Beyond One Million Genomes

D7.3

Catalogue of met and unmet use case WGs requirements - Final

Project Title (grant agreement No)	Beyond One Million Genomes (B1MG) Grant Agreement 951724		
Project Acronym	B1MG		
WP No & Title	WP7 - Support for 1+MG Use Case Working Groups		
WP Leaders	Serena Scollen (ELIXIR Hub), Marco Tartaglia (OPBG), Giovanni Tonon (HSR/ACC), Andres Metspalu (UT), Katja Kivinen / Andreas Scherer (FIMM)		
Deliverable Lead Beneficiary	01 - EMBL/ELIXIR		
Deliverable	D7.3 - Catalogue of meet and unmet use case WGs requirements Final.		
Contractual delivery date	30/04/2023	**Actual delivery date**	27/10/2023
Delayed	Yes		
Authors	Guiselle Kerry (ELIXIR Hub)		
Contributors	Juan Arenas (ELIXIR Hub), Marco Tartaglia (OPBG), Giovanni Tonon (HSR/ACC), Andres Metspalu (UT), Katja Kivinen / Andreas Scherer (FIMM)		
Acknowledgements (not grant participants)			

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

Deliverable type	Report
Dissemination level	Public

Document History

Date	Mvm	Who	Description
01/10/2023	0v1	Giselle Kerry (ELIXIR Hub)	Initial draft written and circulated to WP participants for feedback
19/10/2023	0v1	Nikki Coutts (ELIXIR Hub)	Circulated to OG, Stakeholders and GB for review
27/10/2023	0v2	Giselle Kerry (ELIXIR Hub)	Final comments closed
27/10/2023	1v0	Nikki Coutts (ELIXIR Hub)	Version uploaded to the EC Portal

Table of Contents

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

1. Executive Summary

This deliverable provides an overview of the progress made by the Working Groups (WG8, WG9, WG10, and WG11) within the B1MG initiative in the fields of genomics and infectious disease research.

WG8 - Rare Disease

WG8 successfully completed a Rare Disease Proof of Concept, utilising various GA4GH Standards and existing services. While this achievement marks a significant milestone, critical requirements must be addressed before scaling the infrastructure effectively. These requirements encompass healthcare and research pathways, real-time feedback mechanisms, data standards, ethical considerations, and interoperability. The dedication to user-centric approaches and the recognition of these needs reflect a commitment to long-term sustainability.

WG9 - Cancer

Significant milestones have been reached in integrating genomics into clinical practice for cancer treatment. An essential accomplishment is the agreement on a minimal cancer dataset, forming the foundation for standardised data collection and analysis. Ambitious goals for the future include establishing interfaces for data exchange and analysis, solving Ethical, Legal, and Social Implications (ELSI) bottlenecks, achieving interoperability, and expanding into diagnostic applications. Collaboration and innovation remain key drivers in this group.

WG10 - Common & Complex Diseases

WG10 made significant progress in assessing member states' genomics status, exploring policy, education, IT infrastructure, and ethical considerations. The effort put into harmonisation, establishing standards, and developing cross-border infrastructure highlights the complexity of this challenging field. Their vision encompasses delivering personalised medicine across borders, while addressing ethical, legal, and technical challenges, ensuring widespread participation and collaboration.

WG11 - Infectious Diseases

WG11 established a network of experts and conducted comprehensive assessments of ELSI, standards, and infrastructure needs. Notable achievements include mapping potential datasets related to COVID-19 and submitting SARS-CoV-2 data to the COVID-19 data portal. Future goals involve piloting data sharing for infectious diseases, securing funding for critical phenotypes, and addressing ethical and legal compliance.

Discussion

The progress made by these working groups is commendable, but ongoing unmet requirements emphasise the complexity of genomics and infectious disease research. It becomes of utmost importance that collaboration, innovation, and a commitment to addressing challenges will be crucial for unlocking the full potential of these use cases.

Conclusions

In conclusion, the B1MG initiative's Working Groups have demonstrated dedication, expertise, and collaboration in advancing genomics and infectious disease research. While significant milestones have been achieved, ongoing efforts are required to meet the complex challenges in these fields. Continued commitment to user-centric approaches, harmonisation, and innovative solutions will drive enduring contributions to healthcare and research for the benefit of society.

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

2. Contribution towards project objectives

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

	Key Result No and description	Contributed
Objective 1 Engage local, regional, national and European stakeholders to define the requirements for cross-border access to genomics and personalised medicine data	**1.** B1MG assembles key local, national, European and global actors in the field of Personalised Medicine within a B1MG Stakeholder Coordination Group (WP1) by M6.	Yes
	2. B1MG drives broad engagement around European access to personalised medicine data via the B1MG Stakeholder Coordination Portal (WP1) following the B1MG Communication Strategy (WP6) by M12.	Yes
	3. B1MG establishes awareness and dialogue with a broad set of societal actors via a continuously monitored and refined communications strategy (WP1, WP6) by M12, M18, M24 & M30.	Yes
	4. The open B1MG Summit (M18) engages and ensures that the views of all relevant stakeholders are captured in B1MG requirements and guidelines (WP1, WP6).	Yes
Objective 2 Translate requirements for data quality, standards, technical infrastructure, and ELSI into technical specifications and implementation guidelines that captures European best practice	**Legal & Ethical Key Results**	
	1. Establish relevant best practice in ethics of cross-border access to genome and phenotypic data (WP2) by M36	Yes
	2. Analysis of legal framework and development of common minimum standard (WP2) by M36.	No
	3. Cross-border Data Access and Use Governance Toolkit Framework (WP2) by M36.	No
	Technical Key Results	
	4. Quality metrics for sequencing (WP3) by M12.	No
	5. Best practices for Next Generation Sequencing (WP3) by M24.	No
	6. Phenotypic and clinical metadata framework (WP3) by M12, M24 & M36.	Yes
	7. Best practices in sharing and linking phenotypic and genetic data (WP3) by M12 & M24.	Ye
	8. Data analysis challenge (WP3) by M36.	No
	Infrastructure Key Results	
	9. Secure cross-border data access roadmap (WP4) by M12 & M36.	No
	10. Secure cross-border data access demonstrator (WP4) by M24.	Yes

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

Objective 3 Drive adoption and support long-term operation by organisations at local, regional, national and European level by providing guidance on phased development (via the B1MG maturity level model), and a methodology for economic evaluation	**1.** The B1MG maturity level model (WP5) by M24.	No
	2. Roadmap and guidance tools for countries for effective implementation of Personalised Medicine (WP5) by M36.	No
	3. Economic evaluation models for Personalised Medicine and case studies (WP5) by M30.	No
	4. Guidance principles for national mirror groups and cross-border Personalised Medicine governance (WP6) by M30.	No
	5. Long-term sustainability design and funding routes for cross-border Personalised Medicine delivery (WP6) by M34.	No

3. Methods

This third and final overview of the 1+MG Use Cases WGs has been produced after

Analysis of regular and ad hoc meetings and workshops including:

- 1+MG use cases working group meetings
- B1MG Operational group meetings (1+MG WGs & B1MG WP)
- 1+MG WG F2F (Brussels)
- GA4GH plenary meeting

4. Description of work accomplished

4.1.1 WG8 - Rare Disease

At the end of 2021, WG8 completed a successful Proof of Concept (P.o.C.) for Rare Disease using various GA4GH Standards (Beacon, DUO, SAM, BAM, CRAM, VCF, Crypt4gh, htsget, phenopackets, Passports, TES & WES) and existing services (Matchmaker Exchange, FEGA & ELIXIR/Life Science AAI).

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

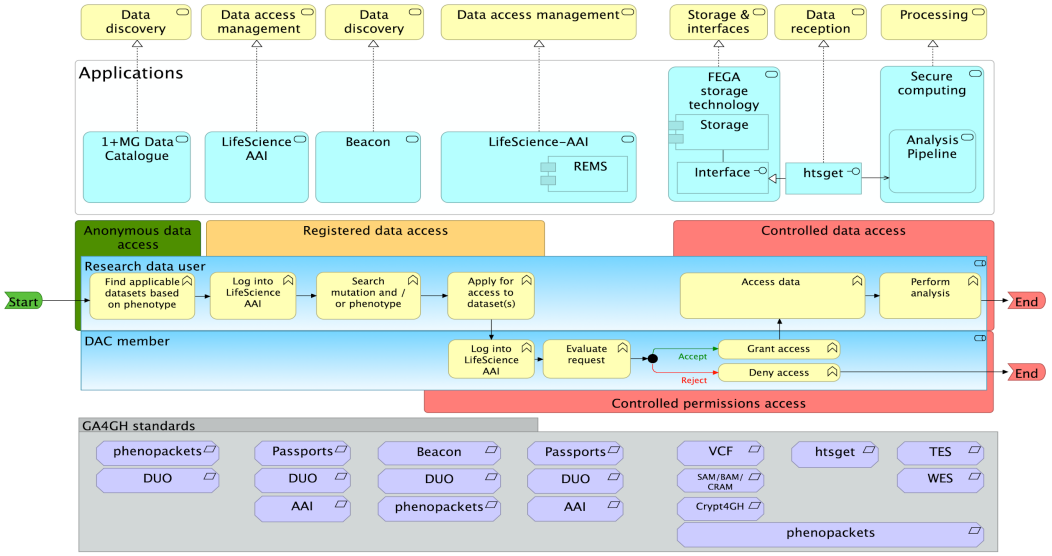

Figure 1: Diagram of the successful P.o.C for Rare Diseases

The successful Rare Disease Proof of Concept (P.o.C.) has marked a significant milestone, with several critical requirements to be further addressed, with the aim to broaden/expand the infrastructure effectively. These requirements encompass various key areas, including the identification of two intertwined pathways: one dedicated to healthcare, particularly in the realm of diagnosis, and the other focused on research, with an emphasis on the discovery of new genes associated with rare diseases. It is imperative to expedite the approval process for basic queries within these pathways to ensure timely access.

Moreover, the need for real-time feedback mechanisms for rare diseases, while considering other potential use cases, is paramount. Registered users should have the flexibility to access multiple queries or opt for case-by-case access, fostering a user-centric approach. Additionally, implementing a notification mechanism for Matchmaker Exchange matches with B1MG data is crucial for seamless data integration.

Sustaining the connection between sequencing records and their respective submitters over time is another essential requirement. This necessitates the establishment of robust Whole Exome Sequence (WES) data standards, which should be seamlessly integrated into the 1+MG IT infrastructure. Bi-directional data discovery mechanisms should also be established to foster connections with external resources, thereby enhancing data accessibility and interoperability.

Ethical considerations, particularly with respect to the longitudinal nature of rare disease data, should guide the development of IT solutions. Furthermore, incorporating diverse data types, such as Electronic Health Records (EHRs), MRI scans, and biobank samples, and recommending standardised formats is critical for comprehensive data integration. Extending the Proof of Concept to enable federated analysis and federated learning will further enhance the capabilities of the infrastructure.

It is, however, crucial to acknowledge that the successful Proof of Concept serves as a foundational starting point for further development.

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

4.1.2 WG9 - Cancer

Currently, WG9 has achieved significant milestones in the field of genomics in oncological clinical practice. They have successfully established a network of experts who share a common vision and alignment regarding the integration of genomics into clinical practice for cancer treatment. An essential accomplishment includes reaching an agreement on a minimal cancer dataset, which forms the foundation for standardised data collection and analysis. Additionally, they have made strides in developing a robust framework with six practical use cases, encompassing three clinical and three research scenarios, along with the creation of synthetic genomic datasets for melanoma and lung cancer, with ongoing work in haematological cancers.

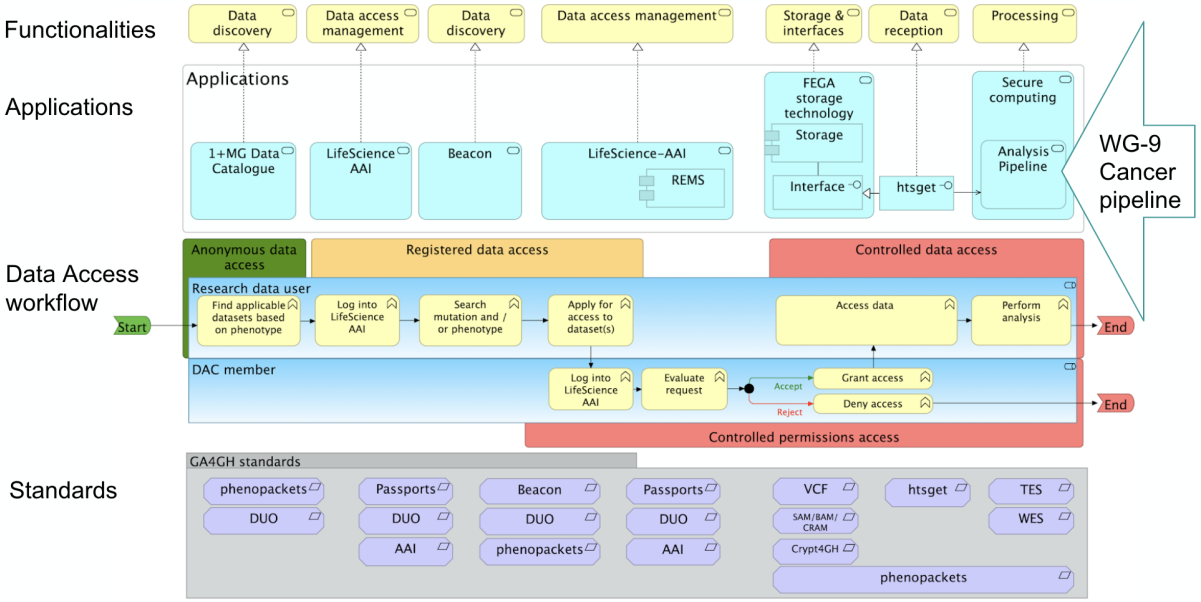

Figure 2: Diagram for the P.o.C for Cancer

WG9's Proof of Concept (P.o.C.) demonstrated the feasibility of their approach, building upon lessons learned from a similar initiative in Rare Diseases, which led to improvements in the Beacon system, and is now implemented in Beacon V2. They have also engaged in valuable discussions regarding the optimal implementation of cBioPortal, exploring options such as Docker and Singularity for efficient deployment.
Furthermore, they have developed a data capture interface that facilitates the generation of JSON-formatted clinical data, leveraging the Cohort Genomics platform.
Looking ahead to the next two years, their ambition is to establish an interface for the exchange and analysis of genomic data for cancer patients across Europe, creating a federated storage system with real-world data that ensures compliance with ELSI requirements. Solving ELSI bottlenecks through a systematic procedure is a top priority, alongside achieving interoperability of clinical data and potentially expanding into diagnostic applications.

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

One of the requirements from WG2 by WG9 is GDPR harmonisation across European countries to facilitate smooth data exchange and the establishment of a clear and efficient procedure for the secure exchange of sensitive, non-anonymized genomic data. In summary, they have made substantial progress, and their vision for the future involves seamless data sharing, ethical considerations, and expanded diagnostic capabilities in the realm of genomic oncology.

From WG3 (Standards & Guidelines), several essential requirements emerge. First, there is a pressing need to establish a comprehensive set of quality measures for genomic data. These measures will play a pivotal role in ensuring the accuracy and reliability of the data, setting the foundation for robust genomic analyses. Additionally, it is crucial to reach a consensus on shared pipelines for the analysis of cancer genomics data. This agreement will streamline data processing and interpretation, promoting consistency and reproducibility across the project.
From WG4 (Federated Cross Border Technical Infrastructure), several critical tasks are on the horizon. The prime objective is to set up a federated infrastructure capable of storing and querying data from cancer patients, accommodating the full complexity level of the defined use cases. This infrastructure should empower users to perform rapid analyses on datasets that match their queries, leveraging existing patient data, and integrating seamlessly with platforms like cBioPortal. Moreover, establishing a reliable authentication and authorization interface is paramount to facilitate secure data exchange within the federated system, ensuring that sensitive patient information is handled with the utmost care and compliance.

4.1.3 [WG10 - Common & Complex Diseases]

During the B1MG period, WG10 has made significant strides in several critical areas. Notably, they established an Expert Group, comprising members from various member states. This Expert Group conducted a comprehensive assessment of each member state's current status regarding genomics, gaining insights into ongoing pilot projects encompassing Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), SNP Microarrays, and Pharmacogenomics. They also scrutinised the activities of national biobanks and assessed funding within the genomics domain.
In a virtual workshop setting, WG10 effectively demonstrated the utility of large-scale genotyping and initiated validation studies in the participating countries. Although they have yet to conduct simulations, they have thoroughly examined their anticipated requirements and raised essential questions across six broad areas, significantly shaping their future endeavours. These areas include policy considerations, focusing on public trust, modernization of education, inclusivity, IT infrastructure and technology, and the legal and ethical framework.
Within the policy realm, they aim to ensure that the general public, researchers, clinical professionals, and policymakers are well-informed about genomics. They emphasise the importance of modernising education and facilitating access to sequencing services while addressing issues related to financing and reimbursement. Additionally, they stress the need for robust IT infrastructure, secure data access, interoperability, and adherence to ethical and legal standards in data management, privacy, and informed choice.

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

Regarding ELSI WG10 grapples with questions about informed consent for personal recommendations based on Polygenic Risk Score (PRS) scores and the dual use of PRS data for healthcare and scientific research.
In the realm of standards and guidelines, they explore the suitability of genotyping arrays, the possibility of developing a new array tailored to the European population, and the inclusion of clinically relevant rare disease markers.
Federated cross-border infrastructure considerations revolve around data storage in standardised formats, with questions about who should mandate these standards and whether a Genotype Imputation server, akin to the Michigan Imputation server, is warranted.
Finally, in the pursuit of delivering personalised medicine across borders, WG10 seeks guidance on attaining medical device status for PRS and conducting cost-benefit analyses to assess the economic impact of PRS in healthcare. These multifaceted endeavours reflect WG10's dedication to advancing genomics and addressing its complex ethical, legal, and technical challenges within the B1MG framework.

4.1.4 WG11 - Infectious Disease

At the current stage, WG11 has achieved remarkable progress within the initiative. They have successfully established a network of experts, spanning 15 active member states, through Working Group collaboration. This network has allowed them to conduct comprehensive assessments of ELSI standards, and infrastructure needs, laying the foundation for future endeavours. Furthermore, they have meticulously mapped potential datasets, especially focusing on COVID-19-related Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) data, fostering open interactions with potential partners and key players in the field, including the COVID-19 Human Genetics Initiative, BY-COVID, ISIDORe, and the COVID-19 data portal. A significant achievement includes the submission of Finland's SARS-CoV-2 WGS data to the COVID-19 data portal, enhancing the collaborative efforts in tackling the pandemic.
Looking ahead to the next two years, their vision is ambitious and impactful. They aim to initiate pilots for sharing critical COVID-19 host and pathogen data across Europe, which will serve as a pivotal step towards establishing a robust data model for infectious diseases, with COVID-19 as a crucial test case. They aspire to secure funding for sequencing (WGS/WES) critical phenotypes, encompassing extreme phenotypes like severe disease and clinical immunity, as well as addressing long Covid and adverse reactions to drugs/vaccines.
From WG2 (ELSI), they require a shared interpretation framework for ethical concerns, a code of conduct to navigate GDPR-related issues, and a mechanism for recognizing data curation and sharing efforts, ensuring ethical and legal compliance.
WG 3 (Standards & Guidelines) plays a vital role in their journey. They aim to reach an agreement on minimum data standards, harmonise and standardise data-sharing practices, build capacity in using community-developed standards, implement machine-readable Data Management Plans (DMPs), develop metadata standards in collaboration with partners, and raise awareness about the importance of sound data management practices.
Finally, from WG4 4 (Federated Cross Border Technical Infrastructure), they need to establish agreements on both minimal and aspirational data levels. They must distinguish between a usable and an optimal data model, particularly in the context of infectious diseases, where collecting and linking host and pathogen data is crucial. Sharing expertise to ensure widespread participation and collaboration is essential to their collective success. In

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

summary, they've achieved significant milestones, and their aspirations are poised to drive meaningful advancements in genomics and infectious disease research over the next two years.

5. Results

The use cases have been mapped into ZenHub, with each Use Case having it's own "Epic".

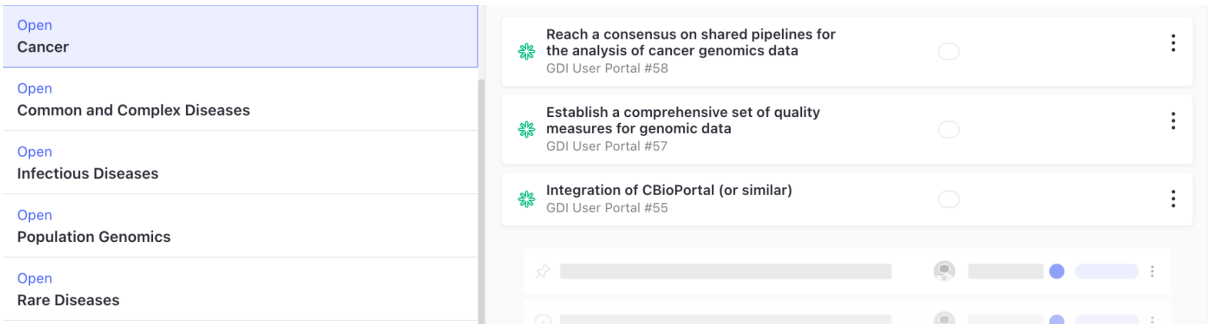

Figure 3: Screenshot of Use Cases as Epics with added high level issues

Within each Epic/Use Case "Issues" have been raised which give a very high level of what needs to be addressed (along with a brief user story) for the infrastructure to function properly for each Use Case.

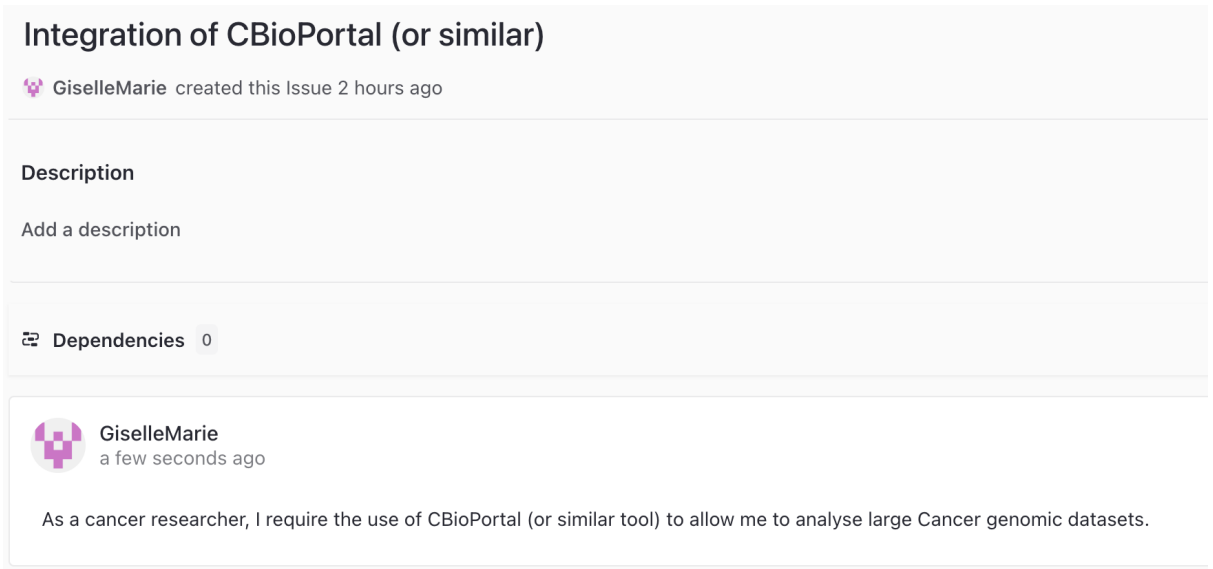

Figure 4: Screenshot of the High Level issue and User Story

These are then split and pruned as required, into either product or functionality epics, where the user story is specific to a product or functionality within the technical Work Packages.

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

6. Discussion

This document highlights the significant progress made by the Working Groups (WG8, WG9, WG10, and WG11) in the field of genomics and infectious disease research within the B1MG initiative. These accomplishments are notable, with successful Proof of Concepts (P.o.C.) and the establishment of expert networks, demonstrating a strong commitment to advancing genomics and addressing complex challenges. However, it is equally crucial to acknowledge the persistent requirements and ongoing efforts required to fully realise the potential of these initiatives.

In the case of the Rare Disease P.o.C., while it represents a milestone, there are critical requirements that must be addressed before scaling the infrastructure effectively. These requirements span various areas, including healthcare and research pathways, real-time feedback mechanisms, data standards, ethical considerations, and interoperability. The recognition of these needs reflects a commitment to the long-term success of the initiative and the importance of user-centric approaches.

Similarly, the genomics in oncological clinical practice initiative has achieved significant milestones, including the agreement on a minimal cancer dataset and the development of practical use cases. However, the journey ahead involves addressing ethical, legal, and technical challenges, achieving interoperability, and expanding into diagnostic applications. The development of a data capture interface and discussions on optimal implementations further highlight the dedication to practical solutions.

Within WG10, the assessment of member states' genomics status and the exploration of policy, education, IT infrastructure, and ethical considerations demonstrate a holistic approach to advancing genomics. The quest for harmonisation, standards, and cross-border infrastructure exemplifies the complexity of the challenges at hand. The aspiration to deliver personalised medicine across borders underscores the far-reaching impact of these endeavours.

Finally, WG11's achievements in mapping potential datasets and fostering collaborations in infectious disease research are commendable. The vision for pilots on data sharing and the pursuit of funding for critical phenotypes reflect a commitment to addressing pressing global health concerns. The requirements for shared interpretation, standards, and technical infrastructure highlight the need for a coordinated and comprehensive approach.

In summary, the progress made by these working groups is a testament to their dedication and expertise. However, the ongoing unmet requirements underscore the complexity of the genomics and infectious disease landscape. Continued collaboration, innovation, and a commitment to addressing these challenges will be essential for realising the full potential of these initiatives and making lasting contributions to healthcare and research.

7. Conclusions

In conclusion, the achievements of Working Groups (WG8, WG9, WG10, and WG11) within the B1MG initiative stand as a testament to their dedication, expertise, and collaborative spirit in advancing genomics and infectious disease research. The successful Proof of Concepts and the establishment of expert networks represent significant milestones,

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

reflecting a strong commitment to addressing complex challenges in these fields. However, it is crucial to acknowledge that these accomplishments are stepping stones on a longer journey. The persistent requirements and ongoing efforts, whether in scaling the infrastructure, addressing ethical considerations, achieving interoperability, or expanding into diagnostic applications, underscore the multifaceted nature of these endeavours. The commitment to user-centric approaches, harmonisation, and innovative solutions will be pivotal in realising the full potential of these initiatives. As we move forward, continued collaboration and a collective commitment to addressing unmet requirements will be essential in making enduring contributions to healthcare and research, ultimately benefiting society at large.

8. Next steps

The 1+MG Use Cases have provided invaluable insight into what is needed in order to create a Genomic Data Infrastructure both from a technical and ethical standpoint. The momentum gained from the Rare Disease and Cancer Use Cases will be translated across to the Genomic Data Infrastructure (GDI) Project - where Cancer, Infectious Disease and the Genome of Europe will provide the main Use Cases as part of this project.

The unmet use cases will continue to be populated into the GDI ZenHub for the technical experts to address in a systematic fashion.

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