

GENOMED4ALL

D2.1

Data processing, Data Management Plan and GDPR compliance report



GENOMED4ALL

Genomics for Next Generation Healthcare

D2.1

Data processing, data management plan and GDPR compliance report

Revision **v1.0**

Work package	WP2
Task	T2.1
Due date	30-06-2021
Submission date	30-09-2021
Deliverable lead	The European Institute for Innovation through Healthcare
Version	V1.0
Authors	Nathan Lea, Dipak Kalra (i-HD); Francesco Cremonesi (Datawizard)
Reviewers	Silvia Uribe Mayoral (UPM); Tommaso Foglia (Datawizard)

Abstract

Monitoring data protection compliance for an initiative like GenoMed4All can only be successful if there is a swift understanding of compliance across all partners and a clear concordance. To that end, the need for a reliable and consistent approach to understand data flows, sources, recipients and roles with regards overall responsibility (including Data Controllers, Processors and Custodians) is clear. Workpackage 2 has therefore agreed that using a rigorous Data Protection Impact Assessment (DPIA) Template is key to help map out not only data flows, but also responsibilities and the prerequisites for ensuring data processing is lawful.



This approach must balance the need to ensure rigour across the Consortium approach to data protection compliance but balance this with an appropriate scope where the partners within the Consortium who are individual legal entities are able to meet their compliance obligations. To that end the proposals and plans put forward in this deliverable are designed to empower and support partners in meeting their obligations. This approach helps to provide some regulatory compliance assurance and helps to assist all partners in achieving their local compliance requirements.

Keywords

Data protection, data management, data processing, data controller, data processor, data subject, information security, compliance, transparency, accountability



Document revision history

Version	Date	Description of change	Contributor(s)
v0.1	02-06-2021	1 st version of deliverable writing for i~HD Review	Nathan Lea (i~HD)
V0.2	04-06-2021	Final pre peer review draft after comments and contributions	Nathan Lea (i~HD); Dipak Kalra (i~HD); Francesco Cremonesi (Datawizard)
V0.3	07/06/2021	Draft completed for Peer Review	Nathan Lea (i~HD), Francesco Cremonesi (Datawizard)
V1.0	14/09/2021	Final Draft Completed after Peer Review	Tommaso Foglia (Datawizard), Silvia Uribe Mayoral (UPM), Nathan Lea (i~HD)

Disclaimer

The information, documentation and figures available in this deliverable are provided by the GENOMED4ALL project's consortium under EC grant agreement **101017549** and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Copyright notice

© GENOMED4ALL 2021-2024

Project co-funded by the European Commission in the H2020 Programme

Nature of the deliverable

R

Dissemination level

PU Public, fully open. e.g., website

CL Classified information as referred to in Commission Decision 2001/844/EC

CO Confidential to GENOMED4ALL project and Commission Services



* Deliverable types:

R: document, report (excluding periodic and final reports).

DEM: demonstrator, pilot, prototype, plan designs.

DEC: websites, patent filings, press and media actions, videos, etc.

OTHER: software, technical diagrams, etc.



Table of contents

1	Executive summary	9
2	Introduction.....	11
2.1	Key requirements.....	11
2.2	Proposed Approach	12
2.3	Tools and Engagement.....	14
2.3.1	Data Protection Impact Assessment Template	14
2.3.2	Local Partner Compliance Questionnaires.....	14
2.3.3	Information Asset Registers	14
2.3.4	Draft Template Data Sharing Agreement.....	15
3	Results to Date and Forthcoming Work	16
3.1	Data Protection Impact Assessment	16
3.2	Engagement and Developing a Data Protection Compliance Position	16
3.3	GenoMed4All Data Management Plan	17
3.3.1	Data summary	17
3.3.2	FAIR data	18
3.4	Information Asset Register	22
4	Conclusions	24
5	Annex 1 – Data Protection Impact Assessment Template	25
5.1	Introduction to Assessment process.....	25
5.2	The Assessment approach	25
5.3	Project Background/Overview	26
5.3.1	Comparison of process steps (simplified): [optional].....	26
5.4	Initial Conclusions.....	26
5.5	Compliance Checks required:.....	26
5.6	GDPR Compliance Checklist – where ‘personal data’ is processed:	27
5.7	Data Subject Rights:	28
5.8	Detailed Transparency Checklist	30
5.9	Security & Access Control Checklist	32
5.10	Information Asset Register Checklist.....	32
5.11	Appendix A – Supervisory Authority ‘High Risk’ Check.....	33
5.12	Appendix B – Broad Privacy Risk Assessment:	36
6	Annex 2.....	38
7	Annex 3.....	40
8	i~HD Data Sharing Agreement Template – GenoMed4All	40
9	Broad Articles for Inclusion	40
1.	Parties to the agreement (Name, designation, organisation, contact information)	40
2.	Context and Description of the parties.....	40
3.	Particulars of the agreement	41
4.	Definitions	41
5.	Aims of the Agreement	41
6.	Roles of the Parties (source and recipient)	41
7.	Data sources	41



8.	Data processing particulars.....	41
9.	Security, confidentiality particulars.....	42
10.	The research purposes covered by this agreement	42
11.	Arrangements for accessing and/or transferring the data	43
12.	Data quality.....	43
13.	Data enhancing and interpretation services to be provided	43
14.	Handling of analysis results	44
15.	Cooperation	44
16.	Data retention	45
17.	Data Breaches	45
18.	Demonstrating compliance - audits and inspections.....	46
19.	Disputes	46
20.	Indemnity	47
21.	Liability	47
22.	Termination and Withdrawal of consent	47
23.	Signatories	47



List of figures

No table of figures entries found.



List of tables

Table 1: Data Summary	18
Table 2: Making Data FAIR.....	18
Table 3: Making Data Openly Accessible	19
Table 4: Making Data Interoperable.....	20
Table 5: Increase Data Re-use (through clarifying licences).....	20
Table 6: Allocation of Resources	21
Table 7: Data Security.....	21
Table 8: Ethical Aspects.....	22
Table 9: Comparison of Processing Steps.....	26
Table 10: Compliance Steps Required	27
Table 11: GDPR Compliance Checklist	28
Table 12: Data Subject Rights	29
Table 13: Detailed Comparison Checklist.....	31
Table 14: Security and Access Control Checklist	32
Table 15: Information Asset Register Checklist.....	33
Table 16: 'High Risk' Assessment.....	35
Table 17: Broad Privacy Impact Assessment	37



Abbreviations

HDs	Haematological diseases
FL	Federated Learning
NN	Neural Network
MDS	Myelodysplastic Syndromes
MM	Multiple Myeloma
SCD	Sickle Cell Disease
DPIA	Data Protection Impact Assessment
IAR	Information Assets Register
AI	Artificial Intelligence
ML	Machine Learning
IRB	Independent Review Board (Ethics Committee)
FAIR	Findable, Accessible, Interoperable, Reusable (Principles)
GCP	Good Clinical Practice
ICMJE	International Committee of Medical Journal Editors



1 Executive summary

Introduction

This Deliverable summarises the proposed approach for addressing GDPR compliance across the GenoMed4All Consortium and its proposed data processing. Successful compliance with GDPR relies on the ability for individual partners to understand their requisite data flows as well as agree with other partners the mode and manner of data provision, receipt and subsequent obligations. This relies on a consistent and clear approach which must be based upon agreed tools as defined by GDPR and supervisory authorities.

For a large collaboration across multiple partners spread throughout Europe within different regulatory jurisdictions, it is clear that a single statement as to compliance and plan to ensure it across all partners is neither possible or appropriate. Individual partners who are acting as providers or recipients will likely themselves have a Data Controller role and will need to meet their own local regulatory and approvals requirements. GenoMed4All must also come to a consensus view of understanding the quality, nature and requirements for the project as a whole in order to ensure its success and assure, at least to the partners, its regulatory compliance.

Proposed Approach

Any approach must be able to support a “bottom up” view from the individual partner perspective in addition to a “top down” view that will allow the GenoMed4All Partners to jointly come to a consensus understanding and view about the data processing particulars. The key tool for understanding data flows, compliance requirements and how they will be met is provided by GDPR itself: it is called the Data Protection Impact Assessment (DPIA). This and the development of a Data Management Plan (DMP) must be informed by outreach across the partners, where we have developed a questionnaire and are planning workshops to understand the data flows. This will inform the development of a Consortium wide Data Sharing Agreement, the Template for which is reported in Annex 3.

Results

The DMP is presented based on preliminary understanding of the data requirements. The template DPIA and questionnaires are provided in Annexes 1 and 2 respectively. It is clear that whilst the broad purpose of reusing health record and omics data to train machine learning algorithms and develop risk stratification models across the disease areas requires some independent ethics review or amendment of existing permissions across partners, important points to clarify are the specifics of data sharing within the consortium and whether GenoMed4All will also be able to openly provide an anonymous data set for other researchers in line with the FAIR principles. To that end Workpackage 2 has commenced the questionnaires and developed the Template Data Sharing Agreement.

Conclusions and Next Steps After receiving preliminary results of the questionnaires, GenoMed4All has started to map the data flows and linkages required (as defined in Deliverable



3.1) and informed the development of the DPIA and Template Sharing Agreement. The next steps will be to formalise the agreement in line with the expectations of the consortium partners. Additionally, GenoMed4All can commence the drafting of a data licencing agreement for parties external to GenoMed4All to use anonymous data in the future. A more pressing matter however will be to define the agreements and governance required for onboarding data from parties external to the Consortium, where such agreements will need to work in line with the consortium agreement as well as reach the same standards as the GenoMed4All Data Sharing Agreement.



2 Introduction

This Deliverable (D2.1) describes the approach and plans that have been agreed within Workpackage 2 to promote and assist GenoMed4All Partners to achieve their local regulatory compliance for GDPR and to help the Consortium represent best practice for meeting these requirements. Successful compliance with GDPR relies on the ability for individual partners to understand their requisite data flows as well as agree with other partners the mode and manner of data provision, receipt and subsequent obligations. This relies on a consistent and clear approach which must be based upon agreed tools as defined by GDPR and supervisory authorities. Additionally, this must include planned or envisaged data processing for parties that express interest in working with GenoMed4All and contributing data, as well as any envisaged sharing of data outside the Consortium Partners.

2.1 Key requirements

For a large collaboration across multiple partners spread throughout Europe within different regulatory jurisdictions, it is clear that a single statement as to compliance and plan to ensure it across all partners is neither possible or appropriate. Even within Member State implementations of GDPR in line with statutory and regulatory authority interpretations, there are variations and a set of secondary legislations. These may be in line with selection of Legal Bases for health research, where one jurisdiction may mandate Consent whilst another will advocate Public Task for Public Authorities. There may be other secondary legislation on the use of certain kinds of data, including genetic data, and certain kinds of processing, notably machine learning and artificial intelligence.

Additionally individual partners who are acting as providers or recipients will likely themselves have a Data Controller role for data that they would provide or and / or receive and will need to meet their own local regulatory and approvals requirements. This mandates the need for a clear mapping of data flows, regulatory checks locally and an understanding of data recipient arrangements and purposes.

GDPR compliance additionally requires lawful purposes for processing of personal data. In the context of scientific research, the need for independent ethical review and a favourable opinion is required as a matter of research integrity and law across Europe. The scope of existing approvals may or may not cover GenoMed4All and in each jurisdiction, the processes and requirements for meeting a reasonable standard of ethical review may differ, at least procedurally.

Contrasting with this, GenoMed4All must also come to a consensus view of understanding the quality, nature and requirements for the project as a whole in order to ensure its success and assure its regulatory compliance both to the partners and any external parties wishing to contribute to any data sets and / or use data gathered by Genomed4All partners. A consensus view on the ethical particulars across the consortium as well as the possibility of an independent ethical advisory board that would operate in harmony with existing approvals is also highly



desirable if not mandated. A consensus view would include basic principles and core values considered essential for an ethical framework for data flows and for sharing genetic data (such as scientific value, user protection, facilitating user agency, trustworthiness, benefit and sustainability).

The overarching requirement will be to develop a governance set of agreements and codes of practice for data handling. The agreements primarily include:

- A data sharing agreement across the consortium partners
- A data processing agreement for consortium partners
- Data sharing and processing agreements for non-consortium partners to share and receive data on behalf of the consortium
- A data licensing agreement for non-consortium parties who may request access to or receipt of data

The proposed approach has been designed to explore the key elements to meet each of these requirements.

2.2 Proposed Approach

It is clear from these challenges that an approach must be able to support a “bottom up” view from the individual partner perspective in addition to a “top down” view that will allow the GenoMed4All Partners to jointly come to a consensus understanding and view about the data processing particulars. In essence, the regulatory compliance position must be ascertained jointly between partners and severally based on their own individual compliance measures that they alone are responsible for.

To that end being able to map out a global view of data flows and ensure that the Consortium can offer an agreed and authoritative treatise of the overall data flows for GenoMed4All whilst being able to clearly establish scoped boundaries of responsibility is clear. To clarify, GenoMed4All as a consortium cannot meet the regulatory requirements of individual partners, but it is very well placed to support and inform partners how they meet their individual legal responsibilities in pursuance of collective goals. This includes agreement on the umbrella data sharing and licensing agreements that meet partner expectations and needs.

The key tool for understanding data flows, compliance requirements and how they will be met is provided by GDPR itself: it is called the Data Protection Impact Assessment (DPIA) and it is a mandated tool for any Data Controller to risk assess against GDPR compliance risk for any high risk data processing activity. Where GenoMed4All cannot take a Data Controller role because it is not itself a *legal person* or entity, it can serve as a vehicle to build a consistent understanding of data protection particulars where individual partners can take a view as to whether they require running their own DPIA. In any event, GenoMed4All itself can develop, nurture and update a DPIA for reference and to support regulatory compliance, transparency and data flow planning and development for the Consortium.



In conducting an impact assessment it is always advisable to engage with experts and risk owners to best understand the data flow and protection requirements. It is usually required by ethics committees within an organisation, and in any case it's always helpful to enable them to arrive at a suitable decision. Workpackage 2 is addressing this through two means. Firstly, a questionnaire that is helping GenoMed4All understand the background compliance context and existing permissions for data flows. Secondly, to supplement answers and add additional detail by means of a workshop or workshops (or sounding boards) for each of the disease areas and collaborations across the Consortium. Essential to the engagement elements is an understanding of what data sources exist and what the expected data flows will be to service the clinical research and machine learning that is anticipated. This suggests the need for an *information assets register* that would define the partners, their role as provider and recipient and the categories of data that they would be processing. This in turn will allow the Consortium to take a view on whether partners are Data Controllers (joint or distinct) or Data Processors, and thereafter deploy the appropriate agreements in line with the requirements defined by the DPIA and questionnaire responses.

To date the materials required to manage these needs have been developed. An agreed DPIA template, questionnaire and template data sharing agreement have been developed and reported in Annex 1. The DPIA has been commenced and the questionnaires have been distributed and responses received. In August 2021 GenoMed4All held its first GDPR sounding board to better understand the data flow requirements and address the particulars of the data sharing agreement template for parties within the consortium. The sounding board identified that the governance of parties external to the consortium as well as the approach to take to define the particulars of the data sharing agreement template using wording that would be acceptable to all partners.

From a compliance perspective however it is clearly important to establish a basis by which these details would be understood. With this in mind, to trial the approach, Workpackage 2 decided to work with one of the disease area groups as a test case in order to develop a deeper dive. This is the Sickle Cell Disease (SCD) group, where the interactions with the teams involved in this area is helping to guide how we engage with the other disease areas. This was also an important step to be able to establish more explicitly the scope of what can be done by GenoMed4All at a Consortium level, and what must be covered locally by partners. GenoMed4All cannot manage site level compliance by definition, but it can help inform that local process. Workpackage 2 may also take the view that a general DPIA may need to be divided into disease areas or otherwise organised for the sake of simplicity and readability.

With this in mind Workpackage 2 have also engaged with the other two disease areas Multiple Myeloma (MM) and Myelodysplastic Syndrome (MDS) which have prompted the need to better define the nature of the data transfers (i.e. whether in some cases pseudonymised data can be shared) as well as the onboarding of parties external to GenoMed4All. This has highlighted the bounds of what the consortium can support and where it requires the specific guidance from



individual partners, including their needs to conduct a DPIA that will inform how they will agree to wording a pan-consortium set of agreements.

2.3 Tools and Engagement

Workpackage 2 agreed several tools and templates to help run the compliance processes from a global perspective.

2.3.1 Data Protection Impact Assessment Template

The DPIA template developed by partner i~HD (WP2 lead) is provided to all partners to serve as an appropriate basis upon which to conduct a consistent, project wide but partner specific, scoped impact assessment. It has the advantage of having been developed by i~HD's GDPR experts, on the basis of experience from other large consortium projects. It has been developed with health data reuse in mind. The current GenoMed4All DPIA is provided in section 3.1 where it should be noted that, as with any DPIA, it is a living document that will be updated periodically and whenever there is any significant change to processing.

2.3.2 Local Partner Compliance Questionnaires

As is clear from the DPIA template, there will need to be a significant amount of intelligence gathering and the involvement of all partners will need to be addressed. Additionally as the DPIA highlights, purposes of personal data processing must be lawful and this can be established through favourable research ethics committee opinion and through participant consent or lawful exemption from consent. The ethics and consent particulars must therefore be established by means of directly engaging with sites. Workpackage 2 has therefore developed a questionnaire as attached in Annex 2 to help establish where approvals and consents are already in place, where they may need to be amended and where they need to be sought.

In addition to this, the questionnaires can serve as a basis for gathering further details from partners around security, transparency and accountability with regards data flows. These can then be used to update the DPIA and to help partners to engage with the local requirements as the project develops. This will help form the basis of workshops

2.3.3 Information Asset Registers

Following a broad approach to incorporate Supervisory Authority templates for Information Asset Registers, the spreadsheet referred to in Section 3.4 provides a list of partners, their data holdings and their proposed roles and processing for the project. This will be used as the basis for supplying the requisite details for the DPIA, including the Data Flow Diagram and particulars around compliance checks. The Information Asset Register lists the partner name, country, whether they are engaged with external collaborators and the data sources. Scheduling details for the work of Workpackage 2 in running the generic DPIA as well as the status or Independent Review Board / Ethics Committees are also provided.



2.3.4 Draft Template Data Sharing Agreement

This is available in Annex 3 as the culmination of the work with the questionnaires, DPIA and soundboard to date.



3 Results to Date and Forthcoming Work

This section provides the results to date for the compliance checks. Section 3.1 provides the DPIA as it stands at time of submission of this Deliverable, 3.2 the Information Asset Register and 3.3 the proposed workshop and questionnaire plans for the coming months of the project. The DPIA and Information Asset Register will be completed on the basis of the questionnaire responses and workshops held in the coming months of the project.

3.1 Data Protection Impact Assessment

Please refer to the DPIA Template in Annex 1 where this template has been used to conduct the DPIA, guide the questionnaires and GDPR sound board workshop and define the Data Sharing Agreement Template. This DPIA is being used to inform individual partners of the workflows required so that they can conduct their own DPIAs and assess their compliance needs. This DPIA is a living document that has been informed by the questionnaires and GDPR Sounding Board, as well as forthcoming workshops and iterations with the partners around specific data flows and Data Sharing Agreement drafts from the Template in Annex 3.

3.2 Engagement and Developing a Data Protection Compliance Position

During the Workpackage 2 sessions, we agreed that we should engage the partners incrementally. Using the SCD VHIR project as a test case, we would be able to better understand the necessary steps to engage each disease area and project partner to ensure that Workpackage 2 achieved the coverage necessary to aid in the data protection and management compliance requirements.

To that end, the steps would include:

- ☐ Engage with the clinical and technical communities to understand the specification of clinical queries and advanced data processing (e.g. for machine learning)
- ☐ Distribute the questionnaire as provided in Appendix 1.
- ☐ Update the DPIA and Information Asset Register with responses.
- ☐ Arrange a workshop (where necessary) to better understand specific details.
- ☐ Update the DPIA and engage with partners as appropriate where they require assistance.
- ☐ Agree a pseudonymisation and anonymisation approach in line with IRB requirements and risk assessment within the DPIA.

By following these steps, data processing can be fully transparent and accountable, partners are able to handle their local compliance and regulatory approvals and update the Data Management Plan as provided in the next section.



3.3 GenoMed4All Data Management Plan

This section is based on the Data Management Plan template provided by the European Commission for Horizon 2020 projects¹.

3.3.1 Data summary

GenoMed4All focuses on the use of de-identified health and omics data to provide a basis for machine learning and development of risk stratification across several disease areas. This Data Management Plan provides a general description of the data processing activities to be undertaken across the Consortium. The responses given in this section refer to the expected generation of data sets to aid in machine learning. These will have underpinned the main research results that will be published at the end of the project, and also have the potential to be reused by other researchers for risk stratification and likely Medical Device certification should their use in care be desirable.

Aspect	Response/explanation
Purpose of the data collection/generation and its relation to the objectives of the project	Health and care data collected from patients with advanced long-term conditions from several healthcare registries and clinical sites across Europe. This data will be used for machine learning to generate artificial intelligence models that provide a basis for risk stratification across several disease areas including Sickle Cell Disease, Multiple Myeloma and Myelodysplastic Syndromes as well as other haematological disorders.
Types and formats of data generated or collected by the project	These will include medical records, research profiles, existing omics data from existing registries, generation of new omics data. Source data held by Data Source Partners in their existing repositories will need to be processed under their existing approvals; Additionally, there will be some clinical record processing to capture data within the health record. Further, there will be handling of existing Genetic Data that had been derived from Samples, and there will likely be additional processing of biological samples to generate genomics and metabolomics data. In conclusion there will be processing of existing and generation of new data. Samples are in the Biobanks and these samples will be processed for the omics data. Parties external to GenoMed4All will also contribute data in line with the standards established for Consortium partners.
Any re-use existing data and how this will be done	Source data held by Data Source Partners in their existing repositories will need to be processed under their existing approvals; Additionally, there will be some clinical record processing to capture data within the health record. Further, there will be handling of existing Genetic Data that had been derived from Samples, and there will likely be additional processing of biological samples to generate genomics and metabolomics data. In conclusion there will be processing of existing and generation of new data. Samples are in the Biobanks and these samples will be processed for the omics data. The precise processes of data sharing are being agreed incrementally, where agreements are being developed to distinguish between where pseudonymised data may be shared and where anonymous data can only be shared including pseudonymisation at source where permissible by existing approvals and local jurisdiction legal requirements and anonymisation for the machine learning.

¹ Full template available here for reference: https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx



The origin of such data	Data Sources have collected electronic healthcare records, national electronic health records and Social Security or welfare system records that feed the registries. These have been or will be combined with existing omics data derived from samples, with the possibility of additional sample processing within the biobanks. The exact combination of which data sources will vary between the data provider sites.
Expected size of the data	To be confirmed by the end of 2021
Likely users of the data	AI and ML Engineers and Developers for the purposes of developing risk stratification models. The Consortium is developing an anonymous data repository for wider access under Open Data and FAIR Principles where this is a matter for the Consortium and its Data Sources to assess and agree subject to the onboarding of parties external to the Consortium and the development of a data licensing agreement to access the anonymous data set. This is likely to be finalised between M16 and M24.

Table 1: Data Summary

3.3.2 FAIR data

3.3.2.1 Making data findable, including provisions for metadata

Aspect	Response/explanation
Are the data produced and/or used in the project discoverable, identifiable and locatable by means of a standard identification mechanism	Where possible FAIR Principles will be followed and Open Access publications will be provided. The Consortium is exploring the feasibility of developing a means whereby the data sets themselves may be made available anonymously where this is being assessed and discussed by all Partners and their assessments are likely to be completed between M16 and M24. The Risk Stratification Models may be released under open source licences (again to be agreed by partners).
What standard identification mechanism used (e.g. persistent and unique identifiers such as Digital Object Identifiers)	The Consortium will consider the use of DOI and leverage existing registry identification mechanisms as appropriate (in part for traceability internally but also in part for any open data / FAIR abiding anonymous data sets).
Is meta-data available through catalogue?	Yes this will be kept consistent with the data source metadata dictionaries and will agree a common dictionary model.
Can meta-data be used for search?	Yes
Naming conventions used	This will be agreed later in the project (where the internal Consortium processing may require a consistent naming convention as clinical variables are agreed)
Clear versioning supported?	It will be as appropriate to the primary data processing purpose for machine learning.
Additional keyword search supported?	It will be.
What metadata will be created using which standards?	The Consortium will agree the most appropriate standard depending on the feasibility of making an anonymous data set available for open, wider use.

Table 2: Making Data FAIR



3.3.2.2 Making data openly accessible

Aspect	Response/explanation
Will data be made openly available as the default?	Subject to satisfactory processes and adherence to a data licensing agreement to maintain the security and integrity of the data and the privacy of anonymous participants through robust anonymisation, we will make this available by default.
Which datasets will NOT be openly available and why?	The Consortium will not make data available if we believe, or we are advised, that it is not possible to robustly anonymise the data, because of distinctive patterns in the data due to some of the population profiles that are included, or where existing approvals expressly prohibit this, or where the expression of genetic data remains uniquely attributable
How will the data & meta-data be made accessible (e.g. by deposition in stated repository)?	Where feasible, by deposition in a chosen repository.
If known repository, what arrangements explored?	Repository not yet identified
If project-specific access, then:	In line with the particulars of the Consortium wide Data Protection Impact Assessment and where appropriate, its Data Sharing Agreements and Data Licensing Agreements.,
Data Access Committee	If an anonymous data set is feasible, a committee comprising some or all of the partners involved in the project will determine the policies for which data will be made open access as well as oversight by an external Ethical Advisory Board.
Any conditions for access (i.e. a machine-readable license)	This will be determined based on the feasibility of developing an anonymous data set.
What methods or software tools will be needed to access the data?	To be implemented later in the project however preliminary decisions point to Secure Fire Transfer Protocol Servers, 2 Factor Authentication, access controls including Roles Based access and resource controls, role definition, running of servers behind firewalls and institutional Demilitarised Zones and logging and auditing of access and activity against a list of predefined actions that are consistent across all partner sites and the central repository. This will be kept under review as the project develops.
Documentation for software	This will be provided.
Availability of software	Likely accessible by GitLab or GitHub.
Institution and researcher vetting process/procedures - describe	In line with the particulars of the Data Protection Impact Assessment where in the first instance data sharing will be assessed by the Data Sources in line with their existing requirements and thereafter once the data is assembled centrally a decision can be made regarding which group can vet.

Table 3: Making Data Openly Accessible



3.3.2.3 Making data interoperable

Aspect	Response/explanation
Are the data produced in the project interoperable	Interoperability standards will be adopted by design, including the use of HL7 FHIR, in order to harmonise the data coming from seven different pilot site. This is necessary for the project itself, for the data analytics work that will be undertaken, but also serves the benefit that any research data we make openly available later will also be standardized should it prove feasible to make a repository available for further research use.
If not, explain why not	N/A
Data and metadata vocabularies, standards or methodologies used	Existing registries' metadata will be used where possible though an overall catalogue should be in line with the existing registries' choices. Standards and state of the art tools such as ERDRI.mdr will be used.
Standard vocabularies used	The vocabularies in use from existing source data are being used and a decision on whether and what standard vocabulary will be used for the project will be finalised most likely towards M24.
Mappings from uncommon or project-specific ontologies or vocabularies to more commonly used ontologies	We do not anticipate the need to adopt uncommon ontologies or vocabularies.

Table 4: Making Data Interoperable

3.3.2.4 Increase data re-use (through clarifying licences)

Aspect	Response/explanation
Will data be available for onward data-sharing/re-use?	Yes – subject to anonymity and adherence to a Data Licensing, Sharing or Processing Agreement
Approach to data licensing for onward use	As encapsulated in a Data Licencing Agreement that will be subject to a per use review by Consortium Partners
Likely date for data availability for onward use	Uncertain – likely towards the end of 2024
Explain any restriction on date of availability	None is anticipated but feasibility must be checked.
Possible restrictions on onward data-sharing	To be determined later in the project, but it will aim to facilitate onward sharing where possible and appropriate
Data retention policy (including availability for data-sharing)	Will follow data retention policies as determined at sites. Evaluation/open access data will be retained according to Consortium feasibility checks and wider EC guidance.
Description of data quality assurance processes	This will be carried out at source when data is submitted from data source sites and then confirmed on receipt within the centralised repository.

Table 5: Increase Data Re-use (through clarifying licences)



3.3.2.5 Allocation of resources

Aspect	Response/explanation
Estimated project costs for making data FAIR	This will be determined later once the feasibility of sharing is confirmed by the DPIA and Compliance processes and will be considered as part of the feasibility of developing an accessible data set.
Data management responsibility across the project	The co-ordinator will take lead responsibility for this, but other technical partners will support.
Resources required for long term preservation (costs and potential value, who decides and how what data will be kept and for how long)	To be determined later in the project, under the responsibility of the co-ordinator. This would include a sustainability business model for data sets that will be made available open access, covering costs of long-term storage, the maintenance of the data sets if necessary, and any human resources required to manage data sharing assuming the development of this data set is feasible.

Table 6: Allocation of Resources

3.3.2.6 Data security

Aspect	Response/explanation
Data security measures used (including data recovery as well as secure storage and transfer of sensitive data)	A number of information governance and data protection and information security instruments are being used in this and forthcoming project deliverables.
Where data will safely be stored (in certified repositories for long-term preservation and curation). Provide detail	This will be under the responsibility of the co-ordinator, where feasible. For the existing modelling and risk stratification work within the Consortium, the security and certification requirements are being assessed as part of the processes described in Deliverable 2.1.

Table 7: Data Security

3.3.2.7 Ethical aspects

Aspect	Response/explanation
Any ethical or legal issues that can have an impact on data sharing	As part of the feasibility of preparing an anonymous data set. there are potentially local research ethics / independent review board prohibitions and data protection issues which we will examine carefully before determining which data items and on which population profiles can be made available as open research data. The data processing for the primary goals of developing risk stratification models within the Consortium



References to ethics deliverables and ethics chapter in the Description of the Action (DoA) – if relevant	D2.1, D2.2, D2.5, D2.6, D2.7, D2.8
Questionnaires dealing with personal data	N/A – only internally for the Consortium, otherwise patients and participants will not be approached.
How is informed consent for data sharing and long term preservation sought in such questionnaires?	N/A

Table 8: Ethical Aspects

3.4 Information Asset Register

This section provides the headings that have so far been defined to account for each of the partners' contributions, data flows and sample transfers across the three disease areas. These headings will be included in a table will be populated as the questionnaires and workshops commence.

The importance of the Information Assets register is to be able to establish the types of data being used for the project and what each of the parties within the project are doing with those data. This could include being a data source providing data, a recipient processing data or indeed both.

This register helps to link data items, activities and responsibilities to the agreements, approvals and oversight requirements (including entry into the DPIA). This way GenoMed4All can account for not only the continued update of proposed data flows, but also isolate where there may be delays or uncertainties around data transfers.

The Asset Register is a tabular representation that includes the following items:

- Disease Area (SCD, MDS or MM)
- Party Name
- Nationality
- External Party?
- Contact Person
- Data Type to be Processed
- Data Source
- IRB approvals in place?
- Questionnaire Completed?
- DPIA Status
- Contact in WP2
- Sample Transfers Included?
- Data Transfers Underway?



- Number of Participants
- Governing Agreement (where applicable)

To illustrate, an entry might look as follows:

Disease Area: SCD

Party Name: VHIR

Nationality: ES

External Party? No

Contact Person: xxx

Data Type to be Processed: Demographic (including Gender, Year of Birth); labs; Oxygen Scan; Clinical; Genome Wide Association Study

Data Source: secondary use of RADeep

IRB Approval in place? working on protocol for submission

Questionnaire completed? Yes (helped develop questionnaire)

DPIA Status: noted in Consortium DPIA, conducting local DPIA

Contact in WP2: xxx

Sample Transfers Included? genetic (blood) transferred to third party for sequencing

Data Transfers Underway? all uploaded to RADeep

Number of Participants: (estimated) 100

Governing Agreement (where applicable): not applicable.



4 Conclusions

This Deliverable has described the approach being taken by Workpackage 2 to understand, map and guide the expression of data flows, management and compliance requirements and help the partners achieve their own local compliance. The outputs of this work are to develop a data management plan in the Horizon 2020 Template, aid in the partners' approvals processes and regulatory compliance and to keep a track of where each of the partners are in terms of progress with their compliance and approvals. Additionally where advice is needed for elements like information security or uncertainty over appropriate data flows to achieve the goals of GenoMed4All, the details learned through the questionnaires, engagements and impact assessments are designed to help clarify these and address them.

A wider contribution of this work and deliverable is to make sure that scope and responsibility around data processing is both respected and well defined. Through this method, the Consortium can be assured that no regulatory compliance measures are falling between jurisdictions and responsibilities. Through a collegiate approach, challenges and uncertainties can be addressed proactively and with a harmonised view on processing particulars and data protection requirements.

The next steps will be to continue to analyse the questionnaires, update the DPIA through ongoing GDPR Soundboards and finalise the Consortium Wide Data Sharing Agreement, specify the governance Agreements that are required for data contributors outside the Consortium and develop the Data Licencing Agreements for external parties wishing to use anonymous data.



5 Annex 1 – Data Protection Impact Assessment Template

Data Protection Impact Assessment Checklist – GENOMED4ALL

5.1 Introduction to Assessment process

Under the General Data Protection Regulation (GDPR), a Data Protection Impact Assessment (DPIA) is only required where proposed data processing is “likely to result in a high risk to the rights and freedoms of natural persons” (Article 35(1)). However, Article 35(3) explicitly requires one where there is ‘large-scale’ processing of ‘special category’ (e.g. healthcare) data then a DPIA is required.

One other possibility is that the data being processed is already anonymised (see Recital 26) so falls outside GDPR altogether so that no DPIA is actually required.

However, good project management and information governance suggests that there should be a general approach to risk assessment for any project or business enterprise – if only to determine whether a DPIA might be required.

Ideally, one should work from a simple initial Checklist (this document) which identifies possible areas of information risk and compliance requirements to a ‘discussion note’ which explores any issues in more depth and may help identify the necessary mitigation methods and mechanisms to offset most if not all risks. Only if risks are unmitigated or remain ‘high’ would you move to a formal DPIA report.

5.2 The Assessment approach

There should be an overview of the proposed project or business change to explain what processing is envisaged as well as the purpose and intended outcome. The ‘purpose’ is important to establish the legal basis for the processing as well as ensuring that any possible mitigations or counter-measures do not undermine the main rationale for the processing.

The next step is to establish what compliance requirements may apply: GDPR, contractual or other regulatory restrictions, consent requirements, or obligations to preserve the data for legal or other reasons (including the benefit of posterity perhaps).

Once the precise range of obligations has been established, then appropriate checks can be made and recorded within the document.

The most obvious of these being GDPR compliance. There must be a ‘High Risk’ assessment (Appendix A) to determine whether the supervisory authority needs to be informed – generally, it is expected that it will not be necessary; if so, then a formal DPIA report will be needed.

Appendix B has a broader Privacy Impact Assessment that may throw up some broader issues. Initial conclusions as to next steps or particular countermeasures to be considered should be detailed below.



5.3 Project Background/Overview

[Explain business background, including any existing processes and procedures; outline the project including stages, deliverables, and timelines]

5.3.1 Comparison of process steps (simplified): [optional]

This allows identification of what processing is new or changed through the project:

Step	Current	Proposed
Project initiation, including any Independent Review Board approval, up to Task Order from client		No change

Table 9: Comparison of Processing Steps

5.4 Initial Conclusions

Describe initial conclusions around the business continuity, special data protection measures and potential risks against benefits here (even if tentative).

5.5 Compliance Checks required:

Tick	Requirement	Notes [replace guide text with response]
<input type="checkbox"/>	Does the project involve processing 'personal data' of any sort?	Note: not just patient data; may need clear assessment of any anonymization to establish outside GDPR
<input type="checkbox"/>	Does the project involve processing 'confidential data' of any sort?	Note: may be 'commercial in confidence', medical confidentiality, or organisational confidentiality (internally sensitive); may need to check contractual limitations
Data Availability requirements		
<input type="checkbox"/>	Does data need to be held for Good Clinical Practice compliance?	
<input type="checkbox"/>	Does data need to be held to meet 'Open Data' requirements?	



Tick	Requirement	Notes [replace guide text with response]
<input type="checkbox"/>	Does data need to be held to meet ICMJE requirements or commitments?	

Table 10: Compliance Steps Required

5.6 GDPR Compliance Checklist – where ‘personal data’ is processed:

Tick	Requirement	Notes [replace guide text with response]
Article 5: Principles compliance checks		
<input type="checkbox"/>	a) Is processing lawful, fair, and transparent?	
<input type="checkbox"/>	b) Is the purpose (or purposes) of the processing clearly defined	['purpose limitation' so should cover any subsequent or later processing]
<input type="checkbox"/>	c) adequate, relevant and limited to what is necessary	['data minimisation']
<input type="checkbox"/>	d) accurate and, where necessary, kept up to date	
<input type="checkbox"/>	e) kept and permits identification of data subjects for no longer than is necessary	['storage limitation']
<input type="checkbox"/>	f) processed securely	
<input type="checkbox"/>	2) can you demonstrate this compliance?	['accountability']
Articles 13 & 14 compliance		[See detailed Transparency Checklist below]
<input type="checkbox"/>	Did the data come from publicly accessible sources?	[if so then transparency requirements may be reduced, but need to ensure data is accurate & up-to-date]
<input type="checkbox"/>	Are data subjects informed before processing starts for any new purpose if incompatible with original purpose where the controller wants to use data for a different purpose to the purpose for which they currently hold data	
<input type="checkbox"/>	Does the Privacy Notice and/or PIL cover this processing?	
<input type="checkbox"/>	What patient choices are available? Are these explained?	[see also Data Subject Rights below]
Articles 6 and 9: legal bases		
<input type="checkbox"/>	What are legal bases under Article 6	
<input type="checkbox"/>	What are legal bases under Article 9 (if 'special category' data)	
<input type="checkbox"/>	Are Article 6 legitimate interests explained where relevant?	[Complete an LIA form]
<input type="checkbox"/>	Are details of statutory obligations for Article 6 explained where relevant.	[Quote statutes or regulation]



Tick	Requirement	Notes [replace guide text with response]
<input type="checkbox"/>	Is this proposed processing compatible with the declared purposes?	[Check against any privacy notices and public information]
Article 89(1) research exemption		
<input type="checkbox"/>	If for research, do we meet Art 89(1) data minimisation	
Articles 15-23: Data Subject Rights		
<input type="checkbox"/>	Do we support data subject rights?	[See detailed table below]
<input type="checkbox"/>	There is no use of automated decision making (e.g. profiling)	[Otherwise need at least a 'discussion note']
Articles 24-43: Controller-Processor		
<input type="checkbox"/>	A28 & 29: What measures are there to ensure processors comply?	[Is there a formal Data Processing Agreement]
<input type="checkbox"/>	A30: Is there an entry for this processing/data held in the register?	
<input type="checkbox"/>	A32-34: Do we ensure appropriate security, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures?	[separate security checklist?]
<input type="checkbox"/>	A37-39: Is there a DPO and have they been or will they be consulted?	[part of sign-off of the DPIA]
Articles 44-50: International transfers		
	What form of data will be transferred to a third country or international organisation	[describe nature of data and whether identified, identifiable, de-identified or anonymous]
<input type="checkbox"/>	Are there safeguards for international transfers?	[e.g. US Privacy Shield, anonymisation, GDPR equivalence, approved contractual clauses, or BCR]
Article 90: Obligations of secrecy		
<input type="checkbox"/>	Do we meet medical confidentiality requirements?	[Note any national case law and statutory requirements that may affect this]

Table 11: GDPR Compliance Checklist

5.7 Data Subject Rights:

Note if supported and what process/procedure applies; if not, then describe the legal justification for not supporting this right.

<input type="checkbox"/>	To be informed: about processing, about choices, about rights, about controller	
<input type="checkbox"/>	the right of access to see or receive a printed copy	
<input type="checkbox"/>	the right to rectification – to correct any material errors in the personal data	



<input type="checkbox"/>	the right to erasure – where appropriate, to ask that all personal data is erased	
<input type="checkbox"/>	the right to restrict processing – to ask that some or all processing ceases [see opt-out]	
<input type="checkbox"/>	the right to data portability – this only applies to data provided directly by individual	
<input type="checkbox"/>	the right to object to and not to be subject to automated decision-making, including profiling	
<input type="checkbox"/>	Right to object to a Data Processing Authority (typically the relevant supervisory authority of each Member State)	
<input type="checkbox"/>	Where consent is the legal basis, the right to withdraw consent	

Table 12: Data Subject Rights



5.8 Detailed Transparency Checklist²

Does privacy information provided to data subjects include:

<input type="checkbox"/>	The name and contact details of our organisation	
<input type="checkbox"/>	The name and contact details of our representative (if applicable)	
<input type="checkbox"/>	The contact details of our data protection officer (if applicable)	
<input type="checkbox"/>	The purposes of the processing	
<input type="checkbox"/>	The lawful bases for the processing	[Art6 for 'personal data' & Art9 for 'special category']
<input type="checkbox"/>	The legitimate interests for the processing (if applicable)	
<input type="checkbox"/>	The categories of personal data obtained (if the personal data is not obtained from the individual it relates to)	[for Art14]
<input type="checkbox"/>	The recipients or categories of recipients of the personal data	
<input type="checkbox"/>	The details of transfers of the personal data to any third countries or international organisations (if applicable)	
<input type="checkbox"/>	The retention periods for the personal data.	
<input type="checkbox"/>	The rights available to individuals in respect of the processing	
<input type="checkbox"/>	The right to withdraw consent (if applicable)	
<input type="checkbox"/>	The right to lodge a complaint with a supervisory authority	
<input type="checkbox"/>	The source of the personal data (if the personal data is not obtained from the individual it relates to)	[For Art14]
<input type="checkbox"/>	The details of whether individuals are under a statutory or contractual obligation to provide the personal data (if applicable, and if the personal data is collected from the individual it relates to)	
<input type="checkbox"/>	The details of the existence of automated decision-making, including profiling (if applicable)	
<input type="checkbox"/>	We provide individuals with privacy information at the time we collect their personal data from them – or where we obtain personal data from a source other than the individual it relates to, we provide them with privacy information	

² Taken from UK Information Commissioner's Office template though applicability to Member State Jurisdictions to be reviewed on a case by case basis



<input type="checkbox"/>	within a reasonable of period of obtaining the personal data and no later than one month	
<input type="checkbox"/>	if we plan to communicate with the individual, at the latest, when the first communication takes place	
<input type="checkbox"/>	if we plan to disclose the data to someone else, at the latest, when the data is disclosed	
<input type="checkbox"/>	We provide the information in a way that is: <input type="checkbox"/> concise; <input type="checkbox"/> transparent; <input type="checkbox"/> intelligible; <input type="checkbox"/> easily accessible; and <input type="checkbox"/> uses clear and plain language.	[Describe how we check is Plain English, etc.]
<input type="checkbox"/>	When drafting the information, we: <input type="checkbox"/> undertake an information audit to find out what personal data we hold and what we do with it. <input type="checkbox"/> put ourselves in the position of the people we're collecting information about. <input type="checkbox"/> carry out user testing to evaluate how effective our privacy information is	[Note: best practice advice]
<input type="checkbox"/>	When providing our privacy information to individuals, we use a combination of appropriate techniques, such as: <input type="checkbox"/> a layered approach; <input type="checkbox"/> dashboards; <input type="checkbox"/> just-in-time notices; <input type="checkbox"/> icons; and <input type="checkbox"/> mobile and smart device functionalities.	[Note: best practice advice]

Table 13: Detailed Comparison Checklist



5.9 Security & Access Control Checklist

Controls need to be appropriate to level of risk: identified special category data needs more protection against potential misuse than non-personal data.

	Data Security classification (above Official)	<input type="checkbox"/> - Official-Sensitive <input type="checkbox"/> - Secret <input type="checkbox"/> - Top Secret <input type="checkbox"/> - Public Domain
<input type="checkbox"/>	Personal Data involved [GDPR]	
<input type="checkbox"/>	Special Category of personal data involved [GDPR]	
<input type="checkbox"/>	Electronic Communications (inc. cookies) [PECR]	
<input type="checkbox"/>	Credit Card data	
<input type="checkbox"/>	Legal enforcement [LED2018]	
<input type="checkbox"/>	Financial data	
<input type="checkbox"/>	Intellectual Property (detail owner)	
<input type="checkbox"/>	Commercial in confidence (detail owner)	
	Data Location (storage or processing) (include any back-up site(s))	<input type="checkbox"/> - UK <input type="checkbox"/> - EU/EEA <input type="checkbox"/> - EU White-list <input type="checkbox"/> - USA <input type="checkbox"/> - Other:
<input type="checkbox"/>	Is data held in secure data centre?	[detail centre and what certification supports assertion]
<input type="checkbox"/>	Is this new supplier, location, or system?	[If so, need specific IS check; also need formal contract]
<input type="checkbox"/>	Is all user access subject to 2-factor authentication?	<input type="checkbox"/> - no control <input type="checkbox"/> - single factor (e.g. just password) <input type="checkbox"/> - 2-factor (e.g. password & fob) <input type="checkbox"/> - biometric [note: GDPR reqs] <input type="checkbox"/> - Other control:
<input type="checkbox"/>	Are there established JML procedures?	[Joiners, Movers, Leavers]
<input type="checkbox"/>	Are there checks that passwords are robust and secure enough?	[]
<input type="checkbox"/>	Are all administrator & user accounts routinely monitored?	[Particularly for redundant or little used accounts]
<input type="checkbox"/>	Are systems protected against malware and other attacks?	[provide details of protection software and procedures]

Table 14: Security and Access Control Checklist

[Need some aspect of CIA/impact-likelihood assessment]

5.10 Information Asset Register Checklist

<input type="checkbox"/>	Are there new IAs being created?	[provide details]
<input type="checkbox"/>	Are old IAs being retired?	[provide details]
<input type="checkbox"/>	Have IAOs & IACs been consulted?	



<input type="checkbox"/>	Has IAR been updated/amended?	[at least create project task to do so]
<input type="checkbox"/>	Data Retention classification & period	
<input type="checkbox"/>	Data retention procedure/functionality in place	

Table 15: Information Asset Register Checklist

5.11 Appendix A – Supervisory Authority ‘High Risk’ Check

If the DPIA shows ‘high risk’ processing which cannot be mitigated, then the DPIA should be sent to the relevant authority for review before any processing starts. Note that their review may take several weeks to process. A ‘High Risk’ assessment represents a ‘risk to the rights and freedoms of individuals’ – so may extend beyond GDPR consideration, including Human Rights.

GDPR Article 35(3) provides three examples:

- a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
- processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 1013; or
- a systematic monitoring of a publicly accessible area on a large scale

Data Protection Authorities cite:

1. Systematic and extensive profiling with significant effects
2. Large scale use of sensitive data [viz. ‘special category’ in GDPR terms]
3. Public monitoring

These being the same as (a)-(c) above. Applicability across local Member State Jurisdiction to be reviewed on a case by case basis where as a general point the UK Information Commissioner’s Office further identified:

1. **New technologies:** processing involving the use of new technologies, or the novel application of existing technologies (including AI).
2. **Denial of service:** Decisions about an individual’s access to a product, service, opportunity or benefit which is based to any extent on automated decision-making (including profiling) or involves the processing of special category data.
3. **Large-scale profiling:** any profiling of individuals on a large scale.
4. **Biometrics:** any processing of biometric data.
5. **Genetic data:** any processing of genetic data, other than that processed by an individual GP or health professional for the provision of health care direct to the data subject.
6. **Data matching:** combining, comparing or matching personal data obtained from multiple sources.
7. **Invisible processing:** processing of personal data that has not been obtained direct from the data subject in circumstances where the controller considers that compliance with Article 14 would prove impossible or involve disproportionate effort.
8. **Tracking:** processing which involves tracking an individual’s geolocation or behaviour, including but not limited to the online environment.



9. **Targeting of children or other vulnerable individuals:** The use of the personal data of children or other vulnerable individuals for marketing purposes, profiling or other automated decision-making, or if you intend to offer online services directly to children.
10. **Risk of physical harm:** Where the processing is of such a nature that a personal data breach could jeopardise the [physical] health or safety of individuals.

‘High Risk’ assessment using ICO criteria (or appropriate Member State criteria as required):

Criterion:	Assessment	Comments
New technologies		
Denial of service		
Large-scale profiling		
Biometrics		
Genetic data		
Data matching		
Invisible processing		
Tracking		



Criterion:	Assessment	Comments
Targeting of children or other vulnerable individuals		
Risk of physical harm		

Table 16: 'High Risk' Assessment

[The assessment can be one of N/A (not applicable), Low, Medium, or High. The comments should explain how the assessment is justified.]



5.12 Appendix B – Broad Privacy Risk Assessment:

#	Risk Description/detail	Discussion
1.	Data accuracy and timeliness	[Is data accurately recorded & kept up-to-date?]
2.	Differential treatment of patients/data subjects	[Might certain categories of people be adversely affected, e.g. children, vulnerable adults]
3.	Data Accuracy and identification	[Is the identification of individual reliable? Is there a danger of mis-attribution or incorrect linkage of data?]
4.	Holding / sharing / use of excessive data within [Company] systems	[Might too much data be held or for long? Is there a clear justification for data retention? Not 'just in case']
5.	Data held too long within [Company] systems	[Is there a clear data retention period specified and are there processes to ensure its deletion when no longer needed? Are copies tracked and deleted as well?]
6.	Excessive range of access in terms of users to personal data (consider new users/change of access privileges)	[Do more users have access than strictly necessary? Are user roles clear distinguished and reflected in the access privileges? Is there a clear process for granting and revoking access privileges?]
7.	Potential for misuse of data, unauthorised access to systems	[What are the likely threats to the data? What countermeasures are or might be applied? Is it possible for access to be granted inappropriately?]
8.	New sharing of data with other organisations, including new or change of suppliers	[Is data being shared from new data providers or with new data users? Are there new suppliers or data processors? What controls will apply?]
9.	Variable and inconsistent adoption / implementation	[How well will this system work end-to-end? How robust is it against partial adoption or system failure?]
10.	Legal compliance, particularly DP transparency requirements and support for data subject rights	[How well does this system meet legal requirements – or appear to meet legal requirements? Does it meet the 'No surprises' rule? What would happen if an individual requests data erasure or ceasing processing, etc.]



#	Risk Description/detail	Discussion
11.	Medical confidentiality	[Are there any addition sensitivities over confidentiality? Might specific approval (e.g. REC) be required to support this processing?]

Table 17: Broad Privacy Impact Assessment



6 Annex 2

Draft Questionnaire for Ethics and GDPR Compliance

The Workpackage 2 team would be very grateful if you could answer the following questions that will help us in supporting you with your regulatory compliance when participating in GenoMed4All.

This questionnaire covers the lawfulness of your processing primarily, including the current state of your ethics approvals, where you may need to seek amendments from ethics committees to approve your participation in GenoMed4All.

We anticipate that we will need to come back to you only once with a request for more detailed information around specific data protection compliance however we hope that the information you provide here will be sufficient in the first instance.

Please contact Nathan Lea or Francesco Cremonesi if you have any questions at all.

<<ADD STANDARD PRIVACY NOTICE AND GUIDANCE>>

1. Do you already have a local research project including SCD patients that you plan to include as part of your participation in GenoMed4All? Yes / No
 - a. If yes,

Does this project have the approval of your relevant Ethics Board(s)? Yes, No

Have you received informed consent? Yes, No

 - i. If you have received informed consent, does it already include the sharing of individual pseudonymized data / samples with other countries in the EEA?
Data yes / no Sample Yes/no
 - ii. If you have received informed consent, does it already include the sharing of individual pseudonymized data / sample with other countries outside the EEA?
Data yes / no Sample Yes/no
 - iii. Comment: please provide any comments on limitations i.e. genetic data
 - b. If yes, do you have a local database/registry different from Hospital EHR including structured patients' data? Yes/No

if yes:

 - i. please provide further details i.e. type of data gathered, software used (it would be helpful if you could provide a data flow diagram and a brief description of data flows including source and destination).
 - ii. what method do you use for protecting patients' data?
Anonymization/Codification/Pseudonymization
 - c. what assurances do you have for your database / registry / EHR infrastructure? e.g. ISO 27001 certification, Penetration Testing, internal or external audits or within a secure, trusted research environment?



- d. Do you have a chief information security officer and responsible individual (e.g. Information Asset Owner) for the repository or your EHR / digital services infrastructure?
 - e. What is the location of your privacy notice and does it cover your repository?
 - f. If you are aware: what are the legal bases for processing the data for your project?
 - g. Do you ever export data from your repository and if so, under what circumstances?
2. Do you plan to amend the protocol currently approved by your local ethics committee, to cover your involvement with GenoMed4All? Yes, no
- If yes please specify the reason
3. Do you need to submit a completely new protocol to Ethics Board for the genomed4all? Yes, No
4. Are you aware of any specific regulation in your country for the secondary use of data for research? Yes, No
 - a. if yes, does it apply for (tick all that apply):
 - i. data gathered in routine in the EHR
 - ii. data generated in the context of a previous research project
 - iii. data generated from genetic samples
 - iv. any other categories of data (please specify)
 - b. if yes, please specify the rules and refer to the related law / legislation



7 Annex 3

Draft Data Sharing Agreement Template for Consortium Partners

8 i~HD Data Sharing Agreement Template – GenoMed4All

This template should be used to guide the development of a Data Sharing Agreement for the GenoMed4All Consortium partners to share and process data as defined in the Data Management Plan and Data Protection Impact Assessment for the project. Broadly the data sharing will entail:

1. Sharing of data from data provider partners to a centralised pseudonymised database
2. Dissemination to recipient partners for the purposes of machine learning and algorithm training for the GenoMed4All Consortium.

The following headings lists the broad articles that typically appear in a multi-party data sharing agreement. It has been based upon templates, materials and experiences i~HD has harnessed through its involvement with several IMI and EU funded projects including EHR4CR, EHR2EDC, EMIF, HELICAL, ADLIFE and FAIRVASC amongst others.

This template should be read in line with the Data Processing Agreement template for when processors are engaged, as well as the Data Licensing Agreements template for providing access to external parties to the consortium.

9 Broad Articles for Inclusion

1. Parties to the agreement (Name, designation, organisation, contact information)

The parties to the agreement should be listed with their full contact details, e.g. “The European Institute of Innovation through Health Data (i~HD), Oude Mechelsestraat 165, 1835 STROMBEEK-BEVER, B-9000, Grimbergen, BELGIUM” etc.

2. Context and Description of the parties

The context of the activities that are covered by the DSA as well as a brief description of the parties plus their contribution to the activities should be listed. E.g. “Where GenoMed4All is a consortium funded under grant number xxx to explore the sharing of health and genetic data collected from care settings and research registries, partner A will provide a centralised platform of pseudonymised records to allow partners b, c and d to conduct machine learning on the data across a Federated network.” The details of the parties could be listed as follows:

- 2.1. Parties providing the data intended for shared access and use: the data providers
- 2.2. Parties wishing to undertake the specified research using the shared data: the recipients



2.3. Parties intended to act solely as a processor (for instance providing a centralised platform, if any)

2.4. Parties who will act as a processor and controller separately.

3. Particulars of the agreement

Including details of the period that the agreement applies (start and end dates), other agreements that relate to this agreement (e.g. the Consortium Agreement) and a list of schedules to the agreement (which can contain data protection particulars, definition of work and any codes of practice that may apply).

4. Definitions

A list of agreed terms, including examples such as “Data Protection Legislation, Research Data, Pseudonymisation, provider, recipient etc. E.g.

“Data Provider: a Consortium Partner that Provides Data for the purposes of the Research” (where Consortium Partner, Provides, Data and Research will themselves be defined).

Note a Data Custodian role may be defined as the partner who holds the centralised repository and maintains responsibility for reasonable access and sharing.

5. Aims of the Agreement

A description of the aims and scope of the agreement – broadly to allow for the creation of a centralised pseudonymised database and the flows of data from that database to consortium partners. This will include the data items permitted, expected behaviours with the data items and their overall control and governance in line with data protection legislation.

6. Roles of the Parties (source and recipient)

In brief what each party is bringing to the activities in terms of materials, what they will be doing (either providing or conducting research on the data, or managing the central repository) and so forth.

7. Data sources

A description of the data sources explicitly stated.

8. Data processing particulars

This can be provided as a separate schedule but must include a specification of controllers, processors, third party recipients, purposes of data processing, legal bases, the categories of personal data to be processed, whether special category data is included and the article 9 exemptions in place, the existence of a DPIA and so forth.



9. Security, confidentiality particulars

Again can be a separate schedule but this is where the minimum security and confidentiality protection particulars will be specified (including encryption at rest, forbidding of re-identification, and security standard compliance or certification required...).

Additionally measures to protect privacy may also be provided here, including:

- 9.1. To what extent have the data been de-identified?
- 9.2. Details of the anonymisation or pseudonymisation processes that have been undertaken, including which fields have been modified and what arrangements have been made for the handling of pseudonym keys
- 9.3. Confirmation of whether or not pseudonyms will be reused for supplementary data releases, enabling linkage between the releases
- 9.4. Details of any particular security conditions defining how the data custodian must protect the data, such as the use of a data safe haven, restrictions on access and/or restrictions on access to query result sets
- 9.5. May the recipients transfer part or all of the dataset to other locations and countries within their organisation (in particular if outside the EU), and/or to other contracted third parties undertaking part of the research? This will be specified here with additional Standard Contractual Clauses (as well as Scherms II compliant SCCs) as required.

It may be the case that a Code of Practice is drafted as a separate schedule (where this will likely be needed for a third party recipient outside the consortium, within a Data Licensing Agreement).

10. The research purposes covered by this agreement

This may be included as part of the roles of the parties or the schedule of Data processing particulars, or as a separate schedule of work, but it must include:

- 10.1. Specification of the research purpose(s), including therapeutic area
- 10.2. Reference to a specific research protocol that has been approved by the sponsoring partner's independent review boards or is under consideration
- 10.3. Confirmation by the data providers and recipients that relevant permissions exist for the data custodian to facilitate data disclosure and intended use by the research user (including



research ethics approval, participant consent if this is the GDPR legal basis, and any specific permissions relating to sample access and analysis)

11. Arrangements for accessing and/or transferring the data

Again these particulars may form part of the Security and Confidentiality Articles or Schedule but must include:

- 11.1. Specification of the subject population to be included in the shared data
- 11.2. Specification of the data items (variables) to be provided
- 11.3. Format(s) in which the data will be made available (e.g. file types, APIs)
- 11.4. Route of access or method of transferring the data
- 11.5. Specification of samples to be provided for the research
- 11.6. Specification of how the samples are to be accessed and analysis undertaken
- 11.7. Details of how analysis results are to be handled and which party will assume long term custodianship and controllership of the analysis results

12. Data quality

If not covered under the Consortium Agreement then particulars of data quality assurance will be provided here

- 12.1. Any metrics or assessments be provided by the data sources and recipients regarding the quality of the data to be shared
- 12.2. Details of the data quality information to be provided

13. Data enhancing and interpretation services to be provided

This will likely occur for external partners and may need to be part of a Data Licencing Agreement

- 13.1. Will any data enriching activities such as cleaning, variable derivation and/or analysis be undertaken by the data recipients on behalf of the Consortium?
 - 13.1.1. Details of the data enriching activities to be performed
- 13.2. Will data linkage be performed by the data recipients or providers prior to access?
 - 13.2.1. Details of the datasets to be linked, how and when



13.3. Will updates to the dataset be provided during the above specified date interval?

13.3.1. Details of the anticipated data items to be updated, the nature of the updates and the frequency with which they will be provided

13.4. Will additional data collection or samples be provided during the specified date interval?

13.4.1. Details of the additional data and/or samples, and when these will be provided

13.5. Will support be available from the data providers for interpreting and/or analysing the data?

13.5.1. Details of the available support

14. Handling of analysis results

Likely these particulars would be required for an external party within a Data Licensing Agreement.

14.1. Details of the intended transparency and/or utilisation of the results, for example if the knowledge derived from the use of the data is intended for direct publication or other form of open access, or is intended to be used within products or services for the public good, but will not necessarily be published (in line with the Data Management Plan)

14.2. Details of the handling of intellectual property, publication, authorship and acknowledgement, including the grounds on which a data custodian team member or members should be included as authors in publications, and how a data custodian should be acknowledged within research outputs irrespective of whether a team member is a co-author. This will be pursuant to the Consortium Agreement but will need to be specified for external parties.

14.3. Is the research user expected to return any cleaned or derived variables?

14.3.1. Details of the variables to be enriched and/or returned

14.4. Details of the procedure to be followed in the event of the discovery of clinically significant findings (research results or incidental findings) that may have implications for the data subjects within the dataset

15. Cooperation

This will be around ensuring that data subject rights can be addressed appropriately including access requests, withdrawal etc. Likely data providers will specify what is required in line with their local consents.



16. Data retention

These will be in line with Member State laws around research governance compliance and will at least need to cover:

16.1. Details of the trigger event (allowing time for the completion of the analysis and publication of the research etc.) for data destruction of any extracted data sets or analysis tables, who should perform this and what evidence the data custodian will require that the data destruction has been carried out to a required standard

16.2. Is a copy of the released data set permitted to be held as an archive for a longer period?

16.2.1. If so, for what duration, in what format, how should it be protected, what will be the grounds for access, are any audit arrangements required?

The Parties shall retain the processed personal data for which they are the Joint Controllers or the Data Controllers and the Data Processors, as set forth in this agreement, in a structured, commonly-used, machine-readable format for the period of time indicated in *[we could indicate the source, for example, "in the research project submitted for the opinion of the competent Ethics Committee"]*.

In any case, the Parties shall retain the processed personal data for a period of not less than ten (10) years in order to comply with tax, accounting and administrative obligations as required by law and possibly for longer, which cannot be foreseen, due to different conditions for the lawfulness of the processing (for example, legal actions that make it necessary to process the data for longer than 10 years).

Furthermore, in consideration of the nature of the activities pursued by the Parties, the data to be kept in *[other archives, such as personnel files or clinical records, etc.]* shall be subject to the retention periods provided for by the applicable national and European legislation.

17. Data Breaches

In broad terms, how parties will respond to any breaches. Usually 72 hours is the limit for a Controller to inform a Supervisory Authority of a serious breach so a recipient will need to ensure they have responded in good time for the Provider to respond as necessary.

The Parties are mutually obliged to inform one another via _____ in the event of any data breach within one (1) business day of becoming aware of said event. Said notification must be accompanied by all the documentation necessary to allow, where necessary, the competent supervisory authority to be informed. It should also indicate the nature of the data breach, the data subject category, the contact details of the person that can provide further information, and any actions taken or planned.



The Parties undertake to cooperate with regard to the internal investigations in the respective organisations and to jointly prepare the data breach notification as well as, where necessary, the notification to be sent to the data subject pursuant to art. 34 of the Regulation.

For the purposes set out in this paragraph, the Parties indicate hereunder their respective certified and non-certified email addresses to which any information regarding any possible data breaches must be sent:

- for contracting party 1: _____
- for the other contracting party: _____
- for the other contracting party: _____

...

18. Demonstrating compliance - audits and inspections

In broad terms each party (especially providers) need to ensure they are able to audit and inspect another partner's processes and logs where the need for audit logs is specified here. This may include:

- 18.1. Details of the records that the data custodian must maintain in order to be able to demonstrate conformance to the terms of this agreement, such as lists of authorised users and audit logs
- 18.2. Details of the records that the recipients must maintain in order to be able to demonstrate conformance to the terms of this agreement, such as lists of authorised users and audit logs
- 18.3. Procedures to be followed in case of a suspected breach of the terms of this agreement, and any sanctions or penalties to be applied

Each of the Parties acknowledges the right of the other to carry out audits on the operations related to the processing of personal data conducted under the remit of the joint project. To this end, each Party has the right to organise - at its own expense - spot checks or specific audit or reporting activities in relation to data protection and security, availing themselves of personnel expressly authorised to perform said tasks, at the premises of the other Party.

Each Party shall make available all the documentation necessary to demonstrate compliance with their obligations, to facilitate the conduct of audits and inspections and to contribute to the same.

19. Disputes

Standard clauses for dispute resolutions across the parties.



20. Indemnity

Indemnity declarations for each of the parties as appropriate.

21. Liability

Specification of insurance liabilities for the parties.

The Parties that qualify as Data Processor and Data Controller are responsible for their own acts and omissions that may cause or result in a financial loss or fine as a consequence of their insufficient compliance with their obligations under this DSA and the GDPR.

The Data Processors are not liable for any incompliance or unlawful act by it if such act derives from instructions given by the Data Controllers, or any act or omission of the Data Controllers, provided that the Data Processor has fulfilled its obligations under Article 28 (3) of the GDPR.

The Parties that qualify as Joint Controllers have joint liability in respect of the data subjects for any damage caused by processing in violation of the applicable data protection laws in force, unless the damage is due exclusively to the conduct of one of the Parties.

If one Party has paid the entire amount of compensation for damages, said Party shall have the right of recourse against the other Party involved in order to obtain reimbursement of the part of the compensation for damages paid corresponding to the latter's share of liability for the damage caused.

22. Termination and Withdrawal of consent

Standard clauses for each.

23. Signatories

23.1. Name and designation of the signatories on behalf of the research user(s)

23.2. Name and designation of the signatories on behalf of the data custodians

23.3. Date(s) and location(s) of signature

Witnesses (if applicable)





GENOMED 4ALL

genomed4all.eu



@genomed4all



/genomed4all



GENOMED4ALL receives funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 101017549