APPROVED BY WG-5 EXPERTS 5.2.2021

Towards a Proof of Concept of Federated Infrastructure for the EU 1+ Million Genomes Initiative

Author: 1+MG working group 5 & B1MG work package 4
Editors: Tommi Nyrönen, Bengt Persson, Ilkka Lappalainen & Sergi Beltran

Summary

Overview presentation

This is a discussion paper to outline an ELSI (ethical, legal, societal, policy) compliant infrastructure (WG-5) for the European Union (EU) 1+Million Genomes (1+MG) data management purposes. The paper is aimed for the signatory member state policymakers. It extends the 1+MG roadmap 'Chapter 6. Infrastructure' in order to achieve the goals set by the declaration "Towards access to at least 1 million sequenced genomes in the EU by 2022". We propose to divide "Infrastructure" into five connected functionalities: *data discoverability*, *data reception*, *storage & interfaces*, *data access management mechanisms*, and *processing*, and analysed 16 scenarios about how member states and the EU could share the responsibility to sustain these functionalities. The initiative intends to build on and extend the existing European e-Infrastructure investments to create 1+MG infrastructure functionalities. We propose that in 2021 we start a Proof of Concept with the rare disease community to explore and find ways to fill in the existing infrastructure gaps. The goal is to eventually document the technical requirements for all the 1+MG use cases in rare disease, cancer, common/population-level diseases and COVID-19 and other emerging infectious diseases, and be able to expand infrastructure activities to broader healthcare and research contexts.

The main outcomes of this report are:

Endorsement request to focus in 2021 to build a Proof of Concept with the rare diseases community use case.

***Recommendation 1 (minimum)*.** Define a data infrastructure based on shared architecture and interfaces supporting transnational access to the existing data sets stored within the signatory states.

***Recommendation 2 (incremental)*.** Move to an infrastructure that provides access to a single, harmonised 1+MG dataset by joining nationally stored data collections together with globally accepted interoperability standards.

***Recommendation 3 (incremental)*.** Provide a 1+MG use case required data discovery, analysis and access to the 1+MG dataset as a service either on national or transnational data processing platforms.

Table of Contents

Summary 1

Table of Contents 2

1. Introduction 3

2. Initial scoping 4

3. Infrastructure functionalities 6

4. Interoperability levels 8

5. Organisational options 8

6. Next steps 12

- **6.1. What will a Proof-of-Concept look like?** 12
- **6.2. How to measure success?** 13

7. Appendix 15

- **7.1. 1+MG Infrastructure timeline (set December 2019)** 15
- **7.2. Gaps identified during writing process** 15
- **7.3. Glossary of Terms Used** 16
- **7.4. Contracting needs** 17

1. Introduction

This paper was written for the European Union (EU) 1+Million Genomes (1+MG) signatory member states policymakers. It describes an European model for managing large amounts of genomes and clinical information collated from the patients and consented for primary care and secondary use in research. The proposed infrastructures, processes and technical solutions for managing various users are meant to be ELSI compliant. During the writing of the first version many ELSI questions are still unsolved. Working group 5 (Infrastructure) and Working group 2 (ELSI) are collaborating to resolve them iteratively during 2021-22 with the assistance of the whole initiative. The discussion on this paper expands the 1+MG roadmap *Chapter 6. Infrastructure*[1].

We assume that existing European e-Infrastructure organisations and relevant clinical use cases representing stakeholders will be leveraged to construct the proposed infrastructure. Therefore, the focus on this paper is on the added functionalities required to meet the demands of the 1+MG use cases. The 1+MG initiative has already defined the following use cases each with research and clinical drivers: rare diseases, cancer, complex/common diseases and infectious diseases with a current focus on COVID-19 pandemia.

One of the key factors for the proposed infrastructure is interoperability within the European framework. Importantly, it must enable discovery of harmonised data sets compliant with common governance rules. The governance rules determine the conditions for the reuse of collated data derived from digitized human samples across the current and future 1+MG use cases.

The future organisation of the collaborators responsible for the 1+MG infrastructure is influenced by data policies that signatory states want to impose on the data or a specific sensitive part of data. The unclear state of practices has been feeding fears of security breaches and data misuse among the general public that has translated into national policies that restrict the access and flow of information from data controllers. A well-defined infrastructure ensures respect for policies defining who may access data, for what purposes, and under what conditions (e.g, processing only within secure computing environments), through access controls, tools to track data processing, and contractual or other safeguards. At the same time, the 1+MG collaboration across technical and legal experts allows the policies to be developed iteratively so that the infrastructure can help in policy development, deployment and impact assessment.

Possible models of data access that countries might follow include bringing 1+MG compute tasks to the data within national (local) computing environments, access the data on a national cloud or streaming data to a secure cross-border (cloud) processing environment. Technically, data can already be streamed from any part of Europe to where it is needed across the European Internet backbone to i.e. secure cloud service. Streaming should ensure an appropriate level of security, also taking into account variation between the chosen level of security in each country. Once a specification for the required technical security measures are documented and agreed on, multiple European computing centres can carry the burden of providing the computing power for the 1+MG.

[1] https://ec.europa.eu/digital-single-market/en/news/roadmap-1million-genomes-initiative-now-clearly-illustrated-new-brochure

Existing European infrastructures and national/regional data hubs are expected to offer compute, data and network backbone for the 1+MG initiative infrastructure. The initial analysis of the 1+MG use cases indicates that a federated data management and analysis framework (technical and contractual) are needed to fill in the gaps of the existing research and heath e-Infrastructures to meet the goals of the 1+MG initiative. Therefore, both development and operations (DevOps) are in scope for the infrastructure 2021--2022.

2. Initial scoping

The minimal outcome of the 1+MG infrastructure is when a “virtual cohort” is discoverable as a data federation across the signatories. Access to these data can be requested from a single point of entry supporting both research and clinical users. The data management operations are governed jointly by the member states but importantly delivered on national data infrastructures (“data hubs” in Figure 1) that may differ from one member state to another. Regardless of the national solution it must comply with the agreed interoperability standards and expose the national data as part of the European data network in a secure manner. While the maintenance and production services are operated at the national level – development of common standards, access policies and governance model are part of the joint European efforts.

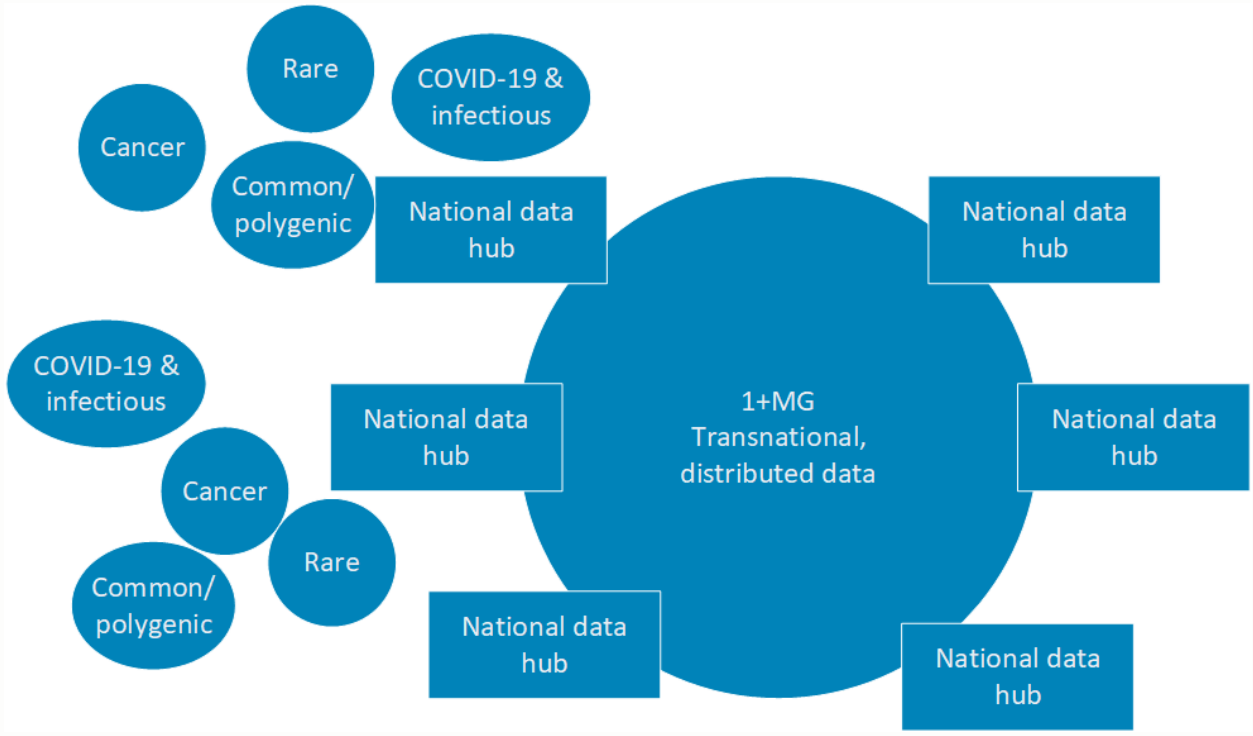


Figure 1. EU 1+ million genomes is a collaboration that is envisioned to happen between distributed national/regional data hubs that will use an interoperable infrastructure solution. Countries are free to organise their data hub, but a single (e.g. national federation) point to connect to the European 1+MG infrastructure is highly desirable.

In short, the data hubs will enable:

(1) A joint **collection of genome sequences and (clinical) phenotypic information** from European subjects that have signed a consent authorizing their data to be used for the 1+MG initiative research and clinical use cases (as part of the national data collection).
(2) Ensure national data collections to comply with 1+MG data analysis and data quality (WG3 and WG4) proposed standards for collecting genomes and clinical information in a way that supports use of these data as part of the European data set.
(3) Provide secure data management services at scale, such as data storage capacity.
(4) Enable standard data access control tools and protocols according to the consent given for the respective data set.
(5) Deployment of shared interfaces (APIs) to support data federation including authorized access and use case-specific semantic tools on data discovery and analysis.

For clarity, we propose the following points to be out of the scope for the 1+MG infrastructure:

(1) A comprehensive European **data-sharing model** that facilities national data collections to be collated, maintained and provided from a single location. We have included this model in our analysis for the purposes of comparison of long-term options. A more realistic model would be based on federated data analysis techniques “bringing compute to data” whereby the member states’ data do not move from the respective national or regional repository.
(2) Semantic interoperability is not in scope for infrastructure expertise as it requires deep knowledge of the structure and meaning of the data being managed and processed (e.g. cancer and rare disease data types). However, technical infrastructure must be able to support the semantic interoperability across the data required by the use cases.

3. Infrastructure functionalities

This report analyses scenarios derived from infrastructure functionalities shown in Figure 2. These functionalities are based on the working principles of the existing European Genome-phenome Archive (EGA) and its federated extension FEGA that ensure secure and consented secondary use of sensitive human genomic data.

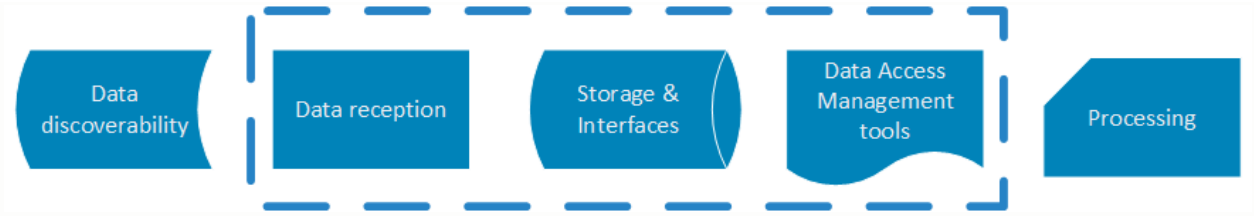


Figure 2. Core infrastructure functionalities inspected in scoping. The 1+MG infrastructure work focuses on developing and integrating the functionalities within the dashed line while keeping the infrastructure interoperable with data discovery and processing services that facilitate access to the 1+MG data. Each of the functionalities can be provided at regional, national, European or even global levels.

These functionalities may be implemented at the regional, national, transnational or global level and therefore it is important to note:

(1) National services must implement their strategy for connecting to the transnational level and apply European principles (e.g. FAIR) and regulations (e.g. GDPR) for data sharing.
(2) Member states may either produce functionalities themselves or rely on services e.g. on the European digital single market
(3) On the global scale, the European Internet area must be controlled so that no sensitive data will leak out from Europe without an appropriate gateway process.
(4) Although the basic requirements for the proposed infrastructure are shared for research and clinical data, the current legal/ELSI requirements are driving separate installations for each use case in many European countries. What is proposed in this document is aiming for a shared infrastructure that is not necessarily used as (direct) part of direct patient care but can facilitate discovery of information defined by each clinical case, and gives means to proceed to use that data for healthcare decision making support functions. E.g. ask a question to identify a rare disease patient with sequenced genomic mutations and observed clinical manifestations. Infrastructure should support implementation of consents, or other legal basis for managing data access according to GDPR and national laws, and suggest a standard for communicating incidental findings communicated to patients through appropriate professionals.
(5) Reliable identity and access management across national borders is essential for success.

Descriptions of the functionalities:

Data Discoverability – Provides the public visibility to the combined 1+MG dataset by making selected descriptive metadata searchable. Data discovery service will collect summary level (but not sensitive) descriptions of all data available from the signatory states. It will link directly to Data Access Management where users can apply for data access.

Data Reception – Uniform processes (such as quality control and standardisation) to receive (download) or access (API) both data and metadata in a consistent way enabling infrastructure to adhere to global standards and principles (e.g. GA4GH, FAIR) for genotypic and phenotypic data. Data reception means logically describing datasets to an extent that they can become actionable on the 1+MG infrastructure even if they are stored nationally or locally.

Storage & Interfaces – Organisations store data and offer interfaces (APIs) following international standards that form the technically interoperable infrastructure backbone. Service building is assumed to leverage national and European investments in e-infrastructure capacities. Storage & interfaces need to provide techniques ensuring data privacy and confidentiality.

Data Access Management Tools – Sensitive data needs specialised services to manage data in an ELSI compliant way. These include e.g. central access portal, central access review process, single collaboration/data use agreement. These tools store user applications for data use and the decisions from the data controllers, i.e. data access authorisations within the legal framework, as well as APIs and standards to communicate access rights to downstream infrastructure services. Together, these tools and processes facilitate and audit secure access for users on a chosen Data Processing service platform.

Processing – Local, high-performance and cloud computing must fulfil appropriate security standards to provide processing capacities for human data coordinated in 1+MG. Distinct processing events happen on the infrastructure: Localisation of the data and code to the appropriate platform (local or distributed) and data analysis by the individual data user who has acquired the access rights for the intended data use.

Further functionalities for infrastructure are actively considered. For example, a catalogue of synthetic datasets that reflect as close as possible the features of the real data is prepared in B1MG WP4 in early 2021.

4. Interoperability levels

A clear distinction between national and European responsibilities on providing the required functionalities for a federated data network described in Figure 2 is vital for a successful European infrastructure. In defining functionalities, we propose to follow the European Interoperability Framework that has recently been adopted by EOSC and is described in Figure 3. This paper focuses on suggesting organisational scenarios that can combine different approaches to each functionality. These scenarios are implemented using different technical and semantic approaches. The implications to legal and social levels will need to be evaluated.

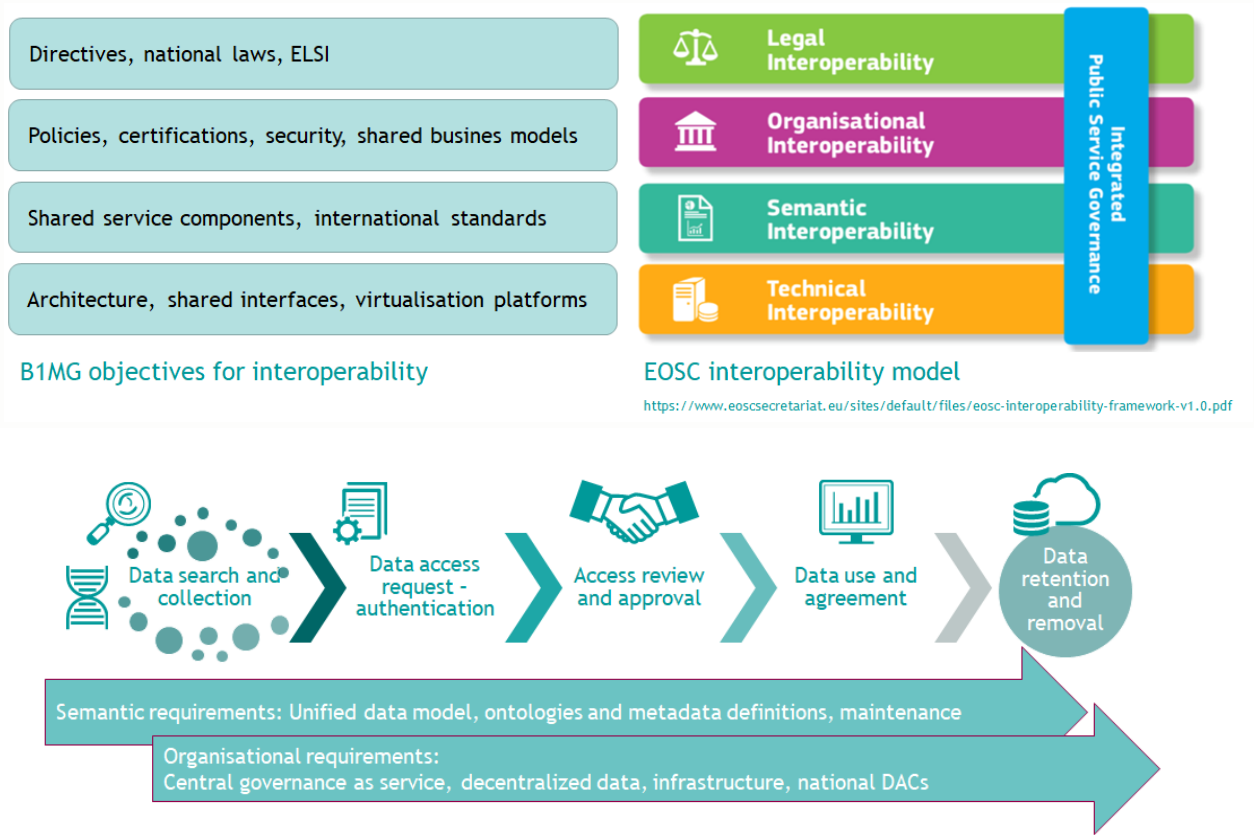


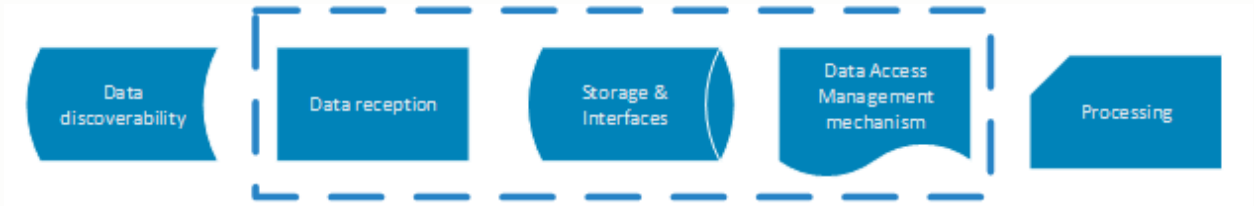
Figure 3. Technical and legal interoperability need to be jointly developed for the cross-border access to European genomes. Semantic interoperability is necessary for the integration of national datasets as part of the European data collection. Organisational interoperability creates a governance structure that is compliant with the legal and ELSI requirements. Data controllers can pass their responsibilities to Data Access Committees (DACs).

5. Organisational options

Using the five functionalities (Figure 2), it is possible to construct 16 different scenarios (Table 1) for collaboration on infrastructure responsibilities for 1+MG. The current focus of infrastructure work on the 1+MG initiative is outlined by the dashed line to support transnational data access processes. A

data discovery and management layer is needed on top of the existing infrastructures to meet the goals of the 1+MG declaration.

European infrastructures are expected to actively offer compute, data and network backbone for the 1+MG Initiative processing. Current generic e-Infrastructures are not yet ready to support 1+MG in sensitive data processing. Furthermore, a sustainable business model for the organisations to work together is a gap currently. Situation may change rapidly. We do not recommend dedicated processing infrastructure to be built for the 1+MG requirements, but we need to find a way to leverage future and existing investments and capabilities (experts) to make generic e-Infrastructures more compatible for sensitive data processing.



Data discoverability	*Data reception*	*Storage& Interfaces*	*Data access control tools*	*Processing service*
National	**National**	**National**	**Transnational**	**National**
National	National	Transnational	Transnational	National
National	Transnational	National	Transnational	National
National	Transnational	Transnational	Transnational	National
National	**National**	**National**	**Transnational**	**Transnational**
National	National	Transnational	Transnational	Transnational
National	Transnational	National	Transnational	Transnational
National	Transnational	Transnational	Transnational	Transnational
Transnational	**National**	**National**	**Transnational**	**National**
Transnational	National	Transnational	Transnational	National
Transnational	Transnational	National	Transnational	National
Transnational	Transnational	Transnational	Transnational	National
Transnational	**National**	**National**	**Transnational**	**Transnational**
Transnational	National	Transnational	Transnational	Transnational
Transnational	Transnational	National	Transnational	Transnational
Transnational	**Transnational**	**Transnational**	**Transnational**	**Transnational**

Table1. Infrastructure scenarios with different scopes and responsibilities. The five scenarios marked in bold have been selected for further discussion, cf. Table 2. Data discovery likely relies on national data hubs in 2021-22 (Green highlight) making data collections discoverable. Data must fulfil 1+MG standards (e.g. quality, data structure/format, semantic interoperability) to be included. Data access federation tools are the ”minimal infrastructure” solution (Red highlight), aligned with ELSI regulations and guidelines. Existing data processing services (e.g. HPC, local, cloud) must meet data privacy standards agreed by 1+MG initiative to be part of the infrastructure.

Not all scenarios are feasible. For example, uploading data to a shared transnational storage was ruled out by discussion between member states, which already eliminates seven or even eight scenarios. Datasets likely are made actionable nationally, e.g. in genome centers, and not by transnational experts, and having fully transnational data management supported by national data processing is also unlikely. This excludes four further scenarios.

Five of these scenarios of different scopes and ambition were chosen for discussion. We propose to evaluate them further (Table 2). Infrastructure can progress in 2021 using a proof of concept to evaluate the scenarios. Work will be coordinated by 1+MG WG-5 and B1MG WP4, and communicated to the 1+MG signatory member states.

Member states can choose their infrastructure engagement strategy with 1+MG using these scenarios depending upon present legislation and technical solutions that fit the national setting. It should be possible to migrate from one scenario to another.

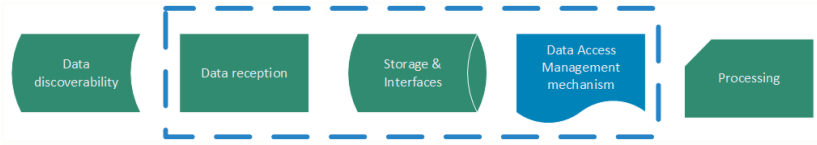
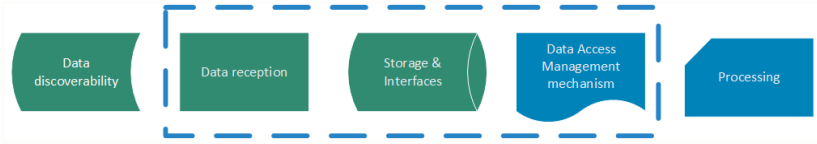
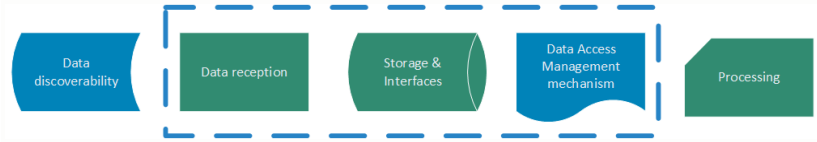
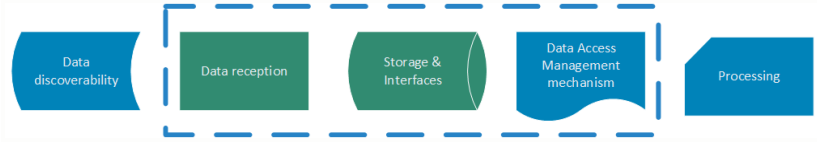
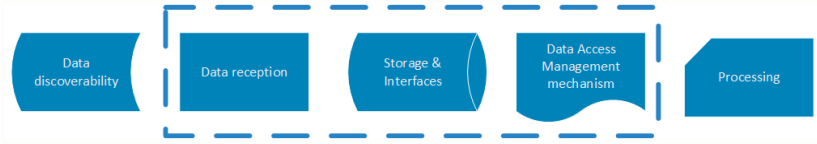
Organisation (**Green** = National, **Blue** = European)	**No.**	**Scenario**
	1	Minimum 1+MG data infrastructure solution is *transnational* data access federation tools to support federation of known *national* data collections listed in a 1+MG catalogue with known standards reported by signatory states. **Minimum success metric**
	2	A federated 1+MG data collection is made discoverable from distributed *national* data collections managed in member states. Access to those data can be made on a secure *transnational* processing platform supported by a *transnational* data access management process.
	3	A single harmonised 1+MG data set is formed from *national* managed collections. The whole dataset is made available as a service on *national* processing platforms. **Recommended success metric**
	4	A single harmonised 1+MG data set is formed from *national* managed collections. The whole dataset is made available on the *transnational* processing platform. **Recommended success metric**
	5	*Transnational* 1+MG data management and processing.

Table 2. 1+MG Infrastructure responsibility scoping scenarios.

6. Next steps

We propose that at the beginning of 2021 we will build a proof of concept (PoC) with 4 - 5 countries at the chosen scope. Finland, Germany, Norway, Spain and Sweden are close to having technical capability and resources to try this. By 2022 we hope to double the number of countries (Figure 4). Initially, we would test functionalities with non-GDPR datasets to avoid jeopardising sensitive data.

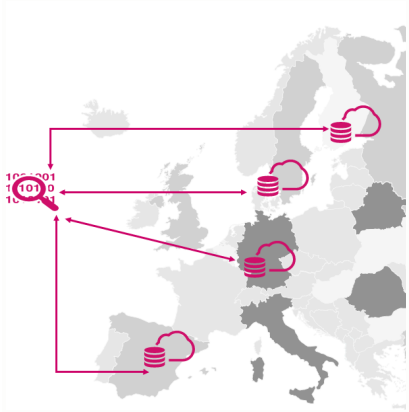


Figure 4. Finland, Germany, Norway, Spain, Sweden are close to having technical capability and resources to build 1+MG infrastructure proof of concept.

6.1. What will a Proof-of-Concept look like?

Once approved and hopefully within the next 6 - 8 months, the 1+MG infrastructure PoC will focus on the rare disease (RD) WG-8 use case. This use case is particularly well suited for a PoC because RD are usually of genetic origin and monogenic, with one or few causative genetic variants. The PoC aims to demonstrate Scenario 4, or Scenario 3 (Table 2) unless a transnational processing service can be made available for 1+MG testing.

Furthermore, the international RD community has long acknowledged the need to collaborate and share data to overcome the bottlenecks posed by the fact that each RD affects less than 1 in 2,000 individuals. In Europe, this need to collaborate and share is at the core of many projects and initiatives, such as the European Reference Networks (ERNs), the European Joint Programme on Rare Diseases (EJP-RD), Solve-RD and Orphanet, among others, which are also key for the development and establishment of key standards and ontologies for the field (e.g. Orphanet Rare Disease Ontology, Human Phenotype Ontology, OMIM, HGVS, etc). The set of Common Data Elements (CDE) for rare disease registries are provided by the Joint Research Centre of the European Commission[2]. At the international level, there are already initiatives like MatchMaker Exchange (MME) that enable anonymised discovery of affected individuals across connected nodes, although in a much less powerful way than what 1+MG aims to enable. Altogether, the RD community is well organised and has the necessary data and use cases available for a PoC to demonstrate federated data discovery and analysis.

[2] https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en (Chapter 6.)

The proposed methodology to design and develop the PoC is the following:

- Arrange meetings and workshops to understand demonstrator requirements, shared responsibilities and identify challenges.
- Work iteratively with other 1+MG use cases to understand their specific requirements and potential synergy with the Rare Disease demonstrator and Infectious Disease demonstrator (with a priority on the COVID-19 use case).
- Federate simulated/synthetic datasets from a number of European countries using the FEGA technology.
- Connect FEGA with tools that increase semantic interoperability to address real clinical and research use case user end-points. Federation of systems like the RD-Connect Genome-Phenome Analysis Platform (GPAP) and/or usage of APIs like MatchMaker Exchange or Beacon might be a realistic starting point to cover some use cases.
- Organise an event during 2021 focusing on security of federated data access.

There are some risks and bottlenecks that might hinder the development of the PoC. The main one is that there is no specific funding allocated for the PoC. In addition, changes and development are most likely necessary in generic European e-infrastructures (e.g. EOSC, EuroHPC, Géant), since they might not be completely adequate for sensitive clinical data; the amount of necessary changes towards compatibility with 1+MG could be a measure for success. However, it is worth noting that the infrastructure capacity may not be the main bottleneck, but rather how it can be set up (e.g. software contracts, data security, processing) to manage the data in the foreseen federation. Once the ELSI&TECH interoperability has been shown, it will be possible to move to real data for RD and/or COVID-19.

6.2. How to measure PoC success?

Several indicators are possible to be used as a measure of the PoC success:

- Number and type of 1+MG use cases and demonstrators enabled by the PoC infrastructure
- Demonstration of the infrastructure and enabled use cases followed by a stakeholders' survey/evaluation
- Number of participating nodes/countries in the PoC infrastructure
- Number of datasets included in each of the nodes/countries
- Data types (e.g. genomes, exomes, phenotypes, clinical data, etc.) and formats (e.g. FASTQ, BAM, CRAM, gVCF, Phenopackets, etc.) included in the PoC. Data must comply with WG3 and WG4 proposed standards.
- Assessment of the GDPR compliance of the PoC infrastructure

Overall by the end of 2021 1+MG infrastructure has moved to “next stage” if

- A consortium of at least three member countries ("spearhead") with expert and technical capacity form a data federation operating suitable technologies

- An active dialogue with the 1+MG governance ensures that the security measures on data privacy are sufficient for the test and production environments hosted in data hubs in the participating member states
- Technical interoperability of federated data access can be claimed to be in pre-production, if a data analysis responding to healthcare questions for rare diseases, cancer, common/polygenic disease or infectious diseases has been successfully executed
- Additional member states individually decide to move towards the proposed coordinated solution. However, it is likely not all member states have reached legal interoperability, and human capacity for the operations yet that would allow them to join the European 1+MG data federation.

More general ambitions to drive discussion between 1+MG infrastructure and use cases

- Confirmed relevance for research and clinical use cases
- Rare Diseases (WG-8) collaboration diagnoses an ultra-rare life threatening condition
- Population level / Common disease (WG-10) or Pathogen COVID-19 (WG-11) data sharing is enabled to recognise risk individuals genetically
- Cancer (WG-9) precision treatment can be initiated by a particularly genotyped cancer type e.g. AML-CML

7. Appendix

7.1. 1+MG Infrastructure timeline (set December 2019)

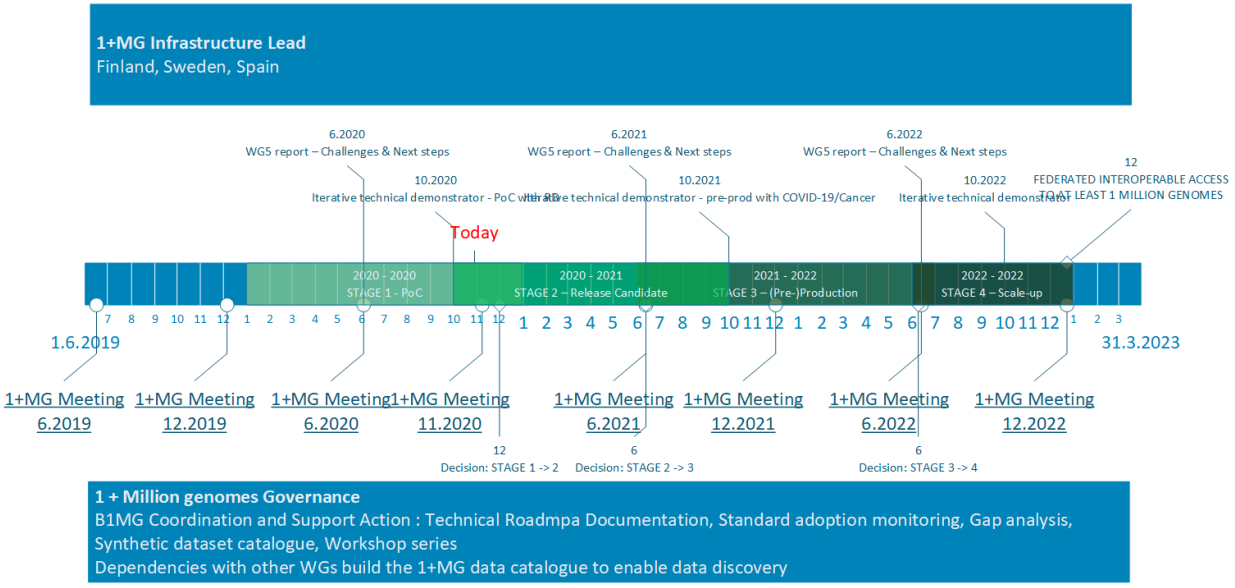


Figure 5. WG-5 Timeline. The objective of infrastructure working group (1+MG WG-5/B1MG WP4) is standards for interoperability to facilitate interfaces for cross-border services bridging national and European infrastructure for data discovery and access of genome, phenotype and clinical sample descriptions. Progress is measured using cross-border pilots from existing infrastructures and use cases that could provide minimum viable product(s) to support the 1+MG requirements. The overall infrastructure service integration is built with security by design as a guiding principle. Non-GPDR simulated/synthetic datasets are required for tests on technical service interoperability, data protection and scalability without the risk of data breach. Infrastructure coordination promotes use of global standards to provide authorised access on national data based on existing user identities, compliant with GDPR and national laws and regulations. Overall process is expected to increase capacity across the European Member States (and beyond) with a support technical expert network and transparent knowledge exchange on best practices at the desired scope subscribed by the national governments and European Commission.

7.2. Gaps identified during writing process

- Who is the owner (in GDPR) of the transnational data catalogue and data access mechanism according to 1+MG ELSI policy
- Define “1+MG Dataset”. Is it a harmonised exome/full-genome dataset with a different data model each of the use cases? Data exposed to compute at different granularities. Enabling “Anonymous/registered data access” to support large-scale queries.
- Infrastructure for both research and healthcare use. The 1+MG infrastructure scoping leverages existing know infrastructures e.g. EGA federated with facilities for the subsequent analysis. The intention of the existing infrastructure components are thus for secondary use of the data and its analysis, not for support of healthcare processes. In the healthcare setting

e.g. ICT service availability requirements are higher than in research. Hospital ICT requirement typical ask for higher SLAs than research infrastructure provide. We have problem how to combine same infrastructure for healthcare and research purposes. Is it possible to serve both use cases (for genomics/omics data services) from a single infrastructure?

- Service/component level responsibility in data sharing between infrastructure and use cases. E.g. service for streaming for specific use cases? What is in the scope?
- Where and on which data to focus first when making data actionable on the 1+MG data federation, sensitive clinical data essential for meaningful genomics research?
- Segmentation of priorities reaching different scenarios into phases/maturity model. What is needed to reach each milestone and the end goal.
- Business model for 1+MG infrastructure service delivery.
- Building scenarios with national / regional data hubs to propose possible solutions to connect national infrastructure solutions to the European 1+MG interfaces.
- Pros and cons for each scenario from legal and technical and business perspectives.

7.3. Glossary of Terms Used

WG-5 infrastructure will work with WG-2 ELSI to produce a glossary of terms[3].

ELSI Ethical Legal Societal and Policy

GDPR General Data Protection Regulation

FAIR Findable, Accessible, Interoperable and Re-usable principles, which are used in constructing data services within the available resources.

Federation is a contract, a framework of trust to share data, technical responsibilities

Semantic i.e. meaning of

Transnational processing. Infrastructure service (operations) can be offered nationally or transnationally. Transnational processing service could be commissioned by the member states or EC to a European legal entity. National processing service could also be transnational ones, but need to have a transnational service access policy appropriate for the 1+MG data processing.

Virtual cohort. 1+MG dataset are expected to exist in a distributed infrastructure setting, but form a virtual whole consisting of data entities covering at least million individuals.

[3] https://docs.google.com/document/d/1jHhgg-G9UYp2M7a15q6T1hXBbg34jOayhVhWlaRnRI0/edit

7.4. Contracting needs

DRAFT Table

Scenario	Contracting needs
1	Between national repository provider and transnational data access federation tool provider
2	Between national repository provider and transnational data access federation tool provider Between national repository provider and secure transnational processing platform provider
3	Between national repository provider and transnational data access federation tool provider Between national repository provider and transnational data discovery provider
4	Between national repository provider and transnational data access federation tool provider Between national repository provider and secure transnational processing platform provider Between national repository provider and transnational data discovery provider
5	Between national repository provider and transnational data access federation tool provider Between national repository provider and secure transnational processing platform provider Between national repository provider and transnational data discovery provider Between national repository provider and transnational FAIRification service Between national repository provider and transnational repository provider