



# Deliverable D3.2

## Node deployment roadmap

<b>Project Title</b> Grant agreement no	<b>Genomic Data Infrastructure</b> Grant agreement 101081813		
<b>Project Acronym</b> (EC Call)	GDI		
<b>WP No &amp; Title</b>	WP3: Deployment of 1+MG National Nodes		
<b>WP Leaders</b>	Bengt Persson (7. UU)		
<b>Deliverable Lead Beneficiary</b>	6. CSC		
<b>Contractual delivery date</b>	31/10/2023	<b>Actual delivery date</b>	21/02/2024
<b>Delayed</b>	Yes		
<b>Partner(s)</b> contributing to deliverable	6. CSC		
<b>Authors</b>	Dylan Spalding (CSC)		
<b>Contributors</b>	N/A		
<b>Acknowledgements</b>	Dilza Campos (VIB)		
<b>Reviewers</b>	Susanna Repo (CSC), Juan Arenas Marquez (ELIXIR Hub)		



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



## Log of changes

Date	Mvm	Who	Description
08/02/2024	OV1	Mercedes Rothschild Steiner (ELIXIR Hub)	Copy circulated to the GDI-MB for review
21/02/2024	1VO	Mercedes Rothschild Steiner (ELIXIR Hub)	Final version submitted to the EC portal

## Table of contents

### Contents

1. Executive Summary	3
2. Contribution towards project outcomes	4
3. Methods	6
4. Description of work accomplished	7
5. Results	7
5.1 GDI Phase Timings	8
5.2 Node aspirations	10
6. Discussion	13
6.1 Phase Deadlines	13
6.2 Landscape Analysis	13
7. Conclusions & Impact	14
8. Next steps	15
9. Glossary	16
Appendix I	17



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



## 1. Executive Summary

This deliverable aims to determine the timelines for each country to deploy a European Genomic Data Infrastructure node, as well as start to understand the expectations and requirements each country has regarding the deployed and operational GDI node. This information will help Pillar II to address specific issues raised by participating countries, and identify common issues and gaps raised by the different countries.

Presented here are the results of two different outreach activities to the participating countries, the first to understand each country's timelines to reach onboarding, deployment, or operational status, and second to get a high level understanding of each country's expectations and requirements of a GDI node.

For the first activity 59% of the countries involved in GDI responded, and for the second 66% initiated a response, and 54% had answered one or more questions. This clearly indicates a risk related to the lack of engagement from some countries, with an associated risk that the development of the technical infrastructure within GDI may not support every country's specific use cases and requirements nor the 1+MG requirements for providing border access to high quality genomic and health data. To address this, a specific face-to-face node engagement workshop is being planned for March 2024 to tackle the technical issues and increase engagement and feedback between the countries and Pillar II. Additional actions would be required to be taken by Pillar I and/or the project coordinator to tackle non-technical issues, e.g. securing long-term national funding for the operations of the infrastructure through the establishment of national genomics plans or equivalent

The responses also indicate a range of ambitions in terms of reaching a specific stage in the deployment process. This needs to be investigated to ensure that the understanding of the requirements for each stage is the same for all countries and also conforms to the 1+MG Data Governance model. This will be explored and completed in association with the quarterly reporting and the maturity level model, as well as a Cross Pillar F2F meeting.



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



## 2. Contribution towards project outcomes

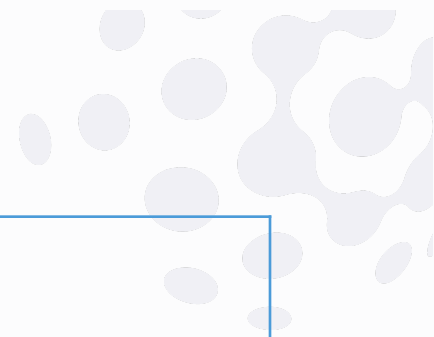
With this deliverable, the project has reached or the deliverable has contributed to the following project outcomes:

[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
<b>Outcome 1</b> 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
<b>Outcome 2</b> 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
<b>Outcome 3</b> 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



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



<p><b>Outcome 4</b></p> <p>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 (e.g., IT and biotech companies), healthcare systems and public authorities at large.</p>	No
<p><b>Outcome 5</b></p> <p>Sustained coordination mechanism for the GDI and for the GoE multi-country project launched in the context of the 1+MG initiative.</p>	No
<p><b>Outcome 6</b></p> <p>Communication strategy – to be designed and implemented at the European and national levels.</p>	Yes
<p><b>Outcome 7</b></p> <p>Capacity building measures necessary to ensure the establishment, sustainable operation, and successful uptake of the infrastructure.</p>	No
<p><b>Outcome 8</b></p> <p>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.</p>	No



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



### 3. Methods

To define the node level roadmaps, the nodes themselves needed to feedback the timelines for their plans and aspirations. Note that in this document a node corresponds to the institution or institutions deploying the infrastructure, on behalf of a country, to allow for a centralised or federated node which enables cross-border data access. These timelines were, minimally, the date on which the node planned to achieve one of the three European Genomic Data Infrastructure (GDI) phases: onboarding, deployment, or operational. These phases are linked to the Quarterly Reporting and Technical Readiness Levels (TRLs) via specific technical and non-technical indicators. Additional information that was deemed relevant was also collected that could also feed into an analysis of the whole landscape of node deployment plans, such as, alignment with Federated European Genome-phenome Archive (FEGA), national and international projects and dataspace seen as important on a national level, e.g. European Health Data Space (EHDS), and high priority use cases, e.g. Genome of Europe (GoE).

Within GDI, WP6 was also working on a node level Data Management Plan (DMP), which covered some of the same information as relevant in this deliverable, as well as additional questions specific for the DMP. WP6 utilise the Data Stewardship Wizard (DSW)<sup>1</sup> to elicit responses from the nodes, and had started gathering feedback from the nodes, so it was agreed that the WP3 questions specific to the node level roadmap would be tagged and added (if necessary) to the questions within the GDI DSW<sup>2</sup>, as detailed in D6.1 Draft data management policy including ELSI best practice<sup>3</sup>.

Responses were invited from the nodes for the DMP related questions with the aim of presenting the responses during the Pillar II workshop in Rome (09.10.23–11.10.23). Additionally, the DMP was also presented at the same workshop, which included some of the questions relevant for the node level roadmap.

Based on feedback from a subset of nodes, WP3 prepared a simplified separate query, which was sent out after the Pillar II workshop. This query asked specifically for the minimal set of dates by which each node aimed to reach a particular GDI phase, and whether this date was based solely on technical or non-technical, or both, indicators as detailed in the Quarterly Reporting spreadsheet.

---

<sup>1</sup> <https://ds-wizard.org/>

<sup>2</sup> <https://gdi.ds-wizard.org/>

<sup>3</sup> <https://zenodo.org/records/7956612>



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



## 4. Description of work accomplished

The plan for outreach to the nodes for input to the node level roadmap, D3.2, was presented<sup>4</sup> for discussion during the August Pillar II meeting on 02.08.2023, and resulted in the proposal to align the work in the DMP with the questions feeding into the node level roadmap. In collaboration with WP6, a set of questions were outlined specific to the roadmap, and mapped to the set of questions already in the DSW, where if a question was novel it was added to the DSW and tagged 'roadmap'.

The DMP and the associated requirements for this deliverable were presented at the September Pillar II meeting (06.09.2023), as well as an outline on how to give feedback on the DMP questions specific to WP6. Feedback was elicited from the nodes, and acted upon within the DSW, and on 27.09.2023 the nodes were requested to complete the questions in the DSW.

In order to complete the DSW questionnaire, an account had to be created for the individual(s) representing each node. This was facilitated by the nodes being asked to nominate up to 3 individuals in a spreadsheet, and an account on the DSW would be created for those nominated. These individuals could then log into the DSW and complete the questions, as well as leave comments. For the roadmap subpart of the DSW, a deadline of 3.10.2023 was proposed to allow time to analyse the results prior to the Rome workshop.

WP6 provided a script to list the responses currently submitted in the DSW ready for further analysis. Node responses in the DSW were monitored and a snapshot taken directly before the Pillar II Rome workshop, and these were presented to the workshop participants as a summary and as a way to engage with additional nodes who had yet to complete the DSW questionnaire.

## 5. Results

Given the different ways that we communicated with the nodes, the results are presented in separate sections.

---

<sup>4</sup> [https://docs.google.com/presentation/d/1j1OSgLmJpoYIW5KwrWKPep\\_EXqAg\\_LFq/edit#slide=id.p1](https://docs.google.com/presentation/d/1j1OSgLmJpoYIW5KwrWKPep_EXqAg_LFq/edit#slide=id.p1)

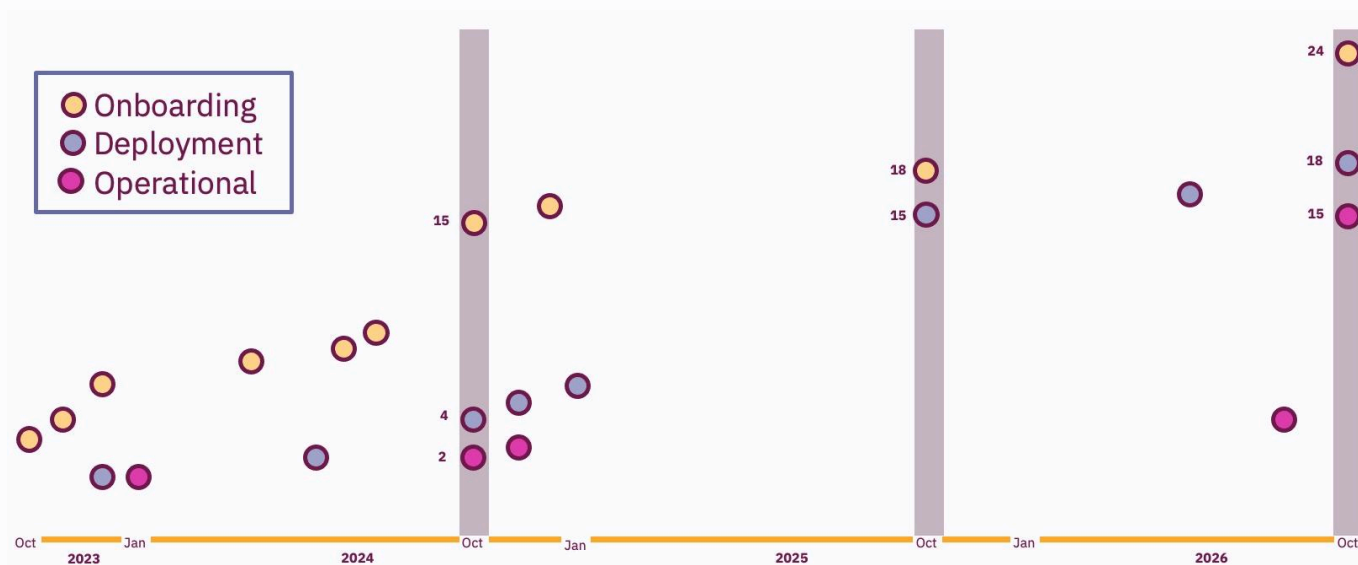


GDI project receives funding from the European Union's Digital grant agreement number 101081813.



## 5.1 GDI Phase Timings

Fourteen nodes completed the spreadsheet<sup>5</sup> with the dates they aimed to achieve each of the GDI phases they aspired to, and this is given in Table 1 (accurate as of 04.1.2024) and Figure 1. All nodes that responded, with the exception of Norway, stated they would achieve onboarding status during 2023. For deployment, only Spain stated it would achieve that in 2023. For deployment, four nodes (Belgium, Finland, the Netherlands, and Portugal) are aiming for 2024, with Norway and Slovenia aiming for 2025. To achieve the operational phase, Spain was aiming for early 2024, with Finland and Portugal aiming for late 2024, while the other nodes (Belgium, Netherlands, Norway, and Slovenia) all aiming for 2026 to become operational, with Slovenia indicating operational status 1 month after GDI finishes.



**Figure 1.** Timeline indicating the number of nodes reaching onboarding, deployment, or operational status between October 2023 and October 2026. A majority of nodes use the default dates of October to hit the next level, indicated by the coloured lines. For those nodes that did not respond, the timelines from the GDI proposal were used.

<sup>5</sup> <https://docs.google.com/spreadsheets/d/1pddAoMqKosjcAykYgJiBXAGoP6Nq1Zq4eSkEpi6lhFw/edit#gid=0>



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





**Table 1.** Node responses regarding the dates at which they aim to achieve onboarding, deployment, or operational phases within GDI. The table shows the date at which a node aspires to achieve each of the applicable phases, the scope of the response, and any comments or justifications for the response. Text in grey indicates the default dates as defined in the proposal and that the node has not actively confirmed these dates. Grey cells indicate a node does not intend to reach that phase during GDI.

Country	Onboarding	Deployment	Operational	Scope of response	Comments and justification
Belgium	2023-12-31	2024-10-31	2026-09-30	Technical	
Bulgaria	2024-12-31	2026-05-31		Technical	
Croatia	2026-10-31				
Cyprus	2026-10-31				
Czech Republic	2024-10-31	2025-10-31	2026-10-31		
Denmark	2024-10-31	2025-10-31	2026-10-31	Technical	
Estonia	2024-10-31	2025-10-31	2026-10-31	Technical	
Finland	2023-10-01	2024-05-31	2024-10-31		
France	2024-10-31	2025-10-31	2026-10-31		
Germany	2024-07-31	2025-10-31	2026-10-30	Technical	
Hungary	2026-10-31				
Ireland	2026-10-31	31/10/2027	31/10/2028	Both	Ireland had committed to the onboarding readiness levels for the duration (48 months) of the GDI project. Genomics infrastructure within Ireland is very minimal at this point in time and there is a reliance on the implementation of our national genetics and genomics strategy and national genetics/genomics office. Therefore given our current status we propose these dates.
Italy	2024-10-31	2025-10-31	2026-10-31		
Latvia	2025-10-31	2026-10-31		Both	



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



Lithuania	2025-10-31	2026-10-31		Both	
Luxembourg	2024-10-31	2025-10-31	2026-10-31		
Malta	2026-10-31				
Netherlands	2023-12-21	2024-10-31	2026-10-31	Technical	National genomics plan is still in the works; national elections in 2023-11.
Norway	2024-06-30	2025-10-01	2026-10-31		
Portugal	2023-10-11	2024-11-30	2024-11-30	Technical	
Romania	2026-10-31				
Slovenia	2024-03-31	2025-10-31	2026-11-30	Technical	
Spain	2023-11-30	2023-12-31	2024-01-31		
Sweden	2023-10-01	2025-01-31	2026-10-21	Technical	

## 5.2 Node aspirations

As of 8.01.24, 16 nodes had registered to and accessed the DSW questionnaire and hence were able to answer the questions in the DSW. Of these, 13 had answered 1 or more questions resulting in a response rate of registered countries of 81%. Their responses are recorded in Appendix 1 as derived from the table in this spreadsheet<sup>6</sup>. Where 'None' is included in a cell no entry was given for that question, despite it being asked. A blank cell indicates the question was not asked as it is a conditional question based on an previous response that did not satisfy the condition for the question to be asked..

The responses for product usage (Table 2) was fairly heterogeneous. 10 out of the 12 nodes who responded indicated they wished to use the Beacon, while 75% wished to use REMS, and 5 countries envisioned the need for additional products at this time. However, for Containerised Computation and Archival Technology, the results were mixed with 5 and 6 or more nodes either undecided and / or aiming to replace that particular product respectively. For federated computation, only 1 node aims to keep the product, with 7 still undecided.

**Table 2:** Summary of responses about the expected product usage per node within GDI. Nodes were given the option to state if they would keep, replace, or were still undecided on the use of a particular product within their node. They were also able to indicate if they anticipated the need for additional products.

<sup>6</sup> <https://docs.google.com/spreadsheets/d/1KJV5eDMJqhhdD5Rvj8InOkozMBfn-BJ6/edit#gid=70597679>



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



	REMS	Beacon	Federated Computation	Archival Technology	Containerised Computation	Additional products
<b>Keep / Yes</b>	9	10	1	5	6	5
<b>Not Decided</b>	3	2	7	4	4	4
<b>Replace / No</b>	0	0	2	2	1	3
<b>Not Answered</b>	4	4	6	5	5	4

For Federated EGA, 50% of nodes aim to host a Federated EGA instance and 50% do not (Table 3). Of those that do aim to host a FEAGA instance all aim to share the same technical infrastructure. Additionally six nodes envision allowing data download, similar to the methodology of data access utilised by FEAGA, however within GDI this would have to be covered by the GDI data governance requirements.

**Table 3:** Summary of the results on the data access model and co-deployment of FEAGA nodes.

	Allow data download?	Also host FEAGA?	FEAGA sharing the same technical infrastructure?
<b>Yes</b>	6	6	6
<b>No</b>	4	6	0
<b>Not Decided</b>	1		
<b>Not Answered</b>	5	4	0

A majority of nodes had important associated projects, and infrastructures or data spaces, but only four nodes indicated they had priority national use cases (Table 4). Where nodes answered the question on priority GDI use cases, all GDI use cases were mentioned at least once (Appendix I).

	<b>European Genomic Data Infrastructure</b>		GDI project receives funding from the European Union's Digital grant agreement number 101081813.
--	---	--	--



**Table 4:** Summary of the responses regarding important associated projects, dataspace or infrastructures, and priority national use cases.

	Important associated projects?	Important infrastructures or data spaces?	Priority national use cases?
<b>Yes</b>	7	7	4
<b>No</b>	4	4	2
<b>Not decided</b>	5	5	5
<b>Not Answered</b>	0	0	5

Only one node had so far decided if it would host non-genomic data, and of the 10 nodes who responded to the question on checking genomic data quality, 5 of nodes indicated they would check the genomic data quality, 2 not checking genomic data quality, and 3 being undecided (Appendix I). However 9 of the 11 nodes who responded aimed to check data integrity once it had been transferred to the node (Appendix I).

There was a range of 2,000 to 15,000 for the number of whole genome sequences each node anticipated to host. This metric appears low relative to the ambition, as well as the number of genomes sequenced by Genome of Europe, however it is possible the question was misinterpreted as whole genome sequences (WGS) can often be interpreted as BAM/CRAM files as opposed to gVCF or VCF files, or that nodes are expecting to host whole exome sequences as well.

Finally no nodes had decided on which ontologies to utilise; however this would be determined in collaboration with the 1+MG WG3 and 4 working groups for the data to be ready for inclusion into the 1+MG.

	<b>European Genomic Data Infrastructure</b>		GDI project receives funding from the European Union's Digital grant agreement number 101081813.
--	---	--	--



## 6. Discussion

### 6.1 Phase Deadlines

There was a 58% response rate to the questions on when a node aims to reach the GDI phases. This lack of engagement is a significant risk to the success of the project, and hence the ability to monitor and provide support to the nodes is subsequently limited. To address this issue, a specific face-to-face workshop is being planned for March 2024, where each node will be invited to present how they plan to reach their operational readiness goal they are committed to as a GDI node, including timelines and support required from Pillar II.

Additionally, based on the questionnaire, it appears that the aspiration of some nodes may be over-ambitious. For example, Spain aims to be operational by the end of January 2024, with both Finland and Portugal aiming for an operational status by the end of 2024. However, all the other nodes aim to reach operational status just by the end of the project (Oct-2026), but before the conclusion of the GDI and is not manageable without nodes pushing their promotion to become operational. Feedback from the nodes indicated clarification on the definitions of operational status should be made by the Pillar leaders and the coordinator in conjunction with the quarterly reporting, to ensure that the aspirations are realistic and achievable for each node. Work is already ongoing, such as the planned workshop in March 2024, to ensure the definitions are better understood.

Some nodes queried if the responses were on whether the aspiration was purely technical, or fully operational, i.e. was actually supporting cross-border data access on real genomic data. This is in part driven by the national situation, for example, the Netherlands stated that the National Genomics Plan was still in progress, so any non-technical deadlines were out of scope. Also as the GDI does not yet have a clearly defined legal framework, and the requirements from the 1+ Million Genomes working groups, as well as the Genome of Europe use case, are still evolving, it may be that the lack of clarity here is affecting the response rate to the phase deadlines. The existence of parallel initiatives, like the EHDS, that could have a potential impact on the final implementation and the fluid regulatory landscape could also be a factor in the engagement and the level of certainty provided by the different nodes when addressing the questions.

### 6.2 Landscape Analysis

The fact that five nodes already expect there to be a requirement for additional products indicates that Pillar II will need to investigate the support needs for these products, and if their use will be isolated to a specific node or represents a common requirement across nodes. Currently the



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



additional products specified in the questionnaire include a data catalogue, Molgenis, Beacon Network, Galaxy, and proxies for APIs. The fact that most nodes agreed to utilise Beacon and REMS also helps with the homogeneity of the GDI. The differences in the storage and compute products may reflect the differences in the underlying hardware and software at each node. These products, as long as they have suitable APIs, are more internal to a node or deployment, so the differences here have less influence on the interoperability of the GDI than other outward facing products. A generic platform, such as Galaxy, can also help provide a more homogeneous computation capability within GDI. Within the deliverable D4.1 - Helpdesk Roadmap, there is the requirement to define the change management processes and these should include how to suggest, evaluate, deploy, and test new products within the infrastructure.

The foreseen technical infrastructure sharing with FEGA is positive for GDI, and was a core aim at the outset to maximise interoperability and reduce unnecessary replication of core infrastructure; however, is not until the final GDI legal framework is in defined that non-technical aspects like data compliance with data governance interoperability and ELSI requirements can be evaluated. The fact that not all nodes aim to co-host a FEGA instance also demonstrates that GDI may help increase the number of countries willing to share genomic data beyond those that are already interested in FEGA. The different methods of data access, i.e. allowing data download or not, indicates that currently the data envisioned to exist within GDI may have different Ethical, Legal, and Societal Issues (ELSI) relating to it depending on the jurisdiction and the source of the data specific to each node. This difference in data access methodology needs to be confirmed with 1+MG WG2 and Pillar I, and if confirmed the current starter kit products will need to be checked to ensure they support this.

## 7. Conclusions & Impact

Improved communication between the Pillar II participants, Work Package 3, and Coordination leads in GDI and the participating nodes is a requirement to help mitigate the identified risks for reaching operational status. The lack of engagement could be due to overloaded communications at the nodes, lack of human resource at the nodes, or the messages not getting through to the intended audience. Understanding the reasons for lack of engagement is crucial. Responses to the node level deadlines for each GDI phase are divergent, and these need to be clarified in terms of the GDI project as a whole.

This deliverable has a significant impact within the GDI. It demonstrates a large diversity in the level of engagement, higher than expected, of the different countries that could have an impact not only on the deployment of the different nodes but also on the readiness of the EU-level infrastructure services, and highlights the need to increase engagement. However it also shows that GDI has the



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

potential to deliver genomic data access in countries which currently are not engaged in FEAGA, increasing the coverage of European Union countries aiming to share genomic and health data.

The volume of data anticipated within the nodes appears to be low, and clarification on these figures needs to be performed in conjunction with the expectations of both national priorities and Genome of Europe, with GoE itself expecting more national genomes than indicated here<sup>7</sup>.

The diversity of products anticipated for use by the nodes means care must be taken when determining the technical performance and conformance of a node to the required service levels and performance indicators. These different products cannot negatively affect the performance of the GDI as a whole, and the ongoing work in WP4 and 5 on the monitoring of the infrastructure must take account of the responses given in this document.

## 8. Next steps

All GDI use cases being mentioned at least once in the responses indicates that the use cases are relevant for all nodes. However, where a priority national use case exists that is not covered by the GDI use cases there should be a way of ensuring the requirements from those use cases are brought into GDI, either via the 1+MG framework or via GDI stakeholder engagement. A procedure for this process will be created in collaboration with WPs 1, 4, and 7.

Further analysis of the landscape question is required, as well as node level clarifications, for example the data access model and how this relates to the 1+MG Data Governance. Presentation of these findings may itself help increase engagement with the nodes, and the planned node engagement workshop in March 2024 will help with this. Additionally the node engagement workshop will enable the countries to feedback on the technical progress and influence the work needed to ensure it meets the national requirements, with the aim of developing a roadmap that aligns the national requirements with those of GDI, both from a governance and technical perspective. Progress should be made on separating the technical requirements out from the operational, legal, and national requirements so the technical monitoring of the progress towards having a genomic data sharing infrastructure can be properly conducted, and also proceed as much as possible while the ELSI and governance model is fully defined.

---

<sup>7</sup> <https://zenodo.org/records/8055610>



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



## 9. Glossary

**BAM/CRAM** - Binary / Compressed Reference-orientated Alignment Map file that contains the raw data from genome sequencing.

**Beacon** - A GA4GH standard allowing a query to be performed on genomic and phenotypic data with a response tailor to the access level of the user.

**Beacon Network** - A network of Beacons which allows a single Beacon query to be sent to multiple Beacons, and the responses returned to the user, allowing a user to find a genomic variant of interest in multiple different genomic data repositories.

**DAC** (Data Access Committee) - A committee that reviews and decides on access requests to controlled access resources.

**FAIR** - Findable, Accessible, Interoperable, Reusable.

**FEGA** (Federated European Genome-phenome Archive) - A federated node of the European Genome-phenome Archive (EGA) that provides controlled access to human genetic and phenotypic data.

**GA4GH** - The Global Alliance for Genomics and Health - A standard setting and policy framing not-for-profit organisation.

**GA4GH Passport** - A GA4GH standard that holds a set of attributes and entitlements for a specific electronic identity, for example the entitlement to access a controlled access resource.

**Galaxy** - An open source platform for FAIR data analysis.

**Molgenis** - Data platform for bioinformatic data analysis.

**REMS** (Resource Entitlement Management System) - Software that allows the management of data access permissions to different resources. It allows users to request access to a resource, DAC to grant or reject these requests, and provides a GA4GH Passport allowing the user to access a particular resource.

**VCF (gVCF)** - (Genomic) Variant Call Format files that contain a list of genomic variants from 1 or more samples.



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





## Appendix I

	Country	PT	BE	BG	FI	SI	LV	NL	ES	SE	EE	NO	IE	FR	LU	DK	DE
Beacon	Beacon	Will keep	Will keep	Will keep	Will keep	Will keep	Have not decided yet	Have not decided yet	Will keep	Will keep	Will keep	Will keep	Will keep	None	None	None	None
	What is the timeline to deploy it?	2023-04-01	2023-09-27	2025-10-31	2024-05-31	2023-12-04			2023-12-31	None	2023-12-15	2023-06-30	None				
	What is the timeline to decide it?						None	None									
REMS	REMS	Will keep	Will keep	Will keep	Will keep	Will keep	Have not decided yet	Have not decided yet	Have not decided yet	Will keep	Will keep	Will keep	Will keep	None	None	None	None
	What is the timeline to deploy it?	2023-06-01	2023-09-27	2025-10-31	2024-05-31	2023-12-04				None	2023-12-15	2023-06-30	None				
	What is the timeline to decide it?						None	None	2023-12-31								



European  
Genomic Data  
Infrastructure



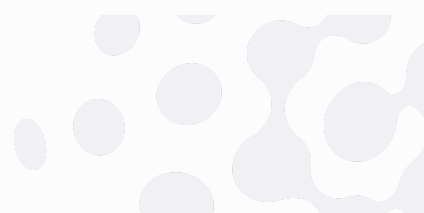
GDI project receives funding from the European Union's Digital grant agreement number 101081813.



	<b>Archival Technology</b>	Will keep	Have not decided yet	Will keep	Have not decided yet	Will keep	Have not decided yet	None	Will replace	Will keep	Will keep	Will replace	Have not decided yet	None	None	None	None
<b>Archival Technology</b>	<b>What is the alternative?</b>								Custom solution based on Spanish node FEGA			FEGA storage and interfaces , currently very aligned with the GDI starter kit component.					
	<b>What is the timeline to deploy it?</b>	2023-09-01		2024-12-31		2023-12-04			2023-12-31	None	2023-12-15	2023-09-21	None				
	<b>What is the timeline to decide it?</b>		2024-09-01		2023-12-31		None										
<b>Federated Computation</b>	<b>Federated computation</b>	Have not decided yet	Will replace	Have not decided yet	Will keep	Have not decided yet	Have not decided yet	None	Will replace	None	Have not decided yet	Have not decided yet	Have not decided yet	None	None	None	None
	<b>What is the alternative?</b>		Galaxy or WfExS						Custom solution based on Galaxy		2023-12-15	2024-12-31	None				



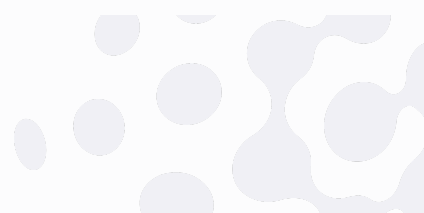
GDI project receives funding from the European Union's Digital grant agreement number 101081813.



									and openVRE								
	What is the timeline to deploy it?		2024-09-01		2024-05-31				2023-12-31								
	What is the timeline to decide it?	2023-11-30		2025-10-31		2023-12-21	None										
Containerised Computation	Containerised computation	Will keep	Will keep	Will keep	Will keep	Have not decided yet	Have not decided yet	None	Will replace	Will keep	Will keep	Have not decided yet	Have not decided yet	None	None	None	None
	What is the timeline to deploy it?					2023-12-22	None		Custom solution based on openVRE			2025-09-30	None				
	What is the timeline to deploy it?	2023-09-04	2024-09-01	2024-12-31	2024-05-31				2023-12-31	None	2023-12-15						
Additional Products	Are you using additional products?	No	Yes	Have not decided yet	Have not decided yet	Yes	Have not decided yet	Yes	Yes	No	No	Yes	Have not decided yet	None	None	None	None
	What are the products?		None	2024-12-31	2023-12-31	None	None	None	None				None				



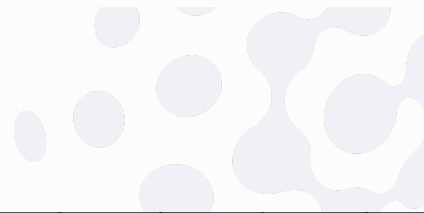
GDI project receives funding from the European Union's Digital grant agreement number 101081813.



What are the products? (1) Name		Galaxy			Molgenis		a data catalogue	Beacon Network			None					
What are the products? (2) Timeline to deploy it		None			4.12.2023		Q3/Q4 2023	31/12/2023			Proxy services for data in and data out endpoints to/from the secure infrastructure.					
What are the products? (3) URL		None			<a href="https://www.molgenis.org/">https://www.molgenis.org/</a>		None				Already deployed for FEAGA. Being used for both FEAGA and GDI currently.					
What are the products? (3) URL											None					



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



Within node Federation	Within-Node Federation	All the data to be accessible through GDI is located on a single centralized data warehouse	All the data to be accessible through GDI is located on a single centralized data warehouse	All the data to be accessible through GDI is located on a single centralized data warehouse	Only part of the data will be centralized	Only part of the data will be centralized	All the data to be accessible through GDI is located on a single centralized data warehouse	All the data accessible through GDI will be federated within the node - no central data warehouse	Only part of the data will be centralized	Only part of the data will be centralized	All the data to be accessible through GDI is located on a single centralized data warehouse	All the data to be accessible through GDI is located on a single centralized data warehouse	Have not decided yet	None	None	All the data to be accessible through GDI is located on a single centralized data warehouse	None
	What is the timeline to decide it?												None				
	Is data transferred in an encrypted format?	Yes	Yes	No	Yes	Yes	Yes		Yes	Yes	Yes	Yes				Yes	

Data Transfer



European Genomic Data Infrastructure



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



	<b>What is the encryption format?</b>	Crypt4GH	We aim to use GA4GH Crypt4GH standard	Data is generated at the same site where it will be stored and will be transferred securely within the LAN segment.	Crypt4GH	None	Data themselves are not encrypted, however, transfer will be performed through https or SSH File Transfer Protocol		crypt4gh	Crypt4gh	Crypt4GH	Crypt4gh				Only on ingestion, not relevant for GDI	
	<b>How will data be transferred to the centralized data storage?</b>	using SDA-cli	The preferred way is Globus, other arrangements will have to be made for partners without Globus	Have not decided yet	Using the htsgit protocol.	Network connectivity.	Through API defined by the storage provider which uses https protocol.		None	None	this is yet to be defined by WP3 and WP6	Using Life Science login authenticated upload endpoints to the Storage and Interface component of GDI. Similar to functionality implemented for				None	



European Genomic Data Infrastructure



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



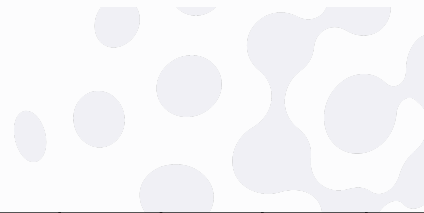
												FEGA Norway.					
	<b>Data users will be allowed to download the data from the node after DAC review?</b>	No	No	Yes	Yes	Have not decided yet	Yes	No	Yes	None	No	Yes	Yes	None	None	No	None
<b>Priority use cases and stakeholders</b>	<b>Will there be prioritization of specific use case or stakeholder needs?</b>	Not decided yet	Yes	GDI use case 7.1 (Genome of Europe demonstrator)	No	No	Yes	Not decided yet	Not decided yet	Not decided yet	Yes	Not decided yet	Yes	None	None	None	None
	<b>Do you have any priority GDI use cases?</b>		GDI use case 7.1 (Genome of Europe demonstrator)	Won't prioritize 1+MG use cases			GDI use case 7.1 (Genome of Europe demonstrator)						GDI use case 7.1 (Genome of Europe demonstrator)				
	<b>Do you have any priority 1+MG use cases?</b>		Yes	Yes			Yes				GDI use case 7.1 (Genome of Europe demonstrator)		Yes				



**European  
Genomic Data  
Infrastructure**



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



			Cancer usecase	Rare Diseases			Use in Polygenic risk score calculations				Won't prioritize 1+MG use cases		Cancer, population genomics, rare diseases				
	Do you have any priority national use cases?		Yes				Yes				No		Yes				
	Please, describe		Belgium Genome Biobank data				Data use for rare disease analysis						Variant lookup, creating a representative reference genome for this lookup and for variant calling				
	What is the expected timeline to your node reach onboarding status?	2023-10-11	2023-12-31	2024-12-31	2023-10-01	2023-12-22	2025-10-31	2023-12-21	2023-11-30	2023-10-01	None	2024-06-30	2026-11-30	None	None	None	None



European Genomic Data Infrastructure



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





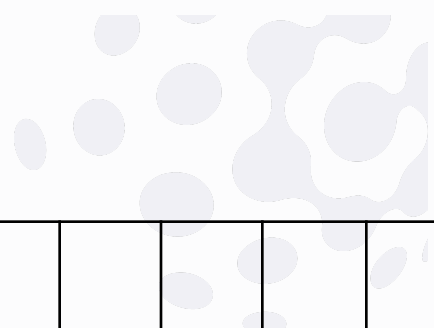
What is the expected timeline to your node reach deployment status?	2024-11-30	2024-10-31	2026-05-31	2024-05-31	2025-10-31	2026-10-31	2024-10-31	2023-12-31	2025-01-31	None	2025-10-01	None	None	None	None	None
What is the expected timeline to your node reach operational status?	2024-11-30	2026-09-30	2027-03-31	2024-10-31	2026-11-30	None	2026-10-30	2024-01-31	2026-10-21	None	2026-10-31	None	None	None	None	None
Will you also host a FEGA instance?	Yes	No	No	Yes	No	No	No	Yes	Yes	No	Yes	Yes	None	None	None	None
Will GDI data and FEGA data be hosted on the same technical infrastructure?	Yes			Yes				Yes	Yes		Yes	Yes				





**European Genomic Data Infrastructure**



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



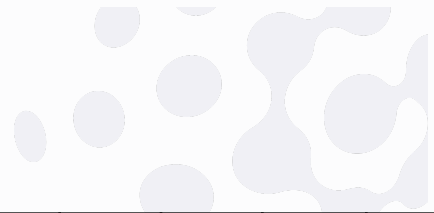
	<p><b>How will the different governance models between FEGA and GDI be handled?</b></p>	<p>They will have separate governance</p>			<p>FEGA service is handled under its own contracts and service ownership. While some of the data holders might coincide, in terms of the DAC process and contracts, the data are separately defined in each case.</p>				<p>We are still defining the governance model for GDI</p>	<p>None</p>		<p>Although FEGA and GDI may be hosted on the same technical infrastructure, with two different governance structures, a level of separation is expected. This could range from two parallel operations/complete deployments, with partly overlapping teams, operating under the different</p>	<p>Unknown</p>				
	<p><b>European Genomic Data Infrastructure</b></p>			<p>GDI project receives funding from the European Union's Digital</p>	<p>grant agreement number 101081813.</p>												



											governance structures (one for EDIC/GDI, one for FEGA), to more integrated solutions with some components in common, and separation at dataset level. Whether some genomes would be available through both governance models has not been explored yet. The					
--	--	--	--	--	--	--	--	--	--	--	---	--	--	--	--	--



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



												Norwegian part of GoE cohort is planned for inclusion in both GDI and in the Norwegian national genome center, and ongoing work on a national genetic variation reference database.					
Do you have any associated projects which are important to GDI?	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	None	Yes	No	Yes	None	None	None	None



European  
Genomic Data  
Infrastructure



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



	Please, list them		None		None		None	None	None		None					
	Please, list them (1) Name of the project		BGB		REMS		None	EOESC4Cancer Operational organisations like HMF (Hartwig Medical Foundation: <a href="https://www.hartwigmedicalfoundation.nl/">https://www.hartwigmedicalfoundation.nl/</a> )	IMPACT ( <a href="https://impact.isciii.es/">https://impact.isciii.es/</a> ). Spanish national project		Genome of Europe		EHDS, National Strategy for Genetics and Genomics Implementation Plan, All Island Cancer Research Institute, Data Access Sharing and Linkage (DASSL), Future Neuro Research Centre			
	Please, list them (2) When it will		None		2024-05-31		None	2024-10-31	2023-12-31		2024-12-31		None			



European  
Genomic Data  
Infrastructure



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



be deployed?															
Please, list them (3) Does it use the same products of the starter kit?		Some of them		Yes		None	No	No		No		None			
What product of the starter kit will be used?		Beacon; REMS					Use cases in EO SC4 Cancer are aligned with the Cancer use case in GDI HMF has its own operational infrastructure (currently using Google Cloud)	Some of the products will be used but they will be customized to have the same function as the starter kit products.		Via 1+MG					



European  
Genomic Data  
Infrastructure



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



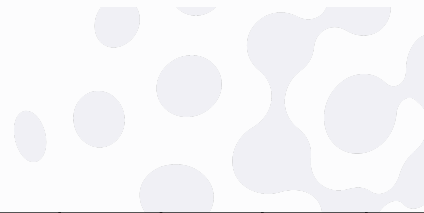
Do you have any associated data spaces or infrastructures important to GDI?	No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	None	None	None	None	None
Please, list them				None	None	None	None		None	None	None					
Please, list them (1) Name of the dataspace/infrastructure				Sensitive data services, SD Connect	ELIXIR-SI data storage	None	Health-RI		BigPicture	National Health Information System (TEHIK)	National Genome Center Norway					
Please, list them (2) When it will be deployed?				2022-11-01	None	None	2023-12-28		None	None	2026-10-31					
Please, list them (3) Does it use the same products of the starter kit?				Some of them	No	None	No		None	No	No					



European  
Genomic Data  
Infrastructure



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

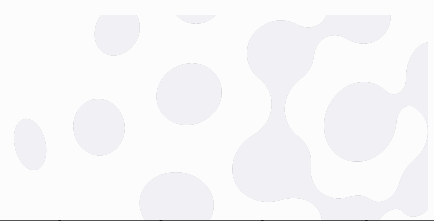


	<b>What product of the starter kit will be used?</b>				htsget; LSAI; REMS	Not dependable on filesystem usage. Used only for data storage.	None	At the catalogue level; same DCAT-AP model should be employed			not decided yet; depends on adjustments of legal framework to allow data flow from national health records to GDI	Not yet decided/explored. Technical infrastructure for the National Genome Center as well as governance model is not decided yet.						
<b>Data Types and Ontologies</b>	<b>What existing genomic data formats/types will you be using?</b>	None	None	None	None	None	VCF files	None	None	None	None	None	None	None	None	None	None	None
	<b>What existing genomic data formats/types will you be using? (1) What</b>	VCF files		VCF files	BAM files	FASTQ files	~50TB				VCF files	FASTQ files	FASTQ files					



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

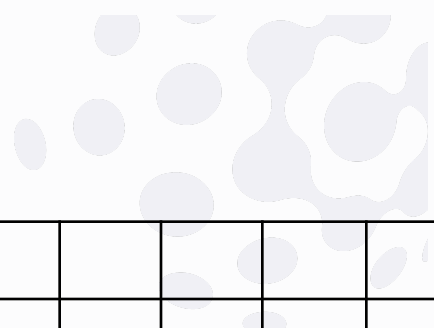




existing genomic data formats/types will you be using?																
What existing genomic data formats/types will you be using? (2) What volume of data of this type will you be working with?	<100GB estimated		1000+	500TB	Smaller portion of data.	BAM files					2000 full genomes	None	Unknown			
What existing genomic data formats/types will you be using? (3) What existing genomic data formats/type	CRAM files			VCF files	BAM files	~500TB					BAM files	VCF files				



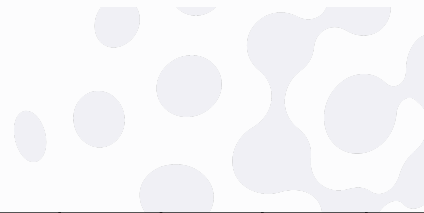
GDI project receives funding from the European Union's Digital grant agreement number 101081813.



s will you be using?																
What existing genomic data formats/types will you be using? (4) What volume of data of this type will you be working with?	~300TB			100TB	Smaller portion of data.						None	Unknown				
What existing genomic data formats/types will you be using? (5) What existing genomic data formats/types will you be using?				CRAM files	VCF files						CRAM files					



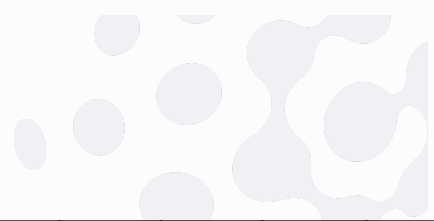
GDI project receives funding from the European Union's Digital grant agreement number 101081813.



<p>What existing genomic data formats/types will you be using? (6) What volume of data of this type will you be working with?</p>				100TB	More than 80% of data will consist of this format.						None					
<p>What existing genomic data formats/types will you be using? (7) What existing genomic data formats/types will you be using?</p>											VCF files					



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



<p>What existing genomic data formats/types will you be using? (8) What volume of data of this type will you be working with?</p>											None					
<p>Has the encodings/terminologies/vocabularies/ontologies will you be using for phenotypic data collection being decided?</p>	Yes	No	No	No	No	No	No	No	None	No	No	No	None	None	None	None
<p>Encoding/terminologies/vocabularies/ontologies</p>	None	2024-09-01	2024-12-31	2024-03-31	None	None	None	2023-12-31		2025-12-31	None	None				



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



<b>Encoding/terminologies/vocabularies/ontologies (1) Name</b>	Minimal Information about a high throughput Sequencing Experiment																
<b>Encoding/terminologies/vocabularies/ontologies (2) Link</b>	<a href="https://fairsharing.org/FAIRsharing.a55z32">https://fairsharing.org/FAIRsharing.a55z32</a>																
<b>Will you use ontologies for user consent?</b>	Yes	Not decided yet	Not decided yet	No	Not decided yet	Not decided yet	Yes	Not decided yet	None	Not decided yet	Not decided yet	Not decided yet	None	None	No	None	
<b>What is the timeline to decide it?</b>		2024-09-01	2024-12-31		None	None	None	2023-12-31		2025-12-31	None	None					
<b>Which ontology will you use?</b>	None																



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



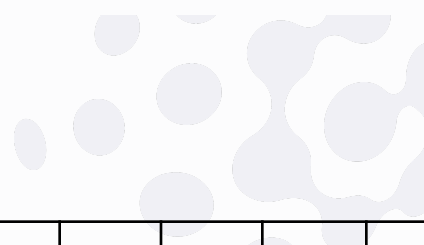
Data Quality	Will you monitor data integrity once it has been transferred?	Yes	Yes	Yes	Yes	Yes	Yes	None	No	None	Yes	Yes	No	None	None	Yes	None
	Will you be keeping a master list with checksums of certified/correct/canonical/verified data?	Yes	Have not decided yet	Have not decided yet	Have not decided yet	Yes	Have not decided yet				Yes	Yes				Yes	
	Will genomic data quality be assessed?	Yes	Have not decided yet	Yes	No	Yes	Yes	None	No	None	Have not decided yet	Yes	Have not decided yet	None	None	None	None
	What is the timeline to decide it?		2024-09-01								2025-12-31		None				



European  
Genomic Data  
Infrastructure



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



	<p><b>How will genomic data quality be assessed?</b></p>	<p>FastQC, BioQC or others defined in the project</p>		<p>We will follow B1MG D3.2 Best practices for Next Generation Sequencing (10.5281/zenodo.7912923) and the recommendations will be implemented in the routine pipeline.</p>		<p>Genomic QC tools (FastQC, Nanopolish, etc.)</p>	<p>None</p>					<p>Routine QC metrics are being implemented in FEQA, as well as in 1+MG dedicated WG. We plan to implement measures from both of these sources.</p>					
	<p><b>If genomic data does not reach a established quality standard, will it still be made available through GDI?</b></p>	<p>No</p>	<p>No</p>		<p>Have not decided yet</p>	<p>Have not decided yet</p>						<p>Have not decided yet</p>					



European Genomic Data Infrastructure



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



Will you use a metadata label to describe genomic data quality?	No		Have not decided yet		None						None					
What is the timeline to decide it?			2024-12-31		Yes	Have not decided yet					Have not decided yet					
Please, describe					overall accuracy (error rate), depth, average read length, number of reads											
What is the timeline to decide it?						None					None					
Will you host non-genomic data?	Have not decided yet	Have not decided yet	Have not decided yet	Have not decided yet	Have not decided yet	Other types of data	Have not decided yet		None	Have not decided yet; Other types of data	Have not decided yet	Have not decided yet	None	None	None	None



European  
Genomic Data  
Infrastructure



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







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