Beyond One Million Genomes

D7.2

Catalogue of met and unmet use case WGs requirements - 2v0

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	D7.2 - Catalogue of meet and unmet use case WGs requirements 2v0.		
Contractual delivery date	30/11/2022	**Actual delivery date**	27/10/2023
Delayed	Yes		
Authors	Giselle Kerry (ELIXIR Hub)		
Contributors	Juan Arenas (ELIXIR Hub), Marco Tartaglia (OPBG), Giovanni Tonon (ACC/HSR), Andres Metspalu (UTARTU), Andreas Scherer (FIMM)		
Acknowledgements (not grant participants)			
Deliverable type	Report		

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

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

The 1+MG Use Case WGs leads and experts have continued to direct the B1MG WP activities in the last 10 months. Their contributions remain critical to determine the resulting infrastructure needs as well as analysing the final solutions and how they support the various scenarios. This deliverable builds on previous informal and formal work, as well as contributions from other project and initiative activities:

- 1+MG use cases working group meetings
- Workshop to identify health care scenarios across 1+MG WGs
- Workshop to identify research scenarios across 1+MG WGs
- B1MG Operational group meetings (1+MG WGs & B1MG WP)
- Stakeholders forum outcomes
- 1+MG Group meetings
- Beacon V2 GA4GH/ELIXIR

This deliverable contains the second formal snapshot of the 1+MG use case WGs' requirements and current implementation state, which was gathered directly as part of the B1MG project. This inventory of met and unmet needs for use case WGs will be used to drive infrastructure development and prioritise future 1+MG infrastructure actions. 1+MG WGs will evaluate successful requirement implementation to ensure that the final 1+MG infrastructure is fit for purpose.

Since the last deliverable, Beacon V2 has now been approved and now supports gene and phenotype/disease/treatment queries which has allowed for the P.o.C. being expanded to include the Cancer Use Case.

Through formal presentations to to the 1+MG CG and B1MG WP leads monthly meeting the WG leads for Rare Disease, Cancer and Infectious Disease have presented their current status, their thoughts for what they would like to achieve in the next two years, and highlighted what assistance they believe they will need from the various working groups in order to attain these. The Common and Complex Disease working group are due to present to this group in January 2023.

Finally, after consultation with WP4 and GDI Pillar II we have chosen to store and track the unmet requirements through the use of both Zenhub[1] and GitHub[2]. Broadly speaking, Zenhub will be utilised to track the high level overview of the unmet use case requirements e.g cBioPortal [3] cannot currently be used in the Cancer P.o.C. Meanwhile, developers will then take these high level issues and break them down into fine grained “issues” and track them within GitHub which will be worked on in a prioritised fashion to either “fix” the unmet requirement or seek alternative solutions. By working in this fashion we can see the multiple steps needed to resolve a high level issue and anticipate likely time to resolve/find alternative solutions.

[1] https://www.zenhub.com/
[2] https://github.com/
[3] https://www.cbioportal.org/

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.	Yes
	8. Data analysis challenge (WP3) by M36.	No
	Infrastructure Key Results	

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

	9. Secure cross-border data access roadmap (WP4) by M12 & M36.	No
	10. Secure cross-border data access demonstrator (WP4) by M24.	Yes
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 second overview of the 1+MG Use Cases WGs have been produced after:

1) 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)
- Stakeholder forum annual meeting outcomes
- 1+MG Group meetings In the case of WG9 the output of the monthly meetings with WP4/1+MG WG5 for the definition of PoC cancer has also been incorporated
- GA4GH plenary meeting

2) Individual meetings with WG leaders
   a) 1+MG WG8 Rare Diseases
   b) 1+MG WG9 Cancer
   c) 1+MG WG10 Common and complex diseases
   d) 1+MG WG11 Infectious diseases

4. Description of work accomplished

The sections below detail the work carried out with the1+MG WGs funded in B1MG

4.1.1 WG8 - Rare Disease

As detailed in the previous deliverable the P.o.C for the Rare Disease Use case has already been realised.

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

However, the following uNmet requirements still persist

- Expected timing to have access to the federated infrastructure and be able to perform individual queries - In Rare Diseases, research and healthcare are often done in conjunction. Consideration still needs to be given to a single access path for Rare Disease diagnosis and novel pathogenic gene identification
- There is existence of large collections of Whole Exome Sequence data from unique patients/diseases - Whole Exome data standards still need to be defined and consideration given on how to include them in 1+MG IT
- The ability to connect the 1+MG Infrastructure and datasets to other initiatives beyond Europe - this could be achieved through the use of bi-directional data discovery mechanisms such as Matchmaker Exchange[4] being leveraged within other resources or infrastructures
- Rare disease is usually longitudinal - this is still yet to be addressed both from an ELSI perspective and an Infrastructure perspective

4.1.2 WG9 - Cancer

Significant steps have been taken since the last deliverable to address some of the unmet use case requirements for WG9, resulting in partial realisation of the P.o.C for Cancer. Figure 1 below details the functionalities and standards that were used in the Cancer P.o.C.

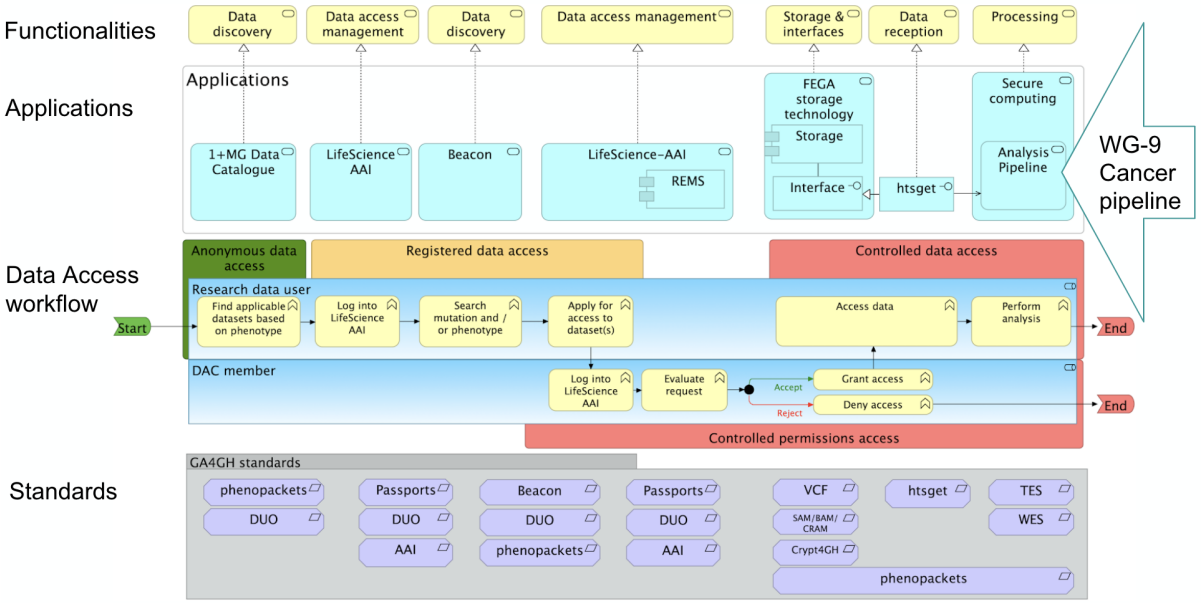

Figure 1: Functionalities and Standards used in the Cancer P.o.C.

The P.o.C for cancer has only been partially completed because, in contrast to the Rare Disease Working Group, which processes data using GPAP, the Cancer Working Group processes data

[4] https://www.matchmakerexchange.org/

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

using cBioPortal because it offers more robust data querying options. cBioPortal, however, only focuses on internal datasets. While cBioPortal is native to Docker[5], Docker is not secure in multi-user HPC environments, so this poses a problem for the 1+MG federated ecosystem. Whilst this is still an unmet requirement for WG9, WP4 are working with the EOSC4Cancer [6] project to investigate a joint solution or best practice in this space.

Despite this, it is critical to understand that once this issue is resolved, the rest of the architecture realised in the Rare Disease P.o.C will remain unchanged for the Cancer use case and it will require no further refinement.

Additional requirements that have been met for WG9 include

- Creation of a synthetic dataset for the Cancer use case
- Beacon V2 implementation in the P.o.C that enables federated discovery queries based on genotype and phenotype (this P.o.C is one of the first demonstrations of the Beacon version 2 standard in a real world use case) - this was a previously unmet requirement of the WG9.
- WG9 have now surveyed a network of experts and agreed upon a minimal Cancer dataset which provides comprehensive detail on what is needed for a dataset to be valuable
- Six use cases (3 clinical, 3 research) and two genomic synthetic datasets (melanoma and lung cancer) were created, with the melanoma one being used in the PoC

Requirements/considerations that still remain unmet are as follows:

- As mentioned above the need to implement cBioPortal as a processing tool
- Cancer has a longitudinal aspect i.e., multiple data points in time - therefore the need to be able to record proper measurements of disease progression and relapse definition is imperative
- The cancer PoC uses a Beacon version 2 implementation, but the ELIXIR Beacon network does not currently support the full Beacon V2 standard, so federated queries based on genotype and phenotype cannot be performed over the network, but only with a specific Beacon V2 deployment
- Beacon Reference Implementation used does not currently support authentication or authorisation via the LifeScience AAI.Once the ELIXIR Beacon Network supports the Beacon V2 standard, this functionality will automatically be supported.

4.1.3 WG10 - Common & Complex Disease

WG10 has not yet had the opportunity to run simulations or consult with their stakeholders. However, the following are documented as previously unmet requirements.

- WG10 differs from the cancer and rare disease use cases in that it focuses on polygenic risk scores composed of many genetic loci rather than working with single SNPs.
- To generate PRS, distributed analysis to different data hubs must be supported, for example, the ability to run common analysis tools on common input sets such as disease loci.

[5] https://www.docker.com/
[6] https://eosc4cancer.eu/

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

WG10 will be presenting their status update early in 2023.

4.1.4 WG11- Infectious Disease

During this period WG11 has been working hard to gain momentum - they have been working on meeting the following requirements in order to realise a P.o.C for infectious disease:

- Whilst still a work in progress the group have now identified a number of datasets that could potentially be used within the COVID use case
- Have sort help to create a synthetic dataset in collaboration with CINECA and GDI Pillar II/WP6
- Have begun to think about other Infectious disease use cases outside of COVID
- Discussions of suggested metadata format have gone ahead with WP3 and are progressing well

5. Conclusions

The creation and development of Beacon V2.0 has added much needed functionality, allowing other use cases to use it as a query mechanism within the infrastructure - this has allowed the Cancer P.o.C in part to be realised. However, it is clear that despite the development of Beacon V2.0 the full ability of this standard cannot be fully realised until the ELIXIR Beacon Network is fully developed allowing support of both Beacon V2.0 and which will in turn allow for authentication via LifeScience AAI.

By working in conjunction with other experts in other projects resolutions to such issues as those being experienced with CBioportal are proving fruitful and also highlight the importance of open dialogue with other initiatives so that a federated European ecosystem can be realised.

Both the Cancer use case and the Rare disease case have again highlighted the need for consideration to be given to the longitudinal aspect of data - Cancer specifically in terms of multiple data points in time and the ability to record disease progression and relapse data and Rare Disease data is generally longitudinal. ELSI consideration for this area should therefore be highlighted.

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