Beyond One Million Genomes

D3.7

Documented best practices in sharing and linking phenotypic and genetic data – 1v0

Project Title (grant agreement No)	Beyond One Million Genomes (B1MG) Grant Agreement 951724		
Project Acronym	B1MG		
WP No & Title	WP3 - Standards & Quality Guidelines		
WP Leaders	Ivo Gut (CRG), Jeroen Belien (VUmc)		
Deliverable Lead Beneficiary	19 - VUmc		
Deliverable	D3.7 - Documented best practices in sharing and linking phenotypic and genetic data - 1v0		
Contractual delivery date	31/05/2021	**Actual delivery date**	27/05/2021
Delayed	No		
Authors	Catia Pinto (1+MG WG3, PT), Jeroen Belien (VUmc, B1MG WP3 & 1+MG WG3, NL), Maarten Ligtvoet (Nictiz, B1MG WP3, NL), Milena Urbini (1+MG WG3, IT), Pim Volkert (Nictiz, B1MG WP3, NL), Wei Gu (1+MG WG3, LU), Michela Tebaldi (1+MG WG3, IT), Attila Patocs (NIO, B1MG WP3 & 1+MG WG3, HU) K. Joeri van der Velde (UMCG, FAIR genomes, NL) Marielle E. van Gijn (UMCG, FAIR genomes, NL) Jan O. Korbel (EMBL, B1MG WP1 & 1+MG WG3, DE) Antonella Padella (IRST, 1+MG WG3, IT) Alfonso Valencia (BSC, 1+MG WG3, SP) Adolfo Muñoz (ISC III, Spanish 1+MG mirror group, SP)		

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

	Miguel Pedrera, Pablo Serrano (Hospital 12 Octubre, Spanish 1+MG mirror group, SP) Carlos Parra (SAS, Spanish 1+MG mirror group, SP) Sergi Beltran (CRG, B1MG-WP3/WP4, 1+MG-WG4/WG5, SP)
Contributors	
Acknowledgements (not grant participants)	
Deliverable type	Report
Dissemination level	Public

Document history

Date	Mvm	Who	Description
05/05/2021	0v1	Jeroen Beliën (VUMC)	Initial draft circulated to WP participants for feedback
13/05/2021	0v2	Jeroen Beliën (VUMC)	WP comments addressed. Version circulated to B1MG-OG for feedback
20/05/2021	0v2	Nikki Coutts (ELIXIR Hub)	Version circulated to B1MG-OG, B1MG-GB & Stakeholders for feedback
24/05/2021	0v3	Jeroen Beliën (VUMC)	B1MG-OG, B1MG-GB & Stakeholders for feedback addressed
27/05/2021	1v0	Jeroen Belien (VUMC) & Nikki Coutts (ELIXIR Hub)	Final version uploaded to the EC Portal
23/06/2021	1v1	Xenia Perez, Juan Arenas & Nikki Coutts (ELIXIR Hub)	Template issue fixed. Reformatted final version resubmitted to EC and Zenodo

Table of contents

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

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

1. Executive Summary

This is the first version of documented best practices in sharing and linking phenotypic and genetic data. It identifies and describes best practices on sharing and linking phenotypic and genetic data in both the healthcare sector and in the research setting. The idea is to, as much as possible, avoid reinventing the wheel, learn from previous/current existing projects to improve performance and avoid mistakes made by others.

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

The documented best practices in sharing and linking phenotypic and genetic data identifies and describes best practices on sharing and linking phenotypic and genetic data in both the healthcare sector and in the research setting to, as much as possible, avoid reinventing the wheel, learn from previous/current existing projects to improve performance and avoid mistakes made by others. The current document is the first version of the best practise recommendations. This document will be updated continuously and an updated version will be published as deliverable D3.10 Q2 2022. This deliverable contributes 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.	No
	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.	No
	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	No
	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

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

	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.	No
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.	Yes
	2. Roadmap and guidance tools for countries for effective implementation of Personalised Medicine (WP5) by M36.	Yes
	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.	Yes
	5. Long-term sustainability design and funding routes for cross-border Personalised Medicine delivery (WP6) by M34.	No

3. Methods

The collection, analysis, use and sharing of genomic data promises major breakthroughs in health research, more specifically for personalized medicine and for population health. Personalized medicine research relies on more than just data generated by genome sequencing; it also entails the study of a patient's overall health, thus the need to link (or match) genomic data with relevant and accurate phenotypic data, such as environmental data, information in medical records and administrative data. As such, to ensure optimal use of genomic datasets for research and development of personalized medicine, linkage of genomic and health related data is a cornerstone for realizing the potential genomic data offers to improve health.

Across Europe there are different data sources of health related data, different taxonomy and ontology codes to label the same condition, making comparisons of different datasets challenging. Moreover, identifying and accessing the relevant datasets is challenging.

Within the 1+MG member states initiative we aim to maximize the impact of explicit and tacit knowledge on people's health characteristics, including their genomes, to deliver effective health and care, through knowledge sharing and application to healthcare services, innovation and research. Member states are expected to benefit from sharing and linking phenotypic and genetic data by exchanging experiences and hard-won solutions with one another. The B1MG WP3 together with the experts of 1+MG WG3 have worked on describing the "best practices" for sharing and linking phenotypic and genetic data so that these best practices can be used as examples to be implemented and scaled up in clinical practices as well as research programs and projects. The identified "best practices" are exemplary practices that have achieved results which could be used for larger scale cross-border initiatives. So the rationale of this document is to identify and describe best practices on sharing and linking phenotypic and genetic data in both the healthcare sector as in the research setting to, as much as possible, avoid reinventing the wheel, learn from previous/current existing projects to improve performance and avoid mistakes made by others. The current document is the first version of the best practise

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

recommendations. This document will be updated continuously and an updated version will be published Q2 2022.

4. Description of work accomplished

4.1 Introduction

4.1.1 Best practices defined

Definition of a best practice

A best practice is a technique or methodology that, through experience and research, has been proven to reliably lead to a desired result [https://whatis.techtarget.com/search/query?q=best+practice[1]].

The common aim of best practices is to be shared and adopted to benefit many more. In the context of the Declaration: 'Towards access to at least 1 million sequenced genomes in the European Union by 2022' and specifically WG3 on Common standards & minimal datasets for clinical & phenotypic data, this best practices document describes the current state of affairs on sharing and linking phenotypic and genetic data. Stated differently, the best practices description may be partial and may be related to only a subset of components being considered of the best practice, or document the lessons learned on what also does not work, so unnecessary mistakes can be avoided by others.

While best practices are well-established programmes proven to be effective through rigorous evaluations, we also included promising or innovative practices which might be still in their infancy but show signs of potential effectiveness in the long run. The goal therefore is to list the level of evidence available to guide decision-makers who are trying to learn from or want to implement (parts of) these practices.

Since we aim to capture also ongoing activities and promising or innovative practices that can evolve into a best practice this document will be regularly updated.

Criteria for identifying a best practice

A general aim of a best practice is to facilitate and improve knowledge sharing. The quality of a documented best practice should be high enough such that implementation of a best practice by others will be successful, to ensure relevant stakeholders trust the documented best practice. Identifying and describing best practices therefore involves judgement of, in case of the 1+MG/B1MG project, and in special the 1+MG WG3, criteria like relevance, effectiveness, efficiency, ethical compliance, sustainability, replicability, community participation and stakeholder collaboration [2]. Best practices also imply the re-use of existing infrastructure where possible, which can lead to better community acceptance, while saving costs by avoiding the "reinvention of the wheel". Below (Table 1) we present those criteria with a description adapted to the task of the B1MG WP3 and 1+MG WG3. The expression "best practice" also refers to promising or innovative practices.

Table 1. Best practice criteria

[1] https://whatis.techtarget.com/search/query?q=best+practice

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

Best practice criteria	Description
Relevance	The best practice must address as well as have a positive impact on sharing and linking phenotypic and genetic data.
Effectiveness	The best practice must work and achieve results that are measurable.
Efficiency	The best practice must be easy to learn, implement and use with a reasonable level of resources and time
Ethical compliance	The best practices must respect the current applicable ethical rules and legal and regulatory frameworks (see also B1MG WP2 & 1+MG WG2 outcome)
Sustainability	The best practice meets current needs, and as carried out, must be implementable/maintainable over a long period with the use of existing resources
Replicability	The best practice must have the potential for replication by others and be adaptable to similar objectives.
Community/citizen participation	The best practice must involve participation of, and describe how citizens and members of the community are involved. It must also empower the community.
Stakeholder participation	The best practice must ensure appropriate representation of, as well as satisfactory collaboration between, relevant stakeholders

Template for describing a best template

Our best practice template is based on a published best practice template [1]. This template (Table 2. Best practice template) outlines all the information stakeholders within the 1+MG initiative might need to consider to make an informed decision, if they want to replicate a best practice. The more information about a best practice is available, the better informed decision making can take place.

The template from [1] is shown in Table 2 below. To improve clarity we used the following abbreviations in the table:

BP: Best Practice
BPD: Best Practice Document
KM: Knowledge Management

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

Table 2. Best practice template

BP Component	BP attribute
Summary of BP	Title: An identifying name for the BPD
	Summary: A short description of the contents of the BPD
BP representation	Pattern Attributes: Contains problem, solution and context
	Reference (URL) or Author Contact Information: Information about the authors of the BPD, including, name, address and email. If available the ORCID[2] should be used.
	Revision Information: Information about all previous versions of the BP
	Reviews Information: Information about reviews of the BPD with URLs or other pointers
Requirement for applying BP	Goal: The intended effect of applying the BP
	Means: The means that are needed for applying the BP, including people and technology
	Skills: The skills and competence required of the end-user for applying the BP
	Cost: An estimation of the costs for implementing the BP
	Barriers: Obstacles or problems that may occur before, during, and after implementing the BP
	Barrier Management: Procedures to follow if certain obstacles or problems are encountered
BP Actor	Community of Practice: Community of practice that may be interested in using the BP
	Champion: The need and role of a champion for the BP
	Owner: The BP owner or responsible who might be an individual, role, department or organization
	Training Needs: The degree to which a person has to be trained in order to use the BP
	Acceptability: The degree of BP acceptance by domain experts - in general and/or in the organization - for resolving the problem addressed by the BP
BP properties	Usability: The degree to which the BP is easy to use
	Comprehensiveness: The degree to which the BP offers a comprehensive and complete view of the problem and solution under consideration
	Relevance: The degree to which the problem addressed by the BP is experienced as significant by practitioners
	Justification: The degree to which evidence shows that the BP solves the problem
	Prescriptiveness: The degree to which the BP offers a concrete proposal for solving the problem

[2] https://orcid.org/

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

	Coherence: The degree to which the BP constitutes a coherent unit, i.e., all parts are clearly related
	Consistency: The degree to which the BP is consistent with existing knowledge and vocabulary used in the target industry sector or knowledge domain
	Granularity: The degree to which the BPD is appropriately detailed
	Adaptability: The degree to which the BP can be easily modified and adapted to other situations
	Activity: The tasks to be carried out in the BP
	Integration: The degree to which the BP is integrated with other BPs and KM components
BP Implementation	Demonstration of Success: A case where the BP is successfully demonstrated
	Installation Time: The time it takes to introduce and implement the BP in an organization
	Application Time: The time it takes to apply the BP in an organization
	Experiences and feedback: Users' opinions, advices and experiences of the BP
	Measurement: Indicators for measuring the quality and performance of the BP

To provide an overview, all of the best practices either identified or for which the template has been completed have been grouped under the best fit category for that BP ("Category of BP" in Table 3). We do realize that a BP could have been assigned to more than only the best fit category. Categories have been chosen (as much as possible) to match categories from the "Data standards and infrastructure" part of the Maturity Level Model (MLM) as being developed by B1MG WP5 (and reviewed by B1MG WP3 and 1+MG WG3: live B1MG WP3 MLM document can be found here[3]).

Since besides best practices we also include promising and/or innovative practices we introduced a classification label "Best Practice classification" and have labeled the practices accordingly (see Table 3. List of BP topics).

Table 3. List of BP topics

Category of BP	Titles of BP	BP classification (Best, Promising or Innovative)
Data model and templates	ART-DECOR	Best
	Phenopackets schema	Promising
	OMOP	Best
	Portal of Medical Data Models	Best
	Maelstrom Data harmonization guidelines	Best

[3] https://docs.google.com/spreadsheets/d/1BfU-RSKDZTBQHOq3tnhyS2BsWrM-hVG3/edit#gid=1699496058

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

	ISO 21838	Promising
Data interoperability, data standards, ontology and controlled terminology	OLS	To be classified
	HPO	Best
	OxO	To be classified
	FAIRplus recipe on how to choose controlled vocabulary	To be classified
	BioPortal	Best
	FAIRsharing	Best
	HL7 FHIR4FAIR FHIR Implementation Guide	Innovative (balloting in September 2021)
	ISO/AWI TR 24305 Health informatics - Guidelines for implement ation of HL7/FHIR based on ISO 13940 and ISO 13606	Promising
	ISO 23903:2021 Health informatics — Interoperability and integration reference architecture – Model and framework	Best
Data infrastructure, data management platforms and tools	REDCap	Best
	ICGC ARGO Data Platform	Best
	European Platform on rare disease registration (EU RD Platform)	Best
	Federated EGA	Innovative
	(Central) EGA (see Federated EGA for description and templated table)	Best
	data.europa.eu	Best
	I2b2-tranSMART	Best
	FAIR4Health	Innovative
	RD-Connect GPAP	Best
	Genomics England Panel App	Best
Data governance, genomics data framework	GH4GH	Best
	FAIR genomes	Promising
	EUCANCan	Promising

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

	CINECA	Promising
	IHCC	To be classified
	Orphanet	Best

The next two paragraphs will present those BPs for which the best practice template has been completed and those BPs that have been already identified and will be worked on in the upcoming period. In each paragraph the best practices are grouped (a separate subsection) as shown in Table 3.

4.2 Templated best practices

This paragraph lists all Best practices which have been identified as such by B1MG WP3 and 1+MG WG3 and of which the BP template has been completed (for the most part). If new information is obtained in the upcoming period the BP templates will be updated accordingly.

4.2.1 Data model and templates

Title: ART-DECOR
Reference: https://art-decor.org/[4]

Summary:
ART-DECOR® is an open-source tool suite that supports the creation and maintenance of HL7 templates, value sets, scenarios and data sets. The tool features cloud-based federated Building Block Repositories (BBR) for Templates and Value Sets. It supports comprehensive collaboration of team members within and between governance groups. It features ontology lookup services that can be used to develop, author and publish health information standards.

Category:
- Standard development and authoring tool
- Ontology lookup service

Topics:
- Use cases and iterative approach

BP Component	BP attribute
Summary of BP	**Title: ART-DECOR**
	Summary: ART-DECOR® is an open-source tool suite that supports the creation and maintenance of data sets, value sets, scenarios and HL7 templates.
BP representation	Pattern Attributes: ART-DECOR is an open-source tool and a methodology for various multidisciplinary stakeholders of healthcare information exchange. It supports comprehensive collaboration of team members within and between governance groups and allows separation of concerns and different views on one single documentation for different domain experts. It supports creation and maintenance data sets, scenarios, HL7 templates, value sets, and more. The tool features cloud-based federated Building Block Repositories (BBR) for Templates and Value Sets. It features ontology lookup services that can be used during the authoring of health information standards.
	Reference (URL): https://art-decor.org/

[4] https://art-decor.org/

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

	Author Contact Information: Maarten Ligtvoet (Nictiz). ligtvoet@nictiz.nl
	Revision Information: This BP is an active project, all updates/revisions can be found under: https://art-decor.org/
	Reviews Information: This BP is an active project, all reviews/issues can be found under: https://art-decor.org/
Requirement for applying BP	Goal: ART-DECOR provides a structured format with metadata annotations that can be converted into various formats. This is useful in research, care and cure. As a healthcare information exchange specification it is used by vendors as a starting point for implementation into their own applications.
	Means: Tools: Basic IT service. People: domain specialist(s), datasteward(s), data specialist(s).
	Skills: Basic understanding of semantics, domain knowledge of use case at hand.
	Cost: personal cost of data curators.
	Barriers: Information not available
	Barrier Management: Information not available
BP Actor	Community of Practice: Researchers, healthcare providers and public health agencies in a wide variety of practices, for example (but not limited to): genetics, rare diseases, oncology, covid-19, medication, vaccination, IHE, lab, discharge, and others.
	Champion: Information not available
	Owner: • ART-DECOR Expert Group: The activities around the tool, its concepts and methodology, development and practice is done by the ART-DECOR Expert Group, a group of acknowledged experts in health IT. • ART-DECOR Open Tools handles the commercial aspects of the ART-DECOR tool suite development and offers/handles support plans for organizations who want ART-DECOR server support for development or production environments. ADOT provides sustainability and thus complements the ART-DECOR Expert Group that drives the development.
	Training Needs: For getting started: no training required besides basic understanding of semantics. Training is available for more advanced topics and use cases.
	Acceptability: High level of acceptance
BP properties	Usability: Easy to medium
	Comprehensiveness: High
	Relevance: High
	Justification: Well documented need to coordinate the semantics in health information exchange and research.
	Prescriptiveness: High
	Coherence: The degree to which the BP constitutes a coherent unit, i.e., all parts are clearly related
	Consistency: High (makes re-use of existing knowledge, standards and vocabulary).
	Granularity: High
	Adaptability: High (open source; and provides output in a structured format which can be adapted to secondary uses).
	Activity: ART-DECOR supports comprehensive collaboration of team members within and between governance groups and allows separation of concerns and different views on one single documentation for different domain experts.
	Integration: High (it is linked to other standards).

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

BP Implementation	Demonstration of Success: VASCA, iCRF generator[5], FAIR genomes (work in progress), IHE, eHDSI, implemented projects in multiple countries. See also https://art-decor.org/[6]
	Installation Time: Relatively short, usually a small number of workshops/training sessions.
	Application Time: Depends on the projects and datasets.
	Experiences and feedback: Information not available.
	Measurement: Information not available.

4.2.2 Data interoperability, data standards, ontology and controlled terminology

Title: HPO
Reference: https://hpo.jax.org/app/download/ontology[7]

Summary:
The Human Phenotype Ontology (HPO) provides a standardized vocabulary of phenotypic abnormalities encountered in human disease. The HPO is currently being developed using the medical literature, Orphanet, DECIPHER, and OMIM.

Category:
- ontology representation language

Topics:

BP Component	BP attribute
Summary of BP	Title: HPO: The Human Phenotype Ontology
	Summary: The Human Phenotype Ontology: provides a standardized vocabulary of phenotypic abnormalities encountered in human disease. The HPO is currently being developed using the medical literature, Orphanet, DECIPHER, and OMIM.
BP representation	Pattern Attributes: HPO project provides an ontology of medically relevant phenotypes, disease-phenotype annotations, and the algorithms that operate on these. The HPO can be used to support differential diagnostics, translational research, and a number of applications in computational biology by providing the means to compute over the clinical phenotype. The HPO is being used for computational deep phenotyping and precision medicine as well as integration of clinical data into translational research
	Reference (URL) or Author Contact Information: https://hpo.jax.org/app/download/ontology Author Contact: Sebastian Köhler sebastian.köhler@gmail.com
	Revision Information: This is an active BP. Last release: April 2021 release of HPO All issues: https://hpo.jax.org/app/news[8]
	Reviews Information: The HPO project has transitioned to a new annotation format in 2019 that is described in Köhler et al (2019) Nucleic Acids Res. Current annotation:

[5] https://www.health-ri.nl/services/icrf-generator
[6] https://art-decor.org/
[7] https://hpo.jax.org/app/download/ontology
[8] https://hpo.jax.org/app/news

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

	http://purl.obolibrary.org/obo/hp/hpoa/phenotype.hpoa[9] Previous annotation: http://purl.obolibrary.org/obo/hp/hpoa/phenotype_annotation.tab: contains manual and semi-automated annotations created by the HPO-team. These are annotations of OMIM-, Orphanet-, and DECIPHER-entries http://purl.obolibrary.org/obo/hp/hpoa/phenotype_annotation_negated.tab[10] : contains negative annotations (i.e. a disease is NOT associated with this HPO-term) This BP has been cited 1338 times in scientific publications. https://hpo.jax.org/app/help/publications[11]
Requirement for applying BP	Goal: The use of a standard vocabulary helpful for computational deep phenotyping and precision medicine as well as integration of clinical data into translational research
	Means: Tools: basic IT service People: data analyst
	Skills: biological, medical and informatics knowledge
	Cost: Information not available
	Barriers: Information not available
	Barrier Management: Information not available
BP Actor	Community of Practice: Information not available
	Champion: Information not available
	Owner: Information not available
	Training Needs: Information not available
	Acceptability: Information not available
BP properties	Usability: Information not available
	Comprehensiveness: Information not available
	Relevance: Information not available
	Justification: Information not available
	Prescriptiveness: Information not available
	Coherence: Information not available
	Consistency: Information not available
	Granularity: Information not available
	Adaptability: Information not available
	Activity: Information not available
	Integration: Information not available
BP Implementation	Demonstration of Success: Information not available
	Installation Time: Information not available
	Application Time: Information not available
	Experiences and feedback: Information not available
	Measurement: Information not available

[9] http://purl.obolibrary.org/obo/hp/hpoa/phenotype.hpoa
[10] http://purl.obolibrary.org/obo/hp/hpoa/phenotype_annotation_negated.tab
[11] https://hpo.jax.org/app/help/publications

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

Title: Concurrent use of open international EHR standards: ISO 13940, ISO 13606, and SNOMED CT terminologies.

Reference: https://www.iso.org/en/contents/data/standard/05/81/58102.html
https://www.iso.org/en/contents/data/standard/06/78/67868.html[12]
https://www.snomed.org/[13]

Summary:
ISO 13940 describes a set of concepts to support continuity of care. ISO 13606 provides a double model strategy (information model-extracts, knowledge model – archetypes) for the modelling and interchange of clinical information. SNOMED CT provides a standard vocabulary to identify clinical concepts. The concurrent use of these three standards facilitates building semantically interoperable clinical information systems.

Lozano-Rubí R, Muñoz Carrero A, Serrano Balazote P, Pastor X. OntoCR: A CEN/ISO-13606 clinical repository based on ontologies. Journal of Biomedical Informatics, 2016, 60: 224-233.

https://doi.org/10.1016/j.jbi.2016.02.007[14]

Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. J Biomed Inform. 2021;115:103697. https://doi.org/10.1016/j.jbi.2021.103697[15]

Muñoz A, Somolinos R, Pascual M, et al. Proof-of-concept design and development of an EN13606-based electronic health care record service. J Am Med Inform Assoc. 2007;14(1):118-129. https://doi.org/10.1197/jamia.M2058[16]

Sánchez-de-Madariaga, R, Muñoz, A, Lozano-Rubí, R, Serrano-Balazote, P, Castro, A L., Moreno, Pascual, M. Examining database persistence of ISO/EN 13606 standardized electronic health record extracts: relational vs. NoSQL approaches. BMC Medical Informatics and Decision Making, 2017, 17:123, 1-14 https://doi.org/10.1186/s12911-017-0515-4[17]

Category:

- Data interoperability, data standards, ontology and controlled terminology

Topics:

- Clinical data sharing
- Secondary Use of Clinical Data for Biomedical Research
- IMPacT. Precision Medicine Initiative. ISCIII[18]

[12] https://www.iso.org/en/contents/data/standard/06/78/67868.html
[13] https://www.snomed.org/
[14] https://doi.org/10.1016/j.jbi.2016.02.007
[15] https://doi.org/10.1016/j.jbi.2021.103697
[16] https://doi.org/10.1197/jamia.M2058
[17] https://doi.org/10.1186/s12911-017-0515-4
[18] https://www.ciencia.gob.es/portal/site/MICINN/menuitem.edc7f2029a2be27d7010721001432ea0/?vgnextoid=22ff08f8ee076710VgnVCM1000001d04140aRCRD&vgnextchannel=4346846085f90210VgnVCM1000001034e20aRCRD&lang_choosen=en

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

BP Component	BP attribute
Summary of BP	Title: Concurrent use of open international EHR standards: ISO 13940, ISO 13606, and SNOMED CT terminologies
	Summary: ISO 13940 describes a set of concepts to support continuity of care. ISO 13606 provides a double model strategy (information model-extracts, knowledge model – archetypes) for the modeling and interchange of clinical information. SNOMED CT provides a standard vocabulary to identify clinical concepts. The concurrent use of these three standards facilitates building semantically interoperable clinical information systems.
BP representation	Pattern Attributes: ISO13940 provides the foundations for organizational interoperability, allowing the creation of a common context between organizations. ISO 13606 provides the mechanism for the modelling of concepts and the interchange of clinical information in a secure way, integrating and normalizing information coming from different sources, allowing its automatic management and processing. It also provides tools (archetypes) for the formal modeling, management and interchange of concepts of the knowledge domain. SNOMED CT provides a standard language for clinical terms.
	Reference (URL) or Author Contact Information: ISO 13940 and ISO 13606 were developed by CEN and ISO under a Vienna Agreement. https://www.iso.org/en/contents/data/standard/05/81/58102.html[19] https://www.iso.org/en/contents/data/standard/06/78/67868.html[20] SNOMED CT is maintained by SNOMED International https://www.snomed.org/[21]
	Revision Information: As any other ISO standards, ISO 13940 and ISO 13606 are revised periodically. A new version of SNOMED CT is released every 6 months.
	Reviews Information: https://doi.org/10.1016/j.jbi.2016.02.007[22]
Requirement for applying BP	Goal: The use of ISO 13940 and ISO 13606 provide the foundations for organizational and semantic interoperability. It creates a way to interchange clinical information (and thanks to separation of information and knowledge, it could be applied to other kinds of information), protects information systems from changes in the knowledge (new concepts, evolution of concepts, integrating new organizations ...). It allows the creation of information repositories keeping all the context and meaning of the original information.
	Means: Tools: Information repositories, Knowledge (archetypes) repositories People: domain specialist(s), technical specialist(s).
	Skills: Knowledge on standards and their use in the building of data repositories
	Cost: to be determined
	Barriers: Scarce dissemination of the model among the scientific biomedical community
	Barrier Management: Dissemination of the model. Evaluation of the proof of concept in IMPaCT
BP Actor	Community of Practice: healthcare providers and public health agencies, primary and secondary use of health information, researchers, public health professionals

[19] https://www.iso.org/en/contents/data/standard/05/81/58102.html
[20] https://www.iso.org/en/contents/data/standard/06/78/67868.html
[21] https://www.snomed.org/
[22] https://doi.org/10.1016/j.jbi.2016.02.007

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

	Champion: Medical Informatics Hospital Clínic-University of Barcelona; Doce de Octubre University Hospital, Madrid, Telemedicine and Information Society Department, Health Institute"CarlosIII"
	Owner: All are open international standards
	Training Needs: Models are relatively simple. First understanding of the strategy requires some training. The separation between information and knowledge isolates both kinds of professionals from training only in their respective field of expertise, what paves the way to is adoption.
	Acceptability: very well acceptability by domain experts.
BP properties	Usability: Once implemented, the use is very natural
	Comprehensiveness: Very high
	Relevance: Very High
	Justification: Well documented need to coordinate the semantics in health information exchange and research.
	Prescriptiveness: Very High
	Coherence: Very High
	Consistency: Very High
	Granularity: Very High. Modelling of the concepts by means of archetypes ranges from very simple concepts to the most complex in a hierarchized way
	Adaptability: Very High
	Activity:
	Integration: High integration level with terminologies. There are archetypes for the integration of genomic information. https://doi.org/10.1016/j.ijmedinf.2018.10.007[23] Integration with other standards is underway.
BP Implementation	Demonstration of Success: Lozano-Rubí R, Muñoz Carrero A, Serrano Balazote P, Pastor X. OntoCR: A CEN/ISO-13606 clinical repository based on ontologies. Journal of Biomedical Informatics, 2016, 60: 224-233. https://doi.org/10.1016/j.jbi.2016.02.007[24] Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. J Biomed Inform. 2021;115:103697. https://doi.org/10.1016/j.jbi.2021.103697[25] Muñoz A, Somolinos R, Pascual M, et al. Proof-of-concept design and development of an EN13606-based electronic health care record service. J Am Med Inform Assoc. 2007;14(1):118-129. https://doi.org/10.1197/jamia.M2058[26]
	Installation Time: to be determined
	Application Time: to be determined
	Experiences and feedback: Sánchez-de-Madariaga, R, Muñoz, A, Lozano-Rubí, R, Serrano-Balazote, P, Castro, A L., Moreno, Pascual, M. Examining database persistence of ISO/EN 13606 standardized electronic health record extracts: relational vs. NoSQL approaches.

[23] https://doi.org/10.1016/j.ijmedinf.2018.10.007
[24] https://doi.org/10.1016/j.jbi.2016.02.007
[25] https://doi.org/10.1016/j.jbi.2021.103697
[26] https://doi.org/10.1197/jamia.M2058

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

	BMC Medical Informatics and Decision Making, 2017, 17:123, 1-14 https://doi.org/10.1186/s12911-017-0515-4[27]
	Measurement: Quality indicators, Success in semantic interoperability testing, Flexibility of the model

Title: FAIRplus recipe on how to choose controlled vocabulary

Reference:
https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/selecting-ontologies.html[28]

Summary: This is a part of the ongoing work of the FAIR cookbook from the IMI-FAIRplus consortium. The recipe listed here aims to provide guidance on how to select the most suitable semantic artefacts given a specific research context in general as well as main themes, i.e. risk assessment, clinical trial, drug discovery or fundamental research.

Category:
- Ontology selecting tool
- Ontology recommendations

Topics:
- Use cases and iterative approach
- Selection criteria
- A set of core terminologies
- Semantic artifacts

BP Component	BP attribute
Summary of BP	Title: FAIRplus recipe on how to choose controlled vocabulary
	Summary: This is a part of the ongoing work of the FAIR cookbook form the IMI-FAIRplus consortium. The recipe listed here aims to provide guidance on how to select the most suitable semantic artefacts given a specific research context in general as well as main themes, i.e. risk assessment, clinical trial, drug discovery or fundamental research.
BP representation	Pattern Attributes: Selecting suitable controlled vocabulary is a core step to standardize a dataset. However, the same "feature" or "variables" can be found in many different semantic resources. What is important in the selection process is to take the context (of why the data was collected) into account. The domain of operation will somehow dictate the semantic framework that makes most sense selecting. This is simply a consequence of the fact that the advances in data standardization in specific fields is such that it is a sound decision to adopt a complete stack of standards, both syntactic and semantic. The "FAIRplus recipe on how to choose controlled vocabulary" aims to provide guidance on how to determine the most suitable semantic artefacts by mapping a dataset to its research context. It includes the approach, selection criteria and a set of core terminologies such as · Organism, Organism Parts and Developmental Stages · Diseases and Phenotype · Pathology and Disease Specific Resources

[27] https://doi.org/10.1186/s12911-017-0515-4
[28] https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/selecting-ontologies.html

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

	· Cellular entities · Molecular Entities · Assays and Technologies · Relations
	Reference (URL): The FAIR cookbook: https://fairplus.github.io/the-fair-cookbook/[29] The recipe of this BP: https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/selecting-ontologies.html#set-of-core-terminologies[30] Author Contact: fairplus-cookbook@elixir-europe.org
	Revision Information: This BP is an active project, all updates/revisions can be found under: https://github.com/FAIRplus/the-fair-cookbook[31]
	Reviews Information: This BP is an active project, all reviews/issues can be found under the Issue Tracker: https://github.com/FAIRplus/the-fair-cookbook/issues[32]
Requirement for applying BP	Goal: Support the choice of controlled vocabulary based on context.
	Means: People: data manager, data analyst (curator), domain experts. Technology: basic IT.
	Skills: Basic understanding of biological and medical ontologies, domain expertise (depending on datasets).
	Cost: personal cost of data curators.
	Barriers: underestimation of resources, incomplete information of retrospective dataset.
	Barrier Management: involve data providers, domain expertise and decision makers (sponsors) in the planning and process.
BP Actor	Community of Practice: Biomedical researchers and healthcare providers.
	Champion: FAIRplus cookbook team
	Owner: FAIRplus cookbook team (in planning of setting up a long-term editorial team for sustainability within ELIXIR).
	Training Needs: minimum if equipped with basic knowledge of bio-ontologies.
	Acceptability: A good level of acceptance among the IMI community and ELIXIR community.
BP properties	Usability: Easy to Medium
	Comprehensiveness: High (comprehensive in translational medical data. Also covers some area in healthcare related data, could be extended)

[29] https://fairplus.github.io/the-fair-cookbook/
[30] https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability/selecting-ontologies.html#set-of-core-terminologies
[31] https://github.com/FAIRplus/the-fair-cookbook
[32] https://github.com/FAIRplus/the-fair-cookbook/issues

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

	Relevance: High (It is directly relevance to the interoperability of linked phenotypic data)
	Justification: To be evaluated.
	Prescriptiveness: High (direct recommendation)
	Coherence: High
	Consistency: High (The BP is built based on existing well-accepted recommendations)
	Granularity: Medium
	Adaptability: High (it can be easily adapted to other situations)
	Activity: Information not available
	Integration: High (it is linked to other standards)
BP Implementation	Demonstration of Success: IMI projects under FAIRification process in collaborations with FAIRplus
	Installation Time: Relatively short (e.g.: a half-day workshop with some following meetings)
	Application Time: Depends on the projects and datasets
	Experiences and feedback: To be collected
	Measurement: To be evaluated

4.2.3 Data infrastructure, data management platforms and tools

ICGC ARGO Data Platform

Reference: ICGC ARGO | Homepage (icgc-argo.org)[33] as well as ICGC ARGO Docs | ICGC ARGO Docs (icgc-argo.org)[34]

Summary: The International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) aims to uniformly analyze specimens from 100,000 donors with high quality clinical data in order to address outstanding questions that are vital to the quest to defeat cancer.

Category:

- Use case cancer
- International consortium

Topics:

- Data dictionary (https://docs.icgc-argo.org/dictionary[35])
  - Sample registration
  - Donor
  - Specimen
  - Primary diagnosis

[33] https://platform.icgc-argo.org/
[34] https://docs.icgc-argo.org/
[35] https://docs.icgc-argo.org/dictionary

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

- Treatment
- Chemotherapy
- Hormone therapy
- Radiation
- Follow up

BP Component	BP attribute
Summary of BP	Title: ICGC ARGO Data Platform
	Summary: The International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) aims to uniformly analyze specimens from 100,000 donors with high quality clinical data in order to address outstanding questions that are vital to the quest to defeat cancer.
BP representation	Pattern Attributes: The collection of high-quality clinical information according to standardised vocabularies is very important to accelerate research into the causes and control of cancer. ARGO is the International Cancer Genome Consortium which can be an example for the classification and annotation of high quality clinical data. The ICGC ARGO Data Dictionary expresses the details of the data model, which adheres to specific formats and restrictions to ensure a standard of data quality. Each clinical field has a data tier and an attribute classification, which reflects the importance of the field in terms of clinical data completion. Thus, a minimum set of clinical data that must be submitted is indicated.
	Reference (URL) or Author Contact Information: The ICGC is a confederation of members: https://www.icgc-argo.org/page/117/icgc-argo-committees[36] Contact: https://www.icgc-argo.org/page/69/contact-us#[37]
	Revision Information: This is an active BP. Dictionary release:December 11, 2020 https://docs.icgc-argo.org/docs/release-notes/dictionary-releases[38] Data release: October 23, 2020 https://docs.icgc-argo.org/docs/release-notes/data-releases[39] Software release: Data Platform v1.55.0 - API v3.2.0 Release Date: June 19, 2020
	Reviews Information: Previous dictionary releases https://docs.icgc-argo.org/docs/release-notes/dictionary-releases[40]
Requirement for applying BP	Goal: Collection of high-quality clinical information
	Means: People: biologist, physician, medical oncologist, data manager, data analyst
	Skills: biological, medical and informatics knowledge
	Cost: personal cost of data curators
	Barriers: missing data on retrospective cohorts; data harmonization

[36] https://www.icgc-argo.org/page/117/icgc-argo-committees
[37] https://www.icgc-argo.org/page/69/contact-us#
[38] https://docs.icgc-argo.org/docs/release-notes/dictionary-releases
[39] https://docs.icgc-argo.org/docs/release-notes/data-releases
[40] https://docs.icgc-argo.org/docs/release-notes/dictionary-releases

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

	Barrier Management: Procedures to follow if certain obstacles or problems are encountered
BP Actor	Community of Practice: biologist, physician, medical oncologist
	Champion: ICGC Executive Board
	Owner: ICGC
	Training Needs: The degree to which a person has to be trained in order to use the BP
	Acceptability: The degree of BP acceptance by domain experts - in general and/or in the organization - for resolving the problem addressed by the BP
BP properties	Usability: medium
	Comprehensiveness: High: it includes several clinical data records, most of which are mandatory in order to submit data
	Relevance: high (The ICGC ARGO Data Dictionary expresses the details of the data model, which adheres to specific formats and restrictions to ensure a standard of data quality.)
	Justification: To be evaluated
	Prescriptiveness: High
	Coherence: High
	Consistency: highly consistent
	Granularity: medium
	Adaptability: medium
	Activity: The tasks to be carried out in the BP
	Integration: The dictionary controlled terminology values were derived from external standards or common terminology used by ICGC ARGO programs. These include: American Joint Committee on Cancer Staging Classifications World Health Organization International Classification of Diseases, 10th Revision (ICD-10) International Classification of Diseases for Oncology (ICD-O)) Cancer Data Standards Registry and Repository (caDSR) Cancer Care Ontario Data Book Reporting Standards RxNorm Common Terminology Criteria for Adverse Events (CTCAE) ECOG-ACRIN Cancer Research Group
BP Implementation	Demonstration of Success: 25k Initiative and the PCAWG
	Installation Time: no installation needed. Upload of TSV template
	Application Time: Depending on the projects and datasets
	Experiences and feedback: not available/ to be collected
	Measurement: Information not available

European Platform on rare disease registration (EU RD Platform)

Reference: https://eu-rd-platform.jrc.ec.europa.eu/_en[41]

Summary: The EU RD Platform copes with the fragmentation of rare disease patients data contained in hundreds of registries across Europe.The information about these patients is spread between hundreds of registries across Europe, at national, regional and local levels.The

[41] https://eu-rd-platform.jrc.ec.europa.eu/_en

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

main objective of the European Platform on Rare Disease Registration (EU RD Platform) is to cope with the enormous fragmentation of rare disease (RD) patients data contained in hundreds of registries across Europe. The Platform makes RD registries' data searchable and findable, thus increasing visibility for each registry, maximising the value of each registry's information and enabling extended use and re-use of registries' data. This is ensured by the European RD Registry Infrastructure (ERDRI), which supports existing registries and the creation of new registries.The EU RD Platform sets EU-level standards for RD data collection and data exchange and provides training on the use of the tools and services offered.In addition to ERDRI, the EU RD Platform includes a data repository composed of the European RD Registry Data Warehouse (under preparation), the JRC-EUROCAT Central Registry and the JRC-SCPE Central Registry.The EU RD Platform is open to all RD registries. Its final goal is to act as a knowledge generation centre benefiting healthcare providers including European Reference Networks, researchers, patients and policy-makers in the common effort to improve diagnosis and treatment for patients living with a rare disease.

Category:

- Use case Rare Diseases

Topics:

- The European Rare Disease Registry Infrastructure (ERDRI) renders rare disease registries' data searchable and findable. This is achieved through the provision of following components: European Directory of Registries (ERDRI.dor), Central Metadata Repository (ERDRI.mdr) and Pseudonymisation Tool (EUPID) https://eu-rd-platform.jrc.ec.europa.eu/erdri-description_en[42]
- Set of common data elements for rare diseases registration: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en[43] and https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/CDS/EU_RD_Platform_CDS_Final.pdf[44]
- https://rd-connect.eu/what-we-do/omics/gpap/[45] The RD-Connect Genome-Phenome Analysis Platform (GPAP) is an online tool for diagnosis and gene discovery in rare disease research

BP Component	BP attribute
Summary of BP	Title: European Platform on rare disease registration (EU RD Platform)
	Summary: The EU RD Platform copes with the fragmentation of rare disease patients data contained in hundreds of registries across Europe.The information about these patients is spread between hundreds of registries across Europe, at national, regional and local levels.The main objective of the European Platform on Rare Disease Registration (EU RD Platform) is to cope with the enormous fragmentation of rare disease (RD) patients data contained in hundreds of registries across Europe. The Platform makes RD registries' data searchable and findable, thus increasing visibility for each registry, maximising the value of each registry's information and enabling extended use and re-use of registries' data. This is ensured by the European RD Registry Infrastructure (ERDRI), which supports existing registries and the creation of new registries.The EU RD Platform sets EU-level standards for RD data collection and data exchange and provides training on the use of the tools and services offered.In addition to ERDRI, the EU RD Platform includes a data repository composed of the European RD Registry Data Warehouse (under preparation), the JRC-EUROCAT Central Registry and the JRC-SCPE Central Registry.The EU RD Platform is open to all

[42] https://eu-rd-platform.jrc.ec.europa.eu/erdri-description_en
[43] https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en
[44] https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/CDS/EU_RD_Platform_CDS_Final.pdf
[45] https://rd-connect.eu/what-we-do/omics/gpap/

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

	RD registries. Its final goal is to act as a knowledge generation centre benefiting healthcare providers including European Reference Networks, researchers, patients and policy-makers in the common effort to improve diagnosis and treatment for patients living with a rare disease
BP representation	Pattern Attributes: In the EU about 30 million citizens in Europe are affected by more than 6000 different rare diseases. The information about these patients is spread between hundreds of registries across Europe, at national, regional and local levels. The EU Rare Disease Platform aims to provide researchers, healthcare providers, patients and policy-makers with a consistent instrument to improve knowledge, diagnosis and treatment of rare diseases. The "Set of common data elements for Rare Diseases Registration" is the first practical instrument released by the EU RD Platform aiming at increasing interoperability of RD registries.It contains 16 data elements to be registered by each rare disease registry across Europe, which are considered to be essential for further research. They refer to the patient's personal data, diagnosis, disease history and care pathway, information for research purposes and about disability. The EU Rare Disease Platform has developed several resources, described below: The European Rare Disease Registry Infrastructure (ERDRI) renders rare disease registries' data searchable and findable. This is achieved through the provision of following components: European Directory of Registries (ERDRI.dor), Central Metadata Repository (ERDRI.mdr) and Pseudonymisation Tool (EUPID).ERDRI supports existing registries in view of their interoperability and the creation of new registries. European Directory of Registries (ERDRI.dor): ERDRI.dor provides an overview of the participating registries with their main characteristics and description.Data input is performed by registry owners. ERDRI.dor consists of eight sections with 38 data fields related to a registry of which 23 are obligatory. Central Metadata Repository (ERDRI.mdr): ERDRI.mdr ensures semantic interoperability between RD registries. It stores all data elements (metadata) used by the participating registries, including the names of the data elements (designations) and their definitions.Within ERDRI.mdr metadata items from any registry can be either uploaded automatically or inserted manually.In case of establishing a new registry or amending an existing registry, a user can select from the metadata contained in ERDRI.mdr. Pseudonymisation Tool (EUPID): The Pseudonymisation tool is provided to all participating registries through the EUropean Patient IDentity (EUPID) Management Services.EUPID is designed to provide distinct pseudonyms for patients in different contexts, prevent duplicate registration of patients, keep a protected link between the different pseudonyms, preserve the possibility for re-identification by a trusted third party. Search broker (ERDRI.sebro):ERDRI.sebro allows any user to retrieve metadata of interest and its hosting registry via ERDRI.sebro's connection to ERDRI.mdr and ERDRI.dor.(In preparation)
	Reference (URL) or Author Contact Information: European Platform on Rare Disease Registration | EU RD Platform (europa.eu) https://eu-rd-platform.jrc.ec.europa.eu/_en[46]
	Revision Information: Information not available
	Reviews Information: Information not available
Requirement for applying BP	Goal: Share standardized data on rare diseases
	Means: Information not available
	Skills: Information not available

[46] https://eu-rd-platform.jrc.ec.europa.eu/_en

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

	Cost: Information not available
	Barriers: Information not available
	Barrier Management: Information not available
BP Actor	Community of Practice: Rare Diseases community: researchers, healthcare providers and public health agencies
	Champion: Information not available
	Owner: European Commission has developed the EU platform on rare diseases
	Training Needs: Information not available
	Acceptability: High level of acceptance among the rare diseases community, often referred as an example
BP properties	Usability: High
	Comprehensiveness: High
	Relevance: High
	Justification: Well documented need to integrate EU registries on rare diseases
	Prescriptiveness: High
	Coherence: Information not available
	Consistency: High
	Granularity: Information not available
	Adaptability: High
	Activity: Information not available
	Integration: High
BP Implementation	Demonstration of Success: Rare diseases data sharing among the EU
	Installation Time: Information not available
	Application Time: Information not available
	Experiences and feedback: Information not available
	Measurement: Information not available

Federated EGA

The Federated EGA will be a resource for discovery and access of sensitive human omics and associated data consented for secondary use, through a network of national human data repositories in Europe, with the aim to accelerate disease research and improve human health. Over the last 10 years, most individual-level human omics data have been generated in the context of research consortia and shared via global repositories such as the European Genome-phenome Archive (EGA). Many countries now have emerging personalized medicine programmes which are generating data from national or regional initiatives. Thus, human genomics is undergoing a step change from being a research-driven activity to one funded through healthcare initiatives. Genetic data generated in a healthcare context is subject to more stringent information governance than research data and often must comply with national legislation. To address this need, the Federated EGA provides a network of connected resources to enable transnational discovery of and access to human data for research. Through its federated model, it is also able to respect jurisdictional data protection regulations. By providing a solution to emerging challenges around secure and efficient management of human omics and

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

associated data, the Federated EGA fosters data reuse, enables reproducibility, and accelerates biomedical research.

The EGA project is currently a collaboration between EMBL-EBI and the CRG, regulated by agreements between the two institutions. The Federated European Genome-phenome Archive (EGA) will be a distributed network of repositories for sharing human -omics data and phenotypes. The GHGA (German Human genome-Phenome Archive) will be the node of the federated EGA in Germany, for example. Typically a node is an organization or project that hosts human genetic data so that sensitive data can remain within a jurisdiction where this is a requirement or otherwise shared across jurisdiction. The federated EGA gathers metadata of -omics data collections stored in national or regional archives and makes them discoverable across the whole EGA network. The EGA is contributing the Federated EGA model, requirements and experiences to several communities and projects like GA4GH, ELIXIR Federated Human Data Implementation Study or ELIXIR Federated Human Data community.

BP concept	BP attribute
Summary of BP	Title: Federated EGA
	Summary: The Federated EGA will be a resource for discovery and access of sensitive human genomics/omics and associated data consented for secondary use.
BP representation	Pattern Attributes: The federated EGA will be a network of national human data repositories in Europe, with the aim to accelerate disease research and improve human health.
	Reference (URL): https://ega-archive.org/federated[47]
	Revision Information: Information on APIs is at https://ega-archive.org/federated[48]
Requirement for applying BP	Goal: FAIR sharing, including access, discoverability across partners
	Means: Adherence to EGA API's (see https://ega-archive.org/federated)
	Skills: Low requirements (the EGA is well-established internationally and has already adapted to the needs of a wide user base)
	Cost: An estimation of the costs for implementing the BP Initial investments costs (staff, IT-resources) for data curation, and costs for data stewardship when consortia make use of the resource as their data repository
	Barriers: Unwillingness to share data consented for research in a timely manner
	Barrier Management: Create incentives for sharing data in a timely manner a requirement
BP Actor	Community of Practice: Researcher to obtain a cohort of individuals/patients to study, or healthcare professionals addressing certain genotype/phenotype/treatment related questions
	Champion: The need and role of a champion for the BP
	Owner: The data submitter acts as controller. Patients can after consent ask for data to be removed.
	Training Needs: Researchers and healthcare professionals need to be trained in their own domain to use the ontologies to describe the data.
	Acceptability: Patients, healthcare professionals and researchers need to realize the potential of data sharing and computer readability of data for their own benefits, which is currently only partly accomplished.

[47] https://ega-archive.org/federated
[48] https://ega-archive.org/federated

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

BP properties	Usability: It is a data model with an underlying IT infrastructure, and needs to be translated to the specific systems used.
	Comprehensiveness: The federated EGA is currently devised with a range of European partners.
	Relevance: Problem addressed by the BP is experienced as significant by practitioners and researchers
	Justification: The degree to which evidence shows that the BP solves the problem. The EGA is one of the most widely used resources for access to sensitive genomic and associated data types globally.
	Prescriptiveness: BP offers a concrete proposal for solving the problem
	Coherence: The BP constitutes a highly coherent unit (i.e., all parts, in this case nodes, are clearly related)
	Consistency: The BP is highly consistent with existing knowledge, with a large portion of the genomics (and associated) data types consented for research in Europe going to the EGA
	Granularity: The BPD is appropriately detailed
	Adaptability: The BP is currently being devised with a range of European partners and at this stage can still be readily adapted to new situations
	Activity: The tasks to be carried out in the BP
	Integration: The degree to which the BP is integrated with other BPs and KM components
BP Implementation	Demonstration of Success: The EGA, currently run at EMBL-EBI and CRG Barcelona, is archiving data from throughout Europe and beyond
	Installation Time: Federated EGA software are made available open source and can be installed within reasonable time requirements
	Application Time: Users will be available to access the resource from all over Europe
	Experiences and feedback: The model is currently being developed with a range of partners in Europe
	Measurement: Development of key performance indicators

4.2.4 Data governance, genomics data framework

FAIR genomes

Reference: https://github.com/fairgenomes[49]

Summary: FAIR genomes: A national guideline to promote optimal (re)use of NGS data in research and healthcare

Category:

- Guideline on NGS
- Dutch consortium, FAIR Genomes is a ZonMw "Personalized Medicine" project, nr. 846003201
- Use cases rare diseases and cancer

Topics:

- Demonstrator : https://fairgenomes-acc.gcc.rug.nl[50]

[49] https://github.com/fairgenomes
[50] https://fairgenomes-acc.gcc.rug.nl

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

- Currently 9 modules with 109 elements:
  - Personal (12),
  - Clinical (20),
  - Material (16),
  - Sample Preparation (9),
  - Sequencing (12),
  - Analysis (11),
  - Leaflet and consent Form (8),
  - Individual Consent (12),
  - Study (9)
- Reusing existing thesauri/ontologies wherever possible

BP Component	BP attribute
Summary of BP	Title: FAIR genomes
	Summary: FAIR genomes: A national (Dutch) guideline to promote optimal (re)use of NGS data in research and healthcare.
BP representation	Pattern Attributes: The FAIR genomes project is a national (Dutch) coordination action to unite currently fragmented guidelines & tools to increase 'FAIR'-ness of DNA data - Findability, Accessibility, Interoperability and Reusability - uniting work from all types of DNA laboratories (rare disease, cancer, research, etc), patients/participants organisations, and has extensive collaborations with (inter)national initiatives, including aligned with NL and international organisations BBMRI, ELIXIR, X-omics, Solve-RD, EJP-RD, GA4GH.
	Reference (URL) or Author Contact Information: URL: https://fairgenomes.github.io/about/[51] Authors ORCIDs: https://orcid.org/0000-0002-7160-5942[52] https://orcid.org/0000-0002-0934-8375[53] https://orcid.org/0000-0003-1615-4197[54] https://orcid.org/0000-0002-1215-167X[55] https://orcid.org/0000-0002-4706-1084[56] https://orcid.org/0000-0002-2440-3993[57] https://orcid.org/0000-0003-1301-5204[58] https://orcid.org/0000-0003-4450-3112[59] https://orcid.org/0000-0002-1073-0539[60] https://orcid.org/0000-0002-0979-3401[61]
	Revision Information: This is an active BP. First release: v0.2[62]. Current release: v0.3[63].

[51] https://fairgenomes.github.io/about/
[52] https://orcid.org/0000-0002-7160-5942
[53] https://orcid.org/0000-0002-0934-8375
[54] https://orcid.org/0000-0003-1615-4197
[55] https://orcid.org/0000-0002-1215-167X
[56] https://orcid.org/0000-0002-4706-1084
[57] https://orcid.org/0000-0002-2440-3993
[58] https://orcid.org/0000-0003-1301-5204
[59] https://orcid.org/0000-0003-4450-3112
[60] https://orcid.org/0000-0002-1073-0539
[61] https://orcid.org/0000-0002-0979-3401
[62] https://github.com/fairgenomes/fairgenomes-semantic-model/tree/v0.2
[63] https://github.com/fairgenomes/fairgenomes-semantic-model/tree/v0.3

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

	All issues: https://github.com/fairgenomes/fairgenomes-semantic-model/issues[64]
	Reviews Information: Several rounds of revisions have taken place using a shared google sheet[65]. This sheet was then transformed to a github repository where via the issues[66] option of github further review took place. Finally it has been converted in one single semantic model for which issue tracking is available, see: https://github.com/fairgenomes/fairgenomes-semantic-model/issues[67]
Requirement for applying BP	Goal: A guideline to promote optimal (re)use of NGS data in research and health - Promote large scale (re)use of all human genomic data in the Netherlands to maximize knowledge extraction for research and healthcare
	Means: Data: the current FAIR genomes semantic model contains 9 modules each of which might have its own source of data. In order to apply or comply to this model, the source data needs to be transformed to the proposed model if it is not semantically annotated yet (either using the provided preferred concept or a mapping towards the preferred concept stating whether that concept match is exact, close, broad, etc as defined at this SKOS site[68]). Tools: depending on your local or national IT facilities you need a form of data warehouse where your dataset, if not in proper semantic format can be processed to the desired model and upon approved request can be delivered (i.e. provided to the requester of data). Your dataset (i.e. (rich) metadata) can also as a first step be listed in a catalogue (linked to if possible a FAIR data point (FDP)) to make your data findable and accessible as a start. The FAIR genomes guideline also contains pointers to tooling that are part of the basic workflow when requesting a NGS test and can be implemented as, or replace, part of your existing workflow. People: domain specialist(s), datasteward(s), data specialist(s) (if data needs to be retrieved from a source system and transferred towards this BP, you need staff to support you in Extracting, Transferring and Loading, so called ETL, it into your target system), basic IT-staff
	Skills: The skills and competence required of the end-user for applying the BP Vocabulary/Ontology expertise for relevant domain (e.g. cancer, rare disease, infectious disease) , basic understanding of semantics, (clinical, genetics, (bio)informatics,..) domain knowledge of use case at hand
	Cost: An estimation of the costs for implementing the BP Initial investments costs (staff, IT-resources) for data curation (i.e. ETL-work as described above) if your data is not fully compliant towards this FAIR genomes model at the source. Maintenance costs: hosting your data set, revisions of data model that require adaptations of your source data or ETL-process
	Barriers: Obstacles or problems that may occur before, during, and after implementing the BP -Getting the non-data professionals engaged in using the recommended ontologies. -Data is fragmented across many hospital departments and institutes and hard to access or change/ harmonize the current practice The MOLGENIS platform has a FAIR Genomes implementation. This platform contains solutions to help harmonize datasets, such as the Mapping Service, to alleviate these barriers.
	Barrier Management: Procedures to follow if certain obstacles or problems are encountered

[64] https://github.com/fairgenomes/fairgenomes-semantic-model/issues
[65] https://docs.google.com/spreadsheets/d/1rnLsmE62t15jCwJfx4mCL5USYSeiXNctCA0XPcgprds/
[66] https://github.com/fairgenomes/information/issues
[67] https://github.com/fairgenomes/fairgenomes-semantic-model/issues
[68] https://www.w3.org/TR/2009/REC-skos-reference-20090818/#mapping

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

	Currently a project is being executed that makes an inventory which barriers the different professionals providing the data (laboratory specialists, clinicians) foresee or encounter when applying the different ontologies recommended by the FAIR genomes project in daily practice to be able to manage the barriers.
BP Actor	Community of Practice: researcher to obtain a cohort of individuals/patients to study and healthcare professionals addressing certain genotype/phenotype/treatment related questions
	Champion: The need and role of a champion for the BP
	Owner: The patient is owner of their data, the hospital has the obligation the data is in such a condition it can be shared Currently the FAIR genomes BP is governed by the FAIR genomes project team and discussions have started to seek for transfer of governance/ownership towards a sustainable/legal body. The licensing of the tangible results, like the codebooks, will be under a CC-BY 4.0 license (Creative Commons Attribution 4.0 International Public License[69])
	Training Needs: Each healthcare professional has to be trained in their own domain to use the ontologies to describe the data.
	Acceptability: The healthcare professionals and researchers need to realize the potential of data sharing and computer readability of data for their own benefits, that is currently only partly accomplished.
BP properties	Usability: It is a data model and needs to be translated to the specific systems used.
	Comprehensiveness: The schema was developed by 14 Dutch institutes dealing with NGS data in research and clinical settings, and should by now cover all essentials.
	Relevance: The degree to which the problem addressed by the BP is experienced as significant by practitioners.
	Justification: We are working towards a national NGS portal to demonstrate that this schema can be used to make these data FAIR.
	Prescriptiveness: The metadata scheme is work in progress and offers a solution for standardizing the exchange of NGS analysis metadata for research and diagnostics.
	Coherence: The FAIR genomes semantic schema consists of multiple layers which together form a simple tree structure. In addition, modules within the schema are (optionally) linked to represent the logical flow of an NGS diagnostic/research analysis.
	Consistency: The FAIR genomes semantic schema reuses existing and often-used ontological definitions and lookup lists (e.g. phenotypes, drugs, tissue types..) wherever possible in order to achieve maximum compatibility with existing systems. Only when definitions are missing are they added as novel ontological terms.
	Granularity: The semantic schema is composed of 4 layers: 1) meta-data about the schema itself, 2) definition of 'modules' which are reusable components concerning a specific topic like 'Material' or 'Clinical', 3) elements within the modules such as 'Age of onset' or 'Medication' for 'Clinical' and 4) lookup lists acting as standardized code systems for elements, for instance ATC-codes for 'Medication'.
	Adaptability: The metamodel of the FAIR Genomes project is currently being adopted for use in EDCs such as Castor/REDCap/OpenClinica etc.
	Activity: The tasks to be carried out in the BP

[69] https://creativecommons.org/licenses/by/4.0/

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

	Integration: The BP makes use of existing ontologies such as HPO etc. ANd is aligned with other European initiatives such as EJP-RD, Phenopackets and Solve-RD.
BP Implementation	Demonstration of Success: The FAIR genomes semantics have been (partially) adopted by the TreCODE system used in the Prinses Maxima Center for Child Oncology, Nictiz (Dutch national health standards) ART-DECOR codebook draft, UMC Groningen 'COSAS' sample database, Solve-RD RD3 sample database,
	Installation Time: The time it takes to introduce and implement the BP in an organization - All resources are freely available and downloadable without restrictions. Using the application ontology or documentation takes no installation time. Using the EDC form templates requires setting up supporting software (e.g. iCRF Generator, MOLGENIS Commander) and may take longer (i.e. hours).
	Application Time:- Application time could be relatively quick (i.e. hours-days), in case of adopting a FAIR genomes generated EDC template or merging the application ontology into an active triple storage system. Redesigning or mapping existing databases and business processes for FAIR genomes compliance may take more time, depending on the number of differences and adoption/FAIRification goals.
	Experiences and feedback: The model has been developed over the course of 2 years as a consensus of 66 people representing 14 Dutch institutes.
	Measurement: A number of quality procedures are in place such as track changes (Git commit log), versioning (Git releases) and release SOP.

4.3 Identified best practices

Within B1MG WP3 and 1+MG WG3 the following best practices were already identified which will be worked on in the upcoming period. The identified Best Practices are listed with at least the following information:

Title:

Reference:

Summary:

Category:

Topics:

4.3.1 Data model and templates

Phenopackets schema

Reference: https://phenopackets-schema.readthedocs.io/en/latest/index.html[70]

Summary:

The Phenopacket Schema represents an open standard for sharing disease and phenotype information to improve our ability to understand, diagnose, and treat both rare and common diseases. A Phenopacket links detailed phenotype descriptions with disease, patient, and genetic information, enabling clinicians, biologists, and disease and drug researchers to build more

[70] https://phenopackets-schema.readthedocs.io/en/latest/index.html

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

complete models of disease. The standard is designed to encourage wide adoption and synergy between the people, organizations and systems that comprise the joint effort to address human disease and biological understanding.

Category:

- Phenotype data model

Topics:

OMOP

Reference: https://www.ohdsi.org/data-standardization/the-common-data-model/[71]

Summary:

The OMOP Common Data Model allows for the systematic analysis of disparate observational databases. The concept behind this approach is to transform data contained within those databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes), and then perform systematic analyses using a library of standard analytic routines that have been written based on the common format.

Category:

- Data model

Topics:

Portal of Medical Data Models

Reference: https://medical-data-models.org/[72]

Summary:

MDM-Portal (Medical Data Models) is a meta-data registry for creating, analyzing, sharing and reusing medical forms. It serves as an infrastructure for academic (non-commercial) medical research to contribute a solution to this problem. It contains forms in the system-independent CDISC Operational Data Model (ODM) format with more than 500,000 data-elements. The Portal provides numerous core data sets, common data elements or data standards, code lists and value sets. This enables researchers to view, discuss, download and export forms in most common technical formats such as PDF, CSV, Excel, SQL, SPSS, R, etc. A growing user community will lead to a growing database of medical forms. In this matter, we would like to encourage all medical researchers to register and add forms and discuss existing forms.

Category:

- Library of system independent medical forms (case record forms)
- CDISC Operational Data Model

Topics:

[71] https://www.ohdsi.org/data-standardization/the-common-data-model/
[72] https://medical-data-models.org/

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

Maelstrom Data harmonization guidelines

Reference: https://www.maelstrom-research.org/about-harmonization/maelstrom-guidelines[73]

Summary:

These guidelines were developed by the Maelstrom Research team to ensure quality, reproducibility, and transparency of the data harmonization process. Based on these guidelines, retrospective harmonization is an iterative process involving a series of closely related, interdependent, and often integrated steps.

Category:

- Data harmonization

Topics:

- Iterative harmonization steps

ISO 21838

Reference: https://www.iso.org/standard/71954.html[74]

Summary: ISO standard under development. Follows a method tested in over 300 ontology-building initiatives and is being documented in ISO 21838, leveraging existing resources wherever possible

Category:

Topics:

- Promoting re-use of existing ontologies and not create a new kid on the block

4.3.2 Data interoperability, data standards, ontology and controlled terminology

OLS

Reference: https://www.ebi.ac.uk/ols/index[75]

Summary:

The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the Samples, Phenotypes and Ontologies Team (SPOT) at EMBL-EBI.

Category:

- Ontology lookup service

Topics:

[73] https://www.maelstrom-research.org/about-harmonization/maelstrom-guidelines
[74] https://www.iso.org/standard/71954.html
[75] https://www.ebi.ac.uk/ols/index

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

OxO

Reference: https://www.ebi.ac.uk/spot/oxo/index[76]

Summary:

OxO is a service for finding mappings (or cross-references) between terms from ontologies, vocabularies and coding standards. OxO imports mappings from a variety of sources including the Ontology Lookup Service and a subset of mappings provided by the UMLS.

Category:

- Ontology mapping service

Topics:

BioPortal

Reference: https://bioportal.bioontology.org/[77]

Summary:

The world's most comprehensive repository of biomedical ontologies

Category:

- Biomedical ontologies
- Mappings

Topics:

- Recommendor: Get recommendations for the most relevant ontologies based on an excerpt from a biomedical text or a list of keywords
- Annotator: Get annotations for biomedical text with classes from the ontologies
- Mappings: Browse mappings between classes in different ontologies

FAIRsharing

Reference: https://fairsharing.org/[78]

Summary: A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies. Anyone can be a user of FAIRsharing. FAIRsharing brings the producers and consumers of standards, databases, repositories and data policies closer together, with a growing list of adopters. Representatives of institutions, libraries, journal publishers, funders, infrastructure programmes, societies and other organizations or projects (that in turn serve and guide individual researchers or other stakeholders on research data management matters) can become an adopter. We also welcome collaborative proposals from complementary resources, we are open to participate in joint projects to develop services for specific stakeholders and communities.

[76] https://www.ebi.ac.uk/spot/oxo/index
[77] https://bioportal.bioontology.org/
[78] https://fairsharing.org/

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

Category:

- Data and metadata standards
- Data policies

Topics:

4.3.3 Data infrastructure, data management platforms and tools

REDCap

Reference: https://redcap.vanderbilt.edu/consortium/library/search.php[79]

Summary: Research Electronic Data Capture

The REDCap Shared Library is a repository for REDCap data collection instruments and forms that can be downloaded and used by researchers at REDCap partner institutions. Curated instruments have been approved for inclusion by the REDCap Library Oversight Committee (REDLOC) after review for research relevance, accuracy in function and coding (see guidelines), and copyright issues.

Category:

- library of clinical form
- curated/approved case record forms

Topics:

---

data.europa.eu

Reference: https://data.europa.eu/en[80]

Summary: The official portal for European data. The portal provides access to open data from international, EU, national, regional, local and geo data portals. It replaces the EU Open Data Portal and the European Data Portal.

The portal addresses the whole data value chain, from data publishing to data reuse.

Going beyond collecting metadata (data about data), the strategic objective of the portal is to improve accessibility and increase the value of open data.

Category:

- Data and metadata standards
- Data policies

Topics:

- Searching data. Here users can find datasets across categories from many different data portals.
- Providing data. This section helps users to understand open data from the perspective of a data provider. There are also instructions for those who wish their data portal to be harvested by the portal.

[79] https://redcap.vanderbilt.edu/consortium/library/search.php
[80] https://data.europa.eu/en

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

- Using data. This section provides details on how open data is being used, as well as its economic benefits.
- Training and library. Here users will find eLearning modules about open data as well as training guides and a knowledge base referencing publications around open data.

RD-Connect Genome-Phenome Analysis Platform (GPAP)

Reference: https://platform.rd-connect.eu[81]

Summary: The RD-Connect GPAP is a key component of EU projects such as EJP-RD and Solve-RD to share and collaboratively analyse and interpret pseudonymised integrated genome-phenome data from Rare Disease patients. The system enables diagnosis and gene discovery. Local instances of the GPAP have been deployed for specific projects (URD-Cat, Nagen1000, MedPerCan) which could be federated. Overall, over 20,000 exome/genomes linked to phenotypic profiles are included in the different instances, with over 500 users.

Category:

- Data infrastructure for Rare Diseases diagnosis and gene discovery
- Genome-phenome data sharing policies

Topics:

- Data collation: the GPAP-PhenoStore module enables phenotypic data submission per disease type through a Graphical User Interface or batch import/export. Phenopackets compatible. Genomic data submission is done through Aspera or SFTP. Metadata is collected through a specific module (batch submission or by experiment).
- Data management and logs: users can manage their submitted datasets and know which other users have specifically analysed it. The GPAP-CohortApp module enables the generation of "in-silico" cohorts based on several criteria to conduct analysis on similar individuals.
- Interoperability: clinical/phenotypic data is collated with standards such as HPO, ORDO and OMIM. Genome data analysis uses standards such as FASTQ, BAM/CRAM and gVCF/VCF. Connection to EGA enabled through GA4GH htsget standard to remotely visualise alignments. API for data access and analysis available. ELIXIR AAI compatibility ready.
- Data discovery: the GPAP is connected to the Beacon Network (GA4GH Beacon 1.0 standard) and MatchMaker Exchange (https://www.matchmakerexchange.org/[82]). Testing implementation of Beacon 2. Enhanced data discovery enabled within the system for authorised users.
- Data sharing: Data sharing policies implemented. Data sharing enabled only for authorised users regulated through Code of Conduct and Adherence Agreement, supervised by a Data Access Committee (DAC). Possibility of embargo period.
- Diagnosis and gene discovery: integrated genome-phenome data analysis and interpretation through the inclusion of many annotations and tools, either included in the system or connected through their web-services.
- Central or local implementations: A central GPAP is available for European clinicians and clinical researchers at https://platform.rd-connect.eu[83]. Local instances have been

[81] https://platform.rd-connect.eu
[82] https://www.matchmakerexchange.org/
[83] https://platform.rd-connect.eu

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

deployed for specific initiatives (URD-Cat, Nagen1000, MedPerCan), with the aim of federating them in line with 1+MG objectives.

Testing/Training environment: https://playground.rd-connect.eu/[84]

Genomics England PanelApp

Reference: https://panelapp.genomicsengland.co.uk/[85]

Summary: Genomics England[86] PanelApp is a publicly-available knowledge base that allows virtual gene panels related to human disorders to be created, stored and queried. It includes a crowdsourcing tool that allows genes and genomic entities (short tandem repeats/STRs and copy number variants/CNVs) to be added or reviewed by experts throughout the worldwide scientific community, providing an opportunity for the standardisation of gene panels, and a consensus on which genes have sufficient evidence for disease association.

Diagnostic-grade 'Green' genes/genomic entities, and their modes of inheritance are used in genome interpretation. Originally developed to aid interpretation of participant genomes in the 100,000 Genomes Project, PanelApp is now also being used as the platform for achieving consensus on gene panels in the NHS Genomic Medicine Service (GMS). As panels in PanelApp are publicly available, they can also be used by other groups and projects.

Category:

-

Topics:

4.3.4 Data governance, genomics data framework

GA4GH

Reference: https://www.ga4gh.org/[87]

Summary: The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework.

Category:

- Data access
- Use cases rare diseases, comm/complex disease and cancer
- Framework(s)

Topics:

- Data use ontology (DUO). A GA4GH-approved Standard The GA4GH Data Use Ontology (DUO) allows users to semantically tag genomic datasets with usage restrictions, allowing them to become automatically discoverable based on a health, clinical, or biomedical researcher's authorization level or intended use. DUO is based on the OBO Foundry principles and developed using the W3C Web Ontology Language. It is being used in

[84] https://playground.rd-connect.eu/
[85] https://panelapp.genomicsengland.co.uk/
[86] https://www.genomicsengland.co.uk/
[87] https://www.ga4gh.org/

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

production by the European Genome-phenome Archive (EGA) at EMBL-EBI/CRG as well as the Broad Institute for the Data Use Oversight System (DUOS).

- Phenopackets[88]. The goal of the phenopacket-schema is to define the phenotypic description of a patient/sample in the context of rare disease, common/complex disease, or cancer. The schema as well as source code in Java, C++, and Python is available from the phenopacket-schema GitHub repository[89]
- Beacon project[90]. Especially the announcement of v2 which allows extending queries to phenotypes, regions, and data use
- Framework for Responsible Sharing of Genomic and Health-Related Data: https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/[91]
- Genomics in Health Implementation Forum (part of GA4GH): https://www.ga4gh.org/community/ghif/[92]

EUCANCan

Reference: https://eucancan.com/[93]

Summary: EUCANCan is a European Canadian cooperation funded by the European Union's Horizon 2020 research and innovation programme and the Canadian Institutes of Health Research. The four-year project aims at enhancing modern oncology, by implementing a cultural, technological and legal integrated framework across Europe and Canada, to enable and facilitate the efficient analysis, management and sharing of cancer genomic data.

EUCANCan is a federated network of aligned and interoperable infrastructures for the homogeneous analysis, management and sharing of genomic oncology data for Personalized Medicine.

EUCANCan proposes to create the EUropean-CANadian Cancer network (EUCANCan), a federated infrastructure whose mission is to enable Personalized Medicine in Oncology by promoting the generation and sharing of harmonized genomic and phenotypic data. EUCANCan builds on work performed by members of the consortium and related projects to align and interconnect existing European and Canadian infrastructures for the analysis and management of genomic oncology data. The EUCANCan network will be composed of reference nodes in Amsterdam, Barcelona, Berlin, Heidelberg, Paris and Toronto which have established strong research and clinical programs in the field of genomic oncology. These reference nodes will work together in an interoperable fashion to provide the genomic oncology community with a uniform computing environment for the processing, harmonization and secure sharing of cancer genome and phenome data in the context of clinical research, enabling the discovery of clinically-relevant patterns of variation in the cancer genome such as biomarkers predictive of therapeutic response. The infrastructure will also provide a proving ground for federated

[88] https://phenopackets-schema.readthedocs.io/en/1.0.0/index.html
[89] https://github.com/phenopackets/phenopacket-schema
[90] https://beacon-project.io
[91] https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/
[92] https://www.ga4gh.org/community/ghif/
[93] https://eucancan.com/

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

genome analysis systems that may one day be integrated into national and regional healthcare systems.

EUCANCan's objectives are: (1) harmonise protocols for the identification and interpretation of germline and somatic variation profiles within cancer genomes; (2) generate strategies for the flow, management, storage and distribution of data within and across EUCANCan nodes; (3) define community standards for data elements, types and formats; (4) develop an open and accessible data portals for the searching and download of EUCANCan data; and (5) define an appropriate ethical and legal frame to ensure the secure sharing of protected individual genomic and phenotypic data across countries."

Category:

Topics:

CINECA

Reference: https://www.cineca-project.eu/[94]

Summary: Common Infrastructure for National Cohorts in Europe, Canada, and Africa. Accelerating disease research and improving health by facilitating transcontinental human data exchange.

Cohorts: CINECA brings together a diverse collection of human cohorts consisting of 1.4M individuals in Canada, Europe, and Africa. CINECA cohorts are selected as they provide a representation of the scales, types, variable consents and ELSI challenges related to global cohorts, thus ensuring a representative set for CINECA's activities. A particular strength of CINECA is that it does not represent a specific disease focus and cohorts are selected to address common diseases, a major worldwide health burden. This will ensure that the federation model and standards are applicable in any disease context and are well tested across our diverse cohorts. CINECA represents a unique opportunity to build one of the world's first transcontinental federated networks of human data discovery and sharing. CINECA's outputs are also immediately applicable to rare disease and will interoperate with rare disease infrastructures such as RD-Connect, Matchmaker exchange and others. This will allow analyses in future that cross rare and common diseases, desirable as rare disease phenotypes inform our understanding of common disease.

Category:

- Use case cohorts with focus on common diseases, and linking to rare diseases
- International consortium

Topics:

- Cohorts
- Harmonised Cohort Level Metadata
- Maelstrom Research data standards
- Metadata model encoded as application ontology: GEKCO (Genomics Cohorts Knowledge Ontology), see http://www.obofoundry.org/ontology/gecko.html[95]

[94] https://www.cineca-project.eu/
[95] http://www.obofoundry.org/ontology/gecko.html

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

IHCC

Reference: https://ihccglobal.org/[96]

Summary:

The International HundredK+ Cohorts Consortium (IHCC) aims to create a global platform for translational research – cohort to bedside and cohort to bench – informing the biological and genetic basis for disease and improving clinical care and population health.

Category:

- International consortium

Topics:

- Atlas, see https://atlas.ihccglobal.org/[97]

Orphanet

Reference: https://www.orpha.net/consor/cgi-bin/index.php?lng=EN[98]

Summary: Orphanet is a unique resource, gathering and improving knowledge on rare diseases so as to improve the diagnosis, care and treatment of patients with rare diseases. Orphanet aims to provide high-quality information on rare diseases, and ensure equal access to knowledge for all stakeholders. Orphanet also maintains the Orphanet rare disease nomenclature (ORPHAcode), essential in improving the visibility of rare diseases in health and research information systems.

Category: ontology development/look up

Topics:

5. Results

The best, promising and innovative practices have been identified and described in this document. Some of the best practices have been (nearly) completely captured in the Best Practice template [1] while others still have to be either identified or to be completed using the Best Practice template.

6. Discussion

Our goal is to identify and describe best practices on sharing and linking phenotypic and genetic data in both the healthcare sector as in the research setting. This is not a static process in that practices come, adapt and go. B1MG WP3 together with the experts from the 1+MG W3 have been able to identify and in part extensively describe a number of best, promising and innovative practices that will be of benefit to all relevant actors within the 1+MG project.

[96] https://ihccglobal.org/
[97] https://atlas.ihccglobal.org/
[98] https://www.orpha.net/consor/cgi-bin/index.php?lng=EN

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

We realise that there might still be relevant unidentified best, promising or innovative practices. We therefore will incorporate new identified best practices in coming versions. We will also apply modifications to already described best practices where deemed necessary.

7. Conclusions

The first version of the documented best practises in sharing and linking phenotypic and genetic data has been established and will be in use by all relevant actors.

The current version of this document will be reviewed to ensure it continues to be fit for purpose and that any changes introduced to the best practices are incorporated into the document.

8. Next steps

This document will be updated in Q2 2022 but in the meantime continuous updates will be made in the working document at the B1MG google drive. Link to live document: Best practices set-up working document[99]

B1MG-WP3 in collaboration with 1+MG WG3 have recently identified a selected number of datasets from the Survey of accessible genomes that could be identified as best practices in e.g. minimal datasets, data dictionaries, data sharing and data linking of genomic and phenotypic and/or clinical data. The initial selected datasets can be viewed via this link[100]. The data owners (listed contacts) of the selected datasets will be approached by B1MG-WP3 & 1+MG-WG3 to discuss their datasets in more detail to help amongst other in identifying and arriving at a minimal generic clinical/phenotype dataset as well as a a minimal specific, for each of the use cases of 1+MG, clinical/phenotype dataset.

9. Impact

This document is to identify and describe best practices on sharing and linking phenotypic and genetic data in both the healthcare sector as in the research setting to, as much as possible, avoid reinventing the wheel, learn from previous/current existing projects to improve performance and avoid mistakes made by others.

10. Glossary of terms, abbreviations and acronyms

API: Application Programming Interface
B1MG: Beyond One Million Genomes
BP: Best Practice
BDP: Best Practice Document
DAC: Data Access Committee
EGA: European Genome-phenome Archive
FAIR: Findable Accessible Interoperable Reusable

[99] https://docs.google.com/document/d/1GOvcR3l3t8T4cJLDVx7kVxbYa9F2oo7deQveMG6Og6c/edit#
[100] https://docs.google.com/spreadsheets/d/1IkpDbH0C0bGKJIDFUMPFuObScmAwFX05/edit#gid=217752475

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

GA4GH: The Global Alliance for Genomics and Health
HPO: Human Phenotype Ontology
ICGC ARGO: The International Cancer Genome Consortium Accelerating Research in Genomic Oncology
IHE: Integrating the Healthcare Enterprise
ISO: International Organization for Standardization
KM: Knowledge Management
ORCID[101]: Open Researcher and Contributor ID
RD: Rare Disease
URL: Uniform Resource Locator

11. References

[1] Meshari Alwazae, Erik Perjons, Paul Johannesson. Applying a Template for Best Practice Documentation. Procedia Computer Science, Volume 72, 2015, Pages 252-260. https://doi.org/10.1016/j.procs.2015.12.138[102] (https://www.sciencedirect.com/science/article/pii/S1877050915035991[103])

[2] Ng E, de Colombani P. Framework for selecting best practices in public health: a systematic literature review. J Public Health Res. 2015;4(3):577 https://doi.org/10.4081/jphr.2015.577[104]

[101] https://orcid.org/
[102] https://doi.org/10.1016/j.procs.2015.12.138
[103] https://www.sciencedirect.com/science/article/pii/S1877050915035991
[104] https://doi.org/10.4081/jphr.2015.577

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