Beyond One Million Genomes

D3.8

Documented best practices in sharing and linking phenotypic and genetic data —2v0

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	D3.8 - Documented best practices in sharing and linking phenotypic and genetic data - 2v0		
Contractual delivery date	31/05/2022	**Actual delivery date**	21/11/2022
Delayed	Yes		
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), Atilla Patocs (NIO, B1MG WP3 & 1+MG WG3, HU) K. Joeri van der Velde (UMCG, FAIR genomes+MOLGENIS, NL) Morris Swertz (UMCG, FAIR genomes+MOLGENIS, 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)		

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

	Adolfo Muñoz (ISC III, Spanish 1+MG mirror group, SP) Miguel Pedrera, Pablo Serrano (Hospital 12Octubre, 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), Harmke Groot (Nictiz, B1MG WP3, NL), Flávio Soares (1+MG WG3, PT) Inês Lourenço (1+MG WG3, PT) Jernej Kovač (1+MG WG3, SI) Ulrika Hermansson (1+MG WG3, SE) Evita M. Lindholm (1+MG WG3, NO) Yanis Mimouni (INSERM, EJP RD coordination, FR) Ana Rath (INSERM, EJP RD Pillar2 Leader, FR) Tala Haddad (INSERM, EJP RD coordination, FR)
Contributors	
Acknowledgements (not grant participants)	
Deliverable type	Report
Dissemination level	Public

Document History

Date	Mvm	Who	Description
14/12/2021	0v1	Jeroen Beliën (VUMC)	Continuation of working document based on 202105 B1MG D3.7 - Documented best practices in sharing and linking phenotypic and genetic data - 2v0 as uploaded to the EU-portal and published at Zenodo
31/10/2022	0v2	Jeroen Belien (VUMC) & Nikki Coutts (ELIXIR Hub)	B1MG-WP3 and 1+MG-WG3 comments addressed. Version circulated to B1MG-OG, B1MG-GB & Stakeholders for feedback
21/11/2022	1v0	Jeroen Belien (VUMC) & Nikki Coutts (ELIXIR Hub)	B1MG-OG, B1MG-GB & Stakeholders comments addressed. Version uploaded to the EC Portal and published at Zenodo

Table of Contents

1. Executive Summary 4

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 **5**
Objective 1 5
Objective 2 5
Objective 3 6
3. Background / Methods **6**
4. Introduction **7**
4.1 Best practices defined 7
4.1.1 Definition of a best practice 7
4.1.2 Criteria for identifying a best practice 7
4.1.3 Template for describing the best template 8
5. Results **17**
6. Discussion **17**
7. Conclusions **18**
8. Next steps **18**
9. Impact **19**
10. Glossary of terms, abbreviations and acronyms **19**
11. References **20**
Appendix I: Templated best and promising practices **20**
Data model and templates 21
Data interoperability, data standards, ontology and controlled terminology 26
Data infrastructure, data management platforms and tools 42
Data governance, genomics data framework 57
Appendix II: Identified best and promising practices **63**
Date model and templates 63
Data interoperability, data standards, ontology and controlled terminology 65
Data exchange standards 66
Data infrastructure, data management platforms and tools 66
Data governance, genomics data framework 70
Appendix III: National implementation of standards **73**

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 second 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 health care 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 listed 'best practices' have been identified by the 1+MG WG3 experts, who are nominated by the Member States, and are exemplary practices that have achieved results which could be used for larger scale cross-border initiatives.

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 health care 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 practice recommendations. This document will be updated continuously and handed over to the EU-GDI project. 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
	Infrastructure Key Results	
	9. Secure cross-border data access roadmap (WP4) by M12 & M36.	No

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

	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. Background / Methods

The collection, analysis, use and sharing of genomic data promises major breakthroughs in health research, more specifically for personalised medicine and for population health. Personalised 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 (defined as the observable characteristics or traits of an organism or a cell line (i.e., the physical manifestation of a genotype[1]), next to the wish to link it to data such as environmental data, social-economical data, information in medical records and administrative data. As such, to ensure optimal use of genomic datasets for research and development of personalised medicine, linkage of genomic and health related data is a cornerstone for realising 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 maximise 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" have been identified by the experts nominated by the Member States and 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

[1] Phenotypic data as defined by NHGRI's Metadata and Phenotypic Data Sharing Expectations FAQ (genome.gov)

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

best practices on sharing and linking phenotypic and genetic data in both the health care 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 second version of the best practice recommendations. This document will be updated continuously and handed over to the EU-GDI project.

4. Introduction

4.1 Best practices defined

4.1.1 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].

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

4.1.2 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". In fact, best practices should where applicable follow the FAIR

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

principles [3]. ELIXIR and the related communities, including the pharmaceutical industry have developed detailed "howtos" towards implementing the FAIR principles. The FAIR cookbook is one of such resources developed via a public-private partnership (IMI-FAIRplus) and becomes a commissioned service (https://faircookbook.elixir-europe.org/content/home.html) of ELIXIR [4]. It contains practical recipes on specific targets to improve the FAIRness of data and other assets. One such example is the recipe on how to choose controlled vocabulary (https://w3id.org/faircookbook/FCB020). The best practices themselves should be findable - we are collecting and publishing them; they should be accessible - they are publically available via Zenodo (open access); they should be interoperable - within 1+MG aim to combine various best practices and therefore the best practices need to be interoperable or at least one should be able to combine them; and best practices should be re-usable - provide enough information such that a best practice can be implemented by others. 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 and description used within 1+MG

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 measurable results.
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.

4.1.3 Template for describing the 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 that 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.

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

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

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

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 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 organisation
	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 organisation - for resolving the problem addressed by the BP

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

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
	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 organisation
	Application Time: The time it takes to apply the BP in an organisation
	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 realise 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 WP6 (and reviewed by B1MG WP3 and 1+MG WG3).

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

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

Table 3. List of BP topics

Category of BP	Titles of BP	Detailed approaches and categories	Applied to data type	BP classification (Best, Promising or Innovative)	Uptake of BP
Data model and templates	ART-DECOR	Standards and Template tool		Best	ART-DECOR is the base for specification, documentation and testing in more than 40 projects and used right now in Germany, Austria, Italy, Poland, Norway, Lithuania and the Netherlands, among other countries. More information see: ART-DECOR Governance Groups / Projects - art-decor.org
	Phenopackets schema	Data model	Phenotype data	Best	
	Observational Medical Outcomes Partnership (OMOP)	Data model	EHR and administrative claims data	Promising	
	Portal of Medical Data Models	Template collection	General	Best	
	Maelstrom Data harmonisation guidelines	Guidelines	General	Best	

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

	Information technology — Top-level ontologies (TLO) (ISO 21838)	Data harmonisation		Promising	
	ISO 23903:2021 Health informatics — Interoperability and integration reference architecture – Model and framework	Model and framework		Best	
Data interoperability, ontology and controlled terminology, ontology collections, mappings	Ontology Lookup Service (OLS)			To be classified	
	Human Phenotype Ontology (HPO)	Ontology	Phenotype data	Best	
	ORPHAcodes	Code system	Rare Disease specific coding system	Best	
	ORDO	Ontology	Rare disease	Best	
	HOOM	Ontological module	Rare disease	Best	
	SNOMED-CT	Ontology	Medical data	Best	41 countries. Actual information.[2]
	International Classification of Diseases (ICD) (e.g. ICD-11 or ICD-10, ICD-O)	Classification system	Morbidity and mortality statistics, reimbursement systems, and automated decision support in health care	Best	Current implementation status in EU member states (2018)[3] and and https://www.who.int/standards/classifications/frequently-asked-questions/icd-11-implementation
	Logical Observation Identifiers Names and Codes (LOINC)	Ontology	Tests, measurements, and	Best	

[2] https://www.snomed.org/our-customers/members
[3] https://webgate.ec.europa.eu/fpfis/wikis/pages/viewpage.action?pageId=912786535

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

			observations		
	Unified Code for Units of Measure (UCUM)	Code system	Units of measures being contemporarily used	Best	
	Ontology Xref Service (OxO)			To be classified	
	BioPortal	Biomedical Ontology collection, mappings	General	Best	
Data standards	HL7 FHIR4FAIR FHIR	Data standards	EHR data, FAIR Health data	Innovative: HL7 Standard for`trial Use (STU) balloted per 2022-09-28	See: HL7.FHIR.UV.FHIR-FOR-FAIR\ FHIR for FAIR Home Page - FHIR v4.3.0[4] and paper: FAIRness for FHIR: Towards Making Health Datasets FAIR Using HL7 FHIR - PubMed (nih.gov)[5]
	FAIRsharing	Data and metadata standards, Data policies	General	Best	
	ISO/AWI TR 24305 Health informatics - Guidelines for implementation of HL7/FHIR based on ISO 13940 and ISO 13606	Data standards	EHR data	Promising	
Data exchange standards	Fast Healthcare Interoperability Resources from Health Level-7 UK (HL7) (FHIR)	Data exchange	standard for health care data exchange	Promising v4.3.0: R4B - HL7 Standard for`trial Use (STU) balloted	See: https://hl7.org/fhir/

[4] http://hl7.org/fhir/uv/fhir-for-fair/STU1/
[5] https://pubmed.ncbi.nlm.nih.gov/35672963/

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

	My Health @ EU - eHealth Digital Service Infrastructure (eHDSI)	Data exchange		Best	EU-wide implementation and use of Patient Summary and ePrescription,
Data infrastructure, data management platforms and tools	CEDAR	Metadata management tool	General	Best	Used by ZonMW and GO-FAIR foundation in COVID-support program: https://www.gofairfoundation.org/m4m/ & https://www.health-ri.nl/initiatives/dutch-covid-19-data-support-programme/workshops-delivering-fair-metadata-covid-19-data The EOSC-Nordic: https://eosc-nordic.eu/metadata-for-machines-workshop-m4m-is-another-great-success/
	openEHR	EHR platform	EHR data	promising	See OpenEHR deployed solutions[6]
	REDCap	EDC tool	General	Best	 See https://www.project-redcap.org/
	International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) Data Platform	Data infrastructure, data platform		Best	

[6] https://www.openehr.org/deployments/provider_deployments/

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

	European Joint Programme on Rare Diseases Virtual Platform (EJP RD VP)	Data platform	Rare disease data	Innovative	The Virtual Platform (VP) is a federated ecosystem, in which resources are enhanced to be amenable to RD research, and made FAIR. Link to VP[7]
	European Platform on rare disease registration (EU RD Platform)	Data platform	Rare disease patient registries	Best	
	Federated EGA	Data repository	Genetic data	Innovative	Oct 2022 launched. 5 countries (Finland, Germany, Norway, Spain, and Sweden) involved. Nr will extend in near future
	(Central) EGA (Note: see federated EGA for description and templated BP table; same technology)	Data repository	Genetic data	Best	Central repositories hosted by EMBL-EBI and CRG, used by many countries
	data.europa.eu	Data repository	General	Best	
	I2b2-tranSMART	Data management tool	General	Best	
	FAIR4Health			Innovative	
	RD-Connect Genome-Phenome Analysis Platform (GPAP)	Data platform		Best	
	Genomics England Panel App			Best	

[7] https://vp.ejprarediseases.org/

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

Data governance, genomics data framework	Global Alliance for Genomics and Health (GA4GH)			Best	
	FAIR genomes			Promising	
	European - Canadian Cancer Network (EUCANCan)			Promising	
	Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA)			Promising	
	International HundredK+ Cohorts Consortium (IHCC)			To be classified	
	Orphanet			Best	
Data dictionaries	ICGC Argo dictionary[8]	Data dictionary	Cancer	Best (WG9 preferred choice)	
	mCODE[9]	open-source structured data elements for oncology	Oncology	Best (WG9 preferred choice)	
	OSIRIS[10]	Minimum Data Set for Data Sharing and Interoperability in Oncology	Oncology	Best (WG9 preferred choice)	
	GDC-NIH[11]	Minimum Data Set for Data Sharing	Oncology	Best (WG9 preferred choice)	

[8] https://docs.icgc-argo.org/dictionary
[9] https://confluence.hl7.org/display/COD/mCODE/
[10] https://ascopubs.org/doi/10.1200/CCI.20.00094
[11] https://docs.gdc.cancer.gov/Data_Dictionary/viewer/

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

		and Interoperability in Oncogenomics			
	CINECA Cohort minimal metadata model[12]	minimal metadata model of the basic set of attributes that should be recorded with all cohorts	Cohorts	Promising	
	European Prospective Investigation into Cancer and Nutrition (EPIC)	EU research project	Prospective cohort	-	

For readability the detailed descriptions of best and promising practices with a (for the most part) completed Best practices template can be found in Appendix I: Templated best and promising practices. Those best practices and promising practices that have already been identified by the experts but still need to be worked on (as in getting additional/relevant information such that it can be templated and taken up in Appendix I) are listed in Appendix II: Identified best and promising practices.

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.

The inventory of status of implementation and use of standards within their country made by the 1+MG WG3 expert group already after first analyses shows that in certain domains an advice for countries not having implemented standards for that domain can be given on preference of a standard being implemented and used in other countries, like SNOMED CT in the domain pathology, or ICD in combination with SNOMED CT in the domain cancer, and an upcoming standard like HPO within the domain of rare diseases..

[12] https://zenodo.org/record/4575460

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

This table (Appendix III) will be further processed and any advice on or preference for standards will be presented in the next version (v2.0) of the Phenotypic and clinical metadata framework.

6. Discussion

Our goal is to identify and describe best practices on sharing and linking phenotypic and genetic data in both the health care 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.

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

This second and within B1MG final version of the documented best practices 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 for the duration of the B1MG be updated via the working document at the B1MG google drive. Link to live document: Best practices set-up working document[13] The identified best practices will be part of the 1+MG Trust Framework and will be considered for implementation in the upcoming Genomic Data Infrastructure project that has an expected starting date of November 1st 2022.

At publication of this document the 1+MG WG9 (cancer use case) in close collaboration with B1MG-WP3 and 1+MG WG3 based on identified best practices on minimal datasets within the field of cancer has released an initial version of the 1+MG WG9 minimal dataset. Currently it is investigated if this 1+MG WG9 minimal dataset for cancer can be implemented within ART-DECOR as suggested in the framework document (B1MG D3.4 Phenotypic and clinical metadata framework — 1v0)[14]. With the datasets enrolled via the EUSurvey - Survey

[13] https://docs.google.com/document/d/1GOvcR3l3t8T4cJLDVx7kVxbYa9F2oo7deQveMG6Og6c/edit#
[14] https://doi.org/10.5281/zenodo.6573853

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

(europa.eu)[15] we will continue to identify best practices (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[16]. 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 as has been done for the 1+MG WG9 use case.

9. Impact

This document is to identify and describe best practices on sharing and linking phenotypic and genetic data in both the health care 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. It will be part of the 1+MG Trust framework and serve the implementation choices within the Genomic Data Infrastructure project.

10. Glossary of terms, abbreviations and acronyms

API: Application Programming Interface

B1MG: Beyond One Million Genomes

BP: Best Practice

BDP: Best Practice Document

CINECA: Common Infrastructure for National Cohorts in Europe, Canada, and Africa

DAC: Data Access Committee

EGA: European Genome-phenome Archive

EPIC: European Prospective Investigation into Cancer and Nutrition

FAIR: Findable Accessible Interoperable Reusable

GA4GH: The Global Alliance for Genomics and Health

GPAP: Genome-Phenome Analysis Platform

HPO: Human Phenotype Ontology

[15] https://ec.europa.eu/eusurvey/runner/1plusMG_Survey2020
[16] https://docs.google.com/spreadsheets/d/1lkpDbH0C0bGKJIDFUMPFuObScmAwFX05/edit#gid=217752475

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

ICD: International Classification of Diseases

ICGC ARGO: The International Cancer Genome Consortium Accelerating Research in Genomic Oncology

IHCC: International HundredK+ Cohorts Consortium

IHE: Integrating the Healthcare Enterprise

ISO: International Organisation for Standardisation

KM: Knowledge Management

LOINC: Logical Observation Identifiers Names and Codes

OLS: Ontology Lookup Service

OMOP: Observational Medical Outcomes Partnership

ORCID: Open Researcher and Contributor ID

OxO: Ontology Xref Service

RD: Rare Disease

UCUM: Unified Code for Units of Measure

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 (https://www.sciencedirect.com/science/article/pii/S1877050915035991)

[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

[3] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[4] Rocca-Serra, P., Gu, W., Ioannidis, V., et al. The FAIR Cookbook - the essential resource for and by FAIR doers. https://doi.org/10.5281/zenodo.7156792

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

Appendix I: Templated best and promising practices

This appendix lists all Best and promising practices which have been identified as such by B1MG WP3 and 1+MG WG3 and of which the BP template has been (for the largest part) completed. In each paragraph the best practices are grouped (a separate subsection) per BP category as shown in Table 3.
If new information is obtained in the upcoming period the BP templates will be updated accordingly.

Data model and templates

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

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

	ART-DECOR can be used to model functional specifications that are agnostic from technical communication standards. Because of this approach these specifications can be implemented in a wider array of use cases, in one example in both HL7v3, FHIR STU3 and FHIR R4. ART-DECOR can also be used as a modelling tool towards implementers and supports the following technical standards: HL7v2, HL7v3, and FHIR (currently DSTU2, STU3 and R4). ART-DECOR supports the following FHIR artefacts: • dataset or transaction as FHIR Logical Model • transaction as FHIR Questionnaire • value set as FHIR ValueSets • OID Registry info as FHIR NamingSystem Support for FHIR Code system and FHIR profile is planned, but not currently supported.
	Reference (URL): https://art-decor.org/ *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), data steward(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

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

	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 organisations 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: High
	Consistency: High (makes reuse 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).
BP Implementation	*Demonstration of Success:* VASCA, iCRF generator[17], FAIR genomes, IHE, eHDSI, and ART DECOR is the base for specification, documentation and testing in more than 40 projects and used right now in Germany, Austria, Italy, Poland, Norway, Lithuania and the Netherlands, among other countries. See also ART-DECOR Governance Groups / Projects - art-decor.org

[17] https://www.health-ri.nl/services/icrf-generator

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

	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.

Title: Phenopackets schema
Reference: https://phenopackets-schema.readthedocs.io/en/latest/index.html
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 complete models of disease. The standard is designed to encourage wide adoption and synergy between the people, organisations and systems that comprise the joint effort to address human disease and biological understanding.
Category:
- Phenotype data model

Topics:

BP Component	BP *attribute*
Summary of BP	*Title*: Phenopackets schema
	Summary: The Phenopackets 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 complete models of disease.
BP representation	*Pattern Attributes*: The goal of the phenopacket-schema[18] is to define a machine-readable phenotypic description of a patient/sample in the context of rare disease, common/complex disease, or cancer. It aims to provide sufficient and shareable information of the data outside of the EHR (Electronic Health Record) with the aim of enabling capturing of sufficient structured data at the point of care by a clinician or clinical geneticist for sharing with other labs or computational analysis of the data in clinical or research environments. The phenopacket schema defines a common, limited set of data types which may be composed into more specialised types for data sharing between resources using an agreed upon common schema. This common schema has been used to define the 'Phenopacket' which is a catch-all collection of data types, specifically focused on

[18] https://github.com/phenopackets/phenopacket-schema

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

	representing disease data both initial data capture and analysis. The phenopackets schema is designed to be both human and machine-readable, and to inter-operate with standards being developed in organisations such as in the ISO TC215 committee[19] and the HL7 Fast Healthcare Interoperability Resources Specification (aka FHIR®)[20]. A semantic datamodel of the phenopackets schema is under construction. This should increase interoperability of phenopackets with other phenotype models such as defined in OMOP.
	Reference (URL) or *Author Contact Information*: https://phenopacket-schema.readthedocs.io/en/latest/index.html https://github.com/LUMC-BioSemantics/phenopackets-rdf-schema
	Revision Information: This BP is an active project, all updates/revisions can be found under: https://phenopacket-schema.readthedocs.io/en/latest/index.html https://github.com/LUMC-BioSemantics/phenopackets-rdf-schema
	Reviews Information: This BP is an active project, all updates/revisions can be found under: https://phenopacket-schema.readthedocs.io/en/latest/index.html https://github.com/LUMC-BioSemantics/phenopackets-rdf-schema
Requirement for applying BP	*Goal*: The phenopacket-schema is designed to harmonise and share phenotype descriptions.
	Means: Basic IT service *People*: domain specialist(s), data steward(s), data specialist(s).
	Skills: Basic understanding of semantics, domain knowledge of use case at hand.
	Cost: personnel costs 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, clinician or clinical geneticist for sharing with other labs or computational analysis of the data in clinical or research environments.
	Champion: GA4GH, X-omics, FAIRgenomes
	Owner: This work has been produced as part of the GA4GH Clinical Phenotype Data Capture Workstream[21] and is designed to be compatible with GA4GH metadata-schemas[22].

[19] https://www.iso.org/committee/7546903.html
[20] http://hl7.org/fhir/
[21] https://ga4gh-cp.github.io/
[22] https://github.com/ga4gh-metadata/metadata-schemas

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

	Training Needs: For getting started: training required for basic understanding of semantics and protobuf, an exchange format developed in 2008 by Google.
	Acceptability: High level of acceptance. Version 1 of phenopackets was approved by GA4GH in October, 2019[23]. Version 2 is currently being finalised by the Global Alliance for Genomics and Health (GA4GH)[24] Clinical & Phenotypic Data Capture workstream.
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: High
	Consistency: High
	Granularity: High
	Adaptability: High (open source; and provides output in a structured format which can be adapted to secondary uses).
	Activity: Information not available
	Integration: High (it is linked to other standards).
BP Implementation	*Demonstration of Success*: Version 1 of phenopackets was approved by GA4GH in October, 2019. Based on initial experiences and feedback from multiple sources, and discussions in the GA4GH Clin/Pheno Workstream and Phenopackets Subgroups, version 1 has been extended to include better representation of the time course of disease, treatment, and cancer-related data. The current document refers to the version 2 of the Phenopackets schema. Version 2 is currently being finalised by the Global Alliance for Genomics and Health (GA4GH)[25] Clinical & Phenotypic Data Capture workstream.
	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

[23] https://www.ga4gh.org/news/phenopackets-standardizing-and-exchanging-patient-phenotypic-data/
[24] https://www.ga4gh.org/
[25] https://www.ga4gh.org/

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

Data interoperability, data standards, ontology and controlled terminology

Title: The Human Phenotype Ontology (HPO)
Reference: https://hpo.jax.org/app/download/ontology
Summary:
The Human Phenotype Ontology (HPO) provides a standardised vocabulary of phenotypic abnormalities encountered in human disease, where each term in the HPO describes a phenotypic abnormality. 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 standardised 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
	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: http://purl.obolibrary.org/obo/hp/hpoa/phenotype.hpoa Previous annotation: http://purl.obolibrary.org/obo/hp/hpoa/phenotype_annotation.tab:

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

	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 : 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
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: personal cost of data curators. All the data is freely available for download and can be browsed online.
	Barriers: Information not available
	Barrier Management: Information not available
BP Actor	*Community of Practice*: clinical diagnostics in human genetics, bioinformatics research on the relationships between human phenotypic abnormalities and cellular and biochemical networks, for mapping between human and model organism phenotypes and for providing a standardised vocabulary for clinical databases
	Champion: the Monarch Initiative[26], the Global Alliance for Genomics and Health[27]
	Owner: The HPO is a product of the Monarch Initiative[28], an NIH-supported international consortium dedicated to semantic integration of biomedical and model organism data with the ultimate goal of improving biomedical research. The HPO, as a part of the Monarch Initiative, is a central component of one of the 13 driver projects[29] in the Global Alliance for Genomics and Health[30] (GA4GH) strategic roadmap[31].
	Training Needs: Information not available
	Acceptability: It is getting accepted more and more. E.g. it is a central component of one of the 13 driver projects in the Global Alliance for Genomics and Health (GA4GH) strategic roadmap

[26] https://monarchinitiative.org/
[27] https://www.ga4gh.org/
[28] https://monarchinitiative.org/
[29] https://www.ga4gh.org/howwework/driver-projects.html
[30] https://www.ga4gh.org/
[31] https://www.ga4gh.org/howwework/strategic-roadmap.html

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

BP properties	*Usability:* Easy to medium
	Comprehensiveness: High
	Relevance: High
	Justification: HPO-based computational disease models are utilised within most, current phenotype-driven genomic diagnostics software
	Prescriptiveness: High
	Coherence: not applicable
	Consistency: High
	Granularity: High
	Adaptability: High
	Activity: Information not available
	Integration: Information not available
BP Implementation	*Demonstration of Success:* Orphanet uses the HPO to annotate rare diseases and has continued to develop annotations to a broad range of diseases The UK's National Institute for Health Research (NIHR) Rare Disease initiatives extensively use the HPO in their RD-TRC (Rare Disease––Translational Research Collaboration) and NIHR BioResource, in wide-ranging studies.
	Installation Time: Relatively short
	Application Time: Depends on the projects and datasets
	Experiences and feedback: Information not available
	*Measurement:*Several publications available on https://hpo.jax.org/app/help/publications

Title: Orphanet nomenclature of rare diseases (ORPHAcodes)
Reference: https://www.orphadata.com/orphanet-nomenclature-for-coding/
Summary:
The Orphanet nomenclature of rare diseases is a unique and multilingual standardised nosological system aimed at providing a specific terminology for rare diseases.
Category:

- Data interoperability, ontology and controlled terminology, ontology collections, mappings

Topics:

- Rare disease

BP Component	BP *attribute*

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

Summary of BP	*Title*: Orphanet nomenclature of rare diseases (ORPHAcodes)
	Summary: RD-specific coding system (ORPHAcodes) representing rare clinical entities (rare disorders, groups of rare disorders, and subtypes of rare disorders) by preferred terms, exact synonyms and definitions. The nomenclature is organised as a classification system facilitating data aggregation and exploitation. Each entity in the nomenclature is assigned a unique and stable ORPHAcode, is semantically aligned with other terminologies for semantic interoperability. It has been translated in 12 languages so far.
BP representation	*Pattern Attributes:* Orphanet nomenclature of RD tackles the under-recognition and inaccurate representation of RD in medical terminologies. RD-specific coding system aimed at improving the visibility of RD in health and research information systems, acting as an interoperability vector between healthcare and research. Progressively implemented in hospitals and registries, ORPHAcodes allow for health data generation for research. The nomenclature is organised as a classification system facilitating data aggregation and exploitation. Each entity in the nomenclature is assigned a unique and stable ORPHAcode, is semantically aligned with other terminologies. It has been translated in 12 languages so far. It is comprehensive, standardised, evidence-based, interoperable, versioned, computable and free (CC-BY 4.0) It is produced according to standard procedures (see: https://www.orphadata.com/references/) based on peer-reviewed publications and in collaboration with expert networks.
	Reference (URL) or *Author Contact Information*: https://www.orphadata.com/orphanet-nomenclature-for-coding/ INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Revision Information: Dynamic resource. Orphanet nomenclature for coding is released in July each year. Versioning and diffs provided https://www.orphadata.com/orphanet-nomenclature-for-coding/ and https://github.com/orphanet-rare-diseases-issues/RD-CODE Different frequency releases possible to attend different use cases.
	Reviews Information: Dynamic resource Review details are published in human readable file with every release. https://www.orphadata.com/orphanet-nomenclature-for-coding/ Review demands can be posted in GitHub https://github.com/orphanet-rare-diseases-issues/RD-CODE/issues

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

Requirement for applying BP	*Goal*: Implementing ORPHAcodes allows for accurate coding rare disease diagnosis, including RD patients remaining undiagnosed after full investigation. It allows for proper statistics about RD, including epidemiology, monitoring hospital activities and performance, amongst others, and to interoperate with other RD data sources as genomic data repositories or disease registries.
	Means: IT: simpe (API consumption; parser for consuming files). Persons: data stewards, coders, data scientists for data exploitation
	Skills: domain knowledge
	Cost: personal cost for data curators/coders. Orphanet nomenclature is free (CC-BY 4.0)
	Barriers: Lack of understanding of the terminological resource structure
	Barrier Management: Tools and Guidelines published for implementation. Trainings provided. Helpdesk available (GitHub)
BP Actor	*Community of Practice*: Hospital managers, hospital information systems, registries and databases; Orphanet national nomenclature hubs
	Champion: Scientific manager (head of the nomenclature team) Caterina Lucano; IT manager (head of the IT team) Marc Hanauer
	Owner: INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Training Needs: need for basic and advanced trainings. Training provided for coders in several European languages, support provided for IT issues
	Acceptability: Widely implemented in hospitals and registries; recommended as best practice by the European Commission and the EU RD Platform for registries.
BP properties	*Usability*: Easy to medium, visualisation tools provided to improve usability. In general, depends on the implementation choices and on the use cases
	Comprehensiveness: fully comprehensive for RD diagnosis: Orphanet maintains the nomenclature following the evolution of knowledge, mappings with other terminologies regularly updated.
	Relevance: High

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

	Justification: No other terminology is specific for RD. Implemented at different levels in health information systems in 12 European countries, part of the mandatory Common data Elements in RD registries. Adopted in EJP RD Virtual Platform, as a Best practice by the EC https://webgate.ec.europa.eu/dyna/bp-portal/transferred.cfm
	Prescriptiveness: High (guidelines)
	Coherence: High (quality assurance management in place)
	Consistency: High (semantic relationships with other terminologies and resources in the domain of health and research)
	Granularity: ORPHAcodes refers to different levels of granularity in a classification tree, from categories and clinical groups to subtypes of RD
	Adaptability: The classification structure is adapted to different use cases (i.e. public health use cases using categories or groups of disorders, geneticists using granular leaves or subtypes)
	Activity: described here • Orphanet nomenclature and classification of rare diseases • Naming rules for the rare disease nomenclature in English • Collaboration with networks of expertise for the revision of the Orphanet nomenclature and classification of rare diseases • Nomenclature production in national language
	Integration: Orphanet produces qualified alignments from Orpanet nomenclature to ICD-10, ICD-11, OMIM, SNOMED- CT UMLS, MeSH, MedDRA and GARD. Orphanet enriches RD entries to SNOMED-CT and to ICD11 through dedicated collaboration agreements. Orphanet is part of the Gene Curation Coalition; Orphanet is the principal contributor with 5 330 entries. There are 112 256 ORPHA-HPO annotations. Many biodata resources utilise Orphanet nomenclature: MONDO, ClinVar, MedGen, the NIH-NCATS Genetic & Rare Disease Information Center, OMIM, GenAtlas, Uniprot, HGNC, LOVD, Reactome, Ensembl, ICTRP, IUPHAR/BPS Guide to Pharmacology, RD-Connect, GPAP, Cellosaurus, DisGeNET, Linking Open Data for Rare Diseases, NanbyoData, PractiKPharma, PubCaseFinder, Radiology Gamuts Ontology Integration of the Orphanet nomenclature into UMLS is planned for end 2022.

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*: French National Database of RD (BNDMR[32]); RD person's card (Portugal) https://doi.org/10.1016/j.procs.2015.08.593
	Installation Time: Depends on the implementation choice. The resource is delivered in different formats: files (xml; json); API
	Application Time: Needs human resources for coding and entering data in routine basis
	Experiences and feedback: http://www.rd-code.eu/wp-content/uploads/2021/12/826607_D5.3_Implementing-countries-Report-on-ORPHAcodes-adoption_VF.pdf
	Measurement: 123,286 downloads in 2021 (incl. nomenclature & classification files)

Title: Orphanet Rare Disease Ontology (ORDO)
Reference: https://www.orphadata.com/ordo/
Summary:
ORDO provides a structured vocabulary for rare diseases capturing relationships between diseases, genes and other relevant features which will form a useful resource for the computational analysis of rare diseases. It is derived from the Orphanet database www.orpha.net , a multilingual database dedicated to rare diseases populated from literature and validated by international experts.
Category:
- ontology and controlled terminology, ontology collections, mappings

Topics:
- Rare disease

BP Component	BP *attribute*
Summary of BP	*Title*: Orphanet Rare Disease Ontology (ORDO)
	Summary: The Orphanet Rare Disease Ontology (ORDO) provides a structured vocabulary for rare diseases capturing relationships between diseases, genes and other relevant features which will form a useful resource for the computational analysis of rare diseases. It is derived from the Orphanet database www.orpha.net , a multilingual database dedicated to rare diseases populated from literature and validated by international experts. It integrates a nosology (classification of rare diseases), relationships (gene-disease relations, epidemiological data) and connections with other terminologies (MeSH, UMLS, MedDRA), databases (OMIM, UniProtKB, HGNC, ensembl, Reactome, IUPHAR, Genatlas) or classifications (ICD-10).

[32] https://www.bndmr.fr/

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

BP representation	*Pattern Attributes:* ORDO acts as a machine-readable interoperability backbone providing reference knowledge to compute relationships between different classes of data: diseases, genes, inheritance, age of onset of disease, incidence, birth prevalence, point prevalence, and their relationships with other terminologies and data source identifiers, so as to ease data exploitation for rare disease research. ORDO is released in 9 languages ORDO is queryable through a SPARQL endpoint https://www.orphadata.com/ordo-sparql-endpoint/ And a Blazegraph triplestore which embed the last version of ORDO and allowing user to play local queries https://www.orphadata.com/data/ontologies/ordo/last_version/blazegraph_ordo.zip
	Reference (URL) or *Author Contact Information*: https://www.orphadata.com/ordo/ INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Revision Information: Dynamic resource. ORDO is released twice a year in July and December. Versioning, release notes and diffs provided The previous versions of the ontologies (ORDO and HOOM) are also provided through Bioportal with differentials: https://bioportal.bioontology.org/ontologies/ORDO
	Reviews Information: Dynamic resource Review details are published in release notes. As ORDO's provenance is the Orphanet database, reviews are made according to Orphanet procedures.
Requirement for applying BP	*Goal*: Implementing ORDO allows to embed the Orphanet nomenclature (ORPHAcodes) together with manually curated annotations and data allowing to be computable with other data sources as genomic data repositories or disease registries in order to facilitate data exploitation by machines.
	Means: IT skills: triple store/SPARQL
	Skills: domain knowledge; RDF/triple store skills
	Cost: Personal costs of datascientists. ORDO is freely available (CC-BY 4.0).
	Barriers: Lack of familiarity with ontologies
	Barrier Management: Documentation and trainings provided. Helpdesk available (GitHub)

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

BP Actor	*Community of Practice*: data scientists, researchers
	Champion: Scientific manager (head of the nomenclature team) Caterina Lucano; IT manager (head of the IT team) Marc Hanauer, Bioinformatics: David Lagorce
	Owner: INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Training Needs: need for advanced trainings, depending on the knowledge handling ontologies
	Acceptability: High
BP properties	*Usability*: Easy to medium. In general, depends on the implementation choices and on the use cases
	Comprehensiveness: High. Fully comprehensive for RD: Orphanet maintains the nomenclature following the evolution of knowledge, mappings with other terminologies, gene-disease association and epidemiological data regularly updated.
	Relevance: High
	Justification: No other ontology is specific for RD. Adopted in EJP RD Virtual Platform, byRD-connect GPAP, amongst others
	Prescriptiveness: High
	Coherence: High (quality assurance management in place)
	Consistency: High
	Granularity: different levels of granularity ; clinical entities classes organised by granularity
	Adaptability: High
	Activity: described here: https://www.orphadata.com/references/ Qualifying associations (gene-disease, phenotype-disease, etc) require an analysis of a mean of 5 publications, depending on the type of data curated.

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

	Integration: ORDO includes qualified alignments from Orphanet nomenclature to ICD-10, ICD-11, OMIM, UMLS, MeSH, MedDRA and GARD. Many biodata resources incorporate Orphanet resources: MONDO, ClinVar, MedGen, the NIH-NCATS Genetic & Rare Disease Information Center, OMIM, GenAtlas, Uniprot, HGNC, LOVD, Reactome, Ensembl, ICTRP, IUPHAR/BPS Guide to Pharmacology, RD-Connect, GPAP, Cellosaurus, DisGeNET, Linking Open Data for Rare Diseases, NanbyoData, PractiKPharma, PubCaseFinder, Radiology Gamuts Ontology Integration of the Orphanet nomenclature into UMLS is planned for end 2022. ORDO is integrated in FAIRSharing https://fairsharing.org/search?q=ordo ORDO is integrated with HPO in an ontological module, HOOM https://www.orphadata.com/hoom/
BP Implementation	*Demonstration of Success*: Projects using ORDO • Harmonising phenomics information for a better interoperability in the RD field • Cellosaurus • DisGeNET-RDF • Linking Open Data for Rare Diseases • NanbyoData • PractiKPharma • PubCaseFinder • Radiology Gamuts Ontology • Rehabilita, Disruptive Technologies for the Rehabilitation of the Future • Solving the Unsolved Rare Diseases
	Installation Time:
	Application Time:
	Experiences and feedback: 46%[33] of Orphanet website users find ORDO very useful/useful https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_survey2021.pdf
	Measurement: 9,000 downloads in 2021 from Orphadata; ORDO is also available on BioPortal (2,682/year) and OLS (EBI)

Title: HPO-ORDO Ontological Module (HOOM)
Reference: https://www.orphadata.com/hoom/
Summary:
Orphanet provides phenotypic annotations of the rare diseases in the Orphanet nomenclature using the Human Phenotype Ontology (HPO). HOOM is a module that qualifies the annotation

[33] http://www.rd-code.eu/wp-content/uploads/2021/12/826607_D5.3_Implementing-countries-Report-on-ORPHAcodes-adoption_VF.pdf

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

between a clinical entity and phenotypic abnormalities according to their frequency and with further annotations (diagnostic criterion, pathognomonic sign) when appropriate.
Category:
- ontology and controlled terminology, ontology collections, mappings

Topics:
- Rare disease

BP Component	BP *attribute*
Summary of BP	*Title*: HPO-ORDO Ontological Module (HOOM)
	Summary: Orphanet provides phenotypic annotations of the rare diseases in the Orphanet nomenclature using the Human Phenotype Ontology (HPO). HOOM is a module that qualifies the annotation between a clinical entity and phenotypic abnormalities according to their frequency and with further annotations (diagnostic criterion, pathognomonic sign) when appropriate. In ORDO a clinical entity is either a group of rare disorders, a rare disorder or a subtype of disorder. The "Clinical Entity" branch of ORDO has been refactored as a logical import of HPO, and the HPO-ORDO phenotype disease-annotations have been provided in a series of triples in OBAN format in which associations, frequency and provenance are modelled.
BP representation	*Pattern Attributes:* HOOM is a machine-readable ontologies association module allowing for using HPO and ORDO together, in order to compute disease-phenotype associations. HOOM provides extra possibilities for researchers, pharmaceutical companies and others wishing to co-analyse rare and common disease phenotype associations, or re-use the integrated ontologies in genomic variants repositories or match-making tools. HOOM is provided as an OWL (Ontologies Web Languages) file, using OBAN, the Orphanet Rare Disease Ontology (ORDO), and HPO ontological models.
	Reference (URL) or *Author Contact Information*: https://www.orphadata.com/hoom/ INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Revision Information: Dynamic resource, based on Orphanet's phenotypic annotations of diseases. Versions in Orphadata. The previous versions of the ontologies (ORDO and HOOM) are also provided through Bioportal with differentials: https://bioportal.bioontology.org/ontologies/ORDO
	Reviews Information: HOOM is a dynamic resource, produced according to standard procedures (to be published)
Requirement for applying BP	*Goal*: Implementing HOOM allows to embed the Orphanet nomenclature in its ontological form (ORDO) together with manually curated phenotypic annotations using HPO, and to exploit added-value information on

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

	frequency of phenotypic abnormalities in the disease population, in order to facilitate algorithms to be used, for instance, in diagnosis of RD.
	Means: IT skills: triple store/SPARQL
	Skills: domain knowledge; RDF/triple store skills
	Cost: Personal costs of data scientists. HOOM is freely available (CC-BY 4.0).
	Barriers: Lack of familiarity with ontologies
	Barrier Management: Documentation provided. Helpdesk available (Orphadata.com)
BP Actor	*Community of Practice*: data scientists, researchers, industry, SMEs
	Champion: Scientific manager (head of the nomenclature team) Caterina Lucano; IT manager (head of the IT team) Marc Hanauer, Bioinformatics: David Lagorce
	Owner: INSERM, US14 – Orphanet Chair: Ana Rath ana.rath@inserm.fr https://orcid.org/0000-0003-4308-6337
	Training Needs: need for advanced trainings, depending on the knowledge handling ontologies
	Acceptability: High
BP properties	*Usability*: Easy to medium. In general, depends on the implementation choices and on the use cases
	Comprehensiveness: Medium-High. Disease annotations with HPO is based on extensive literature review and manual curation, a lengthy ongoing process
	Relevance: High
	Justification: need for added-value RD characterization by their clinical signs and symptoms to be computable for expert systems using algorithms/AI
	Prescriptiveness: High
	Coherence: High (quality assurance management in place)
	Consistency: High
	Granularity: different levels of granularity both in ORDO and HPO
	Adaptability: High

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

	Activity: described here: https://www.orphadata.com/references/ Qualifying associations (gene-disease, phenotype-disease, etc) require an analysis of a mean of 5 publications, depending on the type of data curated.
	Integration: HOOM is used by the Solve-RD project for phenotype similarity-based prioritisation of variants for solving unsolved patients. HOOM is used by many.
BP Implementation	*Demonstration of Success*: Projects using HOOM • Harmonising phenomics information for a better interoperability in the RD field • Solving the Unsolved Rare Diseases
	Installation Time:
	Application Time:
	Experiences and feedback: 89%[34] of Orphanet website users find ORDO-HPO annotations very useful/useful https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_survey2021.pdf
	Measurement: 1,376 downloads from BioPortal; 2,077 downloads in 2021 from Orphadata

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
https://www.snomed.org/
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

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

[34] http://www.rd-code.eu/wp-content/uploads/2021/12/826607_D5.3_Implementing-countries-Report-on-ORPHAcodes-adoption_VF.pdf

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

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

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 standardised 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

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[35]

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 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.
BP representation	*Pattern Attributes:* ISO13940 provides the foundations for organisational interoperability, allowing the creation of a common context between organisations. ISO 13606 provides the mechanism for the modelling of concepts and the interchange of clinical information in a secure way, integrating and normalising information coming from different sources, allowing its automatic management and processing. It also provides tools (archetypes) for the formal modelling, management and interchange of concepts of the knowledge domain. SNOMED CT provides a standard language for clinical terms.

[35] https://www.ciencia.gob.es/portal/site/MICINN/menuitem.edc7f2029a2be27d7010721001432ea0/?vgnextoid=22ff08f8ee076710VgnVCM1000001d04140aRCRD&vgnextchannel=4346846085f90210VgnVCM1000001034e20aRCRD&lang_chosen=en

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

	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 https://www.iso.org/en/contents/data/standard/06/78/67868.html SNOMED CT is maintained by SNOMED International https://www.snomed.org/
	Revision Information: As with 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
Requirement for applying BP	*Goal*: The use of ISO 13940 and ISO 13606 provide the foundations for organisational 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 organisations ...). 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 of 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
	Champion: Medical Informatics Hospital Clínic-University of Barcelona; Doce de Octubre University Hospital, Madrid, Telemedicine and Information Society Department, Health Institute"CarlosIII"

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

	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, which paves the way to its adoption.
	Acceptability: very well accepted 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 Integration with other standards is underway.

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:* 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 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 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
	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 standardised 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
	Measurement: Quality indicators, Success in semantic interoperability testing, Flexibility of the model

Data infrastructure, data management platforms and tools

Title: ICGC ARGO Data Platform
Reference: ICGC ARGO | Homepage (icgc-argo.org) as well as ICGC ARGO Docs | ICGC ARGO Docs (icgc-argo.org)
Summary: The International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) aims to uniformly analyse 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)

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

- Sample registration
- Donor
- Specimen
- Primary diagnosis
- 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 analyse 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 an 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 Contact: https://www.icgc-argo.org/page/69/contact-us#
	Revision Information: This is an active BP. Dictionary release: December 11, 2020 https://docs.icgc-argo.org/docs/release-notes/dictionary-releases Data release: October 23, 2020 https://docs.icgc-argo.org/docs/release-notes/data-releases 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

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

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 harmonisation
	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 organisation - 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 Organisation International Classification of Diseases, 10th Revision (ICD-10)

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

	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

Title: European Joint Programme on Rare Diseases Virtual Platform (EJP RD VP)
References:
1) https://www.ejprarediseases.org/what-is-the-virtual-platform/
2) https://vp.ejprarediseases.org/

Summary: The Virtual Platform aims to open a single door to discover, query and eventually access patient registries, biobanks, genomics & multi-omics repositories, knowledge bases, resources (such as animal models and cell lines libraries), omics deposition & analysis platforms, and translational & clinical research supporting material and services, in a coordinated manner. The Virtual Platform is a federated ecosystem, in which resources are enhanced to be amenable to RD research, and made Findable, Accessible, Interoperable and Re-usable: data stays at the source level but can be queryable at distance from an EJP RD query point. As an ecosystem, multiple query points will be possible, allowing for sending interrogations from one resource to others. Thus, federated discovery, query and analysis are made possible, preserving patient privacy, and respectful of each resource access conditions.

Category:
- Use case Rare Diseases

Topics:
- Resource and data FAIRness and federation
- Discover, query and access resources and data
- Patient registries
- Biobanks
- Animal models
- Cell lines
- Genomics and multi-omics repositories
- Phenome-genome and multi-omics analysis platforms
- Translational and clinical research supporting material and services

BP Component	BP *attribute*
Summary of BP	*Title:* European Joint Programme on Rare Diseases Virtual Platform (EJP RD VP)
	Summary: The Virtual Platform (VP) is a federated ecosystem providing discovery and query and possible analysis capabilities for rare disease

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

	resources, including but not limited to patient registries, biobanks, data repositories, and knowledge bases,genomics & multi-omics repositories, resources (such as animal models and cell lines libraries), omics deposition & analysis platforms, and translational & clinical research supporting material and services, in a coordinated manner
BP representation	*Pattern Attributes:* Nowadays, data sources and valuable resources for research are scattered, unconnected, and do not speak the same languages, both from a semantic and from a technical point of view. The European Joint Programme on Rare Diseases Virtual Platform aims to open a single door to discover, query and eventually access patient registries, biobanks, genomics & multi-omics repositories, knowledge bases, resources (such as animal models and cell lines libraries), omics deposition & analysis platforms, and translational & clinical research supporting material and services, in a coordinated manner. The Virtual Platform is a federated ecosystem, in which resources are enhanced to be amenable to RD research, and made Findable, Accessible, Interoperable and Re-usable: data stays at the source level but can be queryable at distance from any network query point, as well as from a single-entry EJP RD query board. As an ecosystem, multiple query points will be possible, allowing for sending interrogations from one resource to others. Thus, federated discovery, query and analysis are made possible, preserving patient privacy, and respectful of each resource access conditions.
	Reference (URL) or *Author Contact Information*: https://www.ejprarediseases.org/what-is-the-virtual-platform/ https://vp.ejprarediseases.org/
	Revision Information: version 0 reviews undergoing, version 1 will be released on January 2023
	Reviews Information: will be provided following the version 1 release (https://github.com/ejp-rd-vp)
Requirement for applying BP	*Goal*: To build a semantically and technically interoperable network of resources and data sources enabling to discover, query and analyse data in federated manner from any node in the network and in particular through a privileged door to discover, query and eventually access nodes in the network. Nodes include: patient registries, biobanks, genomics & multi-omics repositories, knowledge bases, resources (such as animal models and cell lines libraries), omics deposition & analysis platforms, catalogues, and translational & clinical research supporting material and services.
	Means: Onboarding documentation (in preparation); means are dependent on the readiness of the nodes in terms of data preparation and technical capabilities
	Skills: Onboarding documentation (in preparation); skills are dependent on the readiness of the nodes in terms of data preparation and technical capabilities

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

	Cost: Costs are covered by EJP RD project for partner resources; sustainability plan in preparation. Cost estimates for resource connections were devised.
	Barriers: Depend on the degree of preparedness of nodes; at the node level: technical limitations ; Financial barriers, data access barriers
	Barrier Management: Services (comprising FAIRification services, dockerized metadata and data models , training, etc.) are provided to lower barriers; different levels of connection offered to adapt to the varying technical capabilities and data access possibilities of the nodes
BP Actor	*Community of Practice*: Rare Diseases community: research infrastructures, researchers, healthcare providers and public health agencies, industry (Medicines and devices)
	Champion: Ana Rath , Franz Schaefer (EJP RD Pillar 2 co-chairs)
	Owner: EJP RD
	Training Needs: Onboarding, FAIRification, Exploitation training planned
	Acceptability: High (developments guided by end-users needs and feedbacks)
BP properties	*Usability:* High
	Comprehensiveness: High
	Relevance: High
	Justification: Current lack of interoperability amongst resources/data sources, most of them not being RD-specific, thus less usable by the RD community despite their potential; unequal use of standards and need for harmonisation; lack of legal interoperability pointing the need for federated systems in order that data never leave home.
	Prescriptiveness: High
	Coherence: medium (federated architecture, moving towards more coherence)
	Consistency: High (approved and widely used standards being applied and adopted to the RD community needs when needed)
	Granularity: High (Virtual Platform specifications released and updated; UML diagrams available and updated; Onboarding guidance documentation for different connection level)
	Adaptability: Very High (VP adapting to any new node and data, scalable to other domains)
	Activity: Development, training and community engagement activities
	Integration: High (federated adaptive ecosystem)
BP Implementation	*Demonstration of Success:* Rare diseases data discoverability, queryability

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

	and analysis within the EU
	Installation Time: it varies, several options available per the level of connection chosen (Metadata discovery; Data discovery; Data querying; Federated analysis); possibility to install local components/software or adapt the own resource IT system (e.g., API); or follow requirements and specifications for building a connected node
	Application Time: Immediate (as soon as the node is connected)
	Experiences and feedback: Captured before & during the VP development via a series of surveys. Feedback continuously form available (https://forms.office.com/r/UrgvkD39t8)
	Measurement: number of resources onboarded per time-unit; functionalities released per version; quality and sustainability metrics.

Title: European Platform on rare disease registration (EU RD Platform)
Reference: https://eu-rd-platform.jrc.ec.europa.eu/_en

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 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
- set of common data elements for rare diseases registration: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en and https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/CDS/EU_RD_Platform_CDS_Final.pdf
- https://rd-connect.eu/what-we-do/omics/gpap/ The RD-Connect Genome-Phenome Analysis Platform (GPAP) is an online tool for diagnosis and gene discovery in rare disease research

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:* 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 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

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

	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 and 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
	Revision Information: Information not available
	Reviews Information: Information not available
Requirement for applying BP	*Goal*: Share standardised data on rare diseases
	Means: Information not available
	Skills: Information not available
	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

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

	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

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

Title: European Prospective Investigation into Cancer and Nutrition (EPIC) study
References:
1. https://epic.iarc.fr/
2.https://pubmed.ncbi.nlm.nih.gov/?term=%22european+prospective+investigation%22+AND+cancer
Summary:
The European Prospective Investigation into Cancer and Nutrition (EPIC) study is one of the largest cohort studies in the world, with more than half a million (521 000) participants recruited across 10 European countries and followed for almost 15 years. EPIC was designed to investigate the relationships between diet, nutritional status, lifestyle and environmental factors, and the incidence of cancer and other chronic diseases. EPIC investigators are active in all fields of epidemiology, and important contributions have been made in nutritional epidemiology using biomarker analysis and questionnaire information, as well as **genetic** and lifestyle investigations.
Category:

- Prospective cohorts (research) with focus on common diseases (CVD, cancer), and linking to several exposures (lifestyle, genetic predisposition,nutrition)
- International consortium

BP Component	BP attribute
Summary of BP	*Title: E*uropean Prospective Investigation into Cancer and Nutrition (EPIC) study
	Summary: study is one of the largest cohort studies in the world, with more than half a million (521 000) participants recruited across 10 European countries and followed for almost 15 years. Exposure assessment (lifestyle, genomics, nutrition) as well as several disease outcomes are captured. B1MG project may consult the EPIC study group to evaluate their data infrastructure and codebook; in addition, learn from their bottlenecks.
BP representation	*Pattern Attributes: The EPIC cohort may serve as a use-case for B1MG since it has experience with sharing GWAS data* and interoperable exchange of epidemiological data.
	Reference (URL): https://epic.iarc.fr/ *Author Contact Information: The EPIC study is jointly coordinated by Professor Elio Riboli, Director of the School of Public Health at Imperial College London[36], United Kingdom, and Dr Marc Gunter and Dr Paul Brennan at the International Agency for Research on Cancer[37] in Lyon, France.*
	Revision Information: Information not available
	Reviews Information: Information not available
Requirement for applying BP	*Goal*: The EPIC collection of data and biological samples constitutes an outstanding resource for medical research on chronic diseases. As a publicly funded multi-centre study, EPIC wishes to ensure that those resources are being put to the best possible use. EPIC-Europe includes 521 330 participants. The EPIC data and biospecimens are available for investigators who seek to answer important questions on health and disease in the context of

[36]https://www1.imperial.ac.uk/publichealth/
[37]https://www.iarc.fr/

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

	research projects that are consistent with the legal and ethical standard practices of IARC/WHO and the EPIC Centres (ECs).
	Means: **Tools**: A platform? **People**: domain specialist(s), data steward(s), data specialist(s). **Participating countries**: FR,UK,NL,DK,SE, ES,IT, GR, GE,
	Skills: Basic understanding of semantics, domain knowledge of use case at hand.
	Cost: Information not available
	Barriers: Information not available
	Barrier Management: Information not available
BP Actor	*Community of Practice*: Researchers and healthcare providers in a wide variety of practices, for example (but not limited to): genetics, cardiovascular disease, cancer.
	Champion: Information not available
	Owner: The EPIC study is governed by the EPIC Steering Committee (EPIC SC) and the IARC Ethics Committee (IEC), as well as the institutional review boards of the ECs. The IARC acts as custodian for the EPIC database and the majority of biospecimens (hosted by the IARC Biobank, IBB), whereas biospecimens from the Swedish and Danish centres are stored nationally. A detailed description of EPIC resources is provided in the "Study resources"[38] section.
	Training Needs: n.a.
	Acceptability: n.a.
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: n.a.
	Coherence: n.a.
	Consistency: High (makes reuse of existing knowledge, standards and vocabulary).
	Granularity: n.a.

[38] https://epic.iarc.fr/about/studyresources.php

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

	Adaptability: High (research data).
	Activity: n.a.
	Integration: High (for countries within the EPIC-consortium. (For details on data collection and standardisation (coding), epic@iarc.fr should be consulted).
BP Implementation	*Demonstration of Success: over 1700* academic publications since 1992
	Installation Time: not applicable
	Application Time: not applicable
	Experiences and feedback: Information not available.
	Measurement: Information not available.

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

Federated EGA
Title:
Reference:
Summary:
Category:
-
Topics:
-

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 personalised 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 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 organisation 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 (fEGA)
	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 (fEGA) 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
	Revision Information: Information on APIs is at

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

	https://ega-archive.org/federated
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 realise the potential of data sharing and computer readability of data for their own benefits, which is currently only partly accomplished.
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 (fEGA) 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

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

	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, run at EMBL-EBI and CRG Barcelona, is archiving data from throughout Europe and beyond. The Federated EGA (fEGA) announced the first signings of the Collaboration Agreement between national Nodes and Central EGA in September 2022, with the objective to enable discovery and access to sensitive data across national boundaries. This collaboration agreement initially includes the Finnish FEGA Node[39], the German Human Genome-Phenome Archive (GHGA)[40], the Federated EGA Norway Node[41], the Spanish FEGA (es-FEGA)[42], as well as the Swedish Sensitive Data Archive[43] – many other countries have expressed interest to join in the future.
	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, and more countries have expressed interest; further information on how to join can be found in the FEGA Onboarding Knowledge Base: https://ega-archive.github.io/FEGA-onboarding/.
	Measurement: Development of key performance indicators

Data governance, genomics data framework

Title: FAIR genomes
Reference: https://fairgenomes.org, open source code at https://github.com/fairgenomes.

[39] https://research.csc.fi/-/fega
[40] https://www.ghga.de/
[41] https://ega.elixir.no/
[42] https://fega-test.bsc.es/docs/
[43] https://fega.nbis.se/

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

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 "Personalised Medicine" project, nr. 846003201
- Use cases rare diseases and cancer

Topics:
- Demonstrator : https://fairgenomes-acc.gcc.rug.nl
- 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*: Publication (open access) available at: https://www.nature.com/articles/s41597-022-01265-x. Project URL: https://fairgenomes.org Project authors: https://fairgenomes.org/about/ Authors ORCIDs: https://orcid.org/0000-0002-7160-5942 https://orcid.org/0000-0002-0934-8375 https://orcid.org/0000-0003-1615-4197 https://orcid.org/0000-0002-1215-167X https://orcid.org/0000-0002-4706-1084 https://orcid.org/0000-0002-2440-3993 https://orcid.org/0000-0003-1301-5204 https://orcid.org/0000-0003-4450-3112 https://orcid.org/0000-0002-1073-0539 https://orcid.org/0000-0002-0979-3401
	Revision Information:

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

	This is an active BP. First release: v0.2[44]. Current release: v1.2[45]. All issues: https://github.com/fairgenomes/fairgenomes-semantic-model/issues
	Reviews Information: Summary: several rounds of revisions have taken place using a shared google sheet[46]. This sheet was then transformed to a github repository where via the issues[47] 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. Full details are available in the Methods section of the publication[48].
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 maximise 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 with 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[49]). **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), data steward(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

[44] https://github.com/fairgenomes/fairgenomes-semantic-model/tree/v0.2
[45] https://github.com/fairgenomes/fairgenomes-semantic-model/tree/v1.2
[46] https://docs.google.com/spreadsheets/d/1rnLsmE62t15jCwJfx4mCL5USYSeiXNctCA0XPcgprds/
[47] https://github.com/fairgenomes/information/issues
[48] https://www.nature.com/articles/s41597-022-01265-x#Sec8
[49] 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

	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/ harmonise the current practice The MOLGENIS platform has a FAIR Genomes implementation. This platform contains solutions to help harmonise datasets, such as the Mapping Service, to alleviate these barriers.
	Barrier Management: Procedures to follow if certain obstacles or problems are encountered 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[50])
	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 realise 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.

[50] 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

	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 standardising 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 standardised 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. FAIR Genomes semantics have been applied to enhance and extend the GA4GH Beacon v2 specification as part of MOLGENIS EMX2 'FAIR Data Hub' development (github[51]), which is a rich database template with interconnected FAIR APIs to support fast and seamless FAIRification of life science data.
	Activity: The tasks to be carried out in the BP
	Integration: The BP makes use of existing ontologies such as HPO etc. It 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. The EMX2 'FAIR Data Hub', built on FAIR Genomes, will be used to FAIRify resources including European Reference Networks patient registries, VKGL variant classifications of Dutch genome diagnostic labs, sample catalogues, and more. In EJP-RD, a combination of FAIR Genomes and a bespoke Beacon v2 specification will connect resources to each other and to the Virtual Platform for discoverability and queryability.
	Installation Time: The time it takes to introduce and implement the BP

[51] https://github.com/molgenis/molgenis-emx2

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

	in an organisation - 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.

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

Appendix II: Identified best and promising practices

This appendix lists all best practices and promising practices that have already been identified by the experts but still need to be worked on (as in getting additional/relevant information such that it can be templated and taken up in Appendix I). In each paragraph the identified best and promising practices are grouped (a separate subsection) as shown in Table 3.

The identified best and promising practices are listed with at least the following information:
Title:
Reference:
Summary:
Category:
Topics:

Date model and templates

Title: OMOP
Reference: https://www.ohdsi.org/data-standardization/the-common-data-model/
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:

Title: Portal of Medical Data Models
Reference: https://medical-data-models.org/
Summary:
MDM-Portal (Medical Data Models) is a meta-data registry for creating, analysing, 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:

Title: Maelstrom Data harmonisation guidelines
Reference: https://www.maelstrom-research.org/about-harmonization/maelstrom-guidelines

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

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

Topics:
- Iterative harmonisation steps

Title: ISO/IEC 21838-1:2021 Information technology — Top-level ontologies (TLO) — Part 1: Requirements
Reference: https://www.iso.org/standard/71954.html
Summary: ISO standard under development. Part 1 : requirements (this reference) and Part 2: Basic Formal Ontology (BFO)[52] are published. Part 3: Descriptive ontology for linguistic and cognitive engineering (DOLCE)[53] and Part 4: TUpper[54] are under development. It follows a method tested in over 300 ontology-building initiatives and is being documented in ISO 21838, leveraging existing resources wherever possible.
The Part 1 document specifies:
- required characteristics of a domain-neutral top-level ontology (TLO) that can be used in tandem with domain ontologies at lower levels to support data exchange, retrieval, discovery, integration and analysis.
- the characteristics an ontology needs to possess to support the goals of exchange, retrieval, discovery, integration and analysis of data by computer systems.

Category:
- Data harmonisation

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

Title: ISO 23903:2021 Health informatics — Interoperability and integration reference architecture — Model and framework
Reference: https://www.iso.org/standard/77337.html
Summary: This ISO document enables the advancement of interoperability from the data/information exchange paradigm to knowledge sharing at decreasing level of abstraction, starting at IT concept level (semantic coordination) through business domain concept level (agreed service function level cooperation), domain level (cross-domain cooperation) up to individual context (skills-based end-user collaboration). The document defines a model and framework for a harmonised representation of existing or intended systems with a specific focus on ICT-supported business systems. The Interoperability and Integration Reference Architecture supports ontology harmonisation or knowledge harmonisation to enable interoperability between, and integration of, systems, standards and solutions at any level of complexity without the demand for continuously adapting/revising those specifications. The approach can be used for analysing, designing, integrating, and running any type of systems. For realising advanced interoperability, flexible, scalable, business-controlled, adaptive, knowledge-based, intelligent health and social ecosystems need to follow a systems-oriented, architecture-centric, ontology-based and policy-driven approach.
The languages for representing the different views on systems such as ontology languages like Common Logic (CL) (ISO/IEC 24707[24]) and Web Ontology Language (OWL)[25] – specifically OWL 2[26] (World Wide Web Consortium (W3C®), languages for modelling and integrating business processes like Business Process Modeling Language (BPML) (OMG®), but also OMG's

[52] https://www.iso.org/standard/74572.html
[53] https://www.iso.org/standard/78927.html
[54] https://www.iso.org/standard/78928.html

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

Unified Modeling Language (UML, also specified as ISO/IEC 19505[27]) based representation styles for the different ISO/IEC 10746 (all parts) views are outside the scope of this document

Category:
- Model and framework

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

Data interoperability, data standards, ontology and controlled terminology

Title: OLS
Reference: https://www.ebi.ac.uk/ols/index
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:

Title: OxO
Reference: https://www.ebi.ac.uk/spot/oxo/index
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:

Title: BioPortal
Reference: https://bioportal.bioontology.org/
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

Title: FAIRsharing
Reference: https://fairsharing.org/
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

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

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 organisations 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.
Category:
- Data and metadata standards
- Data policies

Topics:

Data exchange standards

Title: Fast Healthcare Interoperability Resources from Health Level-7 UK (HL7) (FHIR)
Reference: https://hl7.org/fhir/
Release: v4.3.0: R4B HL7 Standard for trial Use (STU) ballote
Summary: Fast Healthcare Interoperability Resources is a next generation standards framework created by HL7. FHIR combines the best features of HL7's v2 , HL7 v3 and CDA product lines while leveraging the latest web standards and applying a tight focus on implementability. FHIR solutions are built from a set of modular components called "Resources". These resources can easily be assembled into working systems that solve real world clinical and administrative problems at a fraction of the price of existing alternatives. FHIR is suitable for use in a wide variety of contexts – mobile phone apps, cloud communications, EHR-based data sharing, server communication in large institutional healthcare providers, and much more.
Category:
- Data exchange

Topics:
- standard for health care data exchange
- software resources, guidelines and procedures

Title: My Health @ EU - eHealth Digital Service Infrastructure (eHDSI)
Reference: OPERATIONS Home - My Health @ EU - eHealth Digital Service Infrastructure (eHDSI) - EC Extranet Wiki (europa.eu)[55] (EU login required
*Summary:*The eHealth Digital Service Infrastructure (eHDSI or eHealth DSI) is the initial deployment and operation of services for cross-border health data exchange under the Connecting Europe Facility (CEF). eHDSI sets up and starts deploying the core and generic services, as defined in the CEF, for Patient Summary and ePrescription. The generic services are the necessary implementation of data exchange at country level, the core services at EU level. These together enable the provision of Cross Border eHealth Information Services (CBeHIS).
Category:
- Data exchange

Topics:
- health care data exchange
- Patient Summary
- ePrescription

[55] https://webgate.ec.europa.eu/fpfis/wikis/display/EHDSI/OPERATIONS+Home

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

Data infrastructure, data management platforms and tools

Title: CEDAR
Reference: https://metadatacenter.org/, Mark A Musen, Carol A Bean, Kei-Hoi Cheung, Michel Dumontier, Kim A Durante, Olivier Gevaert, Alejandra Gonzalez-Beltran, Purvesh Khatri, Steven H Kleinstein, Martin J O'Connor, Yannick Pouliot, Philippe Rocca-Serra, Susanna-Assunta Sansone, Jeffrey A Wiser, and the CEDAR team, The centre for expanded data annotation and retrieval, Journal of the American Medical Informatics Association, Volume 22, Issue 6, November 2015, Pages 1148–1152, https://doi.org/10.1093/jamia/ocv048
Summary: The Center for Expanded Data Annotation and Retrieval (CEDAR) was established in 2014 to create a computational ecosystem for development, evaluation, use, and refinement of biomedical metadata.
CEDAR approach centres on the use of metadata templates, which define the data elements needed to describe particular types of biomedical experiments. The templates include controlled terms and synonyms for specific data elements. CEDAR uses a library of such templates to help scientists submit annotated datasets to appropriate online data repositories.
CEDAR is an end-to-end process that enables:
community-based organisations to collaborate to create metadata templates,
investigators or curators to use the templates to define the metadata for individual experiments,
and scientists to search the metadata to access and analyse the corresponding online datasets.
Category:

- Metadata management

Topics:

- Common Data Elements CDEs
- FAIR metadata

Title: openEHR
Reference: https://www.openehr.org/
Summary:
openEHR is:

- an open standard specification in health informatics that describes the management and storage, retrieval and exchange of health data in electronic health records (EHRs). In openEHR, all health data for a person is stored in a "one lifetime", vendor-independent, person-centred EHR. The openEHR specifications include an EHR Extract specification but are otherwise not primarily concerned with the exchange of data between EHR-systems as this is the focus of other standards such as EN 13606 and HL7.
- openEHR is a non-profit organisation that publishes technical standards for an EHR platform along with domain-developed clinical models to define content.

Picture form About openEHR[56]

[56] https://www.openehr.org/about_us

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

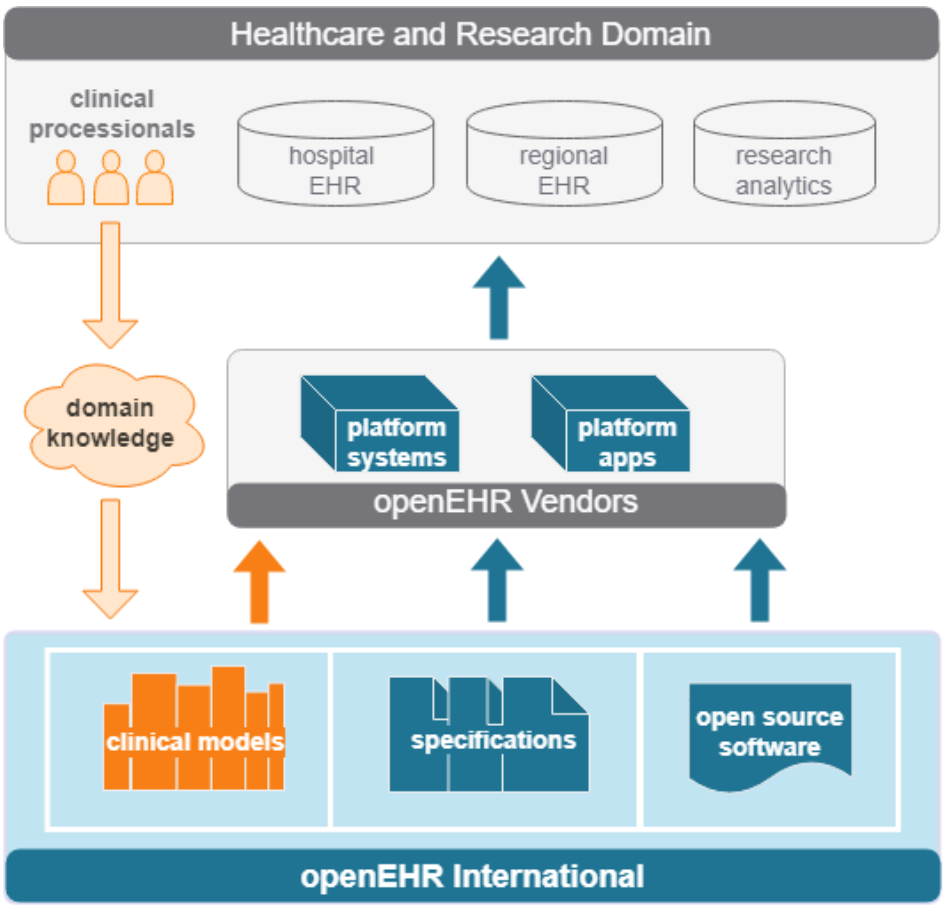

Figure 2: The principal architectural concepts include the lifelong, patient-centric shared health record, future-proof data and clinical process support**.** Picture form About openEHR[57]

Category:
- EHR platform

Topics:
- Lifelong, patient-centric shared health record, future-proof data and clinical process support.
- Open standards and open source software
- Standardised formats and processes, data and software platform

Title: REDCap
Reference: https://redcap.vanderbilt.edu/consortium/library/search.php
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:

Title: data.europa.eu

[57] https://www.openehr.org/about_us

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

Reference: https://data.europa.eu/en
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.
- **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.

Title: RD-Connect Genome-Phenome Analysis Platform (GPAP)
Reference: https://platform.rd-connect.eu
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/). Testing implementation of Beacon 2. Enhanced data discovery enabled within the system for authorised users.

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

- **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. Local instances have been 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/

Title: Genomics England PanelApp
Reference: https://panelapp.genomicsengland.co.uk/
Summary: Genomics England 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:

Title: GA4GH Beacon
Reference: https://beacon-project.io/
Summary: The GA4GH Beacon is an API (sometimes extended with a user interface) that allows for data discovery of genomic data without revealing or accessing the actual data by enabling queries on data stored at data repositories that deploy the Beacon API . Especially the new v2 APIs now allow extending queries to phenotypes, regions.
Category:
- Data discoverability
Topics:

Data governance, genomics data framework

Title: GA4GH
Reference: https://www.ga4gh.org/
Summary: The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organisation, 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:

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

- 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 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).
- 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/
- Genomics in Health Implementation Forum (part of GA4GH): https://www.ga4gh.org/community/ghif/

Title: EUCANCan
Reference: https://eucancan.com/
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 Personalised Medicine.
EUCANCan proposes to create the EUropean-CANadian Cancer network (EUCANCan), a federated infrastructure whose mission is to enable Personalised Medicine in Oncology by promoting the generation and sharing of harmonised 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, harmonisation 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 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:

Title: CINECA
Reference: https://www.cineca-project.eu/

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

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: GECKO (Genomics Cohorts Knowledge Ontology), see http://www.obofoundry.org/ontology/gecko.html

Title: IHCC
Reference: https://ihccglobal.org/
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/

Title: Orphanet
Reference: https://www.orpha.net/consor/cgi-bin/index.php?lng=EN
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:

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

Appendix III: National implementation of standards

This appendix lists per country (in alphabetical order) all information about the level of implementation of (inter)national standards within specific domains or use cases.

Overview_tables_per_country.xlsx

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