



Deliverable D5.2

Second annual report on coordination and outreach

Project Title Grant agreement no	Genomic Data Infrastructure Grant agreement 101081813		
Project Acronym (EC Call)	GDI		
WP No & Title	WP5: Technical Coordination and Outreach		
WP Leaders	Tommi Nyrönen (6. CSC), Dylan Spalding (6. CSC)		
Deliverable Lead Beneficiary	6. CSC		
Contractual delivery date	31/10/2024	Actual delivery date	24/10/2024
Delayed	No		
Partner(s) contributing to deliverable	CSC		
Authors	Dylan Spalding (CSC)		
Contributors	Stefanie Kirschenmann (CSC), Jaakko Leionen (CSC)		
Acknowledgements	N/A		
Reviewers	Dilza Campos (VIB), Melissa Konopko (ELIXIR)		

Log of changes

Date	Mvm	Who	Description
01/10/2024	v0.1	Dylan Spalding (CSC)	Initial version
07/10/2024	v0.2	Stefanie Kirschenmann	Added comments



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



		(CSC), Jaakko Leionen (CSC)	
09/10/2024	v0.3	Dylan Spalding (CSC)	Comments addressed, submitted for review
15/10/2024	v0.4	Dilza Campos (VIB), Melissa konopko (ELIXIR)	Reviewed and comments added
16/10/2024	v0.5	Yaseen Jakhura (ELIXIR Hub)	Quality checked and sent to GDI-MB review
16/10/2024	v0.6	Dylan Spalding (CSC)	Addressed reviewers comments
24/10/2024	v1.0	Yaseen Jakhura (ELIXIR Hub)	Deliverable submitted to EC

Table of contents

Contents

1. Executive Summary	4
2. Contribution towards project outcomes	5
3. Methods	7
4. Description of work accomplished	8
4.1 Technical Coordination	8
4.1.1 Zenhub and GitHub	8
4.1.2 Pillar II meetings	8
4.1.3 Malta WP5 Workshop	9
4.1.4 Brno Workshop	9
4.1.5 Milestones 7 & 8	9
4.1.6 Task forces	10
4.1.7 Pillar II and Pillar III interaction	11
4.1.8 Change management—RFC process.	13
4.2 Outreach	13





**European
Genomic Data
Infrastructure**

4.2.1 GA4GH	13
4.2.2 EHDS	14
4.2.3 EOSC-ENTRUST	14
4.2.4 Genome of Europe	14
4.2.5 ECCB	15
4.2.6 EOSC4Cancer	15
4.2.7 International Infrastructures and projects	15
5. Results	16
6. Discussion	17
7. Conclusions & Impact	17
8. Next steps	18



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



1. Executive Summary

This report describes the coordination and outreach activities performed by work package 5, and collaborating work packages, since D5.1—Report on coordination and outreach, towards the establishment of a federated infrastructure for access to genomic and associated phenotypic data. Over the reporting period (01/10/2023 - 01/10/2024) existing coordination and communication methods have been changed to address issues and risks that have been identified, which includes:

- Establishment of inter-product and inter-pillar task forces
- Adjustment of the Pillar II meetings to help increase engagement
- Holding face-to-face technical and coordination meetings
- Establishment of cross-pillar meetings

During the reporting period milestone 7 was achieved, which required the planning of a demonstrable user story and associated coordination with work package 3 and participating nodes. As of 2nd October 2024, we are actively coordinating and planning the milestone 8 which demonstrates the infrastructure supporting a Pillar III use case.

In terms of outreach, GDI has become a driver project of GA4GH, and as such now has the opportunity to influence the standards that enable genomic data sharing on a global scale. As such GDI has been represented at GA4GH meetings, and in the GA4GH Steering Committee. To maximise the impact of GDI, and realise the vision of interconnected domain specific data spaces, GDI has participated and collaborated with other infrastructures and projects, both at European and global level, to ensure that the solutions being developed and deployed by GDI are interoperable with these other infrastructures and projects.





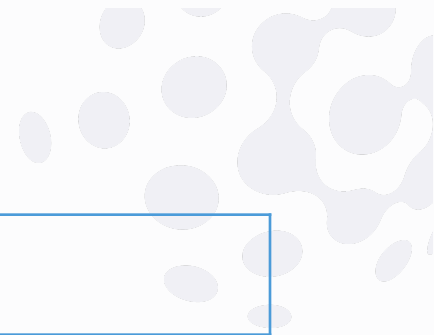
2. Contribution towards project outcomes

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
<p>Outcome 1</p> <p>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.</p>	No
<p>Outcome 2</p> <p>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.</p>	Yes
<p>Outcome 3</p> <p>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.</p>	No
<p>Outcome 4</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</p>	No





(e.g., IT and biotech companies), healthcare systems and public authorities at large.	
<p>Outcome 5</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>Outcome 6</p> <p>Communication strategy – to be designed and implemented at the European and national levels.</p>	No
<p>Outcome 7</p> <p>Capacity building measures necessary to ensure the establishment, sustainable operation, and successful uptake of the infrastructure.</p>	Yes
<p>Outcome 8</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





3. Methods

Pillar II is comprised 4 work packages:

- WP3 to support the progression of nodes through the three phases—onboarding, deployment, and operational as well as software development;
- WP4 for European level operations which defines the SOPs, develops the User Portal, and operations for the operational infrastructure;
- WP5 for technical coordination, training, and outreach;
- WP6 for data management and technical implementation of the data governance

Pillar II is tasked with the technical deployment of the infrastructure as defined by 1+MG, and WP5 coordinates the different WPs in Pillar II to support this, as well as facilitates communication and interactions between Pillar II and Pillars I & II to ensure the technical infrastructure supports the required data governance and sustainability from Pillar I, and the use cases and innovative solutions (such as federated analysis and learning) as defined by Pillar III.

WP5 facilitates the transition of the nodes through the three phases of maturity; from onboarding, through deployment, to operational. As such, WP5 needs to provide the technical coordination to support the progress of the nodes towards operational, provide the training and knowledge transfer as required to support the nodes, and outreach to stakeholders and associated projects, organisations, and infrastructures. WP5 has been arranging meetings, workshops, email and Slack channels, as well as project management and repositories to facilitate this. As the specifics of the core infrastructure due to be deployed become clearer, specific training requirements can be identified and addressed, as well as more general training from production nodes to the deployment and onboarding nodes.

Each node that aims to deploy GDI infrastructure is required to identify their aspirational phase, and the timeline to achieve that. While each node must support the five functionalities (data discovery, data access and management, storage and interfaces, data reception, processing) as defined by B1MG / 1+MG¹, the node may choose how to implement them within the governance framework as defined by Pillar I. This is to support the diversity in the national policies for genomics and health data, as well as funding processes. A challenge for WP5 is to ensure that this national diversity is supported within the GDI and does not negatively affect the FAIRness of the infrastructure as a whole.

The early milestones, such as the release of the GDI Starter Kit² and the deployment of it across 7 vanguard nodes in milestone 7³, help identify the gaps and requirements for the infrastructure and enable the nodes to investigate the technical standards and implementations on their own physical

¹ <https://doi.org/10.5281/zenodo.6089583>

² <https://github.com/GenomicDataInfrastructure/starter-kit>

³ <https://drive.google.com/file/d/1-hj3eM98qZmb5Ew5RiAFz7SyxTk-bySk/view>



infrastructures. WP5 works to ensure that the different work packages and Pillars in GDI support the 1+MG vision, and has over the last 12 months developed new procedures and methodologies to support the communication across them.

4. Description of work accomplished

4.1 Technical Coordination

4.1.1 Zenhub and GitHub

To track and help identify blocking issues relating to the deployment of the GDI infrastructure, a Zenhub⁴ project with 72 team members has been created for GDI, which supports agile project management processes. Zenhub supports workspaces, which can relate to specific tasks, deliverables, or milestones within the project as required, as well as sprints, epics, and product backlogs, however each participating institution maintains control over their own project management procedures. Within each workspace epics can be defined to track the progress of tasks towards a particular goal, as well as roadmaps and reporting tools. Zenhub supports both GitHub and Slack integration, and GDI has 37 GitHub⁵ repositories, with 84 people across 7 different teams, supporting collaboration on the software across the different nodes and with external collaborators.

4.1.2 Pillar II meetings

Pillar II meetings have been held regularly on a monthly basis, on the first Wednesday of each month. These meetings provided a platform for the different work packages and Pillar to give updates, as well as discuss specific issues related to the deployment of the infrastructure, and were aimed to reach the different countries and their national coordinators as described previously⁶. However, it became clear that the meetings, while well attended with an average of approximately 54 attendees over the last 12 months, were not reaching all the necessary participants within Pillar II, especially the nodes. Therefore, the Pillar II meeting was re-formatted to have presentations from three different nodes at each meeting, where the nodes could present their current status, problems, and requests for help, as well as the existing WP and Pillar updates. To try to ensure that all nodes were encouraged to participate, a draft timetable was generated alphabetically for the nodes, with 2 operational nodes selected and 1 onboarding / deployment node selected for each meeting. Nodes could request a different time slot only in exceptional circumstances to try to ensure all nodes

⁴ <https://app.zenhub.com/>

⁵ <https://github.com/orgs/GenomicDataInfrastructure/repositories?type=all>

⁶ <https://zenodo.org/records/10037055>



participated. The first meeting in the new format was held in September 2024 and Belgium, the Czech Republic, and Bulgaria presented.

Each of the four work packages in Pillar II, except WP5, organise their own monthly meetings.

4.1.3 Malta WP5 Workshop

A WP5 workshop⁷ was held in Malta on 13–14 March 2024. This was co-located with a Pillar I meeting, which also enabled the first x-pillar meeting to be held. The main goals of the WP5 workshop were to address the risk of lack of engagement with the nodes, understand how the nodes see themselves and their interaction with GDI, and identify nodes with similar issues to form collaborations that could help address these issues. 22 nodes were represented in person, with 2 other nodes present virtually.

At the x-pillar meeting the minimal viable product (MVP) concept was presented by Pillar I, and this initiated the discussion between all Pillars on the definition of this, and how clinical impact can be demonstrated by GDI. A report of the WP5 workshop and x-pillar meeting was presented to the 1+MG Special Group in Brussels on 30th May 2024.

4.1.4 Brno Workshop

In collaboration with WP3, a technical workshop⁸ was held in Brno between 2.9.24 and 4.9.24 with a total of 85 in person attendees. This provided an opportunity for the recently established task forces to meet in person and start to plan their work. Additionally, during this meeting the nodes were requested to indicate what they could achieve by 2025 and 2026⁹, and this input would be used to help formulate the MVP, as any definition of the MVP would need nodes to agree. As identified in D3.2 - Node deployment roadmap¹⁰ there is a risk of lack of engagement with the nodes and this was further identified at the workshop with the underspending in GDI as a whole. As mentioned above, the Pillar II meeting has been adjusted to try to increase node engagement specifically, and the task forces to try to lower the bar for node engagement in different tasks within GDI. A feedback form was distributed to the participants and as of October 15th 2024 had 28 responses, with 78.6% of those responses indicating that the workshop was useful overall, however further analysis of the results is ongoing.

4.1.5 Milestones 7 & 8

In collaboration with WP3, WP5 helped define and coordinate milestones 7 & 8. Milestone 7 demonstrated seven different nodes, with diverse synthetic datasets, connected to the User Portal to

⁷ <https://docs.google.com/document/d/1iiLA5J7VGOen3RqSj2NuxPxYrvWocbP6ZV8ZE-ciU7Q/edit?pli=1#heading=h.inartoxze8ck>

⁸ <https://docs.google.com/document/d/1QoUkxZDyuBy3moyA22X-yqUR3nSSFy7ZttCzFGhf8lw/edit#heading=h.qjdqxs>

⁹ https://docs.google.com/presentation/d/1GcaF_zm-qUvX9Bvoo21IRuBdoZoevFOASJ1Fa5SVoD8/edit#slide=id.p1

¹⁰ <https://zenodo.org/records/10688135>



demonstrate the complete user journey - from data discovery to data access and analysis. Each of the seven nodes hosted synthetic data, descriptions of which were displayed in the User Portal, and four nodes demonstrated data access and analysis in a Secure Processing Environment (SPE). Milestone 8, in progress as of 2nd October 2024, required coordination between all work packages and Pillars within GDI as it is due to extend milestone 7 by demonstrating a Pillar III use case, and must conform to the data governance as defined by Pillar I, and incorporate a first step towards the federated analytics as required by Pillar III.

4.1.6 Task forces

Responses from the Malta workshop expressed a desire to see more clearly specified APIs within the GDI Starter Kit. The GDI Starter Kit was originally product based, primarily to hit the early milestones within GDI and act as a demonstration tool, or reference implementation, to allow different nodes to deploy and understand how the proposed infrastructure would operate, based on the outcomes of the B1MG project. However, development of the different products has been slow, and the different maturity of these products has not been fully addressed across all products within the Starter Kit. Hence, it was proposed to set up an API task force, whose aim would be to document and link to existing API specifications, where they exist, such a Beacon, REMS, htsgget, or TES, and to identify gaps within existing, or non-existing APIs that need to be addressed within GDI.

Following the identification of the need for this specific task force, a set of additional task forces were set up to address specific technical issues that span across products, or that are not addressed specifically by a particular product. Each task force is self-organising, and has a specific Slack channel which currently is the prime method of communication for each task force. A statement of purpose is created for each task force, and reporting from the task forces will be added to either the Pillar II meeting agenda, or the WP3 meeting agenda (TBC).

Task Force	Number of participants in the task force	Statement of purpose
Productisation	21	Should define what this means in a GDI context, who is responsible for what (e.g. database failover, logging), testing requirements (e.g. code coverage percentage, unit testing), vulnerability monitoring, etc. The goal would be to define the base level requirements that the software needs to meet to be tagged 'production', as well as a report on what nodes need to consider to be labelled 'production' once they have deployed the software.



Persistent Identifiers	6	Map out which contexts within GDI necessitate persistent identifiers (PIDs), for each of these contexts explore exactly which entities require PIDs, and for each entity whether or not the task of defining and maintaining PIDs are sufficiently covered. For those entities not sufficiently covered, an approach for defining, provisioning and maintaining PIDs, such that they can facilitate the relevant context. Depending on the assumed scope of service provision and maintenance, determine viability of within-infrastructure delegation of responsibilities as opposed to necessitating external delegation.
Metadata	36	This squad is tasked with determining the metadata required for GDI, and to try to understand the different requirements and use cases across different nodes and other data spaces / infrastructures. This is very closely linked to T8.2.
Data Protection and Security	17	Identifying the specific responsibilities for each actor (e.g. PO, DevOps, administrator), the requirements which must be met by the productisation task force to conform to DPbDD requirements (e.g. the different products and deployments), with the result that a DPIA is defined for the infrastructure as a whole, and templates for the different nodes where required.
API	24	Define the required APIs across GDI, determine where an existing product or standard does not meet the requirements (or does not exist), and work with existing (or new) POs to generate a roadmap to address these issues.
AAI	41	Identifying the existing issues for the use of LS AAI in GDI, identify a viable solution to address these issues, and a roadmap to deploy this.

4.1.7 Pillar II and Pillar III interaction

As the requirements from the four GDI use cases, as well those for innovative solutions, such as federated analysis, come from Pillar III, a methodology for the interaction of Pillars II & III was created.



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

The decision path¹¹ (Figure 1) details how the requirements and feature requests are communicated from Pillar III to Pillars II & I, and how the feedback is returned where applicable. A template¹² document has been created to facilitate the way such features are requested, and helps track the progress of the decision through the decision path. The formal RFC process, which is being defined and is described below, will ultimately replace this decision path as it allows more detailed version tracking of the decision, and the involvement of a wider variety of stakeholders.

A simple illustration of the decision path for incorporating new features in the reference implementation. (Later, this will be replaced by SOP:s etc.)

Legend



Description of the illustration

- Time runs from left to right.
- The next step occurs only when there is an agreement to continue at the previous step.
- Each step can include discussions and improvements.
- The upper part of the illustration is common, while the lower interactions with dashed arrows occur when applicable.
- The final decision to create a reference implementation is displayed as a handshake in the illustration.
- A rejected idea may be reiterated in pillar III and suggested later as an improved idea.

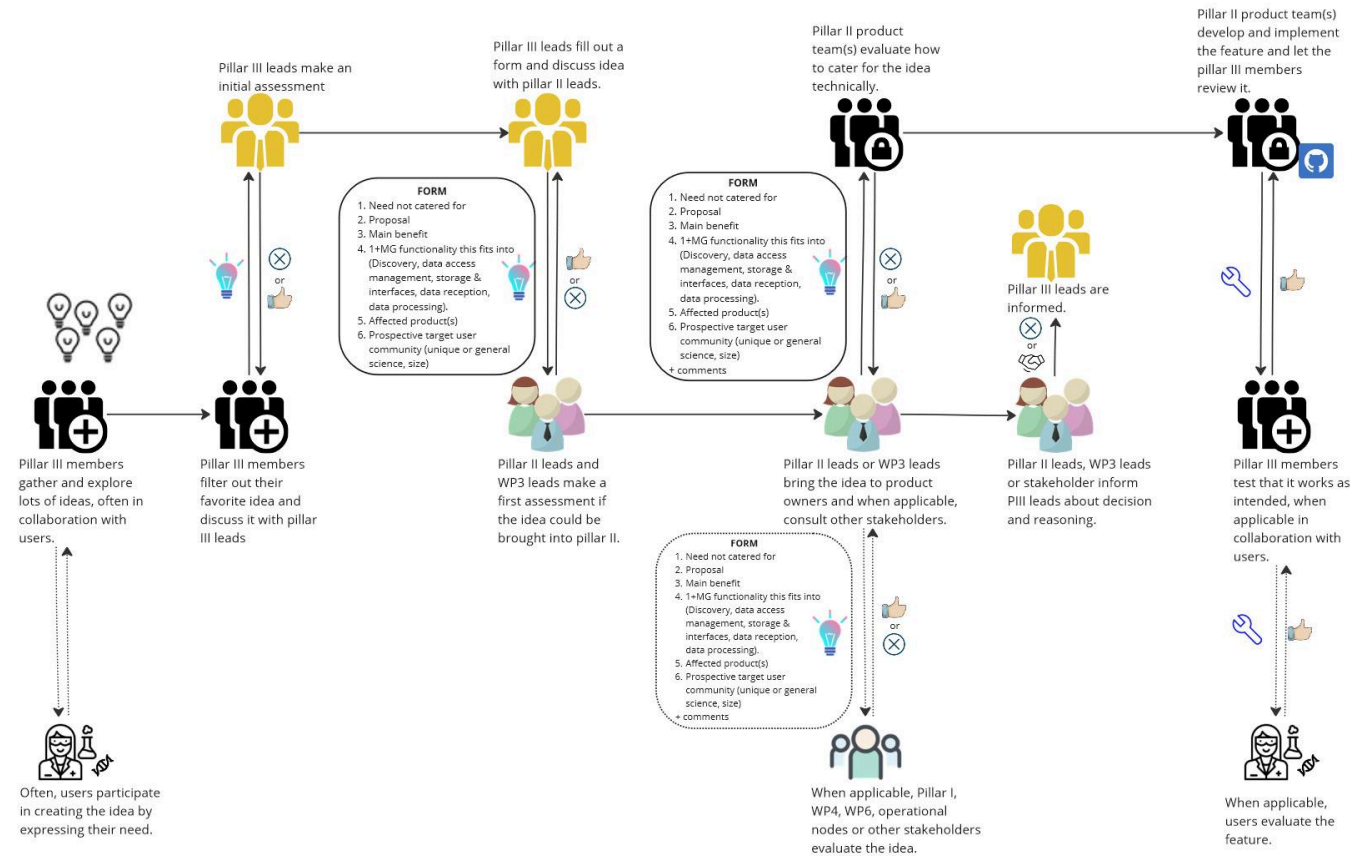


Figure 1: Diagram detailing the decision path for bringing new features and requests from Pillar III into the infrastructure.

¹¹ https://docs.google.com/presentation/d/10BYx6vscDzPxs2qQzS21XFFe0-He8Qk1bXVBxEvBl_pUU/edit#slide-id.g269d753bb91_o_o

¹² https://docs.google.com/document/d/1PCyH_SyGbrccfaJJ_Bpa191cWUXvhysPNLVSQq1aiW8/edit#heading=h.gjdgxg





4.1.8 Change management—RFC process.

A Request for Comments (RFC) process is in the process of being set up¹³. The aim of this procedure is that for every 'substantial' change in the infrastructure a formal process is followed which ensures that all relevant stakeholders have the opportunity to have input and express feedback on the proposal. In general such procedures are a form of change management, which ensure that the requirements that initiated a change, the discussions (including arguments for and against), the decision process, and the outcome are all recorded. This helps to ensure that any further change is done with full understanding of the previous changes, and the reasons for these changes.

The RFC process has to support a wide and diverse range of audiences and stakeholders, and as such has been specifically designed to support different communication mediums, such as GitHub and Google Docs.

4.2 Outreach

4.2.1 GA4GH

GDI was present at the Global Alliance for Genomics and Health¹⁴ (GA4GH) Connect Ascona¹⁵ meeting during 21–24th April 2024, and the joint Connect¹⁶, Plenary¹⁷, and National Initiatives Forum¹⁸ (NIF) meetings in Melbourne, Australia, between 16–25th September 2024. During Connect in Ascona, GDI as a driver project was presented in the plenary session, and GDI contributed into discussions on the GA4GH Passport standard, the Experimental metadata standard, phenopackets, provided feedback on EU ELSI requirements for technical standards, and requirements for data discovery. In Melbourne, presentations were given regarding the Beacon and its deployment in GDI, the cancer minimal metadata model, and GDI requirements for the Cloud workstream, especially regarding secure processing environments (SPEs) and federated or distributed computation, including how access to the resource, as opposed to the data, could be standardised.

During the Plenary, interest was expressed in understanding the variant or allele frequencies in under-represented populations to address the European bias¹⁹ in global genomics, and this has implications for the Genome of Europe work and understanding the genetics of different populations

¹³ <https://github.com/GenomicDataInfrastructure/rfc>

¹⁴ <https://www.ga4gh.org/>

¹⁵ <https://broadinstitute.swoogo.com/ga4gh-connect-2024/agenda>

¹⁶ <https://broadinstitute.swoogo.com/ga4gh-12th-plenary/connect>

¹⁷ <https://broadinstitute.swoogo.com/ga4gh-12th-plenary/agenda>

¹⁸ <https://broadinstitute.swoogo.com/ga4gh-12th-plenary/nif>

¹⁹ <https://www.nature.com/articles/s41562-019-0619-1>



within Europe, and methodologies for representing the allele frequencies (such as the proposal²⁰ by Ireland for demonstrating clinical utility via the Genome of Ireland), as well as access requirements, within GDI. An example described included a Federated GnomAD²¹ which envisions a set of federated nodes that each calculate and make queryable their own aggregated allele frequencies, and includes GA4GH standards such as Variant Representation Standard²² (VRS) and work on cohort allele frequencies within and Variant Annotation²³ (VA) specification, as described here²⁴.

4.2.2 EHDS

GDI Pillar II continues to interact with the European Health Data Space (EHDS) by the HealthData@EU²⁵ and TEHDAS²⁶ projects, and is a participant in the Common Data Spaces²⁷ along with EUCAIM²⁸ and EHDS. Additionally, with ELIXIR GDI inputs into the Data Spaces Support Centre²⁹ (DSSC) which aims to identify commonalities and best practices across different data spaces & domains within the EU.

4.2.3 EOSC-ENTRUST

GDI is monitoring outputs from the EOSC-ENTRUST³⁰ project, and is helping to input into the requirements via both the ELIXIR coordination and the Federated Human Genomics as a catalyst for European TRE provision³¹ (driver 1) in EOSC-ENTRUST WP7 and associated milestone report M7.1. This interaction with EOSC-ENTRUST aims to ensure the alignment between the outputs of EOSC-ENTRUST and the requirements of GDI, especially with respect to the technologies used and the data governance.

4.2.4 Genome of Europe

It is expected that the first data made available within the GDI will come from the Genome of Europe (GoE), and as such close collaboration is required between GoE and GDI to ensure that the GDI supports the requirements of GoE, and that the technical infrastructure, sustainability, and data governance are in place ready to support GoE as the data become available. GoE has provided a concrete use case for the volume of data to initially be managed by the GDI, as well as questions for

²⁰ https://www.canva.com/design/DAGFwO5lFzc/1-lfoA8wt_QaeUoek6bgA/view?utm_content=DAGFwO5lFzc&utm_campaign=designshare&utm_medium=link&utm_source=editor#1

²¹ https://docs.google.com/presentation/d/1XZIBJEKmlC1gqhrNZTssF9--qgeJaSzZmd44g4329Xk/edit#slide=id.g2f529d36a87_o_778

²² <https://vrs.ga4gh.org/en/stable/>

²³ <https://va-ga4gh.readthedocs.io/en/latest/>

²⁴ <https://gnomad.broadinstitute.org/news/2023-11-ga4gh-gks/>

²⁵ <https://ehds2pilot.eu/>

²⁶ <https://tehdas.eu/>

²⁷ <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

²⁸ <https://cancerimage.eu/>

²⁹ <https://dssc.eu/>

³⁰ <http://eosc-entrust.eu/>

³¹ <http://eosc-entrust.eu/drivers>



the harmonisation and interoperability of the data the GoE will provide in GDI, and the metadata associated with the genomic data generated by GoE.

4.2.5 ECCB

A specific GDI workshop³², entitled 'Connecting to the pan-European ELIXIR Genomic Data Infrastructure: Technical Infrastructure and Data Requirements' was held on Tuesday 17th September 2024 during the European Conference on Computational Biology in Turku and was attended by 17 people from Europe and beyond, including representatives from industry. This included a description of the architecture and infrastructure, details of how to connect to the infrastructure, the requirements for data inclusion, details of the Genome of Europe project which is expected to supply data to the GDI, and a session for questions and answers from the audience. Input into a GDI use case was received from the attendees.

4.2.6 EOSC4Cancer

GDI has been collaborating with and monitoring outputs from EOSC4Cancer³³ to help ensure that the federated genomics part of EOSC4Cancer is interoperable with GDI, utilising the same standards. The GDI Starter Kit is intended to be utilised in a EOSC4Cancer demonstrator. There is a high level of overlap of the standards and solutions between the two projects, such as the use of DCAT-AP³⁴ for standardised dataset metadata representation, Beacon³⁵ for data discovery, and LS Login³⁶ for federated identity management. Additionally, EOSC4Cancer has been using cBioPortal³⁷ for a cancer specific platform, and phenopackets³⁸ for a collateral cancer use case, which may have experience and knowledge to share with the cancer use case in WP7.

4.2.7 International Infrastructures and projects

To coincide with the GA4GH Connect and plenary meetings in Australia, a 1.5 day workshop³⁹ was hosted by Australian BioCommons⁴⁰ to understand the landscape of international activities in human genomic data sharing and analysis. GDI Pillar II was presented, with a focus on the GDI Starter Kit and how it aims to address some of the issues identified for allowing cross-border access to genomic data.

³² <https://eccb2024.fi/workshops-tutorials>

³³ <https://eosc4cancer.eu/>

³⁴ <https://semiceu.github.io/DCAT-AP/>

³⁵ <https://docs.genomebeacons.org/>

³⁶ <https://lifescience-ri.eu/ls-login/>

³⁷ <https://www.cbioportal.org/>

³⁸ <http://phenopackets.org/>

³⁹ https://docs.google.com/document/d/1_PdAao3DQ2ZWL_eL_KZY0EXAY4ckp1BmOCVYnS5i5VZAE/edit

⁴⁰ <https://www.biocommons.org.au/>



There is ongoing collaboration between Australian BioCommons and GDI via the Swedish node as product owners for htsget. The tool ELSA⁴¹ which allows a data manager, working on behalf of a DAC, to manage the sharing and access to genomic data, including of different participants within a virtual cohort based on their consent and the proposed research question, was discussed, and it's possible use within GDI to manage and track participants consent and support the data minimisation principle. An opportunity to collaborate with the Pan Canadian Genome Laboratory⁴² (PCGL) on Beacons, especially on how the data are represented within Beacons to ensure interoperability between different Beacons was also discussed, as there are different ways to represent the same object or objects within a Beacon which can cause incompatibility. An additional discussion on supporting phenopackets as both a query and a response from a Beacon was identified.

5. Results

1. Pillar II meetings adjusted
2. Task forces created
3. Interactions between Pillars II and III formalised
4. X-Pillar meetings established
5. Interactions with associated infrastructures, projects, and data spaces
6. GA4Gh driver project
7. In-person workshops held

For point 1, the Pillar II meeting was adjusted to address the identified lack of engagement from some nodes, and diversity in both interpretation of a GDI node, as well as the national landscape in which a node would operate, needed to be addressed to ensure that GDI delivers on having at least 15 nodes operational by November 2025.

For point 2, to address the issues across, and gaps between, products in the GDI Starter Kit, as well as tasks that cross Pillar or need additional engagement, the task forces were created. An example of a cross product task is the Data Protection and Security task force, and an example of a between product task force is the API task force. Both these task forces are open to members from all Pillars.

For point 3, before the formal RFC process is finalised there is a need to bring the requirements from Pillar II, the use cases as well as the innovative solutions, to Pillar II so that Pillar II could investigate the deployment of these solutions or address the use case requirements.

For point 4, related to point 3 but in this case across the whole of GDI, the x-pillar meetings have been established which help bring the governance requirements to Pillars II & III, as well as the possible technical solutions and use case requirements to address these.

⁴¹ <https://github.com/elsa-data>

⁴² <https://genomelibrary.ca/>



For point 5, GDI is continuing to interact with associated projects, and data spaces, such as EHDS, EOSC-ENTRUST, EUCAIM, and GA4GH. This is necessary to ensure the maximum interoperability of the relevant infrastructures, and maximise the FAIRness of the data.

Related to point 5, point 6 extends this influence to a global scale, and in turn identifies collaborations for the development of the infrastructure outside Europe. This helps ensure the European requirements are considered in developing the global standards, and additionally helps maximise interoperability globally.

Lastly, point 7 tries to build a GDI community via face-to-face meetings, and hence engagement within the project. These try to address the identified risk of lack of engagement in GDI along with the adjusted Pillar II meetings (1).

6. Discussion

1. Determination of the MVP
2. Ensure infrastructure suitable for real data
3. Maximise node engagement

Point 1 is a continuing discussion that was initiated by Pillar I at the workshop in Malta. Progress has been made in specifying this, for example by framing it in terms of MS8 and the Genome of Ireland use case, but work is still required to ensure that this can be achieved within GDI, and that the MVP includes the real data that is required within the GDI infrastructure to make it relevant and useful for the end users, such as the GoE data.

Point 2 is an extension of point 1 to data beyond GoE, and the requirement for GDI to learn from other infrastructures that already share real genomic data with their users, such as Federated EGA which already enables the sharing of real genomic and phenotypic datasets.

There is a real risk of lack of engagement from the nodes (3), and this is being addressed in multiple ways, such as face-to-face meetings, node presentations at the Pillar II meeting, identification of underspending at a node level and planning to re-assign some of these funds. Work is also ongoing to try and ensure the intra-node communication is functioning well, for example making sure the node representatives in each Pillar are aware of their specific issues and requirements.

7. Conclusions & Impact

1. GA4GH Driver project
2. Establishment of task forces



GDI has become a GA4GH driver project, and as such this enables GDI to help drive the specifications for evolving and new standards for genomic data sharing, which will have global impact. Already the benefit of this has been seen, for example with collaborations with infrastructures and institutions worldwide. The collaboration with Australian BioCommons by the htsget product owners is one example, as is the prospective collaboration with Ontario Institute of Cancer Research for Beacon data management.

But with potentially greater impact is the influence GDI can have on the new and evolving standards. An example would be the AAI and Passports used for identity management within GDI. GDI has a requirement for these to have a high level of assurance, and to link to the EU eID, and potentially the eWallet. GDI is ideally placed to bring these requirements to GA4GH and influence the definition, and hence adoption of these standards globally.

The task forces focus attention on a specific task of requirement within GDI, and as such encourage participation of different individuals who may find contributing directly to the product code repositories in GitHub daunting in the first instance. Additionally, the task forces can bring in external knowledge, for example via 1+MG experts, which helps broaden the knowledge base from which a solution may be found. The task forces are self organising, with the aim of encouraging as much active participation as possible.

8. Next steps

1. Monitor the task forces, and bring relevant outcomes to GA4GH
2. Define the MVP and ensure it is operational
3. Continue to increase node engagement

The activity in the different task forces varies, and there is a risk that due to the self-organising nature of the task forces that they may not progress. Momentum needs to be established, and once established, maintained, until the relevant task is complete. A method to allow each task force to report back to relevant stakeholders, such as the product owners, needs to be established so work the task forces do can be monitored and acted upon. Where applicable, the results should be brought to GA4GH via GDI's role as a driver project. An example here would be the evolving requirements in the AAI task force for a high level of assurance and zero trust to be supported in the GA4GH AAI and Passport standard.

The MVP needs to be defined including what the nodes can achieve, what is technically achievable, and what is legally achievable within the timeframe of GDI. This should include specifying the real data that can be included in the GDI before it finishes. Consideration should be given to how other operational genomic data sharing infrastructures manage to share their data, and if lessons can be learnt from them relevant to any minimally viable datasets that could be included within GDI, such as GoE data. GDI, by basing itself on global genomic data sharing standards, is well-placed to make data in GDI interoperable with other infrastructures, and as a driver project can influence the



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

development of these standards to support the gaps and issues that may prevent such data entering the GDI. This may mean defining a roadmap to support the 1+MG ambition that extends beyond the end of GDI, if supporting the full 1+MG ambition is infeasible within the GDI timeframe.

Many of the coordination tasks have been aimed at increasing node engagement within GDI, based on D3.2, the underspending in GDI, and the results of the various face-to-face meetings. This lack of engagement remains a risk to the project, and increased engagement is vital. Once milestone 8 has been achieved, it is hoped that the demonstrated technical functionality for a Pillar III use case may help build commitment to the GDI, but this may be limited without a clear path to having real data within the GDI. Identification of real data that GDI can host, such as GoE data, and a process and mandate to host such data is needed to ensure continued commitment of the nodes to GDI.

