European Genomic Data Infrastructure

Deliverable D3.3

Infrastructure status report and updated roadmap

Project Title Grant agreement no	**Genomic Data Infrastructure** Grant agreement 101081813		
Project Acronym (EC Call)	GDI		
WP No & Title	WP3: Deployment of 1+MG National Nodes		
WP Leaders	Bengt Persson (7. UU)		
Deliverable Lead Beneficiary	7. UU		
Contractual delivery date	30/04/2024	**Actual delivery date**	20/05/2024
Delayed	Yes		
Partner(s) contributing to deliverable	UU		
Authors	Anna Hagwall (UU) Johan Viklund (UU)		
Contributors	Bengt Persson (UU)		
Acknowledgements	Dimitris Bampalikis (UU)		
Reviewers	Rob Hooft (HRI) Dylan Spalding (CSC)		

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

Log of changes

Date	Mvm	Who	Description
19/04/2024	0V1	Anna Hagwall, UU	First draft
22/04/2024	0v2	Johan Viklund, UU	Second draft
07/05/2024	0v3	Johan Viklund, UU	Incorporating comments from reviewers.
09/05/2024	0v4	Mercedes Rothschild Steiner (ELIXIR Hub)	Copy circulated to the GDI-MB for review
20/05/2024	1VO	Mercedes Rothschild Steiner (ELIXIR Hub)	Final version submitted to the EC portal

Table of contents

Contents

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

European Genomic Data Infrastructure

Executive Summary

This deliverable documents the current status of the implementation of the roadmap and adaptations to the roadmap defined in B1MG[1], complemented with the national assessments of the GDI deliverable 3.2[2].

In addition, changes in the product portfolio based on an identified gap in the initial solution have taken place.

The project GDI continues along the path defined in B1MG and by 1+MG toward secure cross-border access of genomic and phenotypic data in Europe.

[1] https://zenodo.org/records/10046982
[2] https://doi.org/10.5281/zenodo.10688135

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

European Genomic Data Infrastructure

1. Contribution towards project outcomes

With this deliverable, the project has reached or the deliverable has contributed to the following project outcomes:

[Select 'Yes' (at least one) if the deliverable contributed to the key result, otherwise select 'No'. For more details of project outcomes, see here]

	Contributed
Outcome 1 Secure federated infrastructure and data governance needed to enable sustainable and secure cross border linkage of genomic data sets in compliance with the relevant and agreed legal, ethical, quality and interoperability requirements and standards based on the progress achieved by the 1+MG initiative.	Yes
Outcome 2 Platform performing distributed analysis of genetic/genomic data and any linked clinical/phenotypic information; it should be based on the principle of federated access to data sources, include a federated/multi party authorisation and authentication system, and enable application of appropriate secure multi-party and/or high-end computing, AI and simulation techniques and resources.	No
Outcome 3 Clear description of the roles and responsibilities related to personal data and privacy protection, for humans and computers, applicable during project lifetime and after its finalisation.	No
Outcome 4 Business model including an uptake strategy explaining the motivation, patient incentives and conditions for all stakeholders at the different levels (national, European, global) to support the GDI towards its sustainability, including data controllers, patients, citizens, data users, service providers	No

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

(e.g., IT and biotech companies), healthcare systems and public authorities at large.	
Outcome 5 Sustained coordination mechanism for the GDI and for the GoE multi-country project launched in the context of the 1+MG initiative.	No
Outcome 6 Communication strategy – to be designed and implemented at the European and national levels.	No
Outcome 7 Capacity building measures necessary to ensure the establishment, sustainable operation, and successful uptake of the infrastructure.	Yes
Outcome 8 Financial support to the relevant stakeholders to enable extension, upgrade, creation and/or physical connection of further data sources beyond the project consortium or to implement the communication strategy and for capacity-building.	No

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

2. Methods

Deliverable 3.2 was completed (2024-02-21) by CSC as part of GDI WP3 (and WP5), in collaboration with WP6, by sending a joint survey assessing the national roadmaps and plan.

Deliverable B1MG D4.3 was completed (2023-10-27) by WP4 in B1MG.

Changes to the product portfolio were established after gaps had been identified in preparatory meetings for milestone 7, at WP3 meetings, as well as during a workshop dedicated to analyse and illustrate all steps of the data insertion pathway (still in draft state).

A discussion about the naming of the starter kit has taken place among the Pillar II leads and their deputies. The decision was to continue using this term. It will be used throughout the project for the complete official European Genomic Data Infrastructure (GDI) product portfolio. Thus, the user portal is also regarded as part of the starter kit.

3. Description of work accomplished

The deliverable 3.2 has been documented extensively and the results have been presented both at the GDI pillar II technical workshop[3] in Rome in October 2023 and at the Pillar I / Pillar II cross-pillar meeting[4] in Malta in March 2024.

Deliverable B1MG D4.3 documents the latest version of the 1+MG roadmap and is built on the first version of the roadmap, defined some years earlier. Since then we have identified the following points where the roadmap could be clarified or augmented:

- During the discussions in preparation for milestone 7 "Early adopter production nodes deployed with real data", we identified that it would be advantageous to use the FAIR data point protocol[5] (FDP) for pulling metadata to the 1+MG catalogue that is part of the user portal. There are several ways to do this in practice, including adding information in xml files or using implementations that are part of data management tools such as Molgenis[6]. There was a decision to include FDP (the api and reference implementation) as an additional GDI product in the starter kit. As FDP has been developed in the Netherlands which also is a vanguard node, this node was contacted with the proposal to assign a product owner. The proposal was accepted and a PO now has been assigned and contributes to the project, fully aligned with the other product owners in various meetings, Slack channels etc.

[3] https://docs.google.com/document/d/1BVQ20ap2LOzmonWWIdANWdVf-EBD9gw4G4wuavzEJwE/edit#heading=h.gjdgxs

[4] https://docs.google.com/document/d/1iiLA5J7VGOen3RgSJ2NuxPxYrvW0cbP6ZV8ZE-ciU7Q/edit#heading=h.jnart0x2e8ck

[5] https://www.fairdatapoint.org/

[6] https://molgenis.org/

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

- From discussions between Pillar II and Pillar III it has been decided that there will only be one product called "Computation" that will encompass both containers and processes for federated analysis. This PO will work closely together with pillar III to make sure that the federated analysis use cases can be run.
- The starter kit term will be used throughout the project for the official GDI product portfolio, to signal that the services are provided as a starting point and that national adaptations of the deployment are necessary (rather than fully plug and play). As such, the user portal: access control and the user portal: data catalogue are part of the starter kit. Previously they were not, as the starter kit then referred exclusively to the first version of the collection of tools that could be installed to implement the five functionalities at each of the GDI nodes, released in June 2023, which excluded these.

5. Results

Through D3.2, we have a clearer understanding of the national roadmaps in relation to B1MG D4.3. We are however fully aware that these roadmaps, especially concerning which GDI products to implement, can and will change over time as more countries start exploring the products and how they interact with their national setup, potentially exposing limitations.

In the current setup the user portal harvests data from the FAIR Data Points in each node and will expose the GDI beacon network to authorised researchers. It's also the entrance for applying for access to data. This was deployed and integrated as part of milestone 7. With the addition of the user portal the starter kit now contains the foreseeable necessary products to achieve the goals of GDI. It is possible that we will find more gaps over time and will then act upon these in the same timely fashion.

We are still limited in what we can do with real data because of outstanding issues with the legal framework. However, we are proceeding with the deployment of the nodes and instead aim to test components and routines with "real-like" data (cf MS 7). This way we hope we can have an infrastructure that can conform to the legal requirements once the legal framework is in place. From D3.2 and the Malta F2F in March 2024 it became clear that the infrastructure should evolve from being 'product' based to 'API' based, i.e. the APIs used to connect the different products should be specified, clarifying that the products themselves can be changed or replaced depending on the national use case. The use of Global Alliance for Genomics and Health[7] (GA4GH) standards for interoperability between the products means this work is already started, but additional definition of the APIs needs to be completed, in conjunction with WPs 4,5 and 6 and with input from Pillars I and III.

[7] https://www.ga4gh.org

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.

6. Discussion

The *starter kit* has changed in nature: where this used to be positioned as a set of exemplary tools that could be used to implement the five functionalities needed by every GDI node, and that have an expert product owner inside the GDI project, the addition of the user portal as the central synchronisation point for metadata and access to the other products in the starter kit moves the GDI product portfolio from a set of different software products for a GDI node to a complete service offering for the GDI infrastructure including its planned European operation centre.

It is important to note that the *starter kit* is the GDI product portfolio. There will not be several coherent releases of different versions of the entire starter kit. Instead, each product is extended and improved as necessary, while maintaining interoperability with the other products.

7. Conclusions & Impact

These past months we have obtained a better understanding of the future roadmap, how aligned the nodes are, and have made necessary additions to the starter kit to reach the goals of GDI.

8. Next steps

The work toward a secure cross-border access of genomic and phenotypic data is continuously ongoing and the progress on a technical and node deployment level is monitored by both WPs and GDI coordination, as well as the GDI Management board and 1+MG working group leads.

GDI project receives funding from the European Union's Digital Europe Programme under grant agreement number 101081813.