European Genomic Data Infrastructure

Deliverable D7.4

Report to identify the initial set of relevant data for use cases

Project Title Grant agreement no	**Genomic Data Infrastructure** Grant agreement 101081813		
Project Acronym (EC Call)	GDI		
WP No & Title	WP7: GDI use cases		
WP Leaders	Alfonso Valencia (BSC) Salvador Capella-Gutierrez (BSC) Marc Van Den Bulcke (SC) Oliver Stegle (DKFZ)		
Deliverable Lead Beneficiary	BSC		
Contractual delivery date	31/01/2024	**Actual delivery date**	26/03/2024
Delayed	Yes		
Partner(s) contributing to deliverable	BSC, Erasmus MC, UT, UH, HSR, ELIXIR Hub, UL, HRI		
Authors	Laura Portell-Silva (BSC) Salvador Capella-Gutierrez (BSC) Jeroen van Rooij (Erasmus MC) Helen Ray-Jones (Erasmus MC) Andre Uitterlinden (Erasmus MC) Priit Kleeman (UT) Andreas Scherer (UH) Marco Morelli (HSR)		

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

Contributors	Giovanni Tonon (HSR) Giselle Kerry (ELIXIR Hub) John Hancock (UL)
Reviewers	Jeroen Beliën (HRI) Marc Van Den Bulcke (SC)

Log of changes

Date	Mvm	Who	Description
20/12/2023	0v1	Laura Portell-Silva (BSC)	First draft sent to WP7
09/02/2024	0v2	Laura Portell-Silva (BSC)	Added input and comments from WP7 members
15/02/2024	0v3	Laura Portell-Silva (BSC)	Sent to reviewers
27/02/2024	0v4	Laura Portell-Silva (BSC)	Updated according reviewers and sent to Coordination
26/03/2024	1VO	Mercedes Rothschild Steiner (ELIXIR Hub)	Final version submitted to the EC portal

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

Contents

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

1. Executive Summary

Establishing minimal datasets and standards within GDI facilitates seamless data exchange, driving innovation in healthcare and improving patient outcomes. Deliverable 7.4 encapsulates the collaborative efforts of GDI, 1+MG, and B1MG initiatives towards establishing minimal datasets and standards for genomic data exchange and integration.

Efforts focused on leveraging prior work from 1+MG and B1MG initiatives to define minimal datasets for GDI use cases, particularly in Cancer, Infectious Diseases and the Genome of Europe. Collaborative refinement processes ensured comprehensive standards adherence.

Substantial progress has been made, including the successful submission of the Genome of Europe proposal, advancements in defining minimal datasets for Infectious Diseases and Cancer, and the development of prototypes for Genome-wide association studies (GWAS).

Challenges such as accommodating diverse dataset standards and clarifying basic data standards within GDI require ongoing dialogue and collaboration. Balancing standard adoption with cost considerations remains crucial.

Lastly, standardisation, interoperability, and collaborative frameworks positions are a priority in GDI and efforts towards this goal will continue.

2. Contribution towards project outcomes

With this deliverable, the project has reached or the deliverable has contributed to the following project outcomes:

[Select 'Yes' (at least one) if the deliverable contributed to the key result, otherwise select 'No'. For more details of project outcomes, see here]

	Contributed
Outcome 1 Secure federated infrastructure and data governance needed to enable sustainable and secure cross border linkage of genomic data sets in compliance with the relevant and agreed legal, ethical, quality and interoperability requirements and standards based on the progress achieved by the 1+MG initiative.	**Yes**
Outcome 2 Platform performing distributed analysis of genetic/genomic data and any linked clinical/phenotypic information; it should be based on the principle of federated access to data sources, include a federated/multi party authorisation and authentication system, and enable application of appropriate secure multi-party and/or high-end computing, AI and simulation techniques and resources.	**Yes**
Outcome 3 Clear description of the roles and responsibilities related to personal data and privacy protection, for humans and computers, applicable during project lifetime and after its finalisation.	**No**
Outcome 4 Business model including an uptake strategy explaining the motivation, patient incentives and conditions for all stakeholders at the different levels (national, European, global) to support the GDI towards its sustainability, including data controllers, patients, citizens, data users, service providers (e.g., IT and biotech companies), healthcare systems and public authorities at large.	**No**

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

Outcome 5 Sustained coordination mechanism for the GDI and for the GoE multi-country project launched in the context of the 1+MG initiative.	**Yes**
Outcome 6 Communication strategy – to be designed and implemented at the European and national levels.	**No**
Outcome 7 Capacity building measures necessary to ensure the establishment, sustainable operation, and successful uptake of the infrastructure.	**No**
Outcome 8 Financial support to the relevant stakeholders to enable extension, upgrade, creation and/or physical connection of further data sources beyond the project consortium or to implement the communication strategy and for capacity-building.	**No**

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

3. Methods

The aim of this deliverable is to include a report that outlines the initial set of data types relevant for the GDI use cases. To accomplish this, we have based the preceding efforts on the work done in the 1+MG and B1MG initiatives, specifically focusing on the identification of minimal datasets. This crucial groundwork lays the foundation for initiating the identification of data within the GDI context.

Before GDI, a collaborative effort, between the B1MG project and the 1+MG Working Groups (WGs), defined minimal datasets tailored to specific use cases, that included “Rare Diseases” and “Cancer” domains. The intention is to leverage this foundational work in GDI to ensure harmonisation and alignment with the GDI WP7 use-cases.

The development process adhered to comprehensive standards, engaging B1MG use-cases within 1+MG working groups, fostering a collaborative refinement process for the creation of agreed-upon minimal datasets. The developed datasets include the minimal dataset for:

- Rare Disease[1], comprising 16 agreed-upon Common Data Elements (CDEs).
- Cancer[2], encompassing 140 (37 mandatory, 40 recommended, and 63 optional) items organised into eight conceptual domains for the collection of cancer-related clinical information and genomics metadata.
- The Genome of Europe consists of three mandatory items that are biological sex at birth, age (at the time of retrieving samples for sequencing), and ancestry (most likely determined by country of origin, or equivalent).
- Infectious Diseases is a work in progress (currently COVID use case; being extended to the broader infectious disease domain).

The nature of this work continues within GDI and in collaboration with the 1+MG relevant WGs, where the intention is to validate and adapt the existing datasets, emphasising their applicability to the Genome of Europe initiative and the intricacies of Cancer and Infectious Diseases within the GDI framework. In addition, in the context of GDI, the data collected will use these minimal datasets to ensure coherent harmonisation. This ongoing collaboration ensures that the groundwork laid by the 1+MG initiative integrates into GDI and enhances the evolving landscape of genomic data interoperability.

[1] https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en
[2] https://doi.org/10.5281/zenodo.8239363

4. Description of work accomplished

As outlined in the Methods section, the ongoing effort to enhance minimal datasets persists within GDI. The focus is on validating and adapting the current datasets, with a particular emphasis on their relevance to the Genome of Europe initiative and the complexities associated with "Cancer" and "Infectious Diseases" within the GDI framework.

4.1 Genome of Europe

The Genome of Europe consortium as part of GDI task 7.1 and the WG12 of 1+MG meets monthly to discuss and plan for any issues related to the Genome of Europe. A large part of these discussions last year were related to drafting a proposal for the Digital Europe - Genome of Europe call (DIGITAL-2023-CLOUD-AI-04-GENOME), which opened in March of 2023 and was submitted successfully on November 22 of 2023. The proposal aims to collect the first 100.000 Whole Genome Sequencing (WGS) samples as part of the Genome of Europe, prioritising the drafting of samples from existing datasets, sequencing from existing cohorts and biobanks, or finally recruiting novel samples from scratch where needed. The final proposal contained 29 partner countries (26EU, 3 nonEU), 51 institutes, 200 scientists, and held a budget of 45M, of which 20M would be funded by the EU call, and 25M is matched by the partner states. The call text, and the proposal discuss aspects of the data collection, processing, storage and uptake into GDI solutions. The proposal was written by the Genome of Europe WG12, with input from experts for the various specific tasks, including members from GDI pillars. If the proposal is funded, we expect that it would start around July of 2024, and run for 42 months. The proposal contains 7 work packages, of which WP4 was tasked with "Data Infrastructure", including the alignment with GDI. A number of decisions were discussed during the proposal writing, which we briefly outline here:

- Minimal phenotypic data required for a sample in Genome of Europe is age, biological sex on birth and ancestry (most likely determined by country of origin, or equivalent), following the 1+MG framework.
- The data infrastructure services are operated by the prospective 1+MG nodes currently under establishment in the member states.
- WP4 in Genome of Europe will coordinate with GDI and oversee the flow of Genome of Europe datasets across the 1+MG lifecycle (aligned with the TEHDAS action and utilising 1+MG GDI infrastructure), and provide the data management/stewardship knowledge and capacity.
- GDI will provide the data discovery and access services, and the capacity to analyse the data. Involvement of Genome of Europe WP1 (coordination) and WP4 with GDI Pillar III will allow GDI to monitor and consider any new requirements from Genome of Europe, and make recommendations to Pillar II for upgrades.

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

- Basic use cases of Genome of Europe will be implemented directly in GDI as part of validation of the 1+MG infrastructure, while the most advanced use cases will be scientifically led by Genome of Europe with GID supporting and implementing the technical requirements.

A number of topics were identified which need further deliberation, we outline a non-exhaustive list here:

- Additional data will be collected on the Genome of Europe samples regarding their collection and processing, including elements like the sequencing platform (Illumina, Nanopore) and coverage, among others. This will be done in alignment with the 1+MG framework, but needs to be detailed out for Genome of Europe.
- Exact definition of what "data" means needs to be determined and agreed upon as well. For the Genome of Europe proposal, distinction was made on "raw individual-level data" (to mean FASTQ and BAM files, or equivalent), "processed individual-level data" (to mean vcf, or equivalent), "summary-level or aggregated data" (such as variant allele frequencies in groups of samples, in flexible format such as txt or equivalent). Simply speaking, raw data provides all functionality, but triggers legal and logistical challenges, whereas summarised data provides limited functionality, but may trigger fewer barriers.
- Following on from the previous item, the use cases and pilots within the Genome of Europe are multiple, and designed to request differing requirements of the level of data and subsequent infrastructure available. These can be flexible (to some extent) and should be worked out further. This includes defining typical "uses" of such data and how to handle those in the future.

4.2 Cancer Research

The groups of experts from GDI Task 7.4 and 1+MG WG9 meet every first Monday of the month to discuss advancements of the project, often involving external experts. The main deliverables that were produced in the last year were:

- A minimal dataset for clinical data in cancer, in collaboration with 1+MG WG3: members of the working group were involved and existing models were surveyed, to result in a data model with three tiers of variables: mandatory, recommended, and optional. The values of the fields are chosen from appropriate dictionaries and the use of free text is limited as much as possible.
- 3 clinical use cases for cancer (melanoma, non-small cell lung cancer, and chronic myeloid leukaemia): the group imagined real-life situations in the daily clinical practice, where a physician must deal with a patient with specific genetic mutations conferring, for example, resistance to therapy. By producing a synthetic genetic dataset, together with the corresponding clinical data, a query to a GDI-like infrastructure was imagined, to be passed to

people in GDI Pillar II - 1+MG WG5. These use cases can be used to produce a proof of concept similar to the one presented for the use case rare disease by B1MG.

- With the same partners, a thorough analysis of the existing Starter Kit (and before of the B1MG rare disease PoC) to identify the elements that need to be changed to account for the high complexity of the cancer use cases.

Currently, the working group is identifying possible ways to support the GDI platform. Specifically, as a group, it is framing the use cases from the viewpoint of the final oncology user, with the mission to ensure that the basic needs of the community are covered and the most common questions in oncology addressed.

4.3 Infectious Diseases

The GDI Task 7.3 and 1+MG WG11 experts meet every last Tuesday of the month to discuss their tasks and topic of their areas of focus. Several activities are currently happening:

- A core team of five experts is developing a minimal dataset for infectious diseases. This team is drawing upon input from the Cancer working group and existing publicly available materials. Progress will be shared with the larger group as the project moves on. Also, surveys will be conducted within GDI Task 7.3, Pillar III, and Pillar II to seek consensus on various aspects of the minimal dataset for infectious diseases.
- Efforts are also underway to disseminate collected information. Initial contact has been made with the editors of the Infectious Diseases Toolkit (IDTk) to understand their data requirements and upload procedures. Collaboration with an IDTk member who is also part of GDI Task 7.3 will be valuable in this regard.
- Discussions about the implementation of two use cases developed by the WG11 focus group are happening. These cases involve Covid-19 healthcare and research scenarios, including the creation of a synthetic human genomics dataset with inserted variants that have been published to be present in individuals with severe symptoms of Covid-19. Analysis schemes and a minimal core dataset have been developed by WG11 members from Germany and Norway that are in line with these use cases. The next step involves determining how these datasets and methods can be implemented in a federated analysis scheme.
- Development of a minimal dataset for infectious diseases (MDID), extending beyond Covid-19. A core team is compiling ontologies for key concept areas, such as data submitter, host, pathogen, host sample sequencing, pathogen sequencing/characterization, treatment and environmental variables, to establish standards for data collection. The purpose of this MDID is to serve as a reference scheme that provides guidance for observational data collection based on specific standards, thereby ensuring uniform and harmonised structure and content of the modelled databases and facilitating data analysis. Agreement on a MDID collected

from infectious disease patients is a prerequisite to achieving compatibility and interoperability.

4.4 Building on the experience of the 1+MG/B1MG use-cases

Aside from the three clear use-cases mentioned before, GDI also includes a task to build on the experience of the 1+MG/B1MG use-cases. The work in this task started on the discussion of the minimal dataset based on 2 generic prototypical questions:

1. Why do individuals with certain disease-specific genes not develop the disease?
2. Why do some gene variants cause adverse side effects for medications?

In the discussions, the 1+MG/B1MG deliverables were reviewed. The following standards/terminologies were preliminarily recommended by B1MG-WP3/1+MG WG3 experts[3,4]:

Table 1. Standards used depending on the domain.

Domain	Standard/terminology
Cancer	SNOMED CT (preferred) or ICD10-O
Rare diseases	ORPHAcodes (part of Orphanet)
Phenotypic abnormalities	HPO
Common and complex diseases	Primarily ICD-10, transitioning to ICD-11, and SNOMED CT
Direct and indirect cause of death	ICD-10, transitioning to ICD-11
Cardiovascular diseases or comorbidities	SNOMED CT
Capturing medicinal data	ISO IDMP[5] is recommended

It is important to take into account that for the SNOMED CT standards, a licence is needed. The European Commission currently contributes 60% of the annual base licence fee.

[3] https://zenodo.org/records/10058688

[4] https://framework.onemilliongenomes.eu/data-models-ontologies

[5] https://www.ema.europa.eu/en/human-regulatory-overview/research-and-development/data-medicines-iso-idmp-standards-overview

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

For exposure ascertainment, the following validated questionnaires are recommended (but are not limited to):

- Quality of Life: SF-12, SF-36 or EORTC-QLQ-C30
- PROMS/PREMS
- Smoking: GATS, lifetime smoking status, and pack-years
- Physical activity: IPAQ
- Obesity: BMI and waist circumference

See also ICHOM[6] for patient-centred outcome measures.

We emphasise adherence to the interoperability framework as per the Global Alliance for Genomics and Health (GA4GH) standards, particularly concerning APIs, data use conditions (ADA-M, Data Use Ontology), and phenopackets[7] for standardised phenotype data sharing.

In addition to standards, application-specific ontologies exist, incorporating definitions and terms from existing standards while adding specific missing elements. Examples include FAIR genomes for the human genome and a semantic version of phenopackets[8] that incorporates the ontology standards mentioned above and increases interoperability with other FAIR resources.

Promising minimal dataset schemas to be used in GDI are the FAIR genomes metadata schema[9] and the FAIR genome semantic model[10]. The FAIR genomes project is a ZON-MW funded national coordination action that aims to gather currently fragmented guidelines and tools to increase FAIRness (Findability, Accessibility, Interoperability and Reusability) of DNA data. This work includes data from all types of DNA laboratories (rare disease, cancer, research, etc), patients/participants organisations, and has extensive collaborations with (inter)national initiatives, in alignment with Dutch and international organisations BBMRI, ELIXIR, X-omics, Solve-RD, EJP-RD, GA4GH, B+1MG.

Proposed minimal datasets are using FAIR genome semantics and B+1MG recommendations.

[6] https://www.ichom.org/patient-centered-outcome-measures/
[7] https://github.com/phenopackets
[8] https://github.com/LUMC-BioSemantics/phenopackets-rdf-schema
[9] https://pubmed.ncbi.nlm.nih.gov/35418585/
[10] https://github.com/fairgenomes/fairgenomes-semantic-model/blob/main/generated/markdown/fairgenomes-semantic-model.md

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

Following the Barcelona meeting in February 2024, we propose combining the preliminary prototypical questions to formulate a new question that addresses the main expectation for GDI and covers the main point of both previous prototypical questions:

A user can do the Genome-wide association studies (GWAS) and Polygenic Risk Score (PRS) analyses across federated nodes, phenotypic datasets are changeable and the user can use most relevant GWAS and PRS tools and methods.

The proposed datasets serve as prototypes to demonstrate the feasibility of conducting Genome-wide association studies (GWAS) using federated datasets. Currently, GWAS is conducted in diverse ways, with PLINK and R (SAIGE) being the most common entry-level tools. As research questions evolve, so do the requisite datasets. Data undergoes additions and removals over its lifetime as gene donor consents are granted or withdrawn, and research focuses shift.

GWAS effectively identifies causal SNPs in diseased individuals. Diseases like Alzheimer's and Parkinson's, characterised by defects in multiple genes, have been extensively studied using GWAS. Thus, GWAS can be used to know why some individuals with defective genes do not manifest the associated diseases.

GWAS holds significant importance in pharmacogenomics due to the increasing availability of genotype data linked with drug-response phenotypes. This integration enables GWAS to uncover genetic determinants of drug response, identifying associations between genetic variants and both drug efficacy and adverse reactions.

5. Results

The initiatives undertaken by the various expert groups within GDI and 1+MG have reached significant results:

- Successful submission of the Genome of Europe proposal, laying the groundwork for extensive genomic data collection across Europe.
- Progress in defining minimal datasets for infectious diseases[11] and implementing use cases for Covid-19 healthcare and research.
- Development of a minimal dataset for clinical data in cancer and identification of clinical use cases for specific cancer types. A preprint[12] describing this work was written, and a shorter version has been submitted to a published journal.
- Analysis of existing infrastructure and identification of necessary adaptations to support complex cancer research initiatives.

[11] Worksheets Minimal Dataset Infectious Diseases.xlsx

[12] https://www.biorxiv.org/content/10.1101/2023.10.07.561259v1

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

- A minimal dataset that serves as prototypes to demonstrate the feasibility of conducting GWAS using federated datasets is a work in progress and can be found in the Annex.[13]

[13] X Mappping_WGs_minimal_datasets.xlsx

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

6. Discussion

Building on the experience of the GDI use-cases, efforts to establish standards and frameworks for data interoperability emerge as a central theme. The recommendations for standards such as SNOMED CT and ISO IDMP reflect a commitment to promoting data harmonisation and accessibility. Additionally, initiatives like the FAIR genomes project underscore the importance of promoting FAIR principles across genomic data repositories.

A collaborative effort to harmonise existing minimal datasets has started[14], with the overarching aim of achieving semantic unification. As part of this task, we are exploring the possibility of establishing crosswalks to facilitate interoperability and streamline data integration.

Several questions and challenges arise from the discussions, underscoring the need for ongoing dialogue and collaboration:

- Implications of Introducing FAIR genomes metadata schema: The adoption of standardised metadata schemas like FAIR genomes may require adjustments within existing infrastructure and workflows, warranting careful consideration of interoperability and compatibility issues.
- Support for multiple dataset standards: As GDI includes diverse use-cases and domains, accommodating multiple dataset standards may enhance flexibility and accessibility while ensuring compliance with regulatory requirements and best practices.
- Existence of basic data standards within GDI: Clarifying the existence and scope of basic data standards within GDI can streamline data management processes and promote consistency across projects and initiatives.
- Adoption of community defined standards: Balancing the adoption of standards with cost considerations underscores the importance of prioritising community-driven, open standards while ensuring sustainability and accessibility for all stakeholders.

Currently, it is evident that certain standards, such as the Data Use Ontology, are applicable across all use cases. However, others, like phenopackets, have thus far only received approval or validation for the rare diseases context. Notably, the current version of phenopackets does not align with the requirements of the cancer use case, thereby presenting a challenge in its implementation within that particular domain.

[14] Mappping_WGs_minimal_datasets.xlsx

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

7. Conclusions & Impact

This deliverable illustrates a concerted effort towards standardisation and interoperability within GDI. By establishing minimal datasets across various domains such as “Infectious Diseases”, “Cancer”, and the Genome of Europe, the initiative aims to harmonise data types and facilitate seamless data exchange and integration.

The submission of the Genome of Europe proposal represents a significant milestone in genomic data collection across Europe. If funded, this initiative will gather extensive WGS samples, enhancing our understanding of genetic variations within European populations.

Efforts directed towards infectious diseases and cancer research highlight the importance of tailored approaches in addressing specific healthcare challenges. The development of minimal datasets and clinical use cases underscores the potential for targeted treatments and improved patient outcomes through enhanced genomic insights.

Despite progress, challenges remain, including compatibility issues with standardised metadata schemas like FAIR genomes and ensuring alignment with regulatory requirements and best practices. Ongoing dialogue and collaboration are crucial to address emerging challenges and ensure the sustainability and accessibility of genomic data initiatives.

In summary, by prioritising standardisation, interoperability, and collaborative frameworks, these initiatives are poised to advance our understanding of human genetics, drive innovation in healthcare, and ultimately improve patient care and outcomes.

8. Next steps

Our next steps involve maintaining our collaborative efforts to harmonise the available minimal datasets, with a focus on exploring the possibility of establishing crosswalks, with the ultimate goal to reach semantic unification.

Additionally, efforts will be dedicated by the GDI use-cases to identifying examples of synthetic data that align with the recommendations outlined in this document. Later on, collaboration with Pillar II will involve testing the GDI infrastructure and starter kit components using the datasets provided by the use-cases. This approach aims to offer a holistic understanding of the operational processes within GDI.

9. Annex

Common data elements for GWAS prototype will be:

1. Personal
   - Personal identifier
   - Genotypic sex
   - Year of birth
   - Ancestry
2. Genetic (products of whole genome sequences)
   - WGS;
   - single nucleotide variants (SNVs);
   - copy number variants (CNVs);
   - phased genotype;
   - phased CNV data;
   - sequencing quality data (readme about used method, technology, processing software version, filters, reagents, chemistry);
3. Clinical data (will change over time as research evolving and biobanks can provide)
   - Molecular diagnosis gene - Gene affected by pathogenic variation that is causal for disease of the patient. Genes lookup (19202 choices of type)
   - Unobserved phenotype - Phenotypes or symptoms that were looked for but not observed, which may help in differential diagnosis or establish incomplete penetrance.
   - Phenotype - The outward appearance of the individual. In medical context, these are often the symptoms caused by a disease.
   - Molecular diagnosis other - Causal variant in HGVS notation with optional classification or free text explaining any other molecular mechanisms involved.
   - Age at last screening - Age of the patient at the moment of the most recent screening
   - Medical history - A record of a person's background regarding health, occurrence of disease events and surgical procedures.
   - Medication - ATC codes
   - Medication side effects - ICD-10 or ICD-11 code
   - Drug regimen - dosage (ATC code, package count, mg, how to use)
   - Clinical diagnoses - ICD-10 or ICD-11 code
   - Smoking - (yes/no)
   - BMI - body-mass index number
   - Dates of data collection - date
   - Lab measurements - LOINC codes
   - Education level - pharma PRS uses it as risk parameter for following drug regime
4. Access declaration:
   - Data usage permissions

- i. Open data (public usage)
  ii. Restricted to scientific research
  iii. Needs special permissions (ethics approval) to use
- Allowed to use in certain research types;
  i. No restrictions
  ii. Only medical research