, appropriate fees (or lack thereof), and publication conditions) (see subsequent sections).
- Must transparently label data with any special access and use conditions (see examples in Appendix A) – following a controlled vocabulary – to ensure predictability for data requestors, and to enable the proper functioning of the central 1+MG DAC.
- Must provide transparency about the procedures and timelines required to fulfil any additional access and use conditions, such as project-specific ethics approvals, as well as procedural fairness (e.g., justified refusals, appeals) and regular reporting of performance and outcomes.
- Should strive to remove any unnecessary additional access and use conditions which are not based on legal or ethical requirements to which the Data Provider / the data use is subject, such as those stemming from institutional policies and procedures.
- Must provide clarification in a defined timeframe where descriptions in the 1+MG Data Catalogue are unclear (including descriptions of data-specific access and use conditions). *Note that Data Providers can reduce efforts by providing clear and accurate metadata, refined over time.*

1+MG Central Coordination

- Should provide guidance and compliance management tools to Data Providers on how to ascertain and communicate additional data-specific access and use conditions, e.g., by establishing a standard rights metadata vocabulary, providing checklists on national legal or ethics requirements.

Member Countries:

- Should strive to modify excessively restrictive access and use conditions stemming from national or regional legislation.
- Should strive to harmonise data access and use rights and additional data access and use conditions across Europe, through legislation (e.g., EHDS) and guidance.
- Must establish a National Coordination Point within the country to coordinate between the 1+MG DAC and (multiple) national Data Providers and inform 1+MG as well as the Data Providers about the national requirements and set-up in 1+MG.

2. 1+MG Access Processes and Central Data Access Committee

The federated network of 1+MG includes numerous Data Providers with authority over access to different resources. There may often be more than one 1+MG Data Provider within certain Member Countries. This presents a practical concern where a Data Requestor seeks cross-border access to multiple resources in multiple countries. In such a scenario, the requestor must be able to make a single access request that is reviewed in a coordinated fashion. Multiple different national or institutional Data Access Committees (DACs) reviewing access requests are likely to result in duplicated effort, delays, or inconsistent decisions. The 1+MG must therefore strive to fulfil the "single access principle" for the 1+MG European research cohort. Under this principle, requesting access to multiple data collections in multiple countries would require a single request and a single review by a trusted, independent body.

We recommend that 1+MG establishes an independent, central 1+MG Data Access Committee (DAC). It may in practice consist of one or multiple committees with a mandate to review access requests to ensure the projects fall within the scope of 1+MG purposes, that the data requested are relevant (data minimisation), and that the request respects published access and use conditions associated with the data (purpose limitation). General use of the 1+MG DAC

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

will be **mandatory** for Member Countries and Data providers.[13] The access decisions of this 1+MG DAC will also be **binding** on Data providers, though they would retain discretion to conduct their own review following the documentation of the 1+MG DAC's review and exercise a justified **veto** of the central decision within a limited time (ideally 2 weeks after receiving the decision by the 1+MG DAC, recognizing the need for national extensions where necessary). Following stakeholder consultation, the following approach is recommended in case of conflict: where a veto is not deemed justified by the 1+MG DAC, arbitration should aim to solve the dispute. If no consensus can be found, the respective data should not be disclosed. This approach offers economies of scale, establishes a trustworthy, competent, single point of access, preserves the ultimate authority and responsibility of Data providers, encourages the development of trust through repeated interaction, and provides flexibility for different contexts.[14] The design of the proposed access decision by the 1+MG DAC and the veto by the Data provider would need to be closely considered in relation to GDPR roles. It is suggested that the input from the national Data Providers are treated as part of the 1+MG DAC, which is decisive for the data they have brought into 1+MG if the position is justified. As such, complicated joint controllership arrangement and problems with the legal basis on the national level can be avoided. In this case, 1+MG as a legal entity would take sole controllership for the data disclosure.

The ability of 1+MG DAC to assess the scientific relevance of the data requested as well as the ethical/legal acceptability of the project is dependent on the quality of the scientific and rights metadata provided by the Data provider. Where Data providers seek to take advantage of this central service, they will be incentivised to provide high-quality scientific and rights metadata through the data catalogue. The Data provider should be able to rely more and more on the 1+MG DAC as trust is established through multiple interactions, while both parties can refine rights metadata descriptions and interpretations over time.

To foster the trust of stakeholders, the 1+MG DAC must have the appropriate composition, resources and independence to adequately review scientific, legal, and ethical aspects of submissions. The 1+MG DAC may actually consist of a central office providing administrative services as well as multiple committees with domain specific expertise (e.g., disease areas), which may include representatives from different Member Countries and/or Data providers. The details on the members of the 1+MG DAC, resources, the 1+MG DAC's position within the 1+MG, its working procedure etc. will be defined in an internal policy of the 1+MG. The decisions of the 1+MG DAC should generally be taken in a competitive time frame (suggested: 5 working days). Additional time may be needed where data requesters are asked for clarifications, or where additional consultations are needed, e.g. in case of the involvement of representatives of vulnerable groups). The 1+MG DAC will inform the affected National Coordination Points and Data Providers of the reason for these delays, involve them where needed and update them on the development. The administrative (management) roles of the central office should include the following:

- a central helpdesk as a first point of contact for requesters with questions about the access procedure and scientific and rights metadata.
- a single point of contact for access requests to 1+MG.
- an eligibility check to confirm the datasets and/or record types requested match the variable-level inclusion criteria in the data analysis plan.
- reporting obligations relating to the access process and decisions (see below)

[13] Local review requirements may be required by national law (e.g., a designated REC review or DPA review), which would be special access and use conditions (see Appendix A).
[14] Alternatives include the following: 1) each Data provider has its own DAC – requests are channelled there; 2) each Member Country establishes one or more central, national DACs – requests are channelled there (e.g., similar to EHDS2); ; 3) 1+MG establishes an optional central DAC. Alternative 1 will require substantial capacities at each Data provider to ensure responsible and timely access, which many Data providers may struggle to realise. Alternative 2 requires legislation on the national level to establish a legal entity to host the central DAC. Alternative 3 could be a compromise between the different options but may lead to longer response times as it is more difficult to coordinate and will lead to parallel work.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Transparent and documented IT tools may be adopted to provide a more streamlined (semi-automated) and consistent management of requests and documentation.

The roles of central 1+MG DAC include the following:

- reviewing requests and making access decisions.
- arbitrating inconsistencies between Data Provider vetos and stated rights metadata.

Policies and procedures need to be established for revoking access, both as part of the regular procedure as well as in cases of a detected misuse and the subsequent information flow.

Data Providers

- Must provide in a defined time frame additional information on data if any clarification is required to progress with the data access procedures.
- Are required to transparently state their intention to conduct their own review as an additional access and use condition.
- Local review processes must follow access principles and best practices including transparency, timeliness, internal appeals, and reporting on performance.
- Data Providers must agree on a harmonised scope and template form of information to be obtained by the requestor/project.

Member Countries

- Should strive to simplify the organisational complexity of the 1+MG network and to pool limited access review resources and expertise (e.g., by establishing genome initiatives) and reduce national or regional complexity over data access decisions to the extent legally, politically, and practically feasible.
- Support on the national level the implementation of the standard data access information request form template (see Appendix B), i.e. replacing where necessary existing forms of data access bodies under the condition that the required information will be covered by the 1+MG form.
- Should provide an independent redress mechanism available to any stakeholder affected by an access decision. This includes in particular requesters whose access request was turned down or Data Providers whose veto was challenged. This may have to align with national law where applicable. Mechanisms may also be available under forthcoming data governance legislation.

Data Requesters

- Should provide clear, true and complete information and documentation in the access request for the review procedure, in particular information about the development, training, or application of AI within the project (see Appendix B - Access Request form).
- Must respond within defined time frames in case of questions or requests by the 1+MG DAC in order for the continued review of their request. A no-response will lead to the deletion of the access request.

1+MG IT Infrastructure

The 1+MG IT Infrastructure must provide the following infrastructure to support an end-to-end, efficient and responsible access process:

- Data Management
    - Authentication and access control of authorised individuals from Data provider.
- Data Discovery:
    - A Data Catalogue that allows researchers to find scientifically relevant data. It is the responsibility of Data providers and National Coordination Points to supply

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

this information, e.g., by adopting appropriate interfaces to allow metadata syndication. The Catalogue must include metadata on access and use conditions.

- o (Federated) Record-level Search Portal and System. The 1+MG infrastructure must provide a single portal for a privacy-preserving, federated query system that allows queries at the level of datasets, subgroups, or record-types (according to variable-level inclusion criteria), as a minimum FAIRness standard, and also to enable granular operation of the GDPR data minimization principle. The record-level search must also allow query of access and use conditions.

- End-to-end Access Information Management Tools. The 1+MG infrastructure must provide a common European portal and IT tools:
  - access request portal and common e-request form template (for entering common minimum requestor/project information and for selecting project-relevant datasets, subgroups or records). (See Appendix B) *Data providers or Member Countries may require additional, data-specific modules where they require additional information. These may be specified as "additional access and use conditions".*
  - access review workflow support (presenting the project information and selected data to the 1+MG DAC).
  - access decision management and its implementation (documenting the access decision making; generating a data access/use agreement incorporating the data requestor/project information and selected data; enabling access to requested data).
  - reporting on access and use (a register of standard records of approved, ongoing and completed projects for transparency and reporting purposes).

3. Ethics and Consent

As a general ethical principle, data subjects must be in agreement with their data being made accessible to the research community for precision medicine research purposes through 1+MG. Such agreement is typically captured as research ethics consent, opt-in or opt-out, and must generally be based on minimum core information elements, including the precision medicine purposes, the voluntary nature of participation, the types of research organisations who may be granted access, and associated individual risks and (lack of direct) benefits. **(See 1+MG Consent Policy)** Ethically, data may still be shared without the individuals' informed consent or where (existing) consents do not cover the core elements under certain conditions (e.g., impossibility of obtaining consent) as confirmed and approved by a Research Ethics Committee (REC). Beyond these general ethical principles, there may be additional national legal requirements relating to consent found in data protection laws (often but not always dependent on the legal basis), biomedical research laws, healthcare laws, biobanking laws etc. e.g., requirements for consent content or form, legal instruments to lift medical secrecy around healthcare data. Exceptions to consent requirements may be granted on a national level (e.g., where data sharing from healthcare contexts is based on a law; where a research ethics committee lawfully waives the consent requirement). It is also a general ethical principle that there should be an ethics review of research projects involving human tissue and personal health and genetic data, though this requirement may not apply uniformly in contexts of secondary use.

<u>Data Providers:</u>

- Must generally ensure transparency rules have been observed and consent has been obtained according to the 1+MG core consent elements and national/regional legal requirements.
- Where data are not made available by legislation, must obtain a principal research ethics committee (REC) approval for making data available cross-border for general,

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

precision-medicine research purposes (subject to legally required access and use conditions, which may include a project-specific REC approval).

- The REC in this case shall consider at a minimum the following specific issues:
  - the collection of the data is in compliance with ethical standards
  - appropriate ethics consent or justified waiver to make the data available in 1+MG for the specified objectives under the common governance and oversight mechanisms.
  - risks and benefits and safeguards, including those relevant for special groups,
  - the handling of incidental findings (IF), including those potentially reported back from 1+MG (where applicable; more details below and in the Incidental Findings policy).
  - Must document all relevant ethics consents and/or approvals that were obtained.

Member Countries:

- Should work together to harmonise requirements (e.g., scope of review) for REC approval related to research projects involving the secondary use of genomic and related-health data.
- National ethics committees, where applicable, may provide nationally tailored guidelines.
- Must extend the 1+MG EU-level guidance on transparency, consent and ethics for additional rules mandatory on the national or regional level, where applicable. This can be realised through the National Coordination Point or another suitable body.

1+MG CC

- Should provide EU-level guidance and checklists about transparency, consent and ethics for secondary use.
- May assist Data providers with documenting ethics consents and approvals.
- Should establish 1+MG advisory bodies to anticipate emerging ethics issues.

Data Requesters:

- Must obtain a project-specific local or national REC approval (or an explicit exemption thereof) for their planned research as a safeguard to ensure the fairness of processing under the GDPR.

4. Data Security

[This section only provides an overview of requirements. More detailed requirements for a security framework have to be compiled by B1MG WP4 and/or the 1+MG WG5 on the IT infrastructure]

General

- The 1+MG IT Infrastructure must provide secure storage and processing environments, authentication and authorization protocols, access logging, following 1+MG rigorous security standards assured through appropriate contractual agreements, certifications (where they received data from other centres) and external audits (against 1+MG standards) by suitable independent entities as defined in a separate 1+MG DPbDD infrastructure / IT infrastructure policies.

Discovery

- As variable-level search generally involves processing of personal data to generate statistics, the 1+MG infrastructure for federated search must ensure a robust data protection and security framework to avoid disclosure of personal data as well as inappropriate processing (e.g., user authentication, output controls (e.g., n>5), query

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

budgeting (limiting the number of queries that can be made by a requestor), monitoring for suspicious behaviour, terms of use).

Secure Processing Environment for Data Analysis

- Ensure only selected data are made available to authorised users.
- Allows efficient analysis of data from multiple Data providers/countries, ideally in a federated manner across national secure processing environments.
- Where a federated analysis is not technically possible for a certain analysis, permitting temporary pooling and analysis in one or a few secure processing environments. These IT infrastructure providers allowed to pool data must have obtained an appropriate certification of security and procedures.

5. Data Protection

As part of the governance framework, Member Countries must agree on a general legal GDPR compliance framework.

Controllership

- Generally, the Data Provider will be the sole controller for pre-processing, storing and managing data in advance of any access request unless a legislation is in place that assigns such tasks to another entity (e.g. the 1+MG legal entity).
- The Data User will generally be the sole controller for the research project (or may be a consortium of data users in joint control).
- The 1+MG legal entity will be the controller for the access provision to the Data User where the access decision is made by the 1+MG DAC and the 1+MG legal entity signs the data user agreement.
- Where access decisions are made by the 1+MG DAC but with an absolute time-limited veto exercised by an entity mandated on the national level (e.g. the Data Provider) or an approval to be given by such entity, there may be joint controllership for the associated processing.

Transparency

Data Providers

- Must generally meet GDPR transparency requirements (and where applicable additional national transparency and/or consent requirements), including detailed information about cross-border access and secondary use for research, healthcare and/or policy making purposes before data are made available through the 1+MG legal entity.
- Must document that they have provided this information to the data subjects, or provide a justification that an exception applies in a DPIA (see below).

1+MG CC

- Should prepare a European-level checklist of core information elements, which can be tailored to include additional national requirements by advisory bodies in Member Countries.
- Should publish a transparent register of all authorised projects (see below).

1+MG IT Infrastructure

- Should support the information provision on data use to individuals

Data Users

- Must make all Art. 14 information to the data subject available through the 1+MG to ease the information retrieval for the data subject.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Data Minimization

Data included in 1+MG is pseudonymised personal data and any data access must be in compliance with the data minimisation principle.

Data Providers

- must ensure data are appropriately pseudonymised, removing all direct identifiers, before making data available through 1+MG, with the support of 1+MG Infrastructure.

1+MG DAC and/or Data Providers

- must ensure only relevant data is provided, necessary to conduct the specific research project.

Data Requesters

- must identify granular data-field inclusion criteria as part of their access request, justified by their data analysis plan.

Data Subject Rights

Data Providers

- must ensure data subjects are informed of their rights and any exceptions (e.g., transparency, access, rectification, portability, withdrawal of consent, objection),
- must enable, to the extent possible and required, the exercise of these rights.

1+MG IT Infrastructure

- must enable the implementation of the changes due to the exercising of data subjects' rights

Data Providers and Member Countries

- should seek to harmonise applicable rights across 1+MG.

Individual Withdrawal and Limitations

1+MG

- must establish standard policies and IT tools to ensure respect for individual rights of withdrawal, objection and/or erasure, as well as any standard limitations e.g., once access is granted, or once analyses are completed or archived.
- Ideally, Member Countries should strive towards establishing an EU law limiting rectification in the context of archiving.

Data providers

- must ensure that data subjects are informed on the limitations of rights in the downstream use of data.

1+MG Infrastructure

- must foresee a versioning of datasets where data subjects have exercised their rights of rectification, objection to (certain) processing or erasure to ensure that only those users who cannot use updated datasets still use the original data on which they started their processing.

Accountability

Data Providers

- must comply with any applicable EU and national legal requirements.
- must conduct a data protection impact assessment (DPIA) covering sharing the data through 1+MG verifying and documenting compliance with both European and national data protection rules and respect for the fundamental rights of data subjects. *Given this is*

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

clearly high-risk processing, there should not be any exceptions. This may be a DPIA specific to the Data Provider's 1+MG activities, specific to a collection, a module of part of the Data Provider's broader healthcare/research processing activities. It may also be conducted jointly with other Data Providers across 1+MG or based on overarching solutions based on legislation where the DPIA is performed as part of the legal procedures. A DPIA is a valuable procedural safeguard and accountability tool to ensure documentation that the Data Provider complies with data protection principles. A procedural requirement for data inclusion is also more flexible than having substantive requirements. As data may come with different access and use conditions, they will likely also come with different legal compliance requirements.

- The DPIA must cover the following elements:
    - data sharing plan - description of potential data flows through 1+MG.
    - lawfulness - legal basis to include data in 1+MG; in case of consent as legal basis, confirming consent requirements above where applicable.
    - transparency - confirming transparency requirements (see below).
    - data minimisation - confirming data have been appropriately pseudonymised and de-identified before inclusion to 1+MG. 1+MG certified data hubs and/or national competent bodies under the DGA may assist by further de-identifying data and re-pseudonymising data to apply a 1+MG pseudonym (the resulting data would be "double" coded).
    - respect for fundamental rights - processes and contact points must be in place to ensure organisations in the processing chain coordinate so that data subjects can exercise their rights.
    - Risk identification, assessment, and responding safeguards.

1+MG

- should provide support through generic DPIA templates and risk assessments of common platforms, tools and services.

6. Data Quality / Utility

Data Provider:

- Must ensure data meets 1+MG data quality/utility standards, or transparently tagging data with quality information
- Must provide scientific metadata following standard data models
- Must provide at least the minimum datasets in agreed data models (where applicable)
- For longitudinal cohorts, must ensure (meta)data are up to date and properly versioned.

1+MG (Federated) Infrastructure

- Provides data curation tools and training.
- Due diligence check on minimum meta(data) models.

Member Countries and 1+MG

- Adopt 1+MG data quality/utility standards
- Align with the European Commission on EU-level developments on relevant standards
- Provide support for data curation activities through training and, where applicable, suitable tools.
- Ensure harmonisation of services offered by federated infrastructure.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

7. Re-contacting Data Subjects

Data Providers can additionally support research by enabling the re-contact of data subjects. This may be for the following purposes: collection of additional information or samples; invitation to participate in additional observational studies or clinical trials; return of clinically relevant findings. Ethically, offering these opportunities to be re-contacted may be seen as a way to give back to participants.

Data Providers

- should seek appropriate approvals and/or consents so that individuals can be offered opportunities to be re-contacted in the future for each of the specific purposes listed above.
- must clearly label patient data according to the permitted purposes of recontact.
- where applicable, should establish appropriate recontact processes (e.g., permission management systems, maintaining up-to-date contact details, establishing a process for accurate and secure de-pseudonymisation, re-contact processes with appropriate support, offering translation where needed).

Member Countries

- Should strive to provide central procedures for re-contacting individuals through appropriate channels depending on the purpose.

8. General Conditions for Data Use

A contractual framework will be in place including a Data Use Agreement and a Processing Agreement (between user and the 1+MG Legal Entity) (see Appendix C). The **Data processing agreement** will address the services provided by 1+MG Infrastructure to the Data Users.

1+MG Legal Entity

- must establish **standard Data Permit/Data Use Agreement templates** including any specifics about the requestor, project, and data requested, as well as the general conditions covered in this section.
- will act as processor for the data processing with respect to the use.
- will conclude (sub)processing agreements with all entities providing the federated 1+MG IT Infrastructure for data use.
- The central DAC will have the power to remove data user access where they violate the conditions of the data use agreement (see Appendix C), in particular purpose limitation or confidentiality. The central DAC will also have the power to prohibit future access by the user for a certain period (e.g., 5 years).

Data Providers

- must not withdraw data from use without cause. (see also “Data Subject Rights”).

Data Privacy and Security Conditions

- Contractual Limitations on Data Use (see Appendix C), including at minimum
  - Must respect that processing takes place in 1+MG’s SPEs only
  - Use for approved purposes only.
  - Duty of Confidentiality.
  - No re-identification
  - No individual-level linkage (without permission).

B1MG has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 951724

  - Respect Retention period.
    - Default 1 year limit, with extensions or pauses (sleeping clause) permitted.
  - Personal Data Breach reporting obligations.
- No personal data must be exported by the Data User
- Data Users may bring in their own data and/or own software (including commercial software and open-source) subject to security review by the 1+MG IT Infrastructure and reproducible documentation.

Reporting Incidental Findings of Health Relevance to Individuals and Family Members

[The detailed Incidental Findings Policy can be found in D2.2]

Data User

- Must report certain clinically actionable findings to the Data provider.

1+MG

- Should foresee a technical solution in each country that channels the reporting of relevant findings to the Data provider. The 1+MG NCP will be responsible for establishing this solution in the respective country.
- Should implement the possibility of a "no return" flag is to be foreseen that interrupts the reporting.

Data Provider

- Must establish policies and processes to responsibly handle the findings, subject to national policies, and to the individual's consent. (see 1+MG Incidental Findings Policy).

IP and Commercialisation

Data Providers:

- Must agree not to make any IP claims on results derived from the data.
- Must agree not to pursue IP protections that would prevent or block access to or use of any data or results drawn directly from data.

Data Users:

- May pursue IP claims on downstream discoveries or inventions.
- Must not pursue IP claims that limit the use of 1+MG data itself
- Should report to 1+MG any IP claims that were derived in the context of the data use

Publication

Data Providers

- A default, time-limited publication embargo for pre-publication data, subject to an opt-out by Data Providers.

Data Users

- Respect the 1+MG publication policy, which comprises among others the elements listed below.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Respect any publication embargos for pre-publication data, and not publish during the embargo period without written confirmation of a justified derogation from the embargo (as part of the use conditions in the contract or as amendment to the contract)
- Must Acknowledge all data sources and 1+MG.
- Should report the results of completed projects (e.g., publications) to 1+MG DAC to add to the access register.
- Publish manuscripts on the results from research with 1+MG data as open access.

1+MG

- Ensure compliance of publication with 1+MG rules e.g. open access, acknowledgement, respect for embargos, public reporting of results.
- Regular publication searches and reporting on published findings.
- Provide professional communication on a representative selection of publications for the general public in English, based on the recommendation of a communication advice panel.

National Coordination Point

- Perform translation of selected communication into national languages at least for those cases where data of the country was used.

Archiving

There are ethical, policy, and possibly legal obligations for data users to archive data from research projects for the purposes of reproducibility. Solutions have to be found how this can be realised in a federated environment. (See in this context: DPbDD requirement analysis.)

1+MG Legal Entity

- Must ensure archiving of data to Data Users where necessary in exchange for the prohibition of a download of personal data.

Data Provider:

- Data Providers must agree to archiving of data used for research projects for a certain period of time after the completion of the analysis (e.g., 10 years) even if data are withdrawn for further use from 1+MG.
- Must establish an appropriate GDPR framework to permit this (e.g., information of data subjects on purpose of reproducibility archiving and requirements including limitation of data subjects rights)

IT Infrastructure:

- Must have technical data management systems in place to create project-specific data archives that allow reproducibility. This may be done as a pooled record, federated record, or protocol record (with explicit data versioning).

Data User:

- Must provide information on the required archiving period.
- Must archive any own data used in the project in its own chosen archiving system.

Return of Enriched and Improved Data to Data providers.

Data Provider

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Must agree that enriched or improved data returned after research projects are available in 1+MG to the research community under the same terms as the original data (see also Archiving).

1+MG

- Must include specific details in the data use agreement addressing the return of specific enriched or improved data.

Data User

- Must specify in access request any enriched or improved data (e.g., annotated, harmonised, corrected) that will be generated as part of the research project that could have potential general scientific value.
- Must return enriched or improved data of potential general scientific value to 1+MG and Data providers for further sharing free of charge.
- Must provide suitable metadata that describes the enrichment.
- May have an embargo of maximum N months before enriched or improved data will be made available to the research community to allow time to publish.

Fees

- Not addressed in this framework. This will be part of the future sustainability plan of 1+MG.

9. Requirements for 1+MG Secure Processing Environments

The 1+MG must provide a performant, flexible, sustainable and secure environment for data users. This will include establishing rules around the use of commercial versus open source software, considering aspects of cost, performance, stability, and data protection.

Member Countries must provide an SPE suitable for federation or must agree to use the SPE provided by another party.

Outlook

This high-level framework is the basis for more detailed work within B1MG and beyond. Many tools and documents can only be developed once the data governance is adopted by the 1+MG initiative. We identified a need for the following tools in the future:

- 1+MG Data Access Committee - Standard Operating Procedures
- 1+MG Data Use Agreement template
- 1+MG Data Processing Agreement for data use [to the extent that this is not provided for by law in accordance with Art. 28.3 GDPR]
- 1+MG Data Protection by Design and by Default Guidelines for 1+MG infrastructure
- IT infrastructure Policies (jointly with WG5)
- 1+MG Data Security Framework (by WG5)
- 1+MG Data Interoperability Framework (by WG5)

Further subjects will be identified during the further progression of the 1+MG towards deployment.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Appendix A. "Additional" Access and Use Conditions.

The policy above describes a number of general access and use conditions that apply to providing access to and secondary use of genomic and related health data through 1+MG. The 1+MG permits Data providers to apply certain special access and use conditions where necessary, such as where the conditions are required by Member State law. Some conditions may require significant Data provider effort to fulfil, and the Data provider must be transparent about the process, timelines, and costs associated. Any special conditions must be transparently stated as metadata by the Data provider.

Specific project institutional approvals (beyond approval by a central DAC)
Specific project approval by Data Provider's designated DAC (prohibiting delegation to central DAC).
Specific project approval by Data Provider's (local or national) REC.
Specific project approval by Data Provider's country's DPA.
Specific project review/consultation by Data Provider's DPO (and possibly a DPIA).
Specific project review/consultation by (special) participant representative group.

Individual Data Subject Consent and Transparency
No information of individual data subjects on the data use is possible.
No return of incidental findings

Restrictions on Types of Research / Research Organisations (to be considered by a DAC)
Certain type or category of research, e.g., - disease-specific; - population-specific (e.g., age category); - methods development;
Type of organisation (e.g., not-for-profit organisations only).

Appendix B. Common Data Access Request Form Elements / Access Criteria

Access Request Form template must include the common minimal information:

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Project title
- Name and contact details of the Principal investigator of the project
- Information about the institution
  (Institution needs to be registered including information on signing official and DPO; where several signing officials are applicable, the relevant official is to be chosen)
- Persons who will have access to the data (PI and his/her group)
- Scientific abstract
- Lay summary of project
- Data analysis plan that describes the data types and analysis approaches to be employed to achieve the goal of the project.
- Information on enriched data that may be generated through the project
- Description of any AI employed, developed, trained, validated, or exported by the project.
- Estimated project duration
- Legal basis under Art. 6.1 GDPR and legitimation under Art. 9.2
- Funding source
- Research ethics application and decision for the specific project
- Machine readable characterisation of the project based on controlled vocabulary (e.g. field of research, commercial versus pre-competitive research, data types to be accessed, etc.), (checked for consistency with analysis plan and other project documents by 1+MG Access Office). *This will be necessary for using automated tools to check that access requests are consistent with data availability conditions.*
- Selected data records and/or data types of data to be accessed (relevant records should be automatically selected from the catalogue browsing and/or Beacon search)
- Information if access request is only upheld if ALL selected records are available (or potentially what a critical % of available records is)
- Confirmation by a signing official at the requester's institution (requester could be white-listed at their institution).
- In case of consortia with other researchers requiring access: modules may be added to collect relevant information from additional partners. Ideally the software / portal allows "invitation" of collaborators who can build on the information already provided by the person launching the access request in making a subsequent request to be added to the project.
- Information on required processing environment (required to assess the feasibility of conducting the research in the 1+MG environment, the risk of processing, compile processing agreement and provide a quote on compute costs where applicable)
- Information on own data that may be uploaded for the research
- Once data access is granted, an additional contract will be needed to access the computing infrastructure, including the estimated costs of processing:
  - Specification of accessed data types to be migrated from 1+MG to computing environment.
  - Information if additional own data will be uploaded including specification of data types and amount (required for compiling a contract including processing agreement and assess storage / ingestion cost where applicable)

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Appendix C. Common elements of a Data Use Agreement

- Identification of parties to the agreement and their legal capacities
- Context
- Definitions
- Purpose / object of the agreement
- Obligations of data users
- obligation to process in the 1+MG secure environment [to be extended where needed]
- Obligation of 1+MG entities (including processing agreement)
- Intellectual property
- Financial provisions
- Term and termination
- Breach of agreement
- Consequences of breach of the duties
- Liability
- Applicable law- Jurisdiction
- Arbitration
- Exhibits
    - Identity of participating parties to the processing and description of their activities (e.g. involved data centres)
    - Project description including data analysis plan
    - 1+MG data security framework

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

6. 1+MG Transparency and Consent Guidance

v2.0 June 2023

Authors: Adrian Thorogood, Regina Becker

Main contributors: Miheala Matei, Susanne Rebers, Signe Mezinska, Maria Panagiotopoulou, Michaela Mayrhofer, Ayodeji Adeniran

Additional contributors and/or reviewers: entire WP2 / 1+MG ELSI WG

1. Introduction

1.1. Purpose

The transparency and consent guidance is a document that should help to compile GDPR and ethical compliant information necessary to provide if data are intended for secondary use. The document also includes a consent guidance where consent can be an informed consent under ethics regime and/or also consent as a legal basis under the GDPR. The guidance is therefore largely applicable independent of the chosen legal basis. Where consent as a legal basis is a relevant element, this is pointed out in the text.

This document can serve as a general guidance for transparency and consent in secondary use also beyond the 1+MG initiative. It describes what elements have to be covered. The content of a concrete information and consent sheet can only be compiled if the data governance for the secondary use is agreed.

1.2. Background

The 1+MG initiative aims to promote responsible cross-border access and secondary use of genomic and related-health data across Europe for research, healthcare, and policy-making purposes. This document provides consent recommendations for prospective data collections intending on making data available cross-border through a repository for research (where the exact projects cannot be fully identified at the time of recruitment).[15] The guidance focuses primarily on information and consent *content* elements, as consent models and processes[16] may vary across countries and contexts. Information content is also important for the transparency requirements under the GDPR, which have changed from the requirements under the Data Protection Directive 95/46/EC and where not all information sheets have been adapted to these changes. These content elements can also be used to design information for re-consenting or notifying individuals that their data will be included in such a resource.[17]

1.3. Nature of the Recommendations

Recommendations are made that 1+MG adopt 1) minimum requirements (MUST); 2) best practices (SHOULD); and 3) points-to-consider (non-directive). If a minimum requirement is missing, this may mean that a Data provider cannot legally or ethically make data available

[15] Secondary use for healthcare and policy-making will be addressed separately and outside this deliverable.
[16] This document focuses on consenting adult populations. Additional considerations for minors and other vulnerable populations are addressed in the 1+MG Special Subjects Policy.
[17] A future checklist will be developed outlining minimal requirements when assessing existing consents to determine if legacy data collections can be used for research and accessed cross-border. Making legacy data collections available will be greatly facilitated by the proposed European Health Data Space legislation. However, where data are made available through a separate data resource like the 1+MG, it will still be required that dedicated information is provided and an own legal basis for the data transfers to and from the resource is established.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

through 1+MG, or can only do so subject to special data and access and use conditions. Best practices may also constitute national legal requirements in some countries.

The recommendations are informed by the legal requirements of the European *General Data Protection Regulation* (GDPR)[18], the interpretive guidance of the European Data Protection Board (EDPB), research ethics principles[19] and guidelines, as well as legal data governance principles, such as those outlined in the draft *Data Governance Act*, and implemented in the 1+MG Data Governance Policy. Ethical requirements are in particular based on the *International Ethical Guidelines for Health-related Research Involving Humans* by CIOMS. Justifications and explanations are provided. Legal consent requirements depend on the legal basis selected under the GDPR Art. 6 and the legitimation under Art. 9. The guidance provided is applicable for all legal bases, but always points out where a consent legal basis under the GDPR may lead to additional requirements, a stricter regime with respect to information related to consent, scope of the consent, interpretation of what counts as "freely given" as well as in consequences of withdrawal. [20] Requirements for consent (either as a research ethics consent or as a GDPR legal basis) may also depend on national laws.[21] Some illustrative examples are provided. National advisory bodies (e.g., ethics committees) are expected to provide additional, nationally-tailored guidance. Ultimately, it is the responsibility of the organisations involved in collecting data to identify and comply with all norms applicable to their activities.

This guidance is agnostic to different collection and sequencing contexts across Europe[22], including: population databases, genomic research projects, precision medicine clinical trials, genomic medicine initiatives, as well as clinical care (such as predictive, diagnostic or confirmatory genome sequencing). The guidance is designed for any organisation who plans to make data collected in a primary context available through a repository for research projects, where the details of these projects cannot be fully identified at the time of the data collection (or even at the time of the transfer to the repository). Some practical implementation examples are provided to facilitate application of the guidelines in specific contexts.

As 1+MG has not yet determined all aspects of its organisational structure, data governance and legal framework, some key information elements have not yet been fully defined. These elements are relevant to provide transparency and to obtain a valid informed consent.[23] 1+MG is working to clarify these elements so that concrete wording or even a 1+MG specific part of the information sheet, where applicable, can be provided as an appendix in future versions of these guidelines.

1.4. Context: ethical and legal challenges

The collection/generation of genomic and related-health data and widespread use for research and healthcare raises a number of ethical issues around informed consent. This includes the risk of privacy breaches; psychological distress due to the type and amount of personal data being

[18] Whole genome sequence data and related-health data included in 1+MG will generally be treated as pseudonymised data (which is personal data).
[19] Including respect for persons, beneficence, and justice with a focus on the main ethical concerns raised by the informed consent process in the context of genomics.
[20] E.g., depending on national law or authoritative interpretations, this may include greater specificity of purposes and recipients; more details about the scope of data subject rights (especially if data are accessed by downstream controllers); and potential power imbalances between controllers (public bodies) and data subjects precluding consent.
[21] E.g., National, regional or sectoral data protection law, medical research law, health law, bioethics law, biobanking law, health research regulations.
[22] The territorial scope is primarily focused on national, regional or institutional data collections established in Member States of the European Economic Area (EEA) who are signatories to the 1+MG Initiative. It is possible that other countries outside the EEA (e.g., UK, Switzerland) are also permitted to contribute collections. Data access may be provided to researchers across the EEA and globally, under appropriate conditions.
[23] The 1+MG as recipient, the defined purposes for which data are made available and the decision making processes need to be defined for obtaining an informed consent, not just under the GDPR but also under ethics principles. See CIOMS 2016 (12); WMA 2016 (12).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

processed and shared; risks of harm if data are misused or misinterpreted; the handling of results and incidental findings that have implications for the health of participants and/or their families, including limitations of health care systems to provide adequate follow-up care; and issues of vulnerability (e.g., risk of discrimination) related to factors including cultural, ethnic, linguistic, and socio-economic considerations.[24] Ethical issues in genetic/genomic research include the risk of therapeutic misconception; the risk of misunderstanding the purpose and design of this type of research as compared to clinical trials testing medical interventions; and misunderstanding of the risk–benefit ratio.

Additional legal and ethical issues arise where genomic and health-related data are made available to broad communities of users and organisations for secondary use. Sharing sensitive and potentially identifiable genomic and related-health data raises concerns about increased risk of privacy breaches, affecting the rights and interests of data subjects and their families. Countries outside the EEA may not provide equivalent legal protections or ethics oversight mechanisms. Moreover, the specific purposes, recipients of data, and associated risks cannot be fully specified at the time of an initial consent, raising issues about transparency, and about whether the consent is sufficiently informed and specific. Even where the scope of consent is made clear and understood, there are concerns about the effectiveness of oversight and enforcement mechanisms to ensure data are only used for consented purposes. In short, transparent information is needed to enable individuals to make informed decisions about cross-border access and secondary use of genomic and related health data, combined with robust governance frameworks to ensure data are processed responsibly.

1.5. Methods

We have analysed ethics literature on practical and ethical challenges raised by the process of consent in the context of biobanks, genomic/genetic research, precision medicine, genomic medicine initiatives, and clinical care (see References). In our search in the PubMed databases and other relevant specialised journals, we have prioritised two types of publications:

- findings and recommendations based on empirical studies, where conclusions of the study were drawn from empirical evidence (e.g. qualitative and quantitative studies, such as assessment of research participants' perceptions of research based on the information provided through the consent form/notice, etc.,);
- recommendations, reports, guidelines, models of consent developed by key leaders and initiatives in the field.

We have also used conclusions from project workshops organised in the course of the B1MG project as well as the 1+MG use cases. We have also reviewed GDPR requirements and European Data Protection Board (EDPB) guidance. A sample of national data protection law implementations applicable in the research and healthcare sector were reviewed and are reported in the appendix.

2. General Guidelines – Communicating Complex Information and Documentation

Both research ethics and the GDPR require consent to be informed. For that reason, full information should be provided before consent is sought. In addition, the GDPR has requirements on transparency that apply independent of the chosen legal basis. Consequently, the affirmative action signifying consent is separated from the information provision, either as the final part of the template or – more often – as a separate form, in the following referred to as a"consent form", in particular, as there should be sufficient time foreseen between information provision and consent given. The documents provided to individuals therefore typically include an "information sheet" that clearly describes what an individual can expect when participating in (genomic) research or undergoing a genomic healthcare test or what a secondary use of health and genomic data would entail. The information

[24] See 1+MG Special Subjects Policy.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

sheet contains sufficient information for the individual to make an informed decision. It is then linked with the consent form to record the consent process and individual agreement. A general challenge with transparency and consent in this area is how to meaningfully communicate complex information. The following guidelines address this.

1. The information sheet and consent form SHOULD present the information concisely, using clear and plain language to promote comprehension.[25]
2. Where the information sheet becomes too long and complex, the structure SHOULD be layered, presenting the most important and legally required information first and providing additional information in optional sections.[26] Consider implementing e-consent tools that facilitate comprehension e.g., by including visualisations, hyperlinks, and self-directed review of information.[27]
3. All data protection related information SHOULD be easily found in one place (e.g., one section within the information sheet) and not mixed with other information such as on the research performed.[28] Some information is highly relevant from both a research ethics and a data protection perspective (such as information on purposes, data types, categories of recipients, withdrawal rights). A layered structure permits key ethics and data protection information to be prominently presented first, with references to a subsequent section covering all data protection aspects.
4. Information that is relevant only for some purposes or some data types (e.g. legal basis, international transfer, retention time) SHOULD be provided in relation to these purposes.[29]
5. The cultural, linguistic, and socio-economic context SHOULD be considered when preparing information sheets and consent forms (e.g. religious beliefs that may not accept certain types of genetic/genomic tests, potential for stigmatisation of vulnerable groups, etc.).
6. The information sheet SHOULD explain basic concepts such as genomic versus genetics, biosamples, genomic data, genetic variants, precision medicine, genomic research, biobank, data repository, coding.[30]
7. Community groups MAY be involved to ensure such issues are appropriately communicated in the documents.
8. The content of information sheets and consent forms provided to the participants may change over time. Where such changes happen, both the information sheets and the consent forms should have version numbers or other equivalent identifiers to allow referencing the other. It MUST be possible to easily establish which information was provided to the individual and what consents were obtained.
9. The consent form MUST demonstrate confirmation from individuals that they have read and understood the information sheet, and/or that the content was explained to them (e.g., as a tick box). Where different versions of the information sheet exist, the consent form MUST reference the version of the information sheet relied upon. The consent form SHOULD obtain confirmation that the individual had the opportunity to ask questions prior to signing the form.

[25] GDPR Rec 58; Art 13; Art 7 - where consent is the legal basis; EDPB 05/2020, para 60.
[26] EDPB 05/2020, para 69.
[27] GDPR Rec 58.
[28] Article 29 Working Party, Guidelines on transparency under Regulation 2016/679, paras 8, 11, 33. This includes information on the identity of the controller, data protection officer (DPO) contact, categories of data, purposes, recipients or categories of recipients, etc.
[29] EDPB 05/2020
[30] See e.g., US National Institutes of Health, "Fact Sheets"; Torpy JM, Lynm C, Glass RM. Genetics: the Basics. *JAMA.* 2008;299(11):1388. doi:10.1001/jama.299.11.1388.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

3. General Guidelines – Processing Data for Multiple Purposes

Depending on the context, transparency information and the consent form may need to cover multiple purposes, e.g.,

a. *the initial research project/healthcare test,*

b. *making data available through a repository for research (where specific research projects cannot be fully identified at the time of recruitment); and*

c. *making data available through a repository for healthcare secondary use*

In such cases:

1. The information sheet SHOULD clearly distinguish the sample and data collection, the subsequent sample and data processing[31], as well as risks and ethical issues[32] associated with each purpose for which data are processed[33].
2. The consent form MUST give individuals separate choices to agree on each purpose.[34]
3. Consider using separate consent forms or even consent processes for primary and secondary use purposes. Some overview may be needed to make the big picture clear. If a single consent form is used, consider using separate sections corresponding to the different purposes.

.

4. Information about Making Data Available through a Repository for Future Research Projects

While there may be different collection and recruitment contexts, the central focus of these guidelines is prospective recruitment contexts where the (or one of the) intentions is making data available through a repository to be accessed (cross-border) and used as part of future research projects. The key challenge here is how to satisfy transparency obligations when the details of future research projects cannot be fully identified at the time of recruitment. Considerations specific to consent and to processing for multiple purposes are addressed in later sections.

4.1. Information around the Purpose: the "What", "Why", "Who" and "How" of Data Sharing for Future Research

[31] GDPR Art 13(1)(c).

[32] This approach is consistent with the fact that study subjects will typically have the option to participate in the primary purpose while opting out of making data available in the repository. A full description of the primary purpose should include the identity and location of any partner organisations participating in any project-specific data sharing. Additional considerations for genetic research include the following: In order to avoid any misconception or confusion for the research participant as to the purpose of research and the expected clinical relevance of its results, explain what is different about a genomic research project as compared to a classical clinical trial (e.g., not to evaluate a treatment but to better understand the cause and mechanisms of the participant's condition). Indicate if the research project requires recruitment of family members (e.g., for direct collection of their genomic or health-related data), why this is necessary, and if so, how family members will be recruited (e.g., asking participants' help to identify and/or initiate contact with relatives).

[33] The primary purpose may involve different categories of risks e.g., risk of injury from physical research interventions; privacy breach from the collection, storage, use, and sharing of samples/genomic and related-health data as part of the primary research project (unauthorised access and re-identification, leading to potential discrimination, stigma, or worry). As genetic/genomic information might contain health information about biologically related family members, the privacy risks for members of the family should be explained. In terms of psychological risks, genetic/genomic data may reveal information about possible family relationships, including non-paternity; indicate that some individuals would find this information distressing.

[34] GDPR Art 7(2)(b) - where consent is the legal basis. CIOMS 2016 Guidelines.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. The information sheet MUST state that genomic and related-health data will be **made available through a repository** for health and biomedical research.[35]
2. The data protection section MUST mention the **identity of the controller** (the initial data collecting organisation) and the contact information of its data protection officer (DPO) [where applicable] responsible for the transfer to the repository.[36]
3. The data protection section MUST mention the **legal basis** for the controller's processing (all or part) of the individual's personal data to make it available through a repository for future research projects.[37] This purpose/legal basis MUST be distinguished from other purposes/legal bases.[38] The legal basis should also mention on which basis the processing of genomic and health data is legitimised (i.e. consent or the national / EU legislation).[39]
4. Given that the details of future research projects cannot be fully identified in the information sheet at the time of recruitment, it is justified to use indefinite language (e.g., may) where necessary, as long as this does not undermine the fundamental rights of data subjects.
    a. The information sheet SHOULD describe the scope of research for which the data will be used to answer in the future in a general way, as well as the processing operations already known at the time of consent, such as data transfers, curation, storage in a repository, and making data available to researchers (as outlined in the repository/biobank plan). [MUST where consent is the lawful basis].
    b. The information sheet SHOULD specify a **well-described area of research** (e.g., precision medicine research to better understand, prevent, and treat disease).[40] [MUST where consent is the lawful basis].
5. The information sheet SHOULD explain the reasons for data sharing and the **expected benefit for the society**, (e.g., to enable qualified researchers across the EEA and around the world to collaborate, check each other's results, and ask new questions, which can accelerate research, helping us to better understand and address disease).
6. The information sheet MUST describe **how access is provided** (e.g., in a secure processing environment) and what the **governance mechanisms** are (e.g., review and approval by a data access committee). Please note that the 1+MG governance mechanisms are still under development and recommendations for clauses can only be provided at a later stage. *Note: this requirement cannot be fulfilled without more details on the 1+MG infrastructure and governance.*
7. The information sheet MUST identify the **controller for the access provision** and specify the l**egal basis**.
8. The information sheet SHOULD explain that general research results of future studies will be **published in academic journals** and presented at conferences.

4.2. Information on Recipients

[35] GDPR Art 13(1)(c); Art 6(1)(a) - where consent is the legal basis.
[36] GDPR Art 13(1)(a),(b). For the DPO contact information, a general DPO office is preferable to an individual who may change. To avoid confusion, it is preferable to highlight that the DPO is generally responsible for matters related to data protection, as opposed to scientific aspects of ongoing studies.
[37] GDPR Art13(1)(c).
[38] E.g., legitimate interests (Article 6(1)(f)) for archiving part of the data needed to establish audit trail and ensure reproducibility of research studies already conducted; Compliance with the law (Article 6(1)(c)) or Vital Interest (Article 6(1)(d)), when reporting of certain health-related findings are foreseen; etc.
[39] Art. 29 Working Party, Guidelines on transparency under Regulation 2016/679
[40] CIOMS 2016. Where consent is the legal basis - GDPR Rec 33; EDPB 05/2020, paras 155-160.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. The information sheet MUST state that data will be transferred to and **hosted in a data repository**. The name and location of the entity responsible for providing access to data for the future research projects MUST be stated (e.g., the 1+MG legal entity, if applicable).[41] *Note: this and the following requirements cannot be fulfilled without more details on the 1+MG infrastructure and governance.*
2. The information sheet MUST provide the **repository's full contact details** (legal representative; DPO) and legal basis for providing access to future users if this is not done in a separate information sheet by the repository made available in parallel.
3. Where the repository is based in an **international organisation** or a **third country**, the information sheet MUST mention the existence or absence of an adequacy decision by the Commission or reference the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.
4. The information sheet MUST state future potential **categories of recipients** that data may be shared with through the repository, e.g. *bona fide* researchers at external research organisations including academic research institutions, healthcare institutions, pharmaceutical companies, bioinformatics and health technology companies as users or IT platforms as services providers.[42]
    a. The information sheet MUST NOT limit sharing for secondary use to specific recipients only (e.g., partner organisations).[43]
    b. The information sheet SHOULD mention that users may base their processing on a different legal basis and/or may be subject to a different applicable (national) data protection law. This may curtail the scope of data subject rights (as explained below in the context of each right).
5. The information sheet MUST NOT limit the **geographical location** of recipients to certain countries within the EEA.[44]
6. The information sheet MUST state whether or not there is an intention to share data with **users or IT service providers based in third countries outside the EEA**.[45] Where the data are made available to entities outside the EEA, foreseeable transfer mechanism(s) MUST be mentioned (e.g., adequacy decision, appropriate safeguard). The data protection section MUST indicate where more specific information on the nature of the transfer mechanism can be obtained.[46] Failure to mention the possibility of data use outside the EEA may restrict data sharing to users within the EEA only.

4.3. Categories of Samples and Data

[41] Where data will be transferred to an external organisation who will act as a (joint) controller for storing and making data available for research and who plans to rely on consent as the lawful basis for this processing, the identity of the controller MUST be stated in the information sheet and consent form and the legal basis for access provision specified.
[42] The 1+MG Data Governance Policy requires that access to data for secondary use be provided on a non-discriminatory basis. This recommendation is driven by the general ethical and legal principle of non-discrimination applied to data sharing, as well as the principle of maximising the benefit of research participants' contributions. Especially in the case of publicly funded sequencing and the use of publicly funded infrastructure, access should be made available to any qualified and trustworthy researcher able to advance science. Data protection compliance processes should not be designed to lead to the de facto proprietary treatment of data. Practically speaking, necessary recipients cannot reasonably be predicted in advance for precision medicine research. For public bodies who will be subject to the proposed *Data Governance Act*, non-discriminatory access will be a legal requirement (Art 5).
[43] If you plan to rely on consent as a legal basis, however, ensure your national law or regulations do not require the identity of all recipients to be specified.
[44] Note there is no legal requirement to mention cross-border access within the EEA, though there is nothing wrong with explicitly stating this.
[45] The 1+MG does not foresee enabling access to researchers in third countries outside the EEA without equivalent privacy protections, as derogations (e.g., consent) are exceptional (GDPR Art 49).
[46] GDPR Art 13(1)(f).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. The information sheet MUST describe the **types of data** that will be made available for secondary use (e.g., demographic data, clinical data, genetic data, in case of 1+MG, whole genome, family health history, lifestyle, mobile health data etc.) either by default or additionally on request.[47]
2. The information sheet MUST also describe the **categories of biological samples** that may be made available through a sample repository/biobank (e.g., blood, saliva, tumour tissue), either by default or additionally on request. The information sheet MUST explain that these samples may be used to create additional molecular data.[48]
3. The information sheet MUST clearly describe the categories of existing biological samples and personal data that will be obtained from **existing sources** and explain what those sources are (e.g., samples accessed from existing biobanks, data collected in the context of medical care, and linkage with electronic health records) and specify the **categories of Data/Sample Providers** (e.g., medical centres, government databases). Clarify in particular if linkage to electronic health records will be obtained periodically, and if so, over what period of time.[49]

4.4. Duration

1. The information sheet MUST mention for **how long** the data (and, where applicable) samples will be made available through the data repository for future research. This can be either an absolute timeframe or criteria for how long data will be kept (e.g. relevance for future research). In the latter case, the information sheet SHOULD also provide information how these criteria will be monitored. In addition, a possible time horizon SHOULD be given.
2. The information sheet MUST mention if different timeframes and/or criteria for archiving apply that may require storage beyond the active availability.
3. The information sheet MUST mention what happens to the data at the end of the retention period (i.e. erasure, anonymisation).

4.5. Re-contacting Data Subjects

There are a number of reasons why the data subjects may be re-contacted. Re-contacting in the context of the data reuse is for the interests of the entity providing data to the users' research projects, i.e. 1+MG, as well as the data users themselves. However, contacting channels may be defined on a national or local level and could involve the Data Provider. It is important that the data subject has the possibility to consent, opt-in or opt-out in an informed way to the data transfer to the 1+MG repository but also that if, optional processing in the 1+MG repository might result in re-contacting, a separate informed decision can be taken by the data subject. Therefore, the information sheet MUST explain when and why re-contact may happen.

4.5.1. Re-contacting Related to a Specific Research Reuse

Transparency requirements of the GDPR mean that data subjects must be informed about any purpose that their data are processed for. In the case of data sharing for secondary use, the entity providing data to users as well as the users themselves have information obligations towards the data subject about the specific purpose that the data are used for. This information is essential for data subjects to be able to object to the processing in time. Where the data were not obtained from the data subject directly, Art. 14(5)(b) GDPR permits alternative methods if provision of information directly to the data subject involves a disproportionate effort. However, where re-contacting is foreseen also for scientific reasons, no disproportionate effort may be justifiable as mechanisms and

[47] GDPR Art 14(1)(d), GDPR Recital 61, Council of Europe, Convention 108, Art. 8(1)(c)
[48] GDPR Art 14(1)(d).
[49] GDPR Art 14(1)(d).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

procedures for contacting data subjects are in place. In such cases, 1+MG will provide information on each new purpose following an access decision. Different channels and frequencies / time points to provide this information MAY be possible and SHOULD then be described as well as a choice given between different options where possible.

1. The information sheet MUST explain how detailed **information about future research projects** will be offered/made available to the data subject (e.g., direct communication; website, and/or on request). Where direct communication is possible, the data subject MUST be given a possibility to be informed about each new access in advance. The data subject SHOULD also be able to choose only periodic information or passive information options where information can be found on demand.[50] In case of passive information or aggregated information on data use, the information sheet MUST warn the data subjects that an objection may not be possible in all cases once the research has started.
Note: this requirement cannot be fulfilled without more details on the 1+MG infrastructure and governance.
2. The information sheet MUST (where applicable) mention the possibility of future contact to **collect additional samples/data** (e.g., as part of future research projects) or to participate in clinical trials.
3. The information sheet SHOULD (where applicable) mention the possibility of future contact to **seek renewed consent** (e.g., where necessary because of a substantial change in the scope of research aims supported by the repository or where technologies for analysing samples/data substantially change in an unanticipated and material way).[51]

4.5.2. Re-contacting to Return of Findings of Individual Health Relevance

Organisations who collect and/or generate genomic and related health data are normally required to have a plan in place for handling different kinds of findings with health relevance for individuals or their relatives as part of the primary purpose (e.g., research project/healthcare test). Findings of individual health relevance may include individual research results linked to the aims of a research project, or incidental findings outside the aims of a research project/healthcare test. For recommendations about how to handle findings of individual health relevance, see the 1+MG Incidental Findings Policy. In case of secondary use, it should be clarified to which extent the same policy will also be applied or if incidental findings may be handled differently. In 1+MG, only clinically actionable findings will be reported. This needs to be considered and may deviate from policies applicable to the primary purpose. Here, best practice suggests that these policies SHOULD already be described in the information sheet for the primary purpose of data collection or processing, or, if not applicable there, or if the policy is different to the proposed processing in the 1+MG, then it SHOULD be described in the 1+MG information sheet. These consent recommendations should also be sufficient to cover the handling of incidental findings from secondary use.

The information sheet MUST state if the same incidental finding policy as for the primary purpose will apply. Where this is not the case, the elements MUST be addressed where they differ from the original policy:

1. Explain whether or not findings of individual health relevance found through the secondary use will be reported to participants and/or their families.

[50] EDPB 05/2020 para 161 "A lack of purpose specification may be offset by information on the development of the purpose being provided regularly by controllers as the research project progresses…". Consider if national law or authoritative guidance requires direct notification of data subjects in advance of processing for any specific research project, i.e., in advance of granting access (Art 13(3)).
[51] CIOMS 2016

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

2. Explain where applicable if these findings will be limited to individual results related to the primary aims of sequencing or incidental findings beyond the primary aims of sequencing.[52]
3. Explain the conditions under which findings of individual health relevance will be reported to participants and/or their family members e.g.,
    a. the level of clinical significance (e.g., life-saving, clinically actionable),
    b. the level of validation (e.g., in an approved genetic laboratory), and
    c. the time period (e.g., data use in 1+MG could go over decades).
4. Offer participants the choice not to receive findings of individual health relevance, and explain any situations where preferences may be overridden by professional obligations (e.g., where there is a legal duty to warn family members of life-threatening conditions).
5. Describe the procedure for how findings of individual health relevance may be returned (e.g., reporting through a designated medical professional). Note that procedures in 1+MG may indeed be different to procedures for incidental findings in the original collection context.
6. Explain if findings with shared health implications for biological relatives will be reported to them, and under what conditions (e.g., only with the participant's consent, potential obligation of the participant to inform relatives, or only after the participant's death).
7. To avoid therapeutic misconception, reiterate that the possibility of receiving individual findings of health relevance should not be equated with diagnostic testing or screening.

4.6. General Communication on Research Results and other Information

1. The information sheet SHOULD [where personal communication such as email is used: MUST] explain how the general research results of future studies will be communicated to participants (e.g., list of publications on the data repository website, subscription newsletter, or upon request).
2. Where applicable, data subjects MUST be given a choice if they want to receive newsletters about general research results and other matters related to the 1+MG initiative. In this case, the legal basis for the communication MUST be mentioned.[53] If legitimate interest is chosen as the legal basis for the processing of personal data in the context of information provision, the interest MUST be explained in the information sheet.

Note that communication channels to inform about 1+MG and general research results are still subject to development and a final decision in 1+MG.

4.7. Voluntary Participation and Right of Withdrawal of Consent or Objection to the Processing

1. Individuals MUST be informed that making their data available through the 1+MG repository is entirely voluntary, and that they may discontinue participation at any time, without any penalty or disadvantage by objecting to the processing [where opt-out is the

[52] Note our recommendations for a 1+MG Incidental Findings policy consider all findings of individual health relevance in secondary use contexts as "incidental findings", and that any return of such findings through 1+MG should respect the initial plan and consent established at the time the data were collected.

[53] Directive 2002/58/EC (ePrivacy Directive) only requires consent to unsolicited emails for marketing in Art. 13. The information on 1+MG is not marketing any products or services. Therefore, a consent is not necessarily required. However, if there is no other legal basis under Art. 6 GDPR, GDPR consent rather than an opt-in or opt-out will have to be sought.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

basis of inclusion] or withdrawal of consent [where the legal basis is consent or where an opt-in is obtained].[54]

2. [Where consent is the legal basis for the sharing with and/or from the repository] The information sheet MUST explain that a withdrawal of consent means that the data will not be hosted anymore for future research but that it does not affect the lawfulness of the data sharing that has happened already.[55]
3. The information sheet MUST provide instructions on how to withdraw from participation in the 1+MG repository, to withdraw consent or object to individual purposes or otherwise exercise rights (e.g., contact the research team).[56]
4. The information sheet or information provided at the time of withdrawal MUST offer separate choices to withdraw[57] from (where applicable):
    a. individual research projects;
    b. future active participation/ provision of data (e.g., continuing to undergo physical procedures, site visits, providing longitudinal survey or mobile health data as part of a Cohort study);
    c. future linkage to electronic health records (e.g., as part of a Cohort study);
    d. contribution of data types not required as part of the minimum dataset of the repository;
    e. the continued storage, sharing and use of already collected data and/or biological samples in a repository/biobank; and
    f. contact in the future (per purpose as specified in the consent form).
5. The information sheet SHOULD indicate what would happen to data and/or biological samples should they withdraw from the repository (e.g., samples/data will be destroyed and/or anonymised).
6. The information sheet MUST explain and justify limitations on the right to withdraw from research projects for which data were made available by the research repository, namely where this would jeopardise the integrity of ongoing or archived research projects:
    a. It may not be possible to withdraw data that is already being accessed as part of an ongoing research project where this is not feasible for practical reasons or would seriously impair the research.[58]
    b. it will not be possible to withdraw data from a completed research analysis.
    c. it will not be possible to withdraw data archived after completion of a research project for the duration of the archiving period, to ensure the integrity of completed research projects.

4.8. Other Applicable Rights

[54] GDPR Art 7(3) - where consent is the legal basis. CIOMS 2016 (2). To respect this requirement, any withdrawal process must separate choices to withdraw from the primary purpose, the purpose of precision medicine research, and other purposes (e.g., secondary healthcare use).
[55] GDPR Art 13(2)(d), Art 14(2)(d).
[56] EDPB 05/2020; GDPR Rec 39.
[57] GDPR Rec 43.
[58] The GDPR foresees under Art. 21 that an objection to the processing by the data subject may not have to be followed where the interest of the controller (or indeed a public interest as foreseen in Art. 21.6 prevails. Also Art. 17. 3 foresees that the request for erasure does not have to be followed where this jeopardises the achievement of research processing or for reasons of public health. Even if consent is the legal basis for collection and storage, and for the collecting organisation's research projects, the research organisations accessing pseudonymised data will typically not rely on the same legal basis and thus the right to withdraw consent to processing will not apply.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

1. The information sheet MUST indicate **how** data subjects can exercise their rights (e.g. contacting the research team or through a dedicated tool).
2. [Where there is an **intention to destroy the pseudonymisation table**] The data protection section SHOULD mention that it may not be possible to exercise data subject rights where the link back to the individual's identity is no longer retained.[59]
3. **Right to rectification.** The information sheet MUST explain that data subjects have the right to request correction of inaccurate personal data and the completion of incomplete personal data, but that this right may be limited where necessary to achieve research purposes.[60]
4. **Right of access**
    a. Access to information. The information sheet MUST explain that data subjects have the right to access information on how their personal data is being used and shared, including the purposes of approved, ongoing, or completed research projects, the categories of personal data and the identities of the recipients involved (e.g., names of the institutions and/or principal investigators), and the existence of any international transfers and the associated legal mechanism, etc.. Explain that this information will be available on request, and may be available through additional means (e.g., website, by subscription to a periodic newsletter, regular notifications about new projects). Where applicable, explain if national or EU legislation, the right to access is limited for public sector bodies.
    b. Access to a copy of personal data. Explain that data subjects have the right to access a copy of their personal data for free[61], but (where applicable) that this right may be limited where necessary to achieve research purposes[62], where other laws restrict communicating genetic or health data (e.g., genetic testing laws) or where the data may interfere with the privacy of others (e.g., genetic information of relatives). Mention that information provision on repeated requests may be charged for, if charging is considered in such cases. When assessing the potential applicability of these limitations, it MAY be helpful to distinguish between access to the following types of personal data:
        i. Data directly provided by the data subject. This includes: contact information, past medical history details, answers to questionnaires, measurements done directly on the patient.[63]
        ii. Data generated by virtue of the data subject's participation in research or undergoing the initial medical procedure. This includes raw genomic sequence data and any other data derived from biosamples.[64] Consider mentioning available formats for raw whole genome sequence data include e.g., BAM, VCF.
        iii. Data inferred by analysing the raw data (individual research results).[65]
5. **[Where consent is the legal basis for the data hosting in the repository] Right to data portability**. The data information sheet MUST explain that the individual has a right to have the personal information *they have provided directly* in a structured, commonly

[59] GDPR Art 11.
[60] Art 89(2) derogation where provided by Member State law (in one or more countries).
[61] Art 15(3).
[62] Art 89(2) provides the possibility of a derogation where provided by Member State law. Consider if the limit relating to risks to third parties applies, though this seems unlikely in this context (Art 15(3)).
[63] This data MUST always be made accessible when consent is used as the legal basis. (Art 20 GDPR).
[64] In some cases, there may be national laws in place prohibiting the disclosure of genomic data outside the context of genetic counselling. The consent form must clarify if such additional conditions apply.
[65] This includes the analysis and interpretation of raw genomic data. (Providing access to this type of data is optional regardless of the legal basis, and should be based on the data controller's policy on the return of health-related findings).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

used and machine-readable format to another organisation without hindrance. [66] Data provided directly includes those listed under (i) above, but does not typically include data generated from samples. Mention that the right does not apply beyond the data directly provided by the data subject, to the original controller.

6. The information sheet MUST explain that data subjects have the right to **lodge a formal complaint** with the competent Data Protection Authority (DPA); specify the relevant DPA the data subjects can contact for this matter.

4.9. Risks

1. The information sheet MUST explain that sharing data with researchers from external research organisations may increase the **risk of privacy breaches**, especially considering there is always some risk of being re-identified from genomic and related health information.[67]
2. The information sheet MUST explain the potential **consequences** of such risks.
3. The information sheet SHOULD explain that ongoing progress in science and technology such as artificial intelligence makes it possible to perform **unanticipated forms of research** that may turn out to be controversial.
4. The information sheet SHOULD mention the potential risks of sharing data with **recipients in third countries or international organisations** with potentially lower privacy protections [where applicable]. [This is a MUST where consent is the legal basis.][68]

4.10. Safeguards

1. The information sheet MUST describe in general terms the kinds of **safeguards** that will be adopted to protect personal data and/or biological samples (without being too specific so as to limit changes in the future).[69] E.g.,
    a. data pseudonymisation, meaning all direct identifiers (such as your name, address, data of birth, ID number) will be stored separately and replaced with a unique identifier making it hard to trace the information back to the subject.
    b. only controlled and managed access to secure data repositories is given to qualified researchers, with access being subject to monitoring and auditing.
    c. data access/use agreements for accessing parties limiting their use to pre-approved purposes (e.g., specific studies) and requiring them to refrain from deliberately identifying individuals.
    d. access to data and research projects will be subject to appropriate oversight by a data access committee and/or a competent research ethics committee.[70]
2. The information sheet MAY inform where/how more **precise descriptions of safeguards** and/or the data protection impact assessment can be found.[71]
3. The information sheet MUST mention limitations of safeguards and remaining risks.[72]

[66] Art 20 GDPR.
[67] GDPR Recital (39).
[68] GDPR Art. 49(1)(a).
[69] GDPR Recital 39.
[70] CIOMS 2016 Guidelines.
[71] WP29 Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.
[72] CIOMS 2016 Guidelines; GDPR Recital 39.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

4. Where data are transferred to **third countries or international organisations**, the information sheet MUST reference the appropriate or suitable safeguards and how to obtain a copy of them or where they have been made available.[73]

4.11. Benefits and Commercialisation [ethical requirements]

1. The information sheet SHOULD explain that the main aim of research is **scientific progress**, i.e., to advance our understanding of disease and that the research may range from basic research to applied research. Ultimately, this could lead to new approaches and products to prevent, diagnose and/or treat people with a similar condition.
2. The information sheet SHOULD explain, where applicable, that research aims at general knowledge and that inclusion in the repository is **not likely to benefit the individual and/or their family members directly**. In some rare disease cases, diagnosis of individuals and families may be a clear goal and foreseeable outcome for the inclusion in the repository.[74]
3. The information sheet SHOULD explain that research results may lead to **commercial products**, and that the individual will not have any monetary rights in these products.[75] It may be helpful to provide illustrative examples, such as drugs, clinical decision support systems, etc.

5. Consent to Making Data Available through the 1+MG Repository for Research

1. For data that are obtained in a **research context**, an **opt-in** as informed consent MUST be obtained unless a waiver is given by a competent ethics committee. For data collected in the healthcare context, at minimum an opt-out MUST be offered with respect to transferring data for storage and reuse through the 1+MG Repository. The separate consent MUST repeat key elements of the information as described below.

1. Where the consent form is a separate document from the information sheet, the consent form SHOULD mention that information has been received and refer to the **versioning of the information sheet** where applicable.
2. Consent as opt-in MUST be connected to an **action by the data subject** such as a tick box.
3. The consent form MUST **repeat key elements of the information** material to an informed decision including: the **controller** to whom consent is given for the transfer; the aim of **making data available in a repository** for future research projects in a well-described area, **hosting of data** in a particular repository, **potential access** by researchers at external research organisations, the **types of data** made available (e.g., health and WGS), and the s**cope of potential transfers to third countries**. These considerations are of particular relevance where consent is the legal basis for processing, but may also be relevant where an ethics consent is obtained.
4. The consent form SHOULD consider giving individuals the option to **limit their consent to a narrower area of research** where this choice is likely to allow certain members of the recruitment population to respect important personal preferences (i.e., disease-related research limited to their own disease group).[76]

[73] GDPR Art. 13(1)(f).
[74] CIOMS 2016 Guidelines (11).
[75] CIOMS 2016 Guidelines (11).
[76] If such choices are given, these categories SHOULD be aligned with a controlled vocabulary of administrative (rights) metadata across 1+MG, to ensure both that individual consent and choices can be tracked and enforced, and also that data can be meaningfully integrated and re-used. See e.g. GA4GH Machine-readable Consent Guidance, coding the GA4GH Data Use Ontology.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

5. [Where consent is the legal basis] The consent form MUST mention the **possible risks of transfer** to the repository if it is based in a third country or an international organisation and there are no appropriate safeguards as described in Art. 46 GDPR or adequacy decision in place.[77]
6. [Where consent is the legal basis] The consent form MUST mention that the data subject has the right to **withdraw consent** at any time.[78]
7. Concerning information about specific data use (sharing for defined projects) according to Art. 14 GDPR: ideally data subjects could be given a **choice how** they want to receive information: active messages in each case, regular summaries or passive information only. The possibilities for such a choice will depend on the mechanisms and tools used and may even be specified in such a tool directly. The information provided will therefore have to be aligned with the 1+MG repository.
   Where such choice is applicable, data subjects SHOULD be able to select the preferred way of communication and/or (independently) the timing. A warning SHOULD be added that after the commencement of a research project, an objection may not always be possible.
8. The consent form MUST (where applicable) offer **separate choices for contacting** the individual in the future:
   a. for the purpose of reporting **findings** of individual health relevance;
   b. for the purpose of requesting **additional data** and/or biosamples;
   c. for the purpose of **recruitment into future studies** requiring active participation (e.g., precision clinical trials);
   d. for the purpose of providing **general information** about the repository and the general research results achieved with the data.

References

Regulatory Guidance

EDPB, Guidelines 05/2020 on consent under Regulation 2016/679 at https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

Guidelines and Reports

CIOMS 2016: International Ethical Guidelines for Health-related Research Involving Humans, Fourth Edition. Geneva. Council for International Organizations of Medical Sciences (CIOMS); 2016, at https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/.

WMA 2016: WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks, Adopted by the 53rd WMA General Assembly, Washington, DC, USA, October 2002 and revised by the 67th WMA General Assembly, Taipei, Taiwan, October 2016, at: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

ClinGen: Riggs ER, Azzariti DR, Niehaus A, Goehringer SR, Ramos EM, Rodriguez LL, Knoppers B, Rehm HL, Martin CL; Clinical Genome Resource Education Working Group. Development of a consent resource for genomic data sharing in the clinical setting. Genet Med. 2019 Jan;21(1):81-88. doi: 10.1038/s41436-018-0017-5. Epub 2018 Jun 13. PMID: 29899502; PMCID: PMC6292744.

[77] GDPR Art. 49(1)(a); EDPB Guidelines 04/2020.
[78] GDPR Art. 7(3); EDPB Guidelines 04/2020.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Global Alliance for Genomics and Health's (GA4GH's), Consent Clauses for Genomic Research and Familial Consent Clauses, at https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/

Global Alliance for Genomics and Health: Consent Policy, Sep. 2019, at https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf

Colin Mitchell, Johan Ordish, Emma Johnson, Tanya Brigden and Alison Hall, PHG Foundation, Report The GDPR and genomic data, May 2020, https://www.phgfoundation.org/report/the-gdpr-and-genomic-data.

National Institutes of Health, Special Considerations for Genomics Research https://www.genome.gov/about-genomics/policy-issues/Informed-Consent-for-Genomics-Research/Special-Considerations-for-Genome-Research#families

Examples

FINGEN, Information on the processing of personal data in the FinnGen study, at https://www.finngen.fi/en/data_protection/data-protection-statement.

Genomics ENGLAND, Frequently asked questions about Ethics and Consent at https://www.genomicsengland.co.uk/understanding-genomics/data/ethics-and-consent-faqs/.

Literature

Beskow LM, Weinfurt KP. Exploring Understanding of "Understanding": The Paradigm Case of Biobank Consent Comprehension. Am J Bioeth. 2019;19(5):6-18. doi:10.1080/15265161.2019.1587031.

Cervo S, Rovina J, Talamini R, et al. An effective multi source informed consent procedure for research and clinical practice: an observational study of patient understanding and awareness of their roles as research stakeholders in a cancer biobank. BMC Med Ethics. 2013;14:30. Published 2013 Jul 30. doi:10.1186/1472-6939-14-30.

D'Abramo, F., Schildmann, J. & Vollmann, J. Research participants' perceptions and views on consent for biobank research: a review of empirical data and ethical analysis. BMC Med Ethics 16, 60 (2015). https://doi.org/10.1186/s12910-015-0053-5.

Gaille, M., Horn, R. & The UK-FR GENE (Genetics and Ethics Network) Consortia. The ethics of genomic medicine: redefining values and norms in the UK and France. Eur J Hum Genet 29, 780–788 (2021). https://doi.org/10.1038/s41431-020-00798-2

Halverson CM, Ross LF. Incidental findings of therapeutic misconception in biobank-based research. Genet Med. 2012;14(6):611-615. doi:10.1038/gim.2011.50

Horgan D, Romao M, Hastings R: Pulling the Strands Together: MEGA Steps to Drive European Genomics and Personalised Medicine. Biomed Hub 2017;2(suppl 1):1-11. doi: 10.1159/000481300.

Lidz CW, Appelbaum PS. The therapeutic misconception: problems and solutions. Med Care. 2002;40(9 suppl):V55–63.

Medical Informatics Initiative, Consent Working Group Patient Consent Form Template (English translation 10 Nov 2020). https://www.medizininformatik-initiative.de/sites/default/files/2020-11/MII_WG-Consent_Patient-Consent-Form_v1.6d_engl-version.pdf

Nobile H, Eric Vermeulen, Kristof Thys, Manuela M Bergmann & Pascal Borry (2013) Why do participants enrol in population biobank studies? A systematic literature review, Expert Review of Molecular Diagnostics, 13:1, 35-47, DOI: 10.1586/erm.12.116

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Rotimi, C.N., Marshall, P.A. Tailoring the process of informed consent in genetic and genomic research. Genome Med 2, 20 (2010). https://doi.org/10.1186/gm14; See also Marshall PA: The individual and the community in international genetic research. J Clin Ethics. 2004, 15: 76-86.

Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, Klovins J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. J Epidemiol. 2018 Aug 5;28(8):353-360. doi: 10.2188/jea.JE20170079. Epub 2018 Mar 24. PMID: 29576601; PMCID: PMC6048300.

Thorogood, A., Dalpé, G. & Knoppers, B. Return of individual genomic research results: are laws and policies keeping step?. Eur J Hum Genet 27, 535–546 (2019). https://doi.org/10.1038/s41431-018-0311-3

Tindana P, Depuur C, de Vries J, Seeley J, Parker M. Informed consent in genomic research and biobanking: taking feedback of findings seriously. Glob Bioeth. 2020;31(1):200-215. Published 2020 Feb 23. doi:10.1080/11287462.2020.1717896.

7. Recommendation on practical approach to the management of the generated intellectual property (IP) rights

7 June 2023

Version 1,0

Authors: Olga Tzortzatou, Alexandra Ziaka, Marina Makri

1. Executive Summary

- This deliverable aims to establish a practical approach to the management of the generated intellectual property (IP) rights that emanate from cross-border access and use of personalised medicine data in a pan-European genome initiative.
- It navigates the different IP rights that arise in the context of a pan-European genome initiative, including the copyright on data, patent on inventions and trade secret protection.
- It also critically assesses the Open Innovation scheme, presenting the pros and cons of adopting such an approach.

The deliverable presents a checklist with all the information on IP rights that should be included in data transfer agreements, facilitating researchers who are involved in cross-border research projects.

- It is necessary to reconcile IP rights as a means to encourage research with the public interest which is served through advancing innovation. This could be achieved through the adoption of appropriate governance and contractual access arrangements.

2. Methods

2.1 Deliverable scope

The report is part of Task 2.4 of Working Package (WP) 2, which focuses on the development of a Data Access and Use Governance Toolkit Framework. Such a framework oversees the data

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

linking and data management and checks the compliance with ethical and data protection requirements while it considers the responsibilities of different stakeholders. To achieve this, it is necessary to identify the critical elements for an efficient and transparent governance that allows to set up a digital infrastructure that will enable the cross-border linking of genomic and other health data for research in Europe. In this report, Task 2.4.4 will be addressed. The aim of Task 2.4.4 is to identify a practical approach towards the management of the generated IP rights that emanate from the cross-border access and use of personalised medicine data in a pan-European initiative. This practical approach should be aligned with WP5: Delivering Personalised Medicine cross-borders: Implementation in Healthcare systems and Societal Impact delivery.

It is necessary to align the findings of Task 2.4 with WP1: Framework for Cooperation through Stakeholders Engagement, Awareness & Alignment and WP6: Coordination Office: Project Management, Communication, Governance and Sustainability in order to enable the discussion with stakeholders and consequently the adoption of such a data governance framework. The efficiency of data governance is conditional on the adoption of technical solutions. Therefore, interaction with WP4: Secure Cross-border Technical Infrastructure is essential and should be secured.

2.2 Methodology

The report is based on a mixed method, consisting of findings of the paper "Ethical, Legal and Social Implications in Research Biobanking: A Checklist for Navigating Complexity" which is currently under publication at Developing World Bioethics (Annex 1) and literature review of relevant papers, reports, and international legal instruments. In particular, the outcomes of the paper emerged from a research focusing on biobanking in Africa, which was funded by both the "Beyond One Million Genomes" (B1MG) project and the "B3Africa" Project. The findings of the research enabled the creation of a four-step checklist, which reflects the requirements that researchers should fulfil to ensure the Ethical, Legal and Social (ELSI) compliance of their research project. The research was mainly focused on the African continent. Nevertheless, the paper presents a comprehensive overview that transcends Africa and can be applied in various research settings. The findings of the paper have been complemented by subsequent literature review on IP rights in the context of cross-border access and use of personalised medicine data.

3. Description of work accomplished

3.1 Introduction

Personalised medicine is the treatment that is tailored to individuals or specific groups of patients, who are categorised based on different stratifies, such as genetic variants (European Observatory on Health Systems and Policies, 2020). Despite the absence of a universally accepted definition, the European Commission refers to the Horizon 2020 Advisory Group, according to which, personalised medicine is 'a medical model using characterisation of individuals' phenotypes and genotypes (e.g. molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention' (European Commission, 2023; Council, 2015). It brings together several disciplines aiming to generate new knowledge and address prevalent or rare diseases that threaten human health (Cesario et al, 2022). Its long-term goals include enhancing healthcare, enabling research, discovering therapies and diagnostics, and predicting the predispositions that individuals may have to certain diseases or conditions (Knowles et al, 2017). Personalised medicine has been characterised as "P4 medicine", reflecting its four main characteristics, namely, personalised, preventive, predictive and participatory (Sobradillo et al, 2011). The European Association for Predictive, Preventive & Personalised Medicine is considered to be the key player on personalised medicine on EU level (Kinkorová, 2016).

Scholarship has suggested that precision medicine is a more precise term, as it reflects the development of medicine per category of disease, which is more feasible through the profiling of different categories of patients who appear to have common specific characteristics. More

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

intense use of biomarkers as well as companion diagnostics can play a significant role in shifting from empirical medicine to precision medicine (Seyhan and Carini, 2019).

Personalised medicine is heavily benefiting from digital health and care, which encompasses the 'tools and services that use information and communication technologies (ICTs) to improve prevention, diagnosis, treatment, monitoring and management of health and lifestyle' (European Commission, 2023). Applications of digital health, such as medical mobile apps, software, wearable devices, as well as different AI applications enable the collection of patients' data which further inform the research and, ultimately the development of new treatments (Wong and Breyer Menon, 2022).

According to a study, the biggest challenge to the development of personalised medicine is regulatory uncertainty (Knowles et al, 2017). The lack of regulatory harmonisation is a concerning issue that creates additional barriers that are not always easy to overcome. As we will see below, personalised medicine is heavily dependent on the availability of data. Therefore, the lack of adequate data as well as the lack of optimal use of data, also pose challenges to personalised medicine (Rajam, 2020). Another issue, which requires particular attention, is the digital literacy of healthcare providers and researchers, since their involvement is fundamental in the realisation of personalised medicine goals (Vicente et al, 2019). Other challenges include the integration of Big Data and ICT solutions, the translation of Basic to Clinical Research, the introduction of innovation in the market and the shaping of sustainable healthcare (PerMed, 2015).

One should not forget though that the divide between Global North and Global South countries is still obvious also in this sector. Although the rise of technical uses of genetic resources, such as genome sequencing, gene editing, and computational biology, is expected to have a huge impact across the world, we should not turn a blind eye to the fact that this positive development is only limited to a close circle of elite, high-tech actors, companies, and universities (Atsali, 2020). It has been suggested that the Open Innovation scheme, which promotes data and knowledge transfer, could potentially help break this divide (Joly, 2011).

IP rights constitute a necessary protective mechanism, which accommodates the dynamic nature of life sciences and the high risk investments that it takes to develop new technologies and medicines (Matthews and Zech, 2017). It is an incentive encouraging investments and research. On the other hand though, the outcomes of life sciences can be lifesaving, pointing that way to the fundamental question on how to ensure a balance between the competing interests of IP protection of inventors and the right to health of individuals and society at large.

3.2 Importance of data exchange

As previously said, one of the fundamental drivers of personalised medicine is availability of data. The rise of -omics and the exponential development of Big Data technologies have resulted in the accumulation of huge troves of data about individuals and whole populations, opening the door to personalised medicine (European Observatory on Health Systems and Policies, 2020). The development of pharmacogenomics has led to large-scale investments on building new infrastructures and setting standards that will enable personalised medicine (Song, 2017). The use of data, such as genetics, epigenetics, genomics and molecular profiling data and data on drugs, enables the faster identification of candidate drugs (Seyhan and Carini, 2019). According to Seyhan and Carini (2019), the ideal case scenario would be if researchers had access to data from the entire medical system. Then they would be able to use these data to identify patterns using Artificial Intelligence (AI) and end up benefiting patients by making safer predictions about the most compatible treatment. However, this is not realistic. They propose that all stakeholders, including academia, governments, and industry, should empower patients to consent to the processing of their data to enable the development of precision medicine (Seyhan and Carini, 2019).

3.3 Intellectual Property rights

There are IP rights emanating from the cross-border access and use of personalised medicine data. The national law defines what kind of IP protection will be attributed to an invention or

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

creation. IP frameworks play an increasingly important role as research projects involving biomedical research and health innovation are evolving (Garden et al, 2021). Biomedical research may take a long time to be completed, often requiring high investments without certain results. The attributed IP protection serves as an incentive for investors and researchers to fund and conduct research. The progress of life sciences has been successfully supported by parallel developments in IP, notably patents, through jurisprudence as well as regulatory provisions (Straus, 2017).

The Covid-19 pandemic has brought to the fore unprecedented challenges which have to do with 'exclusive rights, information sharing, affordability of medical treatment and innovation' (Levine, 2020). The need to develop vaccines and medical treatment at the earliest possible time, having the death rates rising in alarming levels, exposed the big challenges that derive from IP rights and the difficulty to achieve a balance between economic gain and public health.

Patents have been criticised for putting barriers to innovation due to the strict rules that govern them (Parthasarathy, 2020). The need to sign licensing agreements with patent holders or the need to find alternative routes in case of patent holders' refusal may slow down innovation. It has been suggested that patents through onerous terms that can be included in the licensing agreements are able to shape entire markets and even industry and technological fields while making it challenging for smaller players to enter the innovation game (Parhasarathy, 2020). Consequently, it has been observed that some inventors have started drifting away from patents (Bessen, 2014). All these concerns have created a movement towards alternative IP regimes that promote openness as a means to increase innovation (Torrance and Kahl, 2013). First, we will go through the different IP rights that emanate from the cross-border access and use of personalised medicine data and then we will examine the Open Innovation scheme.

Copyright

As mentioned before, the requirements for attributing IP protection largely depend on the national law of each State. However, copyright is usually linked with original expression, creativity, or originality. Copyright protection entails the provision of exclusive rights, including the right of reproduction and the right of distribution.

Although scientific publications will practically always be protected under copyright, the answer is not so straightforward when it comes to the decision on whether the research data on which the publication was based, are entitled to the same protection (Hugelier, 2015). Datasets though can be attributed to copyright protection, while at the same time, they enjoy a sui generis database right. As far as biobanks are concerned, it has been suggested that they may fall under copyright protection as a database (Harris and Rosenfield, 2005). Nevertheless, they can be protected also under the database right (Hawkins, 2015). The protection is attributed to the 'person who takes the initiative in obtaining, verifying or presenting the contents of a database, and who assumes the risk of investing in that obtaining, verification or presentation', thus covering multiple actors, including the funders, as well as Universities and Hospitals that might have been involved in the project in question (Hawkins, 2015).

Patent

The right to patent is private and absolute, enjoying only temporary protection (Lefakis, 2004). The patent concerns the intangible property of the invention and not the material integration or application. It is a negative right in the sense that the inventor has the right to exclude everyone else from the exploitation of the invention. Although national IP laws differ, the main requirements that are prescribed in the different legal frameworks are 1) novelty, 2) inventive step, and 3) industrial application (Bently and Sherman, 2018). Harvard oncomouse was the first case where a patent was attributed for a living mammal. In particular, University of Harvard has attributed a patent for a genetically modified mouse, which was very susceptible to cancer (Lefakis, 2004)]

In countries that are members of the European Patent Convention, the patent is attributed either by the national patent office on national level or by the European Patent Office (EPO) through a centralised procedure (Meier et al, 2016). As of 1st of June 2023, the Unitary Patent System is expected to start. This system will introduce a cost-effective patent option for patent

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

protection and dispute settlement, granting patent protection, which will be valid across 25 EU Member States, by submitting only one application before the EPO (EPO, 2023). The Unitary Patent provides a uniform patent protection (Hartmann-Vareilles, 2022), which will be more practical and will entail less cost for inventors (Stevenson, 2022).

The attribution of patent protection is not always an easy task to do, especially when it comes to DNA and human genes that are isolated from their natural environment. Contrary to the US, in EU 'the biological material, including DNA sequences, which is isolated from its natural environment or produced with a technical process' can receive patent protection (Lucchi, 2021). However, the European Patent Convention in art. 53(a) sets some limitations on the patentability of some inventions on grounds of ethical concerns among others, prohibiting the patentability of inventions whose commercial exploitation would contravene "ordre public" or morality. This limitation has been further clarified in the EU Directive on the legal protection of biotechnological inventions, which prohibits the inventions that involve a) processes for cloning human beings, b) processes for modifying the germ line genetic identity of human beings, c) use of human embryos for industrial or commercial purposes, and d) processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes (EU Directive on the legal protection of biotechnological inventions, 1998, art. 6(2); Straus, 2017).

The WTO Trade-Related Intellectual Property Rights (TRIPS) Agreement grants a specialised protection to undisclosed clinical and scientific data, which required considerable effort to collect and were submitted to the competent authority when requesting a market authorisation for a medicinal product (TRIPS, 1995, art. 39). This protection entails a data exclusivity period, 'during which (i) the data submitted may not be relied upon by third parties and/or will not be accepted by the responsible health authority for another Market Authorisation (MA) application, and/or such (ii) an MA will not be granted by the responsible health authority and/or a period of market protection during which an MA, relying on such data, may not be placed on the market' (Meier et al, 2016). This regulatory exclusivity differs from IP rights in the sense that it does not grant legal monopoly, but it rather serves as a reward, preventing third parties from using the same data to get a market authorisation or forcing them to wait for a certain period until they can introduce a generic or biosimilar medicinal product (Meier et al, 2016).

Research in life sciences is closely linked to profit making and patents are the means to achieve it. The freedom allocated to patent holders can have serious repercussions on research and development, as the patent protection allows them to set royalties and negotiate freely the terms of licensing agreements having more power in comparison to candidate licensees (Lucchi, 2021).

Trade secret

Recently, the question whether IP rights and trade secrets arise from data has gained a major momentum (Radauer et al, 2022). In EU level, the Trade Secrets Directive 2016/943 is applicable. According to art. 2 of the Trade Secret Directive, trade secret is defined as 'information which meets all of the following requirements: (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) it has commercial value because it is secret; (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret'. This definition is similar to the one adopted by TRIPS Agreement. Despite the aim of the Trade Secret Directive to ensure the harmonisation of national legal frameworks of the different EU Member States, there are still some divergences in the national implementation (Radauer et al, 2022).

Lately, there has been a shift towards trade secrets. Sometimes it is preferred to protect innovation while ensuring competitiveness using trade secret protection, as the cost of patenting an invention and maintaining a patent is often exorbitant (Arundel, 2021; Crass et al, 2019). This shift became even more obvious during the Covid-19 pandemic, as trade secrets play a significant role in developing vaccines, and creating diagnostics and treatments (Levine, 2020).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

3.4 Open Innovation

Usually, open movement is suggested as an antidote to IP rights access barriers (Walsh et al, 2021). Open science movements have been linked to the "open data", "open software" and "open access" movements and they are promoted as a way to achieve transparency, accountability in research and reusability of outputs in the name of research (Levin and Leonelli, 2017).

Open Innovation assumes that companies can and should use both external and internal ideas as well as external and internal paths to market in their effort to advance their technology (Chesbrough, 2006). It considers that Research and Development (R&D) is an open system. Open Innovation is mostly used in relation to inventions that would normally be entitled to patent protection, while in the copyright context, we usually come across the "open access" term (Walsh et al, 2021). The latter covers everything that can be protected under the copyright regime, such as 'software (open source), data (open data) and other cultural and educational subject-matters (open educational resources)' (Walsh et al, 2021). Open data strategy, according to which 'data should be generated and shared freely', has been inspired by the Open Innovation scheme and open source (Temiz et al, 2022). Scholarship has highlighted the economic potential of open data, but so far it is not considered that open data has successfully realised this potential (Huyer and Knippenberg, 2020).

There are two events that have substantially contributed to increasing the discussion around Open Innovation (Cesario et al, 2022). The first one is the digital transformation that has significantly changed the way that services and systems are integrated and has forced even personalised medicine experts to obtain new skills and literacy. In such an advanced space, keeping a closed regime would be less functional. The second event that brought forward Open Innovation, is the Covid-19 pandemic, that forced actors to move expeditiously and build collaborations to be able to address the unprecedented challenges. The pandemic has highlighted the fundamental role that Open Innovation can play in addressing pressing social needs through knowledge sharing and increasing the prospect of finding innovative solutions (Scotti et al, 2022).

The fact that Open Innovation is solidarity-driven, does not come without challenges though. Indeed, although openness enables the easier acquisition of external valuable knowledge that further leads to advancing innovation, a high level of openness may also elevate the risk of data leakage (Wu et al, 2021). Therefore, it has been suggested that the degree of openness in the Open Initiative scheme should vary depending on the circumstances (Wu et al, 2021). First and foremost, one of the most critical challenges is what happens with the data that can lead to the identification of a person, or when some of the data are protected under IP rights, while others not (Hugelier, 2015). According to Hugelier (2015), there are also ethical challenges, such as the risk of misappropriation or commercialization of data, distribution of research results in an unequal way and disproportionate impact on scientific freedom.

3.5 Information about IP rights

Information about IP rights in the informed consent process

During the recruitment phase, researchers are called to provide research participants with information that will allow them to assess the potential risks arising from their decision to permit the processing of their data and samples and make a thoughtful and well-informed decision. The provision of this information is part of the informed consent process. Researchers can also give additional information about other aspects of the research, such as data commercialisation, a potential conflict of interest as well as information about IP rights. It is important to inform research participants that although they have rights on their data, stemming from data protection law, they do not have IP rights on their data.

Transfer Agreements

Cross-border research projects require sharing of data between the participating entities. This data sharing takes place under specific terms, which are prescribed in the transfer agreements. Transfer agreements are necessary whenever there is a transfer of data to another research

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

entity or within the same institution, as they prescribe the terms by which the parties need to abide by. Transfer agreements include Material Transfer Agreements (MTAs), when the transfer involves samples, Data Transfer Agreements (DTAs), when the transfer involves data, and Data Collaboration Agreements (DCAs) when the data are meant to be used within the framework of a specific collaboration.

The Data Provider must ensure that the data recipient is mindful of the relevant requirements set by the host country, the local government, and the institution. For example, when personal data originate from the EU, the strict rules of the General Data Protection Regulation (GDPR) on data transfers apply. The data transfer must comply with the applicable law, practices, the approval of Research Ethics Committees (RECs), as well as the decisions and practices of other relevant actors, e.g. the national Data Protection Authorities (DPAs) and Data Access Committees (DACs), if applicable in the particular case. In the context of the research routine, the relevant provisions can be merged into one document.

The essential elements to be included in the transfer agreement include information on the contracting parties, a description of the material that is to be transferred, as well as the date of the transfer. The transfer agreement should also include information about the study protocol, the verification of whether legitimate preconditions of the transfer have been fulfilled, information about compensation, the costs of the transfer, the authorised use of data and samples, information about the users, as well as information about the rights of research participants, if applicable. The conclusion of the study marks the duration of the agreement. However, the parties can agree otherwise. The parties also mutually agree on the jurisdiction of courts that would be competent to adjudge their case in case of legal conflict. Nevertheless, they can choose to remain silent, in which case the competent court will be defined in accordance with the general principles of law. The transfer agreement should also include information about warranties, liabilities, and indemnification. Of course, it should also contain information about generated IP rights and ownership, authorship, and publications as well as the return of research results. We will examine more in-depth the last three elements in the next paragraphs, which will be specifically dedicated to data transfer agreements.

Information on IP rights in Data transfer agreements

We will now turn to a specific type of transfer agreement – the data transfer agreement- which is the most relevant for the purpose of this Report. As mentioned before, the data transfer agreement must include information about the IP rights that arise from the cross-border transfer of data. It needs to be stressed that only the rights specifically mentioned in the data transfer agreement are granted since no other rights are conveyed. The data Provider is the one that retains the ownership and/or custody of the data while they retain the right to freely use, disclose or transfer the data to any third party. The data Recipient is not granted any IP right on the data. Additionally, it may be agreed that the Recipient shall hand over a copy of any derived data/variables that may arise from the use of the data to the Provider.

Researchers need to also take into consideration the interest and right of society to benefit from the research results (Knoppers et al, 2006). As far as research results that will come out of the research project, normally the IP rights that will be generated will be owned by both the Provider and the Recipient. If the parties wish to distribute their shares, they need to enter into another separate agreement. The same applies in case each of the parties wants to use the research results for commercial use. However, if they want to use the research results only for their own research/academic use, there is no need for a separate agreement.

The data transfer agreement should also contain information about the procedures for publishing the research results, setting out a framework for arising publications, inventions, or other IP-related issues. If any party intends to submit a manuscript for publication, they should previously send the manuscript to the other party prior to a designated time period. In absence of any objection within this period, the party may proceed with the publication. In cases where the relevant provisions have not been included in the agreement, then they must sign a separate agreement dealing with these issues.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

The issue of identifying the appropriate means for acknowledging the contribution of research conducted in biobanks has been contested. The Bioresource Research Impact Factor (BRIF) is an initiative that allows the recognition of research, but without assigning authorship (Bravo et al, 2015). BRIF has several steps, including 1) citation of BioResources in journal Articles (CoBRA) guidelines, 2) open journal of bioresources, 3) other tools in development. However, there are arising challenges, such as the potential conflict between co-authorship and authorship guidelines or potential disagreement on the interpretation of results on behalf of the researchers involved in the research project. To address these challenges, it is suggested to come up with alternative crediting systems that are promoted by Open Science. For example, it has been suggested that a unique digital identifier, like DOI, should be used for citing and acknowledging the use of bioresources in research projects and publications (Annaratone et al, 2021).

3.6 Need for a balanced protection of ownership

Having discussed the differences between IP rights and Open Innovation in the field of biomedicine, it appears that it is necessary to find a balance between the protection of health on the one hand and the support of R&D on the other. Although it seems reasonable to allow a necessary level of IP protection to encourage private research and development, it is also necessary to ensure the protection of the patients' and the public's interests (European Observatory on Health Systems and Policies, 2020). Also, IP rights need to be reconciled with the protection of the rights to privacy and data protection of research participants. It is necessary to always assess the potential harm that may arise from breaching data, which may end up interfering even with the right to non-discrimination and equality. Therefore, it is important to ensure the enjoyment of both IP rights and the right to data protection, as otherwise there may be a disincentive to conduct research or provide personal data, and this could potentially undermine data-driven innovation (OECD, 2019).

It has been suggested that collaborative platforms can control the use of IP rights through adoption of contractual provisions prohibiting the patenting of upstream technologies which could potentially lead to the restriction of further research (Garden et al, 2021). Although it has been supported that it is more practical if those accessing collaborative platforms, such as biobanks, would have their own IP, there are collaborative platforms, such as the Global Alliance for Genomics & Health (GA4GH), that adopt an open source scheme with a view to enable free use for the public good (Garden et al, 2021). It is necessary to find a balance between the protection of IP rights, which clearly serve as an incentive for researchers and the public good which can be promoted if access to innovation is enabled. This could be achieved if collaborative platforms use appropriate governance and contractual access arrangements, e.g., excluding exclusive licences, which will aim to achieve this highly wanted balance (Garden et al, 2021).

4. Conclusions

The objective of this study was to examine the management of IP rights that may arise from cross-border access and use of personalised medicine data in a pan-European genome initiative. The study is based on the findings of a research that came up with a checklist that aims to help researchers fulfil the different legal and ethical requirements in the context of biobanking research covering all stages of the research process, starting from the protocol design, moving to the participants' recruitment, handling of samples and data and ending up to the communication of research results. The findings of this Report have been also based on a literature review of academic papers, relevant reports, and guidelines. Taking into consideration the findings of this Report, it can be deduced that there needs to be a balance between research and development and the protection of individuals and the public in general. IP rights can be an incentive for research and can accelerate investments, but this should not entail a burden on public health and public good at large. Governance and contractual arrangements that balance both IP rights and innovation for public good would be a good way to go.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

5. Impact

The results of this Report increase the understanding of the management of IP rights that may arise from cross-border access and use of personalised medicine data in a pan-European genome initiative. These results can contribute to the development of a framework of core elements that need to be considered when deciding to embark on a research project, in close collaboration with WP1, WP4 and WP6.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

6. Glossary of terms, abbreviations, and acronyms

AI: Artificial Intelligence

B1MG: Beyond One Million Genomes

BRIF: Bioresource Research Impact Factor

CoBRA: Citation of BioResources in Journal Articles

DCA: Data Collaboration Agreement

DAC: Data Access Committee

DPA: Data Protection Authority

DTA: Data Transfer Agreement

ELSI: Ethical, Legal and Social Implications

EPC: European Patent Convention

EPO: European Patent Office

EU: European Union

GA4GH: Global Alliance for Genomics & Health

GDPR: General Data Protection Regulation

IP: Intellectual Property

MA: Market Authorisation

MTA: Material Transfer Agreement

OECD: Organization for Economic and Cooperation Development

R&D: Research and Development

REC: Research Ethics Committee

TRIPS: WTO Agreement on Trade-Related Aspects of Intellectual Property Rights

WTO: World Trade Organisation

7. References

Instruments

Agreement on Trade-Related Aspects of Intellectual Property Rights 1994 (WTO)

Convention on the Grant of European Patents (European Patent Convention) 1973

Directive (EU) 2016/943 of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secret Directive) 2016, OJ L 157/1

Directive 98/44/EC of the European Parliament and of the Council on the legal protection of biotechnological inventions 1998, OJ L 213/13

Academic papers/ Books/Chapters in Edited Books/Reports

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Annaratone, L., De Palma, G., Bonizzi, G., Sapino, A., Botti, G., Berrino, E., Mannelli, C., Arcella, P., Di Martino, S., Steffan, A., Grazia Daidone, M., Canzonieri, V., Parodi, B., Paradiso, A.V., Barberis, M., Marchiò, C., on behalf of Alleanza Contro il Cancro (ACC) Pathology and Biobanking Working Group. (2021). Basic principles of biobanking: from biological samples to precision medicine for patients. Virchows Archiv, 479, 233. https://doi.org/10.1007/s00428-021-03151-0

Arundel, A. (2021). The relative effectiveness of patents and secrecy for appropriation. *Research Policy*, 30(4), 611. https://doi.org/10.1016/S0048-7333(00)00100-1

Atsali, S. S. (2020). Community-based access and benefit-sharing platform and its role in biodiversity, culture, and intellectual property rights. *The 1st International Conference on Genetic Resources and Biotechnology, IOP Conf. Series: Earth and Environmental Science*, 482**,** 012009. doi:10.1088/1755-1315/482/1/012009

Bently, L., Sherman, B., Gangjee, D., and Johnson, P. (2018). *Intellectual Property Law*. (5th ed.). Oxford University Press. https://doi.org/10.1093/he/9780198769958.001.0001

Bernhard, H.P., and Cookson, C. (eds). (1995). *Genethics – Debating Issues and Ethics in Genetic Engineering*. CIBA-GEIGY Ltd.

Bessen, J. (2014, June 13). History Backs Up Tesla's Patent Sharing. *Harvard Business Review.* https://hbr.org/2014/06/history-backs-up-teslas-patent-sharing

Bravo, E., Calzolari, A., De Castro, P., Mabile, L., Napolitani, F., Rossi, A.M., and Cambon-Thomsen, A. (2015). Developing A Guideline to Standardize the Citation of Bioresources in Journal Articles (Cobra). *BMC Medicine*, 13, 33. https://doi.org/10.1186/s12916-015-0266-y

Cesario, A., D'Oria, M., Simone, I., Patarnello, S., Valentini, V., and Scambia, G. (2022). Open Innovation as the Catalyst in the Personalized Medicine to Personalized Digital Medicine Transition. *Journal of Personalised Medicine,* 12, 1500. https://doi.org/10.3390/ jpm12091500

Chensbrough, H. (2006). Open Innovation: A New Paradigm for Understanding Industrial Innovation. In H., Chesbrough, W., Vankhaverbeke, and J., West, (Eds). *Open Innovation: Researching a New Paradigm*. Oxford University Press.

Council. (2015). Council Conclusions on personalized medicine for patients. OJ C421/2. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2015:421:FULL&from=EN

Crass, D., Garcia Valero, F., Pitton, F., & Rammer, C. (2019). Protecting Innovation Through Patents and Trade Secrets: Evidence for Firms with a Single Innovation. *International Journal of the Economics of Business*, 26(1), 117. https://doi.org/10.1080/13571516.2019.1553291.

European Commission. (2023). Personalised Medicine. https://research-and-innovation.ec.europa.eu/research-area/health/personalised-medicine_en

European Observatory on Health Systems and Policies, Williams, Gemma A, Liede, Sandra, Fahy, Nick, Aittomaki, Kristiina. et al. (2020). Regulating the unknown: a guide to regulating genomics for health policymakers. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/338975

European Patent Office. (2023). Unitary Patent & Unified Patent Court. https://www.epo.org/applying/european/unitary.html

Garden, H., Hawkins, N., and Winickoff, D. (2021). Building and sustaining collaborative platforms in genomics and biobanks for health innovation. *OECD Science, Technology and Industry Policy Papers*, No. 102. https://doi.org/10.1787/11d960b7-en

Harris, R.K., and Rosenfield, S.S. (2005). Copyright protection for genetic databases. *Jurimetrics,* 45(2), 225.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Hartmann-Vareilles, F. (2022). The Unitary Patent System: what is this light at the end of the tunnel?, *ERA Forum*, 23, 143. https://doi.org/10.1007/s12027-022-00711-y

Hawkins, N. (2015). Intellectual Property and Biobanks. In D. Mascalzoni (Ed.), *Ethics, Law, and Governance of Biobanking: National, European and International Approaches* (pp. 39-51). Springer. DOI 10.1007/978-94-017-9573-9

Hugelier, S. (2015). Publishing Open-Access Biomedical Data: Legal Challenges. *Biomed Data Journal*, 1(1). 43. http://dx.doi.org/10.11610/bmdj.01108

Huyer, E., and van Knippenberg, L. (2020). The Economic Impact of Open Data: Opportunities for Value Creation in Europe. *European Data Portal*. https://data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf

Johns Hopkins Medicine. The Legacy of Henrietta Lacks. https://www.hopkinsmedicine.org/henriettalacks/

Joly, Y. (2011). Personalized Medicine in Developing Countries: A Roadmap to Personalized Innovation. Current Pharmacogenomics and Personalized Medicine. 9, 156. https://yannjoly.openum.ca/files/sites/57/2016/01/Personalized-Medicine-in-Developing-Countries-A-Roadmap-to-Personalized-Innovation.pdf

Kinkorová, J. (2016). Biobanks in the era of personalized medicine: objectives, challenges, and innovation. EPMA Journal, 7, 4. DOI 10.1186/s13167-016-0053-7

Knoppers, B.M., Joly, Y., Simard, J., and Durocher, F. (2006). The Emergence of An Ethical Duty to Disclose Genetic Research Results: International Perspectives. *European Journal of Human Genetics*, 14, 1170. DOI: 10.1038/sj.ejhg.5201690

Knowles, L., Luth W., and Bubela, T. (2017). Paving the road to personalized medicine: recommendations on regulatory, intellectual property and reimbursement challenges. *Journal of Law & Biosciences*, 4(3), 453. doi: 10.1093/jlb/lsx030

Kuchinke, W., Krauth, C., Bergmann, R., Karakoyun, T., Woollard, A., Schluender, I., Braasch, B., Eckert. M., and Ohmann, C. (2016). Legal assessment tool (LAT): an interactive tool to address privacy and data protection issues for data sharing. *BMC Medical Informatics and Decision Making*, 16, 81. DOI 10.1186/s12911-016-0325-0

Lefakis, L. (2004). *Biotech inventions*. Sakkoulas Publishing. [IN GREEK]

Levin, N. and Leonelli, S. (2017). How Does One "Open" Science? *Questions of Value in Biological Research, Science, Technology, & Human Values*, 42(2), 280. DOI: 10.1177/0162243916672071

Levine, D.S. (2020). Trade secrets and the battle against Covid. *Journal of Intellectual Property Law*, 15(11), 849. https://doi.org/10.1093/jiplp/jpaa164

Lucchi, N. (2021). A New Era for Biotech Patents? Empirical and Theoretical Considerations on the current Patent Dilemma. *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 289. https://doi.org/10.15168/2284-4503-787

Matthews, D., and Zech, H. (Eds.). (2017). *Research Handbook on Intellectual Property and the Life Sciences.* Edward Elgar Publishing.

Meier, A., Faulkner, S.D., Schoonderbeek, C., Jong, B., Kung, J., Brindley, D., and Barker, R. (2016). An Assessment of Implications of Adaptive Licensing for Pharmaceutical Intellectual Property and Regulatory Exclusivity Rights in the European Union. *Clinical Pharmacology & Therapeutics*, 100(6), 743. doi:10.1002/cpt.511

Morin, S., Barron Moak, H., Bubb-Humfryes, O., von Drehle, C., Lazarus, J.V. and Burrone, E. (2022). The economic and public health impact of intellectual property licensing of medicines for low-income and middle-income countries: a modelling study. *Lancet Public Health*, 7, e169. https://doi.org/10.1016/ S2468-2667(21)00202-4

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

OECD. (2019). *Enhancing Access to and Sharing of Data*. OECD Publishing, Paris, https://doi.org/10.1787/276aaca8-en

Parthasarathy, S. (2020). Approaching intellectual property scholarship differently: A qualitative research review and agenda. *Science and Public Policy*, 47(5), 627. doi: 10.1093/scipol/scaa010

Parthasarathy, S. (2020). Approaching intellectual property scholarship differently: A qualitative research review and agenda. *Science and Public Policy*, 47(5), 627. doi: 10.1093/scipol/scaa010

PerMed. (2015). Shaping Europe's Vision for Personalised Medicine: Strategic Research and Innovation Agenda (SRIA). https://www.icpermed.eu/media/content/PerMed_SRIA.pdf

Rajam, N. (2020). Policy strategies for personalising medicine "in the data Moment". Health Policy and Technology, 9, 379. https://doi.org/10.1016/j.hlpt.2020.07.003

Scotti, F., Pierri, F., Bonaccorsi, G., and Flori, A. (2022). Responsiveness of open innovation to COVID- 19 pandemic: The case of data for good. PLOS One, 17(4), e0267100. https://doi.org/10.1371/journal.pone.0267100

Seyhan, A.A. and Carini, C. (2019). Are innovation and new technologies in precision medicine paving a new era in patients centric care? *Journal of Translational Medicine, 17, 114* https://doi.org/10.1186/s12967-019-1864-9

Sobradillo, P., Pozo, F., and Agustí, A. (2011). P4 Medicine: the future around the corner. *Archivos de Bronconeumología*, 47(1), 35. doi:10.1016/j.arbres.2010.09.009.

Song, C. H., Han, J.-W., Jeong, B. and Yoon, J. (2017). Mapping the Patent Landscape in the Field of Personalized Medicine. *Journal of Pharmaceutical Innovation*, 12, 238. DOI 10.1007/s12247-017-9283-z

Straus, J. (2017). Intellectual property rights and bioeconomy. *Journal of Intellectual Property Law*, 12(7), 576. doi:10.1093/jiplp/jpx059

Stevenson, A. (2022). Essays on the Economics of Patent Rights. *Publications of the Hanken School of Economics Nr 362.* https://helda.helsinki.fi/dhanken/bitstream/handle/10227/511511/362%20-%20Alexis%20Stevenson.pdf?sequence=1&isAllowed=y

Temiz, S. Holgersson, M., Bjorkdahl, J., and Wallin, M.W. (2022). Open data: Lost opportunity or unrealized potential? Technovation, 114, 102535. https://doi.org/10.1016/j.technovation.2022.102535

Torrance, A. W., and Kahl, L. J. (2013). Bringing Standards to Life: Synthetic Biology Standards and Intellectual Property. *Santa Clara High Technology Law Journal*, 30(2), 199. https://digitalcommons.law.scu.edu/chtlj/vol30/iss2/3

Vicente, A.M., Ballensiefen, W., Donertas, D., Eklund, M., Ivask, A., Jönsson, J.-I., Kuhlmann, K., Lawrence, A., O'Driscoll, M., Richer, E., and Trivella, G. on behalf of the Executive Committee of ICPerMed. (2019). The ICPerMed vision for 2030: How can personalised approaches pave the way to Next-Generation Medicine? https://www.icpermed.eu/media/content/Vision_Paper_2019.pdf

Walsh, K., Wallace, A., Pavis, M., Olszowy, N., Griffin, J., and Hawkins, N. (2021). Intellectual Property Rights and Access in Crisis. *International Review of Intellectual Property and Competition Law*, 52, 379. https://doi.org/10.1007/s40319-021-01041-1

Wong, H.K., and Breyer Menon, A. (2022, July 2018). The Trouble with IP for Digital Health and Precision Medicine. *Center for Biosimilars*. https://ramaonhealthcare.com/contributor-the-trouble-with-ip-for-digital-health-and-precision-medicine/

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Wu, H., Han, Z., and Zhou, Y. (2021). Optimal degree of openness in open innovation: A perspective from knowledge acquisition & knowledge leakage. Technology in Society, 67, 101756. https://doi.org/10.1016/j.techsoc.2021.101756

10. Annex

Paper: Tzortzatou-Nanopoulou, O., Akyüz, K., Goisauf, M., Kozera, L., Mežinska, S., Mayrhofer, M.T., Slokenberga, S., Reichel, J., Croxton, T., Ziaka, A., and Makri, M. (2023). Ethical, Legal and Social Implications in Research Biobanking: A Checklist for Navigating Complexity. Developing World Bioethics (Under publication)

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Beyond One Million Genomes

8. IT infrastructure requirements based on a data protection by design and default approach

V1.0 – 4. August 2022

Author: Regina Becker

Main contributors: Sarah Peter, Christophe Trefois, Wei Gu, Davit Chokoshvili

Additional contributors and/or reviewers: entire WP2 / 1+MG ELSI WG

1. Definitions

The documents are built on the 1+MG Glossary. In particular, the following expressions are used:

- **Metadata:** a set of data that describes and gives information about other data. Metadata do not include any data that are processed to produce any results of the data use such as health or genetic data. Metadata can be personal data where they contain information on the subject level (e.g. consent decisions of an individual data subject)
- **Data collections:** Data collections are data that come from the same collection context from the same controller(s) and can be characterised with non-personal metadata. A data collection can e.g. refer to data of a cohort or a public registry.
- **Records:** Data related to individual data subjects
- **Data set:** Data that are grouped in a certain context, e.g. for a user's access.
- **Enriched data:** data where additional information is added, e.g. annotations
- **Derived data:** data that have been created by alteration of the original data (e.g. through data curation)

2. Introduction: Data Protection by Design and Default

Data protection by design and default (DPbDD) was introduced in the EU General Data Protection Regulation (GDPR) and is anchored in Article 25. Organisational and technical safeguards have to be taken by the controller to implement the data protection principles for all processing of personal data.

It is important to notice that DPbDD is more than ensuring privacy, i.e. the protection of the individual against unauthorised disclosure of the data, DPbDD covers all data protection principles. These are described in Art. 5.1 GDPR:

- Transparency
- Lawfulness
- Fairness

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Purpose Limitation
- Data Minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality

DPbDD means that the compliance with these principles must be considered already when the processing is planned and not mapped afterwards ("by design"). The "by default" means that the default state of a system should be "closed" or "protected" and only those data necessary for the purpose should be processed.[79] Disclosure should be an active step that has to be planned in compliance with the above principles.

The European Data Protection Board (EDPB) has published the Guidelines 04/2019 (latest version 2.0 from 20 October 2020) on how to implement DPbDD.

It is the responsibility of the data controller to ensure the DPbDD. However, the data processor has the obligation to assist the data controller in the implementation of Articles 32 to 36 GDPR, which deal with the security of the processing and the data protection impact assessment. When the design of the processing and the IT systems are in compliance with the data protection principles, the demonstration of the sufficiency of the chosen organisational and technical safeguards is part of the accountability obligation of the data controller and a first step for the data protection impact assessment. Even where the providers of the IT infrastructure are only acting as data processors in the 1+MG initiative, they should provide expertise input in how the IT infrastructure should be designed with respect to technical and organisational measures (TOMs) to comply with the data protection principles, in particular on aspects of data integrity and confidentiality as well as to realise purpose limitation and data minimisation.

The so-called "Five Safes" developed in the UK in the context of access to statistical data is an approach that can be used to support the DPbDD exercise.[80] The Five Saves cover:[81]

- Safe projects : Is this use of the data appropriate, lawful, ethical and sensible?
- Safe people: Can the users be trusted to use it in an appropriate manner?
- Safe data: Does the data itself contain sufficient information to allow confidentiality to be breached?
- Safe settings: Does the access facility limit unauthorised use or mistakes?
- Safe outputs: Is the confidentiality maintained for the outputs of the management regime?

The "Five Safes" approach is increasingly used to design the solutions for data sharing by statistics offices. A general guidance on how the Five Safes can be used to create secure data environments for research was published recently by the UK Health Data Research Alliance: "Building Trusted Research Environments" (version 1.0 from 8 December 2021).

This document provided an excellent starting point for the 1+MG DPbDD exercise. We subsequently adapted the "Five Saves" approach into a framework tailored to 1+MG, following envisaged data flows within the cross-border initiative, and taking into account distinct characteristics of the 1+MG forthcoming data governance. The analysis of the different stages will be organised according to the data protection principles of the GDPR, not the Five Safes to allow an easier demonstration of compliance, performance of a DPIA and auditing of 1+MG.

Please note that the current data governance for research use on which the DPbDD recommendations are based do not yet cover access by users outside the European Economic Area (EEA). In addition, the

[79] See EDPB Guidelines 04/2019 subsection 2.2.1
[80] For the background of the Five Safes, see: https://en.wikipedia.org/wiki/Five_safes
[81] http://www.fivesafes.org/

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

data governance for access in the context of rare diseases, which falls between research and healthcare, is also not yet decided.

3. DPbDD considerations for the IT infrastructure

"IT infrastructure" in the following covers all information and communication technology support of the operations of 1+MG. This goes beyond the management of data access and the provision of an analysis platform for data use and also includes information and workflow management of 1+MG. In this context, the data governance of 1+MG, implementing responsible and DPbDD driven procedures and the DPbDD-driven considerations on information management in 1+MG are relevant background documents. The current document recommends a list of requirements that the 1+MG IT infrastructure should fulfil. It is a first draft and may further develop based on the development of the data governance policy or use case requirements as pursued in the subsequent deployment projects such as the Genomic Data Infrastructure (GDI) project. It will have to be complemented with scientific requirements on the IT infrastructure (i.e. use case needs; interoperability requirements). 1+MG Working Groups for Interoperability and Secure IT environment (WG5) and ELSI (WG2) must work closely together to see how the requirements can be translated into workable solutions and define a roadmap for the implementation.

4. DPbDD requirements for the IT infrastructure in 1+MG along the data journey

Data inclusion

- Data transfer tools that are suitable for biomedical researchers and clinicians that allow secure transfers to the 1+MG IT infrastructure.
- Data integrity check after transfer has to be ensured (e.g. use of checksums, data scrubbing)
- Data format validation (e.g. for genomics data, associated metadata)
- Open to discussion: Secondary pseudonymisation of genome and phenotypic data separately and/or physical separation of genomic and phenotypic data on different servers
- Collection / uploading of structured metadata into suitable tools relevant for ingestion into the data management system (e.g. ELSI metadata relevant for data use) and for display in the data catalogue

Data hosting

- Storage that is secured against external and internal attacks and or accidental disclosure (e.g. no direct internet access);
- Protection against data loss (e.g. back-ups and disaster recovery)
- Open to discussion: Encryption of data where the IT environment allows fast decryption / encryption on the fly [criteria to consider: relevance / efficiency / cost as data are almost constantly in use]
- Access restriction to the data: default is no access, fine-grained access control lists (ACLs)
- Data management with possibility to send automated reminders in view of upcoming deadlines (end of access, end of retention time for a dataset version)
- Possibility for record-level metadata on version, use conditions, data use (see below) and retention time
- Ability to remove data from the active system (i.e. no longer accessible)
- Ability to restrict processing to certain users, purposes and subsets of data (at variable level, at data subject level) only (i.e. permit) from a certain time point onwards
- Ability to remove data from back-ups or restrict their restoring

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Data discovery

- Central data catalogue that seamlessly comprises information from national data catalogues where applicable
- Data catalogue should offer sufficient characterisation of datasets for requesters to allow assessing the usability of the data for their purposes; ideally, also synthetic datasets are made available, also for purposes beyond data discovery but to test usefulness of the data. [Comment: synthetic data here would only be on the "lowest" requirement level - reflecting the structure of the data, not mimicking the content distribution and correlation]
- Option for data requesters to specify their type of access request (research, healthcare, policy development, other - still to be defined based on use case input)
  -> such specification influences user/requester interface offered
- Possibility for requesters to select relevant data subjects based on genotype, phenotype and/or other features such as ethnic origin
- For subject level data discovery, the requester must also specify the area of research based on controlled vocabulary (e.g. disease specific, legal basis)
- Authentication and authorisation infrastructure (AAI) that allows authentication of requesters by their employing institution and/or role (relevant roles are researcher, healthcare professional etc.).
- The AAI must be technically interoperable to allow for federated identity management.
- Data requesters must be able to log into one AAI implementation that gives access to data across all 1+MG sites.
- Possibility to block persons from access if they appear on a "blocked list"
- Offer suitable "terms of use" that need to be confirmed before the requester can start the search on a subject-level (i.e. data on individuals)
- Data management system that allows to limit data searches to relevant data only (tagging of data for use restrictions: offer only datasets for search that can be processed under a particular legal basis of the requester [if applicable], by the category of requester for the purposes of the project)
- Ensure that such focussed search cannot be bypassed and/or that incorrect behaviour is monitored / flagged (e.g. where the requester changes the purpose specification to change search)
- Data discovery tool that allows subject-level searches for genotype and/or phenotype without releasing information that can be used to readily identify subjects and/or reconstruct the data by single search or combination of multiple searches by built in protective features
- Logging of subject level searches
- Possibility to download / view synthetic datasets of relevant data collections

Access request: Support data governance procedures

- Data user portal that allows seamless transfer from data discovery to access request on the selected datasets, i.e. selection of data collections and/or a number of individual records on subject level data and launching the access request on that selection
- Web form for the collection of relevant information for access review and approval, depending on access request type. This includes the characterisation of access request based on controlled vocabulary matching the ELSI metadata (still to be defined) and the possibility to upload files such as ethics approvals or evaluation reports by funders.
- Possibility to launch the same access request by several institutions as joint controllers.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Ideally: possibility for electronic signature for registered organisations' signing officials.
- Ideally: possibility for users to be flagged as whitelisted, so that no separate confirmation to confirm the validity of the request is needed
- Possibility for documented approvals / rejections with justification and exchanges on request in the system.
- Access request management system that acquires the necessary access request information (from the user portal), notifications to the relevant actors, ideally validation of submissions by detecting obvious mistakes (e.g. missing elements; inconsistencies (where relevant))), allows communication between actors; API to retrieve information stored in the system.
- Archiving of access requests

Data use:

- Offering non-personal data to users for their purposes where reasonably possible (e.g. existing knowledge on variants; synthetic data)
- Possibility to issue, store and consume data permits; ideally also from different Data Providers for a joint analysis by the user
- Link approved access request to relevant data collection(s) or relevant data types and data subjects (individual records); includes handling of persistent identifiers for datasets
- Allow user operation in a virtual environment that allows operations over a distributed system or offer temporary data pooling as alternative
- Separating individual users accessing from the same user institution;
- Separating virtual workspaces allocated to the same user for different projects
- Strong (multi-factor) authentication ensuring verification of user identity, contribute link to EU eIdenties where these become available
- Only personal, no shared accounts
- Allow several users from the same controller (e.g. via user groups) as well as the possibility to include processors in the permit
- Possibility to give access only to the relevant records including only the data types needed of the relevant data subjects
- Ideally linking of genomic data and phenotypic data on the fly for data processing (from data on separate servers and/or different pseudonym), also decryption / encryption on the fly only when access is needed
- Possibility that external data, which is not held in 1+MG but parallel with the Data Provider or another Data Provider (e.g. public registry) can be linked to the respective records held internally for joint use.
- Possibility for collaborative approaches for data analysis where several users work on the same project
- Offer different user interfaces depending on the access request (healthcare professionals different to researcher; different rights / possibility to operate on the data)
- Open for decision: offer virtual desktop interface or "algorithm to the data" only
- Open for decision: allow 1+MG-held IT infrastructure only or include commercial cloud solutions
- Offer analysis environment for the deployment of containers for research analyses
- Only properly documented, versioned and secure software (e.g. through gitlab) can be offered for users; a comprehensive set of analysis tools / libraries has to be offered.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Users should be allowed to run analyses from a list of approved workflows/pipelines. This should also include search and analysis tools with graphics / visualisation tools that are easy to use for users such as clinicians

- Where users bring in their own software, these need to be checked for malwares and, where applicable, security vulnerabilities. There should be functions and processes to validate the results if they stem from algorithms brought-in by the users.
- Establish a central container registry and the possibility to archive containers.
- Allow user management of access: suspend and resume work
- Possibility to archive and restore the project specific work environment during times of suspended activities
- Logging of data operations
- Monitoring for suspicious behaviour (e.g. certain access patterns; mismatch of stated datatypes / diseases of interest and data accessed)
- Airlocked system, only supervised methods for taking aggregated data / software / results out of the protected environment
- Possibility of data versioning in case of data deletion or rectification or updates with additional data points
- Possibility to archive processed datasets and allow referencing for publications
- Archiving of versioned datasets with information which projects rely on which version to allow reproducibility together with archived containers; a re-running of the analysis should work 10 years later still
- Information portal to provide information on all data use for general public in English, based on requester's information and publications based on data use as well as dedicated communication around data use; API to allow countries to display only projects relevant for the data collection in the country
- Common standards for the above requirements that allow deployment in a federated environment
- Where data use requires data pooling and/or transmission to a different compute environment, secure transfer mechanisms must be established; where data are temporarily stored in another IT infrastructure, safe delete must be ensured
- Automated search for publications based on mandatory reference to 1+MG
- Information management that allows messaging between users and Data Providers, e.g. for reporting incidental findings, requests for additional data / biosamples or recruitment of data subjects in new studies (including automatic management of "no contact" information)

General requirements

- A clear assignment of which considerations (requirements) should be covered during the software development phase and which are to be handled at deployment/operation phase (e.g. via SOPs).
- Implementation in certified compliance with ISO 27001 and ISO 27701.
- At minimum external audit, ideally certification (only certified data centres / compute environments are allowed to pool data across countries)
- Regular testing of effectiveness of safeguards
- Staff must be trained in data protection and IT security (focus depends on role)

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Staff must have secrecy clauses in their contracts where not already subject to professional secrecy by law.
- Data protection policy framework must be in place (minimum policies to be covered has to be agreed jointly)
- Joint agreement on requirements of what makes “secure systems”
- GDPR compliant log information management (how long are logs kept, how are they stored, who has access)
- Management of time-dependent processing (data, metadata, log data) including safe delete at the end of the retention time
- Contract management system with API to data information system
- Information management system that can link access requests to contracts to log files, users, containers, versioned dataset used etc. that is auditable

B1MG has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 951724

9. Information requirements based on DPbDD

V1.0 – 4. August 2022

Author: Regina Becker

Main Contributors: Sarah Peter, Christophe Trefois, Wei Gu, Davit Chokoshvili

Additional contributors and/or reviewers: entire WP2 / 1+MG ELSI WG

1. DPbDD considerations for the information management

Data protection compliant data management includes the management of relevant information on how data can be used but also requires the information on the actors along the data life cycle with their responsibilities and contacts, information on the data use itself, information about organisational and technical safeguards, including the management of such safeguards such as for pseudonymisation and secondary pseudonymisation. Accountability means that the measures taken must be auditable, which again sets up certain requirements for the documentation around the entire life cycle of data in the 1+MG.

The considerations on information management build on the mission of 1+MG, the data governance implementing DPbDD workflows, requirements of the GDPR and practical considerations that link the various needs. The definition of relevant structured and (where applicable) machine readable information is an important input for the design of the IT infrastructure. The current version is a first draft that will further develop with e.g. the decisions on the data governance. It should also be considered to be complemented by information requirements relevant from the user's perspective. The 1+MG Working Groups for Standards (WG3), Interoperability and Secure IT environment (WG5) and ELSI (WG2) must work closely together to define how such information management can be implemented in 1+MG.

2. DPbDD driven information requirements in 1+MG along the data journey

Data inclusion

- Data provider who decides on inclusion and is responsible for specifying the use conditions
    - Legal entity
    - Scientific / clinical contact
    - Administrative contact
    - Legal representative contact
    - Data Protection Officer (DPO) contact
    - Information if Data Provider can be contacted for additional information
- Data use conditions metadata as derived from data subject consent and/or legal requirements [both collection-level and record-level information]
    - Categories of purposes for data use with respect to recipients
    - Limitations on territorial scope for users (only outside EEA can be limited)

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

    - Data subject's consent that Data Provider can disclose additional information where needed by data user
    - Information on data subjects where applicable, e.g. relevant vulnerabilities such as minors and anything that may have implications on processing
- Information around recontacting subjects [data subject-level information]
    - Purposes for re-contacting with communication channels (including [permanent] contacts) and language
        - Participation in research studies
        - Obtaining consent / active information on new purposes
        - General information provision
        - (Incidental findings)
    - General "no return" flag (initially collection level but will be specified on the level of data subjects)
- Information on data retention criteria / requirements / time
- Legal basis for inclusion
- Scientific metadata for all 1+MG data types as defined in 1+MG
- Metadata on changes in datasets (exercise of data subjects rights: deletion, rectification, changed use conditions; changes due to enrichment; changes due to additional data added etc.)
- Information on additional data beyond the 1+MG defined datasets or samples available at Data Provider [or potentially already stored in the 1+MG secure processing environment]
- Integrity check information (after data transfer and periodically during storage)
- Time stamp of upload
- Registration / versioning of datasets (creation of new dataset version 1.0 or adding to existing datasets creating new version 2.0)
- A universal unique identifier (UUID) of the same data and version (identical copy) assigned and used in all federated systems

Data hosting

- Information on data location of the different versions
- Information on retention for individual dataset versions as derived from inclusion data and use data
    - Data archived for project xy; duration of retention
    - Data under deletion request (to be deleted at the end of use or archiving; data not to be restored from backup)
- Information on changes of dataset on subject level
    - Versioning following update of data for new data points
    - Versioning of datasets for changes (rectified)
    - Dataset current version "no more use for N / at all" (objection, withdrawal of consent, delete request)
- Pseudonym matching table (if applicable)
- Controller for hosting

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Legal basis for hosting (if not defined by 1+MG)
- Encryption keys

Access request:

- Type of access request
- Information provided by requester
    - For research project
    - For healthcare use
    - For policy development
- Log information of subject-level searches by requester
- Information on (institutionally) whitelisted and blocked researchers
- Information on signing officials for each institution
- Association of records with defined access requests, i.e. the possibility to keep an auditable track of which records are / were used for which access and the easy information, e.g. where a data subject asks in which projects their data was used.
- Suggested decision by central DAC, where applicable: feedback of National Coordination Points, final decision on request as suggested in the data governance policy (per collection / subset if necessary)
- Objections by individual data subjects
- Logging all interactions until approval and access provision (who, when, what)
- (Contract)

Data use:

- Store permits / contract including expiry dates
- Information on data use
    - For projects / use application: uses version xy of dataset
    - For Data: booked or used for project / use application xy (on record level)
- Documentation on any changes to permit with time-stamped versioning
    - Regarding datasets
    - Regarding controllers and personnel with access
    - Regarding duration
- Research results
    - Information on publication linked to use request (and due to that, data subject)
    - Information on communication on publication (where applicable)
    - Information on patents / products / services derived from use request
- Healthcare results [still subject to healthcare reuse data governance]
    - Feedback on outcome
- Policy development results [still subject to policy development data governance]
    - Outcome report (if or if not intended outcome was achieved)
    - Link to established policies (where applicable)
- "No return" flag for incidental findings

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Log information for data use
- Documentation of exports
- Archiving requirements
- Data stability requirements by data user, i.e. the user is able to indicate at a certain time point when it should not be possible to delete any records from the dataset used in his analysis; may refer to dataset still in use for analysis or for a dataset to be frozen for archiving; needs to be complemented with a written justification
- Versions of datasets
- Reference numbers of workflows in a workflow registry where detailed information on software, data used and IT systems are stored
- Information on any data transmissions (where, when, why)

General

- Information on audits / extension of certification
- Documentation of the testing of the efficiency of safeguards
- Documentation of system changes / upgrades / incidents
- Documentation of data breaches
- Documentation of data and metadata definitions
- Recording of decisions on the different levels
- Documentation of DPIAs
- Documentation of exchanges with DPO and CISO
- Documentation of trainings
- Documentation policy framework with versioning

Information to be provided in research access request [minimum]

- Project title
- Name and contact details of the Principal investigator of the project
- Information about the institution
  (Institution needs to be registered including information on signing official and DPO; where several signing officials are applicable, the relevant official is to be chosen)
- Persons who will have access to the data (PI and his/her group)
- Scientific abstract
- Lay summary of project that can be provided to the data subjects whose data are used
- Data analysis plan that describes the data types and analysis approaches to be employed to achieve the goal of the project
- Description of any machine learning algorithms / AI employed, developed, trained, validated, or exported by the project.
- Information on enriched data (e.g. annotations) and derived data (e.g. curated data) that may be generated through the project
- Estimated project duration
- Legal basis under Art. 6.1 GDPR and legitimation under Art. 9.2
- Funding source(s)
- Peer-review of project where applicable

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- Research ethics application and decision for the specific project
- Machine readable characterisation of project based on controlled vocabulary (e.g. field of research, commercial versus pre-competitive research, data types to be accessed, etc.), (checked for consistency with analysis plan and other project documents by 1+MG Access Office).
  Justification: this will be necessary for using automated tools to check that access requests are consistent with data availability conditions
- Selected data or types of data to be accessed (should be automatically selected from the catalogue browsing and/or Beacon search)
- Information if access request is only upheld if ALL selected records are available (or potentially what a critical % of available records is)
- Status of researcher as white-listed / black-listed or confirmation of validity of request by a signing official at the requester's institution
- Information on required processing environment (required to assess the feasibility of conducting the research in the 1+MG environment, the risk of processing, compile processing agreement and provide a quote on compute costs where applicable)
- Software needed; if applicable: own software to be introduced
- Information on own data that may be uploaded for the research [data types, volume, data origin]

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

10. Addendum: European Health Data Space: Recommendations on a 1+MG - EHDS alignment based on the comparison between the proposed data governance of 1+MG and the draft EHDS Regulation

v.2.0 12 June 2022

Author: Regina Becker

Additional contributors and/or reviewers: entire WP2 / 1+MG ELSI WG

1. Background

The 1+MG ELSI WG has developed under the lead of B1MG WP2 a first proposal for a potential data governance framework to operate 1+MG as a pan-European genomic cohort for research use as envisaged in the Declaration of Cooperation "Towards access to at least 1 million sequenced genomes in the European Union by 2022" of 2018. This proposal still needs to be extended towards other secondary use applications and also to be evaluated in the context of concrete use cases, feedback from the Signatory Countries and in the context of the European Health Data Space (EHDS).

A draft regulation setting the legal framework for the EHDS was published at the beginning of May 2022. Also this text will likely evolve during the negotiations with the Council and the European Parliament.

However, these drafts make an interesting comparison, in particular because they are still flexible and can be further developed according to the strategies chosen by the Signatory Countries. The following analysis is meant to support such strategy development.

2. Comparison of EHDS and 1+MG

Similarities in the set-up of EHDS and 1+MG

- Both aim at secondary use of health-related data for a better understanding of health and better healthcare.
- Both envisage a cross-border processing environment where data remains on the national level as much as possible; this means both aim at technical interoperability of the supporting IT infrastructure.

Some basic differences that originate from the different contexts

- EHDS wants to give access and allow secondary use of all electronic health data held by all organisations involved in the health, care and research sector, whereas 1+MG aims to give access and allow secondary use of a pan-European genome cohort including other health-related data.
- EHDS will be established through European law whereas 1+MG must operate within the existing heterogeneous national legal frameworks in addition to the EU legal framework applying to all structures such as GDPR and Data Governance Act among others.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

This leads to the following basic set-up:

- EHDS must work for all data and therefore rather characterises than curates data whereas 1+MG is selective and aims for semantic interoperability of the data and joint approaches to quality.
- The draft EHDS Regulation provides a legal basis for providing access to data and harmonises national differences for all processing of health-related data in the EHDS including data use whereas 1+MG must establish its own legal toolset and needs to adapt to national legal differences. Once the EHDS is established, 1+MG may benefit to a certain extent from the EHDS legal setup.

The comparison between the EHDS and 1+MG is made based on the current status of the discussion in 1+MG (Declaration, agreed scope, drafts on data governance policy and related policies on incidental findings, vulnerable subjects, communication of research results as well as the transparency and consent policy) and the first draft of the EHDS Regulation, that is expected to undergo further revision. Also some aspects of this analysis are attempted interpretations of the text of the Regulation and will need further clarification in exchange with the European Commission.

3. Conclusion of the comparison

The following considerations are made:

- The EHDS scope of data types covered is wide and 1+MG will currently only comprise a subset of the data. Whereas the data that 1+MG will build on may be largely coming from the same sources in research and healthcare as the EHDS, 1+MG may, according to the current design of the EHDS, not qualify as a data holder. This assumption is built on the definition of a data holder in the EHDS Regulation, where a data holder is a direct stakeholder in health or care, or an entity pursuing research.

  1+MG could rather fall under “Data sharing infrastructures” as referred to in Rec 60 of the EHDS Regulation. These can become an “authorised participant” of the EHDS, a connection node in the EHDS ecosystem that makes data available in the EHDS and that needs to comply with the requirements of Chapter IV of the Regulation.
- Data in 1+MG will come from organisations that will also technically be considered data holders under the EHDS. Therefore, theoretically, all data available through 1+MG could also be requested through the EHDS data request mechanisms. However, it should also be considered that 1+MG could offer alternative procedures of value, such as streamlined and timely data access decision making, well-described and curated high-quality data, easing the obligation of data holders to repeatedly provide access, supporting research reproducibility as well as providing additional transparency and ethical best practices around data collection, sharing and use.
- The EHDS establishment depends on the adoption and implementation of the Regulation. Far reaching implications for primary and secondary health data use are proposed, meaning the final implementation may take several years. 1+MG on the other hand would start earlier based on its own legal instruments such as the establishment of a legal entity, contractual arrangements etc. Therefore, it could also become a first test-ground and pioneer for aspects that are overlapping between EHDS and 1+MG and where they follow the same principles.
- The EHDS and 1+MG have different levels of detail in their current data governance plans. As an example, 1+MG has so far only addressed the data governance for research use However, early case studies have revealed that a healthcare related data use case (clinical reference) will require different mechanisms. On the other hand, EHDS has already worked out plans that could influence the 1+MG decision on currently open

B1MG has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 951724

policy considerations such as dealing with failure to comply with obligations. EHDS and 1+MG could cross-fertilise each other in the coming years.

4. Suggested strategic positioning

We derive from these considerations that EHDS and 1+MG have overlapping, complementary and independent elements. The goal could be to leverage the similar and complementary elements to achieve an economy of scale for Member States and to allow building up capacities already in advance to the EHDS. 1+MG could pioneer here for a potential implementation and serve even as a forerunner for the EHDS by deploying infrastructures and standards that can provide EHDS with a flying start on overlapping elements. According to the current proposed set-up, 1+MG could become an authorised participant within the EHDS. EHDS and 1+MG can also act as sparring partners for each other. These different approaches are not mutually exclusive and could be pursued in parallel.

5. Review of the 1+MG data governance draft in the context of the draft EHDS Regulation

In the following we compare the provisions of the EHDS Regulation with the planned 1+MG Data Governance framework to see if an integration as an authorised participant or a role as pioneer and/or forerunner is possible.

Data & metadata

- *Data: 1+MG as potential EHDS Node*
  1+MG comprises a subset of the EHDS data. However, in 1+MG, the data will be curated into high value datasets and made interoperable within the 1+MG cohort. As such, 1+MG could become a valuable data resource for cross-border integrated and interoperable data in the EHDS as an authorised participant.

- *Scientific metadata: 1+MG as potential pioneer for EHDS*
  1+MG and EHDS both foresee a thorough labelling of data with scientific metadata. The same is taking place in the European Open Science Cloud (EOSC). Here, it would be important that no conflicting developments are pursued. Ideally, the EC should work closely with 1+MG to see how, through 1+MG, metadata models for genomic metadata but also other health data related metadata could be implemented, their practicality and usefulness tested as well as tools for their administration provided.

- *ELSI metadata: 1+MG as potential pioneer for EHDS*
  1+MG will need extensive metadata related to ethical and legal aspects. Some of these may be less applicable for EHDS because in 1+MG, such metadata are used to bridge between legal and cultural differences, and capture different collection and consent contexts. The situation in the EHDS is different. National legal differences are intended to be harmonised. Currently, it is also foreseen that data inclusion and applicable purposes for secondary use are defined by law (the EHDS Regulation) and therefore homogeneous. However, the rights of data subjects under the GDPR may modify the range of purposes. The level of influence of citizens on the purposes for which their data are used has still to be clarified. It is likely that countries that have traditionally built on consent for secondary use may want to modify the current approach towards an opt-in for data use or, at minimum, an opt out. Such an option would also be in line with respecting the essence of the fundamental rights and freedom of the data subjects. Where this results in a more complex data use condition management in the EHDS, again, 1+MG could pioneer the use of such ELSI metadata.

Purposes

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

- *Potential conflict between 1+MG and EHDS*
  The scope of purposes foreseen in 1+MG is more limited as compared to the EHDS, where e.g. also education and training is considered. It was an explicit decision by the Signatory Countries to stick to the purposes defined in the Declaration to avoid a function creep.
  Integrating 1+MG into the EHDS could be ethically problematic if the data availability for all purposes becomes mandatory.
  The EHDS justifies the broadness of scope with the strict safeguards: data minimisation, transparency, and secure processing environments. However, there are concerns in the 1+MG Group that overwriting existing consents based on which data were included into the 1+MG may lead to an erosion of trust.
  This is a dilemma that will not only apply to 1+MG but also to other existing collections. This may therefore become rather a general topic in the negotiations between the EC and the Member States. Nevertheless it could also be addressed in the further discussions between the EC and the 1+MG Signatory Countries.

Data governance

- *Governance bodies: 1+MG to allow building of capacities*
  Both 1+MG and EHDS foresee a central coordination point in each country. For the EHDS, these bodies will have to be established by law based on the Regulation. In the meantime, 1+MG could already lead to the establishment of capacities in existing entities or entities that are being created following the Data Governance Act. Administration and communication channels with data holders can already be developed through 1+MG. Within 1+MG, countries can also start to establish and operate data access bodies who are responsible for data access decision making. Overall, there will be more flexibility in 1+MG with respect to the responsibilities of the various entities (central coordination point, permit authorities and ethics bodies, data holders, IT infrastructure providers), allowing countries to gradually move between different responsibilities and decision making structures. 1+MG may also allow countries to keep some review and decision making power with the local data holders in research who collected the data while at the same time have a high quality review process on the central level as an alternative to EHDS procedures.

- *Data inclusion: 1+MG pursues independent procedures*
  In 1+MG, data will be included on a more heterogeneous legal background than in EHDS where all data are included under the same legal requirements. This makes their legitimate inclusion and administration in the 1+MG more difficult, which will also be reflected in the procedures and contracts required. Here, 1+MG will pursue its own policies, which will be necessary because of the different national legal contexts and the lack of an overarching legal framework.

- *Data access management: 1+MG offers suitable tools for and a streamlined procedural alternative to the EHDS as well as suggestions for relevant features still to be considered in the EHDS governance*
  A main feature of 1+MG will be a centralised support of Member Countries and their stakeholders for the access management, including data selection, access workflow and documentation tools. These tools will also be largely relevant for the EHDS and could be adapted to the EHDS setup. The 1+MG data access administration tools would save many efforts if they will not be duplicated but also deployed for the EHDS data access administration.
  The proposed central access decision making will provide data holders / permit authorities with documentation of the analysis of the request and a decision that the data holder / permit authority can follow or veto. This will avoid the pursuit of the same analysis across several decision making bodies. This feature may be more relevant and advantageous for 1+MG due to the more heterogeneous access and use conditions.
  1+MG data governance builds on experience in the research field and considers practical

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

aspects but also ethical considerations, e.g. around vulnerable individuals or groups. This experience could be valuable for the further elaboration of the EHDS access governance.

- *Data use: 1+MG as pioneer for and source of a secure data processing framework and tools for data use*
  Both 1+MG and EHDS need permits and/or contractual tools, where 1+MG contracts may have to be more elaborate as they cannot rely on a law which regulates major aspects of data availability and responsibilities. However, there are overlapping elements and for those, joint clauses should be introduced where feasible. Working these out together rather than having to change and adapt would be valuable and save efforts.
  A technical implementation of an authentication and authorisation infrastructure (AAI), suitable to provide targeted access will be relevant for EHDS and a close collaboration on such AAI should be pursued. The same applies to the secure processing environment alongside an audit and certification framework. 1+MG could ultimately provide EHDS with an already tested and evaluated system.
  A potential concern is around the EHDS approach on transfer to third countries. EHDS may impact 1+MG's possibility for international cooperations. This will need further exploration with the EC.

- *Results and enriched data: 1+MG to pioneer enrichment and give inspiration*
  Both 1+MG and EHDS consider keeping enriched data for future use. However, not all enriched data may be useful for keeping. Where data is worth keeping, this needs versioning alongside suitable labelling with metadata on the enrichment. Managing an increasing number of datasets modified by longitudinal growths, feedback on enriched data from projects and further adjustment to data subjects' rights exercise will be a complex and non-trivial undertaking. Therefore, 1+MG requires the user to inform about enriched data but will decide case by case if it is worth keeping and administrating the enriched data. This could provide valuable input for an "Enriched Data Policy" in EHDS, where the management of enriched data will be even more complex due to the involvement of data holders that may not even have an interest in the enriched data and their availability.
- *Data archiving: EHDS may take inspiration from 1+MG*
  Whereas EHDS currently only foresees saving the assembly request for the data provided to data users, 1+MG plans a scientific data management with versioning of datasets that consider changes due to longitudinal updates or rectification and erasure following data subjects' exercise of their rights. Where data users pursue a containerised analysis of the data, this could easily be reproduced where required. The possibility for such reproducibility is actually required for research ethics and integrity and should be possible for at least 10 years following a publication.

- *IT infrastructure: 1+MG as a forerunner for the EHDS*
  1+MG aims to implement at least the same technical and organisational safeguards for the IT infrastructure as EHDS. Ideally, the 1+MG IT infrastructure should be developed and set-up in close exchange with the EC to avoid double investment by Member States. 1+MG will then allow testing the IT infrastructure and could provide EHDS with already deployed and usable systems.

Legal setup

- *Legal basis: 1+MG will be impeded by heterogeneous national legislations and may eventually benefit from the EHDS implementation*
  The EHDS will provide a legal basis under Art. 6.1 GDPR for the data availability and provision by the data holders and the Health Data Access Bodies (DABs) as well as a legitimation under Art. 9.2, which will overwrite limitations on the national level based on Art. 9.4 GDPR. These provisions will not cover the 1+MG case and own solutions have to be found.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

While EHDS users will have to provide their own legal basis under Art. 6.1.e or 6.1.f GDPR, the EHDS provides them also with the legitimation under 9.2 and thus enables a straightforward situation with a higher legal certainty. Also 1+MG will benefit from these provisions once the EHDS is established. Until then, the data use in 1+MG will not be as efficient or as complete as it would be ideally because not all countries have implemented Art. 9.2.j GDPR for all relevant stakeholders. Also the requirement for consent or cumbersome authorisation procedures introduced under Art. 9.4 GDPR by the Member States will limit the efficiency and impact of 1+MG without the EHDS.

- *GDPR roles of stakeholders: conceptual differences that will require resolution*
  In the EHDS Regulation, the entity deciding on the access request is seen as joint controller with the user and, where the data are processed directly by the DABs to generate aggregated data for users, the DAB is even seen as sole controller.
  1+MG has a different perception of the legal situation: in the processing chain, the disclosure is a separate element from the processing by the users for their purposes. Subsequently, the entity or entities deciding on the access request are controllers for the data disclosure. The users on the other hand determine the data use purpose and are therefore sole controller for the processing for their purpose. For more details, see the corresponding publication.
  This role assignment can also not be changed by the Regulation because, following the EDPB, the assignment of roles is factual, not based on legal assignments. More alignment between the EC, the regulators, the Member States and 1+MG will be needed on this topic.

- *Data subjects' rights: 1+MG may provide inspiration for EHDS*
  1+MG has performed considerations of how data subjects' rights can be complied with but also how these may have to be limited. Currently, a more active information provision to data subjects is considered, subject to their agreement. This includes information on relevant clinical studies they could participate in, information on data use as well as stakeholder adapted communication on selected research results.

  In particular for the information under Art. 14, it needs to be considered that, where data are shared according to data minimisation, it will not be possible for individuals to find out if or if not their data was used for certain research projects or other secondary use. However, transparency is key for building trust, which is why the 1+MG ELSI WG recommends investing into a more active information systems for citizens whose data are made available for secondary use.

  Limitations of data subjects' rights have to be foreseen though where data are included in already very advanced research projects or after results have been obtained and published. Reproducibility requirements mean that an exercise of objection to the processing or a request for erasure or rectification may only be exercised prospectively. Data subjects need to be made aware of such limitations. To be able to limit the rectification in previous dataset versions, 1+MG may have to rely on legislation being introduced through the EHDS, which is why an exchange on these aspects with the EC will be important.

A more detailed analysis of the differences and similarities between 1+MG and EHDS can be found below.

6. Recommendation by 1+MG ELSI Group

The 1+MG ELSI Group recommends taking an active approach to optimise the interaction between 1+MG and EHDS for a joint pursuit of common goals. In particular, 1+MG should offer to contribute valuable components for EHDS. There are many opportunities for the EHDS to advance its goals and plans by utilising the 1+MG initiative and its work. On the other hand, the 1+MG pan-European genomic cohort will only reach its full potential when it can make use of the

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

legal framework offered for secondary use. Ultimately, the Member States will want to optimise the efforts in this context and prefer integrated and joint solutions rather than duplicate investments and operating costs.

There are only very few real conflicts between the current 1+MG and EHDS approaches. These could likely be resolved in joint explorations. Nevertheless, there is still a lot of conceptual work needed on both sides. Here, both groups can benefit from each other as 1+MG and EHDS come from complementary backgrounds (1+MG with a stronger research and implementation component, EHDS with healthcare and regulatory experience). An active and close cooperation will be key for the success of both endeavours. For this, EHDS and 1+MG will have to be linked on three levels:

- The operational level, bringing together the different working groups and projects, leading to jointly agreed division of tasks and interaction on their accomplishment.
- The political level in the Member States, making sure that activity lines that may be in the responsibility of different departments or ministries will aim for a joint implementation.
- The conceptual level where the three main driving forces DG SANTE for the European Health Data Space (EHDS), DG CNECT for the 1+MG initiative and DG RTD for European Open Science Cloud (EOSC) are ready to pursue joint solutions.

While some need for further alignment along the data governance side have been indicated here, the main work will be on the scientific data and metadata models as well as the technical implementation level, going far beyond the work of the ELSI Working Group. The message and recommendation of the ELSI Working Group is that the possibility for synergistic and integrated developments of 1+MG and the EHDS is possible and real and should be leveraged for efficient use of resources and for maximum impact. We recommend to give a mandate to the relevant stakeholders on the operational level from the 1+MG Coordination Group to JA TEHDAS as well as other projects funded by SANTE, CNECT and RTD relevant for the implementation to develop a joint work programme where solutions will be pursued jointly and synergistically. From that work programme, also the need for additional targeted R&D could be derived and lead to the filling of currently still existing gaps in a targeted manner. Only if the efforts across the different activity lines are brought together, Europe will be able to accomplish the full potential that is in the secondary use of health-related data.

7. Detailed analysis of the 1+MG draft Data Governance Policy and the draft EHDS Regulation from 3 May 2022

Subject	1+MG	EHDS – draft regulation	Comment
Scope of data holders	Data holders voluntarily wanting to contribute data	All data holders in the health or care sector, or performing research in relation to these sectors, as well as Union institutions, bodies, offices and agencies holding data as defined in Art. 33.1 except micro-enterprises [Art 2.2.y / Art. 33].	Definition may not include infrastructures as they are neither a healthcare stakeholder nor performing research. On the other hand, data from infrastructures have often been obtained by data holders falling under the EHDS. That may create different access routes to these data

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Scope of data	Genomic data in combination with clinical phenotyping; additional datatypes possible and encouraged. Minimum datasets will be defined depending on context. Data must be curated into defined data models and ontologies before inclusion.	Defined list including healthcare and health administration data, regulatory data, health research data, molecular data, viral data, environmental, social and lifestyle data. More data can be included through EC delegated act or by national law or voluntarily by data holders. [Art. 33] The EC may define minimum datasets for cross-border datasets. [Art. 58] No harmonisation and/or curation of data is required.	1+MG data will fully be covered by EHDS but comes with the advantage of being already curated to interoperability
Metadata	Detailed harmonised scientific metadata on data quality, data provenance and data content [precise scope still under development]; ELSI metadata providing information on permitted data use and other relevant aspects (vulnerability of subjects; no-return of incidental findings etc.)	Obligatory metadata on information concerning the source, the scope, the main characteristics, nature of electronic health data and conditions for making electronic health data available. [Art. 55] In addition, quality and utility label is introduced that covers elements of data documentation, technical quality, data quality management processes, coverage of the data, information on access and provision of data, and data enrichment. [Art. 56]	At current stage, 1+MG seems to have a more detailed metadata framework.
Mode of inclusion	Voluntary submission or inclusion through governmental decision within the scope of ethical and/or GDPR consents given and, where applicable, national or regional law	Based on Regulation; all listed data types from all listed data holders (healthcare stakeholders, public sector bodies, public and private research stakeholders etc.)	Inclusion in 1+MG only where there is incentive to share data for 1+MG and if the data fulfil the inclusion criteria; heterogeneous use conditions. Much more streamlined in EHDS

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

			but may meet with resistance of data holder / Member States. However, the different approaches are compatible as the setup of 1+MG and EHDS is different.
Data catalogue	Central data catalogue through 1+MG that allows data discovery across the entire catalogue. National data catalogues are only optional.	National data catalogues as well as an EU data catalogue that connects all the national data catalogues. No record-level search currently foreseen.	**Data catalogues will have to be interoperable** **EHDS will have to solve the challenge of data filtering for data minimisation**
Scope of purposes	Only positive list with explicit purpose limitation and enforcement of this list through centralised oversight for harmonisation: scientific research with the scope as defined in GDPR for better health and innovation in healthcare; policy making including quality management; reference for healthcare (secondary healthcare use).	Research, development and innovation, including training, testing and evaluating AI applications, education and teaching, official statistics, public health purposes and broad support of policy development and general support of the mandate of public sector bodies around health – currently all applicable without exception [Art 34] Some purposes that can lead to harm of people, groups or society as well as onward transfer of data is explicitly forbidden. [Art. 35]	Concern that purposes are too broad and overwrite consents given by data subjects; re-purposing of existing collections may erode trust and may lead to requests for data erasure
Access governance [purposes]	Different access governance for different purposes; so far only research access governance compiled.	Access governance for all purposes; details still to be defined. [Art 45]	**Currently, there is a serious conflict as 1+MG data will often be included based on consent where an overwriting of such consent may be ethically difficult.** **There is also currently a "fast track access" for secondary use in healthcare planned**

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

			in 1+MG that may not meet the EHDS access governance.
Bodies involved in access governance	Definition of bodies on the central level: • 1+MG Access Office to support users • 1+MG Data Access Committee (DAC) with a composition determined by 1+MG Members and in accordance to area of research On the national level: • Central coordination point for coordination and support • Data holders • National or local DACs or Ethics Committees • Definition of functions that countries have to fulfil to make data available and decide on access; each country is free to define how competences are distributed among these actors.	Definition of Health Data Access Bodies (DABs) by MS that assume by law (inseparably) both the functions of permit authority and competent body under the DGA, which includes e.g. supporting and supervising data holders, linking data, making data available, analysing data, interacting with the users, providing information on data use and performance etc. A DAB must also support the development of AI systems. A country can nominated several DABs with a responsibility for different data collections. [Art 36/37] Data holders as defined above.	1+MG data access bodies could in principle be independent of the national structures defined in the EHDS. However, the flexible setup in 1+MG may allow to prepare the EHDS bodies that have to be established but it could also offer alternatives where data holders should remain involved in access decisions.
Obligations of data holders	Data holders in 1+MG… Must ensure that they can legally make their data available through 1+MG and can be used in line with the limitation of data subjects rights of 1+MG. Must curate and characterise their data where this is not done through other actors on the national level.	Data holders in EHDS… Must communicate data description. Must provide information for the quality label. Must make data available within 2 months if requested through DAB. Must make non-personal data available to all users through trusted open data portal with robust, transparent and	1+MG will require less of a continuous involvement of data holders in the access provisions than EHDS. In both scenarios, however, information provision on the data to be shared is key. 1+MG will fund data curation and annotation whereas in EHDS only the work

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	Must agree to the terms and conditions under which data are made available in 1+MG (including hosting in 1+MG IT infrastructure, long-term archiving, IP rules etc.,). Must provide additional information on data if clarification is needed for data access request. Must make data enriched through 1+MG use available under the same conditions. Must not withdraw their data unreasonably.	sustainable governance and a transparent model of user access. In place. [Art. 41] Must provide information on compliance with obligations [Art 43]	for data provision may be reimbursed. In 1+MG, data holders may have a more active role in the decision making without the full scope of services and requirements that are associated with such decision making in the EHDS. 1+MG could therefore allow to keep data holders in the decision making process to the extent that they can participate. This allows high quality access decision making and data provision in 1+MG without "disowning" data holders. This could be an attractive option for Member States (and data holders).
Access decision making	Decision by 1+MG central DAC with possibility for veto by data holder and/or national permit body. Access decisions in 1+MG are more complex because of the varied conditions under which data are included (national legal and ethical rules, consents given etc.) Includes ethics reviews.	Decision by DAB is mandatory. [Art. 45] DAB needs to check if the request falls into the purposes listed for the EHDS, if the requested data are suited and if the conditions as required by the Regulation are fulfilled. [Art 46] Only where data exclusively by one Data Holder is requested the Data Holder may decide if it agrees to assume such function. [Art. 49] Where data from cross-border registries and databases is asked, the DAB where the data holder is registered, decides. In case of joint controllers, the DAB in one of the countries of the data holders decides.	As an authorised participant of the EHDS, 1+MG could have its own decision making procedure. The 1+MG decision making procedures may be attractive for the Member States (see also the row above).

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

		In a network of registries, the DAB where the coordinator of the network is located, will decide [Art.53]	
Access request procedures	Central 1+MG unit to serve as one-stop-shop for users, organisation of information provision across 1+MG, central data catalogue etc. 1 access request form for research; different forms for other purposes foreseen. Central Coordination Point acts as a single information point for activities in the respective country with a coordination of national activities, optional national data catalogue, general information provision for countries, coordination of response on data access requests etc. Timeframes are planned to be shorter than in EHDS There is a central support by 1+MG Access Office for all access related user interactions. Detailed description of access procedure for research Currently only provision of data in a secure processing environment is foreseen, no generation of anonymised (aggregated) data according to user requests are planned.	User can address any concerned DAB who then distributes the access requests to the other concerned access bodies. [Art. 45] Where a data holder decides directly on an access request, it needs to inform the DAB within 3 months on the request and the permit issued. [Art. 49] Once access decision is positive, the data provision is requested from the data holders. The DAB is in charge of any linking of data, reduction according to data minimisation and pseudonymisation where applicable. It must also provide aggregated data and/or anonymised data depending on the request [Art. 47, 48] 1 access request form for all purposes [Art. 45] Public sector bodies and EU institutions can for the pursuit of their mandates ask for updated data. [Art. 45] Timeframes for procedures: 2 weeks for request forwarding 2 months + 2 months for access decision 2 months + 2 months for data provision by holder	**Likely conflict for scope of information required if every authorised participant has to use the same form in EHDS: 1+MG will ask more information from research users, less information from healthcare users.** Advantage of 1+MG: may offer faster and cheaper access.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

		2 months + extension for pre-processing [Art. 46] For "data requests", i.e. anonymised, also aggregated data, the DAB can have 2 months for assessing the request and 2 months if possible for the provision of the result	
Data location	Data stored in 1+MG secure infrastructure	Data stored by data holder	Different but compatible; there does not seem to be an obligation that data remains with the data holder.
Data preparation and linkage	Data are "ready to be shared". Preparation of relevant subsets of data takes place nationally through IT infrastructure provider and are already largely preselected by the data user through data selection tools in the application process.	DABs collect, combine, prepare, pseudonymise and disclose data for each specific request [Art. 37]	Advantage of 1+MG
IT infrastructure	Secure IT infrastructure has to be established by each country. This may or may not be part of a National Coordination Point. Contractual arrangements (data processing agreements) are made between 1+MG and the IT IS providers for the data use. 1+MG IT infrastructure must at minimum be externally audited (to be organised by National Coordination Point). If data are pooled in a 1+MG infrastructure, this IT infrastructure must be certified.	Secure IT infrastructure is to be provided by DABs. Standard technical safeguards are suggested. Regular audits are to be "ensured" by DAB (unclear if this means it's internal audit or external audits have to be organised) [Art. 50] Where data from more than one DAB is requested, the EC may provide the secure IT infrastructure for the processing.	Principal approaches seem to be the same; synergistic deployment should be possible
Contractual framework for data access	Non-negotiable contract, signed by 1+MG Central Unit on behalf of the data holders, which	1 data permit issued that specifies the types and formats of data accessed including sources,	Contractual situation in EHDS is not clear and it is not yet foreseeable how this

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	defines the rights and duties of all parties as well as a processing agreement in accordance with Art 28 GDPR with the IT infrastructure providers.	purpose for which data can be used, duration of the permit, information on tools available, fees to be paid as well as any specific condition for access. [Art. 47] In case of cross-border access, one DAB can issue a permit on behalf of all concerned DABs following their access decision [Art. 54] In addition, joint controllership arrangements are established [Art. 51] Contractual commitments of users and data holders are mentioned in Art. 39 as part of the reporting. Art. 53 refers to contractual arrangements for data access (in addition to data permits).	will impact 1+MG; as authorised participant (not DAB) it may not fall under the same legal tools, in particular as it needs to arrange more aspects contractually than EHDS stakeholders.
Request for aggregated statistics	So far not foreseen	DAB must assess request within 2 months and produce result within 2 months where possible. No limitation / information on when this is not possible. [Art. 47]	**Function currently not planned by 1+MG; not clear if it would be required from authorised participants.**
Services for public sector bodies and EU institutions etc.	No special status	Such public entities will get access to data without a permit, without providing information on results, without time limit, without having to provide incidental findings etc. Timelines of data access procedures are unchanged though. [Art. 48]	**1+MG does not foresee exceptions for public sector bodies or EU institutions; such favouring does not seem justified and may create a serious rift between 1+MG and EHDS.**
Data minimisation	Data minimisation is aimed at with record-level search function to allow targeted access requests and information in the	The DABs need to ensure that only relevant data is given access to. DABs must provide anonymised data where the users can achieve	Creation of anonymised data for research is not foreseen in 1+MG because genomic data are difficult to anonymise.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	data access request form. Necessity of requested data is checked by the 1+MG Central DAC. Only those data elements required are given access to by the 1+MG IT infrastructure providers. It is not planned to generate anonymised data for research purposes as genomic data is difficult to anonymise.	their purposes with anonymised data. [Art 44]	
Data use	Data can only be accessed in a secure environment provided in the respective member country. Requirements for the processing environment will be established by 1+MG and all IT infrastructure providers must go through external audits. Where the data analysis requires pooling, data can be pooled in one of the 1+MG IT infrastructures that is certified. Users can bring in their own data into the 1+MG IT infrastructure. 1+MG central office needs to ensure that only non-personal data is exported.	Data can only be accessed in a secure environment provided by the DAB. Users can upload own data to the system. EC implementing acts will defined the precise IT IS requirements. DAB needs to ensure that the user only downloads non-personal data. [Art. 50]	Compatible
Data access temporal limitation	Data access period is limited to the user's needs. There is a possibility to cease and resume access. A regular interim extension may be foreseen as safeguard.	Data access period is limited to 5 years maximum. An extension for up to another 5 years is possible upon justification. [Art. 46} There is a possibility for cheaper interim storage in a storage system with reduced capabilities. Data in the secure processing environment are deleted 6 months	While not being custom, it may happen that a research project will continue over a longer time period. **There will be a conflict between 1+MG where the access is depending on the requirements for the purpose and the EHDS where there is a**

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

		after the expiry of the permit.	**purpose-independent cut-off.** **However, it may be reconciled by the 1+MG not deleting the data and results and simply asking for a renewed application.** Any ceasing and re-establishing of data and analysis environment seems more straight forward and cheaper in 1+MG.
Results	Obligation to acknowledge 1+MG and the data sources (data holders) No IP on data that limits data use	Data user must make results public no later than 18 months after completing the data processing. [Art. 46] Obligation to acknowledge data sources [Art. 46]	Making results public within 18 months may not be compatible with all access requests and does not fit to 1+MG approaches.
Incidental findings	Users must inform on any clinically relevant finding. Member countries and/or data holders must have procedures established to inform affected individuals. Individuals must have a right to know or not to know. It can be decided locally or on the national level according to legal and/or ethical practice if / which findings will be communicated.	Data users must inform the DAB of any clinically significant findings that may influence the health status of the natural persons whose data are included in the dataset.[Art 46] The DAB should inform the concerned person [Rec 49] or the healthcare professional [Rec 44]	1+MG seems to have more elaborate plans around incidental findings at the moment and could inspire EHDS
Data archiving	1+MG must guarantee that datasets are versioned and kept for reproducibility of research results for a guaranteed time frame to be agreed with the user.	Data retentions can foresee longer storage for peer-review procedures but do not allow the long-term coverage. After 6 months following the expiry of the permit the data are deleted. However, a user can ask for the formula on the	1+MG foresees data archiving, which is an advantage for researchers and an incentive to go through 1+MG

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

		creation of the requested dataset to be stored by the DAB. [Art.46] It is not likely though that the data holders will be able to restore exactly the same data again considering changes to the dataset over time.	
International users	Only research use is open to international users. However, for non-members the use may be limited or subject to special conditions	Transfer of non-personal data provided by the DABs based on selected data types are deemed highly sensitive and may be subject to special protective measures that will be defined by the EC. [Art. 61] Limitations of non-personal data transfer or governmental access to non-personal data in the EHDS that could create a legal conflict. [Art. 62] For international transfer of personal data, MS are allowed to introduce their own limitations.	**Implications still unclear but there are concerns that the EHDS rules will reduce the utility of 1+MG for necessary international collaborations and data use.**
Reporting and information provision	Reports on data use including Art. 14 GDPR transparency requirements, performance and results from data use as well as general information on 1+MG are provided by the central support office and made also available to relevant actors on the national level through suitable APIs where necessary. For the data research use, only lay summaries will be made available. The national coordination points will make relevant information available on the national level and,	DABs must report on data use statistics, various performance parameters, audits, revenues and quality labels assigned and Art. 14 GDPR transparency requirements as well as their general role. [Art. 38, 39] For the data re-use the full data access request (application, permit, data request), has to be made public. [Art. 37]	Mostly compatible if infrastructures as authorised participants have the same obligations as DABs. **However, 1+MG would not make the full access request available for confidentiality reasons. Here, 1+MG and EHDS are not compatible.**

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	where applicable, in the relevant national languages. The national coordination points can further complement information beyond the centrally provided information.		
Penalties	Access request deletion in case of no-response by user. Exclusion of users from 1+MG access in case of severe violation of contractual obligations or abuse of search tools. **[no details worked out yet: severity of violations; duration of exclusion; effects on natural person versus institution.** **Also missing: any actions in case of non-performance of 1+MG central support office, national coordination points or data holders]**	[Art 43] Penalties where • Data holders do not declare their data • Data holders do not provide accurate information on data quality • Data holders withholding data from requested access • Users aim to re-identify data subjects [Art. 44] • Users do not comply with obligations Penalties are exclusion from EHDS for up to 5 years and may for data holders include financial fines for delayed information or data provision	**Further work by 1+MG on penalties still needed.** **Clarification of scope of applicability of EHDS rules for 1+MG still needed.**
Roles under the GDPR	The 1+MG and the decision maker on the national level (data holder; data permit authority) are joint controllers for the data disclosure (in the veto model). In general: the entity deciding on access is controller for the disclosure. The user is sole controller for its data use. The provider of the IT infrastructure is	The DAB is seen as joint controller with the data user and liable until the completion of the processing. [Art. 46, 51]. Where the data holder decides on the data permit, the data holder is joint controller with the user. [Art 49] Where the DAB produces anonymous, aggregated statistics for the user, the DAB is sole controller for the processing. [Recital 49] Where several DABs are involved in cross-border	**1+MG and EHDS have different interpretations of the GDPR that needs a resolution**

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	processor for the data use.	processing, they are joint controllers. [Art. 52] Where the EC provides a secure IT IS for cross-border processing, it acts as a processor. Art. 52]	
Legal basis under the GDPR	If 1+MG is created by a legal act, then it could operate to achieve its mandate under Art. 6.1.e. The legitimation may have to be established through the national situation where the coordination is based. Data holders must establish their own legal basis for making the data available in the 1+MG; this could be consent if no other legal option is available. The data access decision on the national side needs to be covered through 6.1.e or 6.1.f together with a legitimation under 9.2.j (for research), 9.2.h (for healthcare) and 9.2.i (for policy development) The user must establish its own legal basis under Art. 6.1.e or 6.1.f. Once the EHDS is established it may benefit from the 9.2.h/i/j provision [to be confirmed]. Without that, not all users will be able to access the 1+MG for research use and/or users may have to go through special approval procedures on the national level.	Users need to provide justification for their legal basis under 6.1 GDPR. [Art. 45]. The possible options are 6.1.e or 6.1.f [Recital 37] The EHDS Regulation provides the legitimation for processing special categories of data for purposes of secondary use under Art. 9.g/h/i/j Data holders must make their data available under Art. 6.1.c in conjunction with Articles 9(2) (h),(i),(j) DABs operate under Art. 6.1.e in conjunction with Article 9(2)(h),(i),(j) [Recital 37] No legal basis for authorised participants, only for data holders and DABs How about the Art. 9 may be available for data users applying for access through authorised participants [tbc]	1+MG is struggling without a European law that overcomes the Art. 9.4 GDPR heterogeneity. 1+MG will therefore benefit from EHDS
Transparency data subjects	Full information provision as required by	Only generic information provision on data use according to Art. 14 but derogation from the	1+MG foresees more transparency and could inspire EHDS

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

	Art. 14 on the webpages on the level of collection. Individual information through automated electronic communication envisaged where data are pseudonymised and suitable contact data exists.	provision of advance, specific information. [Art. 38] Art. 15 information has not been derogated though.	
Data subjects' rights	Explicit foreseeing that some rights may be limited within 1+MG, most namely: • Erasure of data • Objection to the processing Where the data use has gone through a "point of no return". The right to rectification may only apply to data actively used, not archived.	The right to specific information to the individuals is derogated from [Art. 38] Otherwise no limitations are mentioned and in the obligation of the DAB to inform the data subjects about their rights [Art. 38]	1+MG foresees a limitation or rights to ensure research integrity, which is currently not foreseen in EHDS. However, that should not cause a problem where the derogation is in the GDPR. Derogation from rectification through EHDS would be useful though.
Re-contacting data subjects	Mechanisms must be in place in each member country, either through data holder or through national coordination point organised service Re-contacting can be done only for purposes agreed with the data subjects beforehand. Purposes are to • Inform about incidental findings • Gain more information and/or biosamples • Include individuals in new research studies / clinical trials • Information about data use	Pseudonymisation procedure in in the hand of the DABs. Only DABs can reverse the pseudonymisation. [Art. 44] The only purpose mentioned where the DAB will reverse the pseudonymisation is incidental findings. [Rec 49]	1+MG foresees a broader scope of re-contacting; should ideally inspire EHDS as such possibilities are important for research

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724

Overall participation / membership	Only countries signing for the legal framework are members. Data use for policy making and secondary healthcare use is limited to members.	Authorised participants of the EHDS are National Contact Points of the MS, EU institutions / bodies etc, EU research infrastructures, third countries and international organisations where they provide access to health-related data and comply with Chapter IV and where the EC can oversee their compliance.	Overlapping pool of data holders and data; 1+MG allows for alternative decision making models for data
Sustainability	A legal framework will be decided by the members and may include a joint legal entity. An analysis of different models for the legal framework with advantages and disadvantages is underway. A decision is expected as part of the deployment. The financial sustainability is still under discussion but the aim is to allow free data access as much as possible. No reimbursement / fees possible to charge for data enrichment in the context of data use.	The legal framework and the bodies are established by law and obligatory. Sufficient funds need to be made available to the Member States for DABs to fulfil their functions. [Art. 36] Fees can be charged for services by both DABs and Data Holders where an EC delegated act will provide a framework for determining such fees. Fees can also be charged by users for data enrichment. [Art. 42]	A legal personality will allow 1+MG to act as an authorised member of the EHDS. As fees are not obligatory, the different financial approaches are compatible; where 1+MG decides to operate based on fees, these would have to be following the rules to be established by the EC.

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724