

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

Deliverable D7.3

First recommendations for implementation in IT Framework (Incl. a Data Management Plan)

Work Package:	WP7 - Ethical and legal governance framework for transnational data-sharing
Lead Beneficiary:	European Molecular Biology Laboratory
WP Leader(s):	Emmanuelle Rial-Sebbag (INSERM) Melanie Goisau (BBMRI-ERIC)
Contributing Partner(s):	-
Contractual Delivery Date:	31st October 2022
Actual Delivery Date:	31st October 2022
Authors of this Deliverable:	Melanie Goisau (BBMRI-ERIC), Lisa Fériel (INSERM), Emmanuelle Rial-Sebbag (INSERM)
Contributors:	INSERM, CGO, UMCG; WP1, WP2, WP5
Reviewed by:	Michaela Th. Mayrhofer Thomas Keane
Approved by:	Thomas Keane
Dissemination Level:	Public
Type of Deliverable:	Report
Grant agreement:	No. 825775 Horizon 2020 (H2020-SC1-BHC-2018-2020)
Type of action:	RIA
Start Date:	1 Jan 2019
Duration:	54 months

Table of contents:

Executive Summary	4
1 Introduction	5
2 Project objectives and scope of the deliverable	6
3 CINECA cohorts and tools	7
3.1 Overview of CINECA cohorts and use cases	7
3.1.1 CINECA Cohorts	8
3.1.2 CINECA Synthetic Cohort Datasets	8
3.1.3 CINECA use cases	9
3.2 CINECA data discovery in a wider ecosystem	9
4 Methodology	10
5 State of the art: Ethical, legal, and societal implications of further use of genetic/genomic data in CINECA	12
5.1 Further use of genetic/genomic data – societal dimensions	12
5.2 Legal norms	14
5.2.1 Data protection rights	14
5.2.2 Legal framework for health and genetic data further use	18
5.2.3 Security and IT systems: requirements from GDPR	18
5.3 International policies	19
5.3.1 OECD Recommendation of the Council on Health Data Governance	19
5.3.2 GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data and ELSI Toolkit	20
6 Recommendations for the implementation in the IT framework	23
6.1 Engagement and benefit sharing as prerequisites for data sharing	23
6.2 Informed consent and reuse of data	25
6.3 Safeguards and respect for privacy	26
6.4 Further uses and data-access	28
7 Outlook	30
8 Abbreviations	32
9 Delivery and schedule	32



10 References	33
11 Appendices	37
11.1 Appendix 1: Poster “Genetic data sharing for research in Europe: consent not only a legal basis under GDPR”, European Society of Human Genetics Conference 2022, Lisa Feriol, Gauthier Chassang, Emmanuelle Rial-Sebbag	37



Executive Summary

Health data collected in cohort studies are valuable sources for knowledge generation and the advance of biomedical research. However, the use of these data for research projects beyond the initial purpose raises several ethical, legal, technical, and societal questions.

This deliverable addresses these challenges regarding reuse of health/genomic data in CINECA from the ethical, legal, and societal issues (ELSI) perspective and in the light of Open Science and FAIR principles. In responding to the requirements for an appropriate IT and governance framework for CINECA and beyond, the research presented in this document builds on a review of GDPR provisions and their institutional sources of interpretation as well as national laws, the corresponding legal, ethical, and social science literature, as well as stakeholder engagement workshops with patient representatives, African researchers, and co-creative exercises with CINECA technical experts.

The core of this deliverable are the ethical and legal recommendations that take societal implications into account for data access to European, Canadian, and African cohorts. The recommendations address four key areas: (1) Engagement and benefit sharing as prerequisites for data sharing, (2) Informed consent and reuse of data, (3) Safeguards and respect for privacy, and (4) Further uses and data-access.

The deliverable concludes with an outlook on relevant projects such as the European Health Data Space (EHDS) and upcoming ELSI developments regarding data reuse and artificial intelligence (AI).



1 Introduction

Health data collected in cohort studies are valuable sources for knowledge generation and the advance of biomedical research. However, the use of these data for research projects beyond the initial purpose raises several ethical, legal, technical, and societal questions. Sharing and reuse of health data for scientific research purposes is challenged with fulfilling both: (1) utilising the full potential of existing data sets and promoting open science and research practice following the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles (Wilkinson et al. 2016), a key criteria in the EU's open science policy¹; (2) as well as complying with legal frameworks to protect personal data while ensuring research participants' autonomy. It has been emphasised that FAIR is not enough and that standards, policies, and infrastructures to organise metadata are needed (Musen 2022). Federated data infrastructures have become a key approach to make population-scale genomic and biomolecular data accessible across international borders.

Considering these challenges, IT solutions need to be embedded in a comprehensive governance framework that adheres to ethical and legal standards as well as societal values and ensures appropriate implementation to foster data reuse. It has been highlighted that “a good governance framework should specify the scope of research for which data may be used, including any restrictions based either on the original consent or on guidelines generated for the repository, and specify measures that will be used to mitigate or prevent unintended harms and misuses, including transparent decision-making and oversight processes” (O’Doherty et al. 2021, p. 5). This is even more important given the sensitivity of genetic data, especially regarding the vulnerability of certain groups and against the background of a history of unethical use, misuse, and commercialization of, for instance, African genetic data and material. Hence, equitable and ethical use of data, as well as equal partnerships with African countries for fair sharing of data and returning of benefits among all involved stakeholders is vital for good governance. This vision also complies with the requirements not to major discriminations in particular when using genetic information (Joly et al. 2017). The H3Africa guidelines for consent note the importance of a “robust governance framework that should seek to promote global health and research equity and take into account five key elements: respect, authentic community engagement and trust building, the preservation of privacy and confidentiality, feedback of results, and capacity strengthening” (H3Africa Working Group on Ethics and Regulatory Issues for the Human Heredity and Health in Africa (H3Africa) Consortium 2017, p. 10).

While the FAIR principles have become a guiding technical resource for data sharing, legal and socio-ethical considerations are equally important for a fair data ecosystem for further uses of genomic data. As it has been highlighted, FAIR data should be FAIRER, including also ethical and reproducible as key components (Austin 2020). Following these introductory considerations, this deliverable addresses these challenges from the ELSI perspective in connection with the CINECA project and concludes with recommendations for the implementation of an IT framework, especially regarding access and further processing of health and genomic data. To begin with, we will specify

¹ https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en



the projects' objectives and the scope of the deliverable (section 2), followed by an overview of CINECA's technical framework (section 3). Further, we describe the methodology of WP7 and the outcomes of the analysis (section 4). Finally, in section 5, we will conclude with recommendations from the ethical, legal, and societal perspective regarding data access and further processing in CINECA and beyond as well as suggestions for tools. Conclusions and an outlook on further challenges of genomic data uses complete this deliverable.

2 Project objectives and scope of the deliverable

The aim of CINECA is to develop a federated cloud-enabled infrastructure to make population-scale genomic and biomolecular data accessible across international borders, accelerating research, and improving the health of individuals across continents. CINECA will leverage international investment in human cohort studies from Europe, Canada, and Africa to deliver a paradigm shift of federated research and clinical applications. CINECA represents a unique combination of scientific excellence with experience of eleven diverse cohorts and scientific projects such as the European Genome-phenome Archive, CanDIG, and H3Africa. CINECA has assembled a virtual cohort of 1.4M individuals from population, longitudinal and disease studies. Eventually, the CINECA consortium develops one of the largest cross-continental implementations of human genetic and phenotypic data federation and interoperability with a focus on common (complex) disease.

CINECA will not generate novel data from human data, rather it integrates existing resources for federated analyses to deliver new scientific knowledge. Furthermore, it delivers harmonisation strategies and the necessary ELSI framework supporting data exchange across legal jurisdictions. The rationale for sharing and reusing data in public health research is deeply rooted in the promotion of a fair distribution of research risks and benefits, and it has become an essential and powerful tool for public health research.

WP7 provides assessment and guidance on ethical, legal, and societal issues (ELSI), the exchanges of samples and data, resulting in a set of recommendations and governance structure for cohorts from across EU, Canada, and parts of Africa. Over the lifespan of CINECA, WP7 provided a comprehensive report on mapping ELSI (D7.1), contributed to the development of the Data Management Plan (DMP) (D7.4²), identified ethical and legal gaps based on legal analysis of relevant normative instruments in ethics and law, based on a literature review as well as on the analysis of CINECA cohorts' governance frameworks and policies, including findings from stakeholder engagement exercises (D7.2³). In building on the results from these deliverables, D7.3 completes WP7 in providing recommendations for the long-term cooperation between European, African, and Canadian cohorts to intensify collaboration and reuse of data. Together with awareness-raising activities in collaboration with WP6,

² <https://zenodo.org/record/4683415#.Y1vbEoLMJOp>

³ <https://zenodo.org/record/6256296#.Y1vbIoLMJOp>



the outcomes of WP7 target and benefit a wide range of stakeholders within the project consortium and beyond (e.g., data users, researchers, cohorts, participants, etc.).

In conclusion, this deliverable contributes to the following objectives:

- To provide the project with accurate and well-grounded ethical and legal recommendations, taking societal implications into account, for the implementation of an appropriate data-sharing flow between European, Canadian, and African cohorts
- To collaborate with the technical WPs in CINECA to ensure practicality and feasibility of the recommendations and to contribute to training activities
- To address practical and strategic ELSI challenges in international research related to this project

To achieve these objectives, this deliverable builds on a review of GDPR provisions and their institutional sources of interpretation as well as national laws, the corresponding legal, ethical, and social science literature, as well as stakeholder engagement workshops with patient representatives, African researchers, and co-creative exercises with CINECA technical experts (see [Methodology](#)). The recommendations presented in this deliverable are a synergy of key ethical, legal, and societal aspects that need to be considered in research projects including transnational data uses following FAIR principles. While the recommendations have been developed in CINECA, they are intended to also provide guidance for the wider research field. Hence, they are intended as a reference for CINECA consortium partners, future consortia performing research including different cohorts and transnational data uses, especially involving partners from low-income countries and vulnerable populations.

The scope of this deliverable and the recommendations for the implementation of the IT framework are responding to the current state of the technical development in the project. Thus, the focus of this deliverable is on ethical, legal, and societal requirements for equitable and fair data access and reuse as essential prerequisites for a data-sharing flow between Europe, Canada, and Africa.

3 CINECA cohorts and tools

3.1 Overview of CINECA cohorts and use cases

The next sections provide an overview of the data used in CINECA, which build the basis for the following discussion and recommendations. Ethics, legal, and societal implications arise in connection with the specific health data that are accessed and shared for research. Information on which data is further used defines the applicable legal norms, ethical requirements, and societal considerations.



3.1.1 CINECA Cohorts

CINECA brings together a diverse collection of human cohorts consisting of 1.4M individuals in Canada, and European and African countries to facilitate data discovery addressing common diseases. Access to population scale genomic data and use for research has some ethical, legal, and societal implications. Table 1 provides an overview of the CINECA cohorts and data characteristics (more information can be found [here](#)⁴):

Cohort/Resource name	Number of participants	Location	Longitudinal	Diseases	Gender	WGS	WES	RNASeq	Epigenetics	Genotyping
CHILD	3.5k	CA	X	Population based developmental health and disease	M & F	X		X	X	X
CARTaGENE	43k	CA	X	Population based cohort	M & F	X		X		X
CLSA	50k	CA	X	Population based cohort	M & F					X
H3Africa	75k	SA		Multiple communicable and non-communicable diseases in multiple African countries	M & F	X	X			X
BIOS	4k	NL		Population based cohort	M & F	X		X	X	X
Estonian Biobank	51k	EE	X	Population based cohort	M & F	X	X	X	X	X
CoLaus	6.1k	CH	X	Cardiovascular diseases	M & F			X		X
PsyCoLaus	3.6k	CH	X	Mental disorders	M & F			X		X
EGA	700k	UK+ES		Multiple diseases and healthy cohorts	M & F	X	X	X	X	X
UK Biobank	500k	UK	X	Population cohort and disease; cancer, heart disease, stroke, diabetes, arthritis, osteoporosis, eye disorder, depression and form of dementia	M & F	X	X			X

Table 1: CINECA Cohorts

3.1.2 CINECA Synthetic Cohort Datasets

The CINECA project has produced a set of synthetic cohort datasets based on the phenotypic data from four of the participating cohorts: [UK Biobank](#)⁵, [CoLaus](#)⁶, [H3Africa](#)⁷, and the [CHILD Cohort Study](#)⁸. All synthetic cohort datasets are open access and fully accessible under the Creative Commons Licences as specified with each dataset. They were developed to increase accessibility to cohort data for standards development, whilst mitigating ethical and legal privacy concerns. Table 2 provides an overview of the synthetic cohort datasets (more details can be found [here](#)⁹):

⁴ <https://www.cineca-project.eu/cohorts>

⁵ <https://www.ukbiobank.ac.uk/>

⁶ <https://www.colaus-psycolaus.ch/professionals/colaus/>

⁷ <https://h3africa.org/>

⁸ <https://childstudy.ca/>

⁹ <https://www.cineca-project.eu/cineca-synthetic-datasets>



Synthetic Cohort Dataset	Phenotypic data	Genomic data	Generated by	Publication Status
Synthetic Cohort Europe UK1	2521 samples derived from UKBiobank, relating to cancer, diabetes and cardiac	Genetic data based on 1000 Genomes data	TOFU, a tool developed in-house for generating Synthetic Cohort UKBiobank data	European Genome Archive (EGA): https://ega-archive.org/datasets/EGAD00001006673
Synthetic Cohort Africa H3ABioNet	100 samples that have synthetic subject attributes and 47 phenotypic data based on the Human Heredity and Health in Africa	1000 Genomes project phase 3 data, randomly selected 2M variants in chr 22 for 100 samples of African ancestries	Nextflow pipeline that uses a modified version of TOFU	Zenodo: https://zenodo.org/record/4955933
Synthetic Cohort Europe CH SIB	6733 samples using 21 attributes selected from the CoLaus cohort	1000 Genomes project phase 3 data, selected 100 most-common variants in chr 22	DataSynthesizer was used for generating both randomly and statistically correlated synthetic data	Zenodo: https://zenodo.org/record/5082689
Synthetic Cohort NA Canada CHILD	100 select variables for 150 participants, plus COVID and other key variables for CHILD	1000 Genomes project phase 3 data, selected most 100 common variants in chr 22	DataSynthesizer for synthesizing correlated anthropomorphic variables, other variables uncorrelated	Zenodo: https://zenodo.org/record/5122832

Table 2: CINECA Synthetic Cohort Datasets

3.1.3 CINECA use cases

To implement and demonstrate the technical framework for federated analysis, CINECA develops four use cases: PRS (Polygenic Risk Score) (WP4), eQTL (Expression Quantitative Trait Loci) (WP4), FAIR data analysis (WP5), Clinical Decisions and Patient Diagnosis (WP5). These use cases are using mainly synthetic or open access datasets. Data is stored centrally and with controlled access (DACs) or in open repositories (e.g., Zenodo). Depending on the cohort, genotype, sequencing, RNA-sequencing data, and phenotype/clinical data are available.

As mentioned, access to and use of health and genetic data for research purposes comes with ethical, legal, and societal implications that need to be thoroughly addressed for a good governance structure. In CINECA, the development of the tools and use cases is accompanied with considerations regarding the implementation of ethical values and legal norms, especially respect for privacy, which are reflected in the recommendations of this deliverable.

3.2 CINECA data discovery in a wider ecosystem

Utilising the full potential of existing data sets and developing tools to enable data discovery and access cannot happen in silos. Rather, it needs to be addressed as a process of collecting data and building an infrastructure that enables access and sharing. CINECA's federated infrastructure aims to make genomic and biomolecular data accessible across international borders – again with considering



the ethical, legal, and societal implications regarding tools and data flows. To achieve this goal, the consortium has so far achieved the following, which is connected in Figure 1:

- The publicly available (CC-BY) [Genomics Cohorts Knowledge Ontology \(GECKO\)](https://www.ebi.ac.uk/ols/ontologies/gecko)¹⁰, a cohort metadata mapping model developed to enable cross cohort discovery queries with a searchable registry that is updated from cohorts (more information [here](https://www.cineca-project.eu/wp3?rq=gecko)¹¹). It has been adopted by the [International Hundred-K Cohorts Consortium \(IHCC\)](https://atlas.ihccglobal.org/)¹².
- A set of synthetic cohorts based on the phenotypic data from four of the participating cohorts – [UK Biobank](#), [CoLaus](#), [H3Africa](#), and the [CHILD Cohort Study](#). These synthetic cohorts have no identifiable data and are open access.
- Contributed to the development of the [Data Use Ontology \(DUO\)](https://github.com/EBISpot/DUO)¹³, a Global Alliance for Genomics and Health (GA4GH) standard. DUO is a hierarchical vocabulary of machine-readable data use terms which allow the consistent and unambiguous representation of data use conditions to discover, access and integrate diverse datasets (see [D3.5](https://zenodo.org/record/5795449#.Y1vbw4LMJOp)¹⁴).
- Contributed to the development of key genomics standards such as [Beacon](https://beacon-project.io/) v2¹⁵. A Beacon is a genomics variant discovery tool that enables data discovery of genomic and phenoclinic data without compromising the privacy of the dataset. The Beacon uses a [3-tiered access model - anonymous, registered, and controlled access](https://docs.genomebeacons.org/security/)¹⁶.

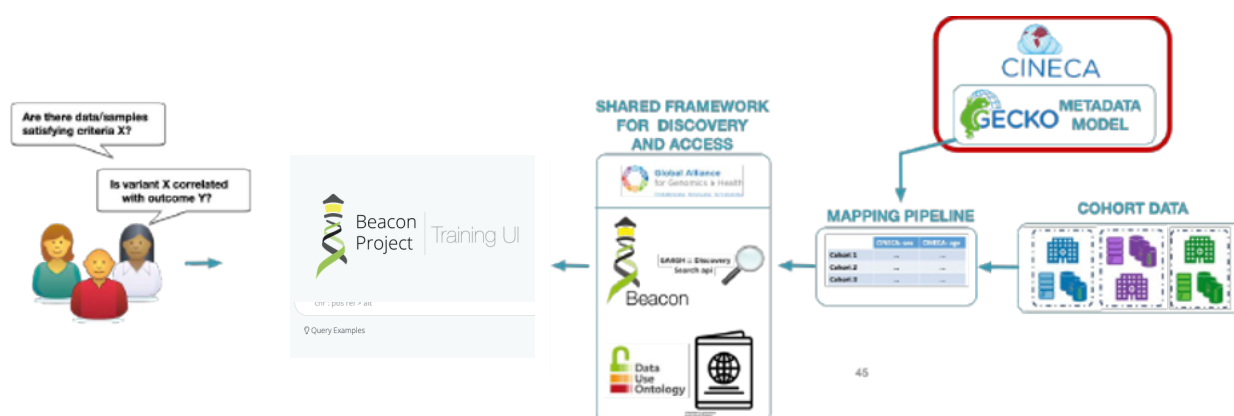


Figure 1: CINECA in the wider data discovery ecosystem

¹⁰ <https://www.ebi.ac.uk/ols/ontologies/gecko>

¹¹ <https://www.cineca-project.eu/wp3?rq=gecko>

¹² <https://atlas.ihccglobal.org/>

¹³ <https://github.com/EBISpot/DUO>

¹⁴ <https://zenodo.org/record/5795449#.Y1vbw4LMJOp>

¹⁵ <https://beacon-project.io/>

¹⁶ <https://docs.genomebeacons.org/security/>

4 Methodology

Considering ethical, legal, and societal issues and engaging with stakeholders throughout the project's lifespan are vital for the establishment of appropriate governance structures (Gottweis and Kaye 2012). This final deliverable of WP7 combines the knowledge gained from its various tasks. This was achieved through a co-creative design that combined research on the ethical, legal, and societal aspects of transnational data sharing, engagement with project internal and external experts and stakeholders, as well as training activities for CINECA consortium partners and the wider field. Figure 2 gives an overview of the co-creative workflow in WP7:

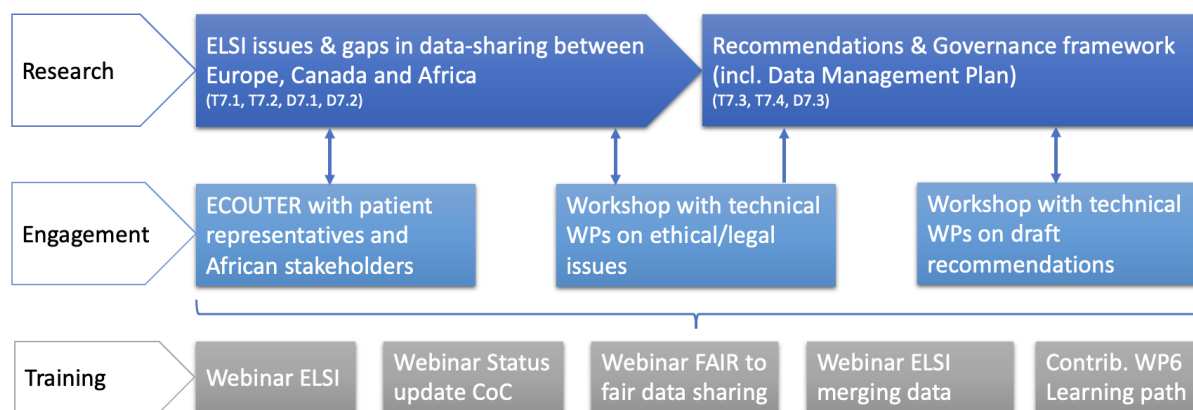


Figure 2: WP7 Co-creative workflow

In detail, the empirical work in WP7 that constitutes the co-creative process and builds the basis for the recommendation consist of:

- An analysis of relevant GDPR provisions, their institutional sources of interpretation as well as national laws, the corresponding legal, ethical, and social science literature. The outcomes are described in Deliverable 7.1 Catalogue on ELSI Issues and Deliverable 7.2 Catalogue of Canada, European and African ethical and legal gaps.
- Stakeholder engagement exercises with African stakeholders during the 5th African Conference on Emerging Infectious Diseases in Abuja, Nigeria, 7-9.8.2019 and members of the BBMRI-ERIC Stakeholder Forum during the Europe Biobank Week 2019 in Lübeck, Germany, 8.-11.10.2019. Furthermore, with members of the technical WPs in CINECA, conducted in three online/hybrid workshops on ethical/legal gaps and requirements described in D7.2, and on the draft recommendations for the implementation in the IT framework. These exercises used the ECOUTER stakeholder engagement methodology (Murtagh et al. 2017). ECOUTER meets the aim of considering ethical, legal, and social questions as well as engaging various stakeholders before the translation of research into

practice, in order to develop an appropriate governance framework as a co-production. The method uses interactive mind-mapping in a group setting. Specific aspects were deepened in semi-structured interviews with selected CINECA experts. All data were collected based on the informed consent of the participants (the empirical study was approved by the INSERM ethics committee, opinion number 19-605) and has been analysed using thematic analysis (Clarke, Braun, and Hayfield 2015).

5 State of the art: Ethical, legal, and societal implications of further use of genetic/genomic data in CINECA

Using health data for different purposes to those originally planned poses several questions such as if the purpose of the additional research study is in the scope of original consent and the involvement of data processors other than the primary data collectors, as well as the preservation of privacy of participants in shared genetic datasets (Schlegel and Ficheur 2017). So far, anonymization is seen as minimum requirement necessary to protect data subjects' privacy in aggregating data, despite the possibility of re-identification through cross-referencing with data concerning ethnic background, locational data, other metadata, health records or even small pieces of identified genetic data (Mittelstadt and Floridi 2016). Special caution is needed with regard to genomic data, as these are highly distinguishable: with only 30 SNPs an individual can be identified (Dankar, Ptitsyn, and Dankar 2018). There is a risk of matching anonymised genetic data with data from other datasets which increases the possibility of re-identification (Shabani and Borry 2018).

The FAIR principles provide guidance for the management of data as well as tools and workflows. The institutional conditions and organisational challenges associated with data sharing need to be considered to ensure responsible, and fair data practices. This requires considering the context of legal requirements, for instance the principle of fairness and transparency in GDPR, expectations of research participants/data subjects, societal aspects and the "ethics work" that is an integral part of data flows, but also fairness, equity and benefit sharing within transnational collaborations.

WP7 empirical investigations within CINECA point to the need of balancing de-identification/anonymization vs. an ethical re-identification for returning results to patients, the data minimisation principle (GDPR) vs. rich data to enable re-use (FAIR), and the level of granularity that is needed vs. granularity that could be harmful (vulnerable/ethnic groups). The following sections address these challenges in highlighting key aspects identified in the recent scientific literature in conjunction with empirical outcomes from stakeholder engagements, as well as the key legal norms. These findings are a central component for the recommendations in section 6.



5.1 Further use of genetic/genomic data – societal dimensions

Besides technical requirements guided by the FAIR principles and legal requirements such as the GDPR, also social and societal implications need to be considered for fair and responsible health data sharing practices and good governance. Risks related to further processing of genomic data in multiple ways and across countries have raised concerns about potential harms related to the protection of privacy, which is also a topic in biobanking (Akyüz et al. 2021). The potentials of research with genomic data, particularly improvements in genetic analysis, has led to a change in the risk profile, especially regarding re-identifiability, stigmatisation, and discrimination on both the individual and group level (Kasperbauer et al. 2018, Helgesson 2012).

These issues can be illustrated with two broadly discussed cases regarding the ethical reuse of genetic information. For example, the case of the Havasupai tribe shows how data collected for genetic research on diabetes were also used for research on ethnic migration and schizophrenia. The latter was generated without informing the research participants about the additional research, eventually opening the possibility of negative consequences for the tribe, such as stigmatization (Garrison 2013, Helgesson 2012). This case emphasizes the risk of stigmatization and discrimination of ethnic groups, even when individuals have given consent for the use of their data in research (O'Doherty et al. 2021). Another example raises the question, whether categorizations used in a genome-wide association study (GWAS) on same-sex sexual behaviour may result in normatively problematic claims on homosexuality. For instance, the operationalization of sexuality, the exclusion of certain groups from the study, and the framing of the research question as a “health issue” bear the risk of stereotyping or stigmatization (Holm and Ploug 2019, Goisauf, Akyüz, and Martin 2020). These two cases challenge the dialectic between anonymization and re-identifiability, and consent as relevant to the individual, whereas group-level harms from analysis of aggregated data are also possible (Mittelstadt and Floridi 2016). While these cases point to critical ethical and societal issues regarding access and further processing of genetic data and its possible consequences, they also help to reflect on the practices and standards used to promote fair and responsible research with genetic data, to inform good governance structures and data management.

The (re)use and further processing of genetic data sheds new light on ethical as well as social and societal issues in genetic research, especially in relation to the multiplicity of stakeholders, research settings and contexts, regulatory bodies and health systems involved in current and future uses. Regarding informed consent, it can be observed that there is a need for information on both sides, researchers and research participants, concerning the use and sharing of health data. For example, findings from the large-scale “Your DNA, Your Say” project (Middleton et al. 2020) provide insights into public (including research participants) attitudes towards sharing of genomic data and health information, and draw on the responses from 36,268 individuals across 22 countries. Results show that the willingness to donate such data as well as trust in data processing including multiple actors are relatively low. In addition to this finding, previous studies have concluded that most research participants want to have a say in how their samples and data are used (Mester et al. 2015, Hemminki et al. 2009). Furthermore, assessments of consent models are shaped by concerns about



the “appropriateness” of research practice and developments, especially regarding research with genomic data (Goisauf and Durnová 2018). Similarly, patient advocates who participated in the CINECA ELSI ECOUTER exercise discussed how/if legal tools can protect research participants whose (personal) data will be shared and if opt-out and withdrawal is even possible in today’s datafied world. (Re)consent was mentioned as an issue and the role of information and understanding of what data actually is for participants to make informed decisions. This was identified as an area of risk together with the question of how data is used and shared by whom for which purposes, and who is in control. Another aspect was the importance of returning results to participants. Concerning researchers, findings from a survey among biobank professionals have shown that informing participants about data sharing and multiple uses of data in the informed consent needs to be improved, together with participant engagement (Goisauf et al. 2019).

Regarding low and middle income countries, some authors argue that broad consent should only be permissible with supplementary safeguards elaborated after “genuine” engagement with the community (Tindana and Vries 2016) and after explicit discussions, notably potential re-identification of genomic data, benefit sharing and commercial use of research results (Moodley and Kleinsmidt 2021). Tindana and Vries (2016, p. 387-388) point out five elements for a good governance framework for genomics in low and middle income countries: respect, community engagement, privacy and confidentiality, feedback of results, and capacity strengthening. The perspective of African stakeholders gathered in the empirical study in CINECA highlights the importance of equal partnerships with African countries for fair (and in a broader sense FAIRER) sharing of data and benefits among all involved stakeholders. Respective mechanisms and safeguards need to be put in place to ensure privacy and to protect patient data from misuse and exploitation. The contributions of all research partners must be acknowledged in terms of credits for the local researchers involved, especially authorship, funding, and training, as well as fair sharing of the research outcomes, knowledge, technologies, and other benefits. Solid governance structures, continuous stakeholder engagement and joint development of common standards are needed to build trust and to establish a fair, transparent, and functioning data-sharing flow between all partners.

Alongside the benefits of accessing and sharing large amounts of health data for research, challenges arise in connection to big data research and the associated ethical and societal implications, for instance regarding informed consent, privacy and the possibility of re-identification, ownership and commercial use as well as the risk of group-level harms (Mittelstadt and Floridi 2016). Additionally, discrimination could appear in the research design, especially in terms of unequal representation of certain groups, which could result in a lack of diversity and biased findings. To that effect, new approaches like polygenic risk scores (PRS) will shed new light on some existing ethical, legal, and societal issues regarding secondary and incidental findings, the relevance of results for family members, determination of health risk and access to risk-stratified care, the representation of various ancestry groups to represent genetic diversity as well as race-based disparities in health care (Slunecka et al. 2021, Knoppers et al. 2021, Lewis and Green 2021).



5.2 Legal norms

5.2.1 Data protection rights

- *The legal regime of health and genetic data processing under GDPR*

What are the specific rules applicable to the collection and processing of health and genetic data under GDPR? Health and genetic data are considered by the GDPR as special categories of personal data (Article 9 GDPR). Indeed, because of the potential information on the health status of a person they are likely to reveal, their sensitivity has led to the establishment of a general principle of prohibition of their processing (Art. 9(1) GDPR). There are several exceptions to this prohibition principle, allowing the collection and processing of data in very limited cases, including the explicit consent of the data subject (Article 9(2)a), the public interest of the processing (Article 9(2)i), and scientific research purposes (Article 9(2)j).

What are the specific rules applicable to the processing of personal data for scientific research purposes? The purpose of scientific research is one of the exceptions provided by the GDPR (Art.9(2)j) allowing the processing of sensitive data. Concerning the processing carried out for scientific research purposes, the GDPR provides for specific provisions. Indeed, the exception provided by the GDPR for the processing of particular data for research purposes (Art.9(2)j) must comply with certain requirements: the processing must be subject to appropriate safeguards for the rights and freedoms of the data subjects such as the implementation of technical and organizational measures, in particular to ensure compliance with the principle of data minimisation (Article 89 GDPR). In particular, the pseudonymisation technique is explicitly referred to in the GDPR. These additional measures required are justified by the fact that the collection and processing for research purposes benefit from certain exemptions which mainly concern the rights of the data subjects (articles 15, 16, 18, 21), insofar as these rights make it impossible or seriously hinder the achievement of the research purpose (Article 89(2) GDPR).

How to ensure the protection data subjects rights to the processing of their personal data under the GDPR? The GDPR recognises several rights of data subjects regarding the processing of their personal data. The data controller shall take appropriate measures to provide the information referred to in Articles 13 and 14 of the GDPR to the data subjects in order to ensure transparent processing of their personal data. This information must explain how to exercise the data subject's rights in order to guarantee the effectiveness of these rights. This information must be concise, transparent, understandable, and easily accessible in simple and clear terms. In addition, the information must be adapted to the target audience, such as children.

These rights are listed in Chapter III of the GDPR and include:

- the right to information: respect for the principle of transparency (Article 13 and 14),
- the right of access by the data subject (Article 15)
- the right to rectify personal data that are inaccurate (Article 16)
- the right to erasure or otherwise called the right to be forgotten for certain reasons (article 17),



- the right to restriction of processing in certain situations (article 18),
- the right to data portability (Article 20)
- the right to object at any time to the processing of data (Article 21)

When a processing of personal data is likely to result in a high risk for the rights and freedoms of data subjects, a data protection impact assessment (DPIA) must be carried out according to articles 35 and 36 GDPR. As genetic and health data are considered to be special data due to their sensitivity, their processing will require this impact analysis. This impact assessment ensures that the processing will comply with the GDPR and will respect the rights of the individuals concerned.

How to carry out a Data Protection Impact Assessment? The Data Protection Impact Assessment (DPIA) must be carried out before the processing is implemented and must be reviewed during the processing, especially if major changes occur in the way the data is processed. The participants in carrying out the DPIA are the data controller, the data protection officer, any subcontractor(s), the IT staff, and the persons concerned by the processing. The DPIA consists of three parts:

1. A detailed description of the processing carried out, including the technical and operational aspects of the processing.
2. An assessment of compliance with the fundamental principles of data protection, namely: an examination of the necessity of the processing and compliance with the principle of proportionality (the data collected and processed are strictly necessary for the purpose of the processing) as well as a description of the measures put in place to guarantee the rights of data subjects.
3. A more technical study of the risks to data security (confidentiality, integrity, and availability) and their potential impact on privacy. This study must be completed by a description of the technical and organisational measures envisaged to deal with these risks and protect the data.

The CNIL (French national authority for personal data protection) has developed a guided and pedagogical tool to carry out this DPIA which is available in English and one of the recommended resources by BBMRI-ERIC^{17,18}.

- ***The different types of consent and consent for what?***

Is it mandatory to obtain consent from data subjects prior to the collection and processing of their health and genetic data? Consent to data processing is not always required. It depends on the legal basis for the processing. If the processing is based on the consent of the data subjects, then yes, the collection of consent prior to the implementation of the processing is mandatory. This consent must

¹⁷ <https://www.cnil.fr/en/privacy-impact-assessment-pia>

¹⁸ For more information: <https://ec.europa.eu/newsroom/article29/items/611236> - Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.



comply with the requirements of the GDPR, namely: free, specific (given for one or more purposes), informed (information provided to the person on the processing) and unambiguous (given by a clear positive act without ambiguity). The persons concerned can change their mind at any time and withdraw their consent. In addition to these conditions, for the processing of **sensitive data** (including health and genetic data), the criterion of explicitness is added, i.e., the data subject must expressly declare his/her consent (for example in writing). The collection of consent must be documented by the data controllers, they must be able to prove that consent in accordance with these requirements has been sought.

However, in the context of scientific research, often the legal basis of the scientific research purposes will be preferred (Art.9(2j)) because it is sometimes complicated to find the person concerned and to solicit his/her consent. Using this legal basis does not require the prior consent of individuals for the processing of their data. However, the obligation to provide information must always be respected by allowing data subjects to object to the processing (right to object – article 21 GDPR).

Nevertheless, it is necessary to recall that Member States may introduce additional conditions, including limitations, regarding the processing of health and genetic data (Article 9(4) GDPR). Thus, a national law may provide that the processing of these categories of data may require the collection of consent even if this is not required by the GDPR. The controller will then have to ensure compliance with the national laws in force in the Member States in which the data is collected and/or processed.

What is the difference between consent to personal data processing for research and consent to participate in research? It is necessary to distinguish between consent to research and consent to data processing for research. Indeed, participation in research is governed by national laws that may require consent under specific conditions. Moreover, it is necessary to recall that informed consent to research is an ethical requirement under the Oviedo, Taipei and Helsinki Conventions. Moreover, this informed consent has been described by the EDPB as a potential “appropriate safeguard” (provided for in Article 89(1) of the GDPR) to be put in place to safeguard the rights and freedoms of individuals in the context of data processing for scientific research purposes. Thus, even if the legal basis chosen is that of scientific research purposes and there is no legal obligation to collect the consent of data subjects, in an ethical approach it can be collected as informed consent. This position is also defended by the CINECA Deliverable 7.2 as a recommendation in order to process the personal data concerned within the project with the most ethical lawful legal basis.

Indeed, **as the GDPR offers rooms for innovating in consent practices when processing data for scientific research in line with recognised ethical standards (Recital 33 GDPR)**, even if consent is not chosen as a legal basis, it should continue to be envisaged as an **ethical requirement** for expressing clear choices on data reuses, when possible. A more flexible practice of consent (i.e., broad consent, dynamic consent, tiered consent, layered consent) would allow both to improve data subjects’ involvement in the reuse of their data and to facilitate accountable practice from the scientific community. In a context where trust of individuals in the re-use of their data remains a continuous challenge, reconsidering consent rather than waiving it could reinforce the data subject autonomy and active contribution to research (See Poster [Appendix 1](#) “Genetic data sharing for research in Europe: consent not only a legal basis under GDPR”, European Society of Human Genetics Conference 2022, Lisa Feriol, Gauthier Chassang, Emmanuelle Rial-Sebbag).



5.2.2 Legal framework for health and genetic data further use

What does “further processing” mean according to GDPR? Scientific research today relies heavily on the further use/processing of data. Further processing refers to the processing of previously collected data for a purpose different from that originally intended at the time of collection. This further processing potentially includes data controllers other than the one who collected the data. We can find other expressions that refer to this situation in the literature, such as “further use”, “further processing”, “secondary use” and “reuse”.

Specific provisions are provided for the framework of further use for scientific research purposes within the GDPR: According to Article 5(1)b of the GDPR, further processing of data for scientific research purposes is not considered incompatible with the initial purposes (presumption of compatibility). That is, the processing will be considered a priori compatible with the initial purposes of the processing provided that appropriate safeguards for the rights and freedoms of the data subjects are put in place (implementation of technical and organisational measures to comply with the data minimisation principle, including pseudonymization - Article 89(2) GDPR). However, as the European Data Protection Board reminds us, this presumption is not a general authorization for further use of data for all cases of research purposes, each case must be considered according to its context.

What are the personal rights to be respected in case of further use? Data subjects must be informed of this further processing before it is carried out, unless one of the exceptions in Article 14(5)b applies. The rights of the data subjects shall be guaranteed unless one of the exceptions of Article 89(2) of the GDPR applies.

It seems necessary to recall that Member States may introduce special provisions for the processing of health and genetic data, including limitations (Article 9(4) GDPR). Indeed, it will be necessary to ensure compliance with any special national rules for the re-use of health and genetic data.

What are the challenges related to health and genetic data further use? The rules surrounding further use of health and genetic data for research purposes have yet to be clarified. Indeed, it is still too subject in practice to divergent interpretations between the various Member States. The European Data Protection Board has responded to some of the questions raised in this regard. Guidelines focusing on the processing of personal data for scientific research purposes are expected from the European Data Protection Board.

5.2.3 Security and IT systems: requirements from GDPR

What are the security requirements foreseen by GDPR? Article 32 of the GDPR designates the controller and the processor as being responsible for implementing appropriate technical and organisational measures to ensure a level of security appropriate to the risk. Among these measures, we find:

- pseudonymisation and encryption of personal data,
- means to guarantee the confidentiality of the data,
- means to restore access to data in the event of a physical or technical incident,



- a procedure to regularly evaluate the effectiveness of the security measures implemented to ensure the security of the processing.

In order to best assess the security measures to be implemented, the controller and the processor must take into account the risks that the processing operation pose to the rights and freedoms of the data subjects with regard to potential destruction, loss, disclosure or unauthorized access of personal data.

In order to attest compliance with these security requirements, it is possible to use **an approved code of conduct** (Article 40 GDPR) or **an approved certification mechanism** (Article 42 GDPR) to demonstrate compliance with the security requirements of the GDPR.

5.3 International policies

5.3.1 OECD Recommendation of the Council on Health Data Governance

The [OECD Recommendation on Health Data Governance](https://www.oecd.org/els/health-systems/health-data-governance.htm)¹⁹ (OECD 2019) “recommend[s] that governments establish and implement a national health data governance framework to encourage the availability and use of personal health data to serve health-related public interest purposes while also promoting the protection of privacy, personal health data and data security. The Recommendation aims to support greater harmonisation among the health data governance [...]”.

As part of WP7, we have assessed the OECD Recommendation of the Council on Health Data Governance and identified the following recommendations as most critical for health (research) data:

- Engagement and participation, notably through public consultation, of a wide range of stakeholders
- Encourage common data elements and formats; quality assurance; and data interoperability standards
- Encourage common policies and procedures that minimise barriers to sharing data for health system management, statistics, research, and other health-related purposes that serve the public interest while protecting privacy and data security
- Clear provision of information to individuals
- Informed consent and appropriate alternatives
- Review and approval procedures, as appropriate, for the use of personal health data for research and other health-related public interest purposes
- Transparency, through public information mechanisms which do not compromise health data privacy and security protections or organisations’ commercial or other legitimate interests

¹⁹ <https://www.oecd.org/els/health-systems/health-data-governance.htm>



- Maximising the potential and promoting the development of technology as a means of enabling the availability, re-use and analysis of personal health data while, at the same time, protecting privacy and security and facilitating individuals' control of the uses of their own data
- Monitoring and evaluation mechanisms
- Establishment of appropriate training and skills development in privacy and security measures for those processing personal health data, that are in line with prevailing standards and data processing techniques
- Implementation of controls and safeguards
- Robust identity verification and authentication of individuals accessing personal health data.

5.3.2 GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data and ELSI Toolkit

The Global Alliance for Genomics and Health (GA4GH) is an international initiative “formed in 2013 to accelerate the potential of research and medicine to advance human health. Bringing together 600+ leading organizations working in healthcare, research, patient advocacy, life science, and information technology, the GA4GH community is working together to create frameworks and standards to enable the responsible, voluntary, and secure sharing of genomic and health-related data”²⁰. All of their work builds upon the [Framework for Responsible Sharing of Genomic and Health-Related Data](#) firstly published on December 9th, 2014, and reaffirmed on September 3rd, 2019. A dedicated group on ELSI has been set up since the very beginning of the Alliance to work on international policies with representatives from all the continents which has ended up with the constitution of a Foundational Work Stream dedicated to the Regulatory & Ethics challenges. This group “focuses on the ethical, legal and social implications of international data sharing. Building on a novel human rights framework, the REWS aims to create and harmonize forward-looking consent and privacy policies, and anticipatory data governance models”²¹.

The Framework for Responsible sharing of genomic and health data²²

This document is considered to be one of the drivers for the research activities to be conducted in the GA4GH environment. It is rooted in research ethics principles and fully dedicated to genetic and health data. In order to spread it at the global level the statement has been translated in 17

²⁰ see <https://www.ga4gh.org/about-us/>

²¹ see <https://www.ga4gh.org/how-we-work/2020-2021-roadmap/2020-2021-roadmap-part-ii/rews-2020-2021-roadmap/>

²² see <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/>



languages including English. This statement “is guided by the human rights of privacy, non-discrimination and procedural fairness. At the same time, it considers all human rights principles relevant, complementary and interrelated, founded as they are on respect for human dignity. Since science proceeds only with the broad support of society, respect for all persons is a primary driver underlying all other derived principles. In particular, this Framework establishes a set of foundational principles for responsible research conduct and oversight of research data systems in the realm of genomic and health-related data sharing. It interprets the right of all people to share in the benefits of scientific progress and its applications as being the duty of data producers and users to engage in responsible scientific inquiry and to access and share genomic and health-related data across the translation continuum, from basic research through practical applications. It recognizes the rights of data producers and users to be recognized for their contributions to research, balanced by the rights of those who donate their data. In addition to being founded on the right of all citizens in all countries to the benefits of the advancements of science, and on the right of attribution of scientists, it also reinforces the right of scientific freedom.” (Statement Preamble). It is based on 4 Foundational principles:

1. Respect Individuals, Families and Communities;
2. Advance Research and Scientific Knowledge;
3. Promote Health, Wellbeing and the Fair Distribution of Benefits and
4. Foster Trust, Integrity and Reciprocity.

It also identifies core elements to be considered for responsible data sharing such as: Transparency, Accountability, Data Quality and Security, Privacy, Data Protection and Confidentiality, Risk-Benefit Analysis, Recognition and Attribution, Sustainability, Education and Training and Accessibility and Dissemination.

Even though not binding for the GA4GH members, this document provides for an agreed framework and a common vision for sharing health and genetic data at a global scale where the legal requirements can be very different from one jurisdiction to another. Thus, it can be considered as a common standard for international research as well as for CINECA activities.

Regulatory and ELSI Toolkit²³

The main strength of the GA4GH ELSI group is to work beyond the only achievement of the common Framework and to propose concrete tools for the research teams that are willing to exchange health and genetic data. Part of these tools are not relevant for the CINECA activities as they are related to the practice of medical genetics or to specific diseases such as rare diseases. However, most of them can be of interest for CINECA as they are dealing with consent in genetics, ethics review and data access committees or patients and public engagement. For the needs of this deliverable the

²³ see <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/>



respective tools that could be useful for CINECA members are listed under the recommendations provided in this document (§6).

To conclude, initial collection of health and genetic data must comply with the legal requirements coming from the enforcement of GDPR and from national laws. In that sense several rights must be respected towards privacy and data confidentiality. Several technical solutions exist to mitigate the remaining risks to breach health/genetic data confidentiality such as pseudonymisation as implemented in Beacon or the deployment of synthetic data. However, regarding the later, even though the legal qualification of synthetic data is still under discussion (Fontanillo López and Elbi 2022) the synthetic data is aiming at overcoming the issues of GDPR compliance regarding privacy when they are shared but several additional legal and ethical challenges are remaining (Bhanot et al. 2021). Thus, developers of synthetic data cohorts must pay attention to ensure robustness of machine learning algorithms, fairness in not increasing biases and not creating new discriminations towards (vulnerable) groups.



6 Recommendations for the implementation in the IT framework

Based on the OECD Recommendation of the Council on Health Data Governance (OECD 2019), the GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data (Global Alliance for Genomics and Health 2019), FAIR principles (Wilkinson et al. 2016, Holub et al. 2018), as well as findings of the WP7 analysis, we identified four key areas for recommendations for future projects like CINECA. The recommendations also include and build upon the Data Management Plan (see version 1 DOI²⁴, version 2 DOI²⁵). To increase practicability, we highlight some tools at the end of each recommendation.

6.1 Engagement and benefit sharing as prerequisites for data sharing

- Encourage common policies and procedures that minimize barriers to sharing data for health system management, statistics, research, and other health-related purposes that serve the public interest while protecting privacy and data security.
- Establish appropriate training and skills development in privacy and security measures for those processing personal health data, that are in line with prevailing standards and data processing techniques to improve data quality.
 - o Raise awareness on ELSI and fair data sharing.
 - o Promote and improve the accessibility to training courses (considering factors such as locations, resources, and language).
- Facilitate engagement and participation of a wide range of stakeholders at early stages of the research process and throughout a project's lifecycle, and consider ethical, legal, and societal implications (ELSI) for the establishment of appropriate governance structures for fair data sharing.
 - o Identify and document all relevant stakeholders (internal/external), including their interests, impact, and potential influences.
 - o Develop jointly common standards to build trust and to establish a fair, equitable, transparent, and functioning data-sharing flow between all partners.
 - o Establish equal partnerships with African countries for fair sharing of data and benefits among all involved stakeholders.
 - o Provide due credit and acknowledgement of all who contributed, especially in view of authorship and publication outputs.

²⁴ [10.5281/zenodo.3909576](https://doi.org/10.5281/zenodo.3909576)

²⁵ [10.5281/zenodo.4683415](https://doi.org/10.5281/zenodo.4683415)



- o Consider harms and benefits for data sharing with individuals, families, and communities as well as impact on vulnerable people and data subjects from low resource countries.
- Encourage common data elements and formats, documentation, quality assurance, and data interoperability standards.
 - o Implement interoperable standards for cohort discovery and access.
 - o Process, use, and transfer of data that is accurate, verifiable, unbiased, proportionate to enhance interoperability and replicability, and based on comprehensive documentation.
 - o Consider commonly used ontologies (such as the GA4GH Data Use Ontology (DUO)) to represent machine-readable data use terms to support discovery of datasets, increase interoperability, and support data linkages for secondary analyses.
 - o Adhere to FAIR principles as a guideline for data sharing and reuse for research purposes, also considering institutional conditions and organizational challenges associated with data sharing to ensure responsible and fair data practices. Doing FAIR(ER)/fair data sharing: consider resources to do the work, manage the process, mediate between tool building and approval, assign responsibilities.
 - o Consider a practical working method on how to implement reproducible, FAIR Open Science for deep health research. Sample information or metadata and documentation of the procedures are important requirements for quality, reproducibility, and sustainability in biomedical research and crucial for the implementation of FAIR principles.
 - o Specify quality control measures and responsibility for data management, including resource management
- Data access governed by a Data Access Agreement between the cohort owner, the Principal Investigator, and the Principal Investigator's institute.
- Robust identity verification and authentication of individuals accessing personal health/genetic data. Clear procedures involving Data access Committee should be adopted for controlled access personal health/genetic data.

Tools:

- [CINECA Webinar on Ethics/ELSI considerations - From FAIR to fair data sharing](https://www.cineca-project.eu/news-events-all/ethics/elsi-considerations)²⁶

²⁶ <https://www.cineca-project.eu/news-events-all/ethics/elsi-considerations>



- [CINECA Webinar on International Data Sharing: Fostering Engagement, Transparency and Accountability](#)²⁷
- [Beacon](#)²⁸
- [FairPlus Cookbook](#)²⁹
- [Regulatory and ethics toolkit GA4GH](#)³⁰
 - o [Framework for Responsible Sharing of Genomic and Health-Related Data](#)
 - o [Genetic Discrimination: Implications for Data Sharing Projects \(GeDI\)](#)
 - o [Global Alliance for Genomics and Health: framework for involving and engaging participants, patients and publics in genomics research and health implementation](#)
 - o [Global Alliance for Genomics and Health: Data Access Committee Guiding Principles and Procedural Standards Policy](#)
- [BBMRI-ERIC ELSI Knowledge Base](#)³¹

6.2 Informed consent and reuse of data

- Clear provision of information
 - o Provide clearly defined, accessible, and understandable information on the purpose, collection, use and exchange of data (third parties, international transfer, terms of access and exchange, identifiability and limits to anonymity or confidentiality), processes, procedures, and governance frameworks for data sharing.
 - o Inform how/if legal tools can protect research participants whose (personal) data is shared and if opt-out and withdrawal is possible, how data is used and shared by whom for which purposes, and who is in control, as well as possibilities for returning results to participants.
- Informed consent and appropriate alternatives
 - o Relying on the legal basis for research, as provided for in Article 9 2(j) GDPR, might be the more realistic and appropriate lawful basis in accordance with Article 6 GDPR

²⁷ <https://www.cineca-project.eu/news-events-all/international-data-sharing>

²⁸ <https://beacon-project.io/>

²⁹ <https://faircookbook.elixir-europe.org/content/home.html>

³⁰ <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/>

³¹ <https://www.bbMRI-eric.eu/elsi-knowledge-base/>

to choose for CINECA, while still developing broad/tiered consent as a more ethical approach and as an “appropriate safeguard” (as understood by Article 89 GDPR) to data subjects’ rights, together with increased efforts to comply with the principles of “fairness and transparency” e.g. by increasing efforts to inform data subjects or by limiting research activities to those promoting public interest.

- o Reflect upfront on broad consent as a model that requires research participants to consent to the use of their donations in yet unknown future research projects, especially further processing. Consider risks related to future use and sharing of health data in multiple ways and across countries and potential harms related to the protection of privacy (especially re-identifiability and stigmatization of vulnerable groups and communities). Develop ways on how informed decision-making could be supported by engagement and transparency. This should also be considered in a tiered consent model. Overall include a continuous ethical review process, evaluations of whether the research is in the scope of the broad consent, and continuous information to participants.
- o Consider ethical acceptability of broad consent use in low- and middle-income countries, notably in Africa, with supplementary safeguards elaborated after engagement with the community and after explicit discussions about potential re-identification of genomic data, benefit sharing and commercial use of research results.

Tools:

- [Regulatory and ethics toolkit GA4GH](#)
 - o [Consent Clauses for Large Scale Initiatives](#)
 - o [Data Use Ontology \(DUO\)](#)
 - o [GA4GH Machine-Readable Consent Guidance: How to Map Data Sharing Consent Language to the GA4GH Data Use Ontology](#)
- [CINECA webinar on Webinar: Ethical, legal and societal issues in international data sharing](#)³²
- [BBMRI-ERIC ELSI Knowledge Base](#)

³² <https://www.cineca-project.eu/news-events-all/ethical-legal-and-societal-issues-in-international-data-sharing>



6.3 Safeguards and respect for privacy

- Implement respective mechanisms and safeguards to ensure privacy and to protect patient data from misuse and exploitation, as well as acknowledge the contributions of the research partners.
- Maximise the potential and promote the development of technology as a means of enabling the availability, re-use, and analysis of personal health data while, at the same time, protect privacy and security.
 - o Consider that cohort data are often sensitive and potentially re-identifying individual participants and ensure compliance with ethical requirements, regulatory frameworks, and informed consent conditions.
 - o Be aware of and balance de-identification/anonymization vs. ethical re-identification (e.g., incidental findings), data minimisation principle (GDPR) vs. need for rich data to enable re-use (FAIR), level of granularity needed vs. granularity that could be harmful (vulnerable/ethnic groups).
 - o Specify data security measures and protection of sensitive data to mitigate the risk of unauthorized access, data loss, and misuse.
- Implement monitoring and evaluation mechanisms.
 - o Track chain of data access and/or exchange to its source.
 - o Continuous ethical review process, evaluations of whether that research is in the scope of the consent model.
 - o Adapt to changing conditions (organizational, technological, risk profile etc.).
 - o Assess benefits and risks of harm in data sharing.
 - o Mitigate risks and establish mechanisms for handling complaints related to data misuse, for identifying reporting and managing breaches, and for instituting appropriate sanctions.

Tools:

- [Regulatory and ethics toolkit GA4GH](#)
 - o [Consent Clauses for Genomic Research](#)
 - o [Global Alliance for Genomics and Health: Consent Policy](#)
 - o [Data Use Ontology \(DUO\)](#)



- [CINECA webinar on Webinar: Ethical, legal and societal issues in international data sharing](#)
- [CNIL DPIA](#)
- [BBMRI-ERIC ELSI Knowledge Base](#)

6.4 Further uses and data-access

- Ensure the data included in the catalogue are updated.
- Implement clear procedures for the assessment of the data users' protocols (e.g., through the appointment of a Data Access Committee).
- Review and approval procedures, as appropriate, for the use of personal health data for research and other health-related public interest purposes.
 - o Understand lawful data requests based on legal basis and public health.
 - o Be aware that health data sharing refers to first, the rules to be respected when the health data are collected and used for research (including genetic research) according to the national health law contexts, second the data exchanged under the umbrella of GDPR for scientific purposes in Europe, and third, to health data transfers to third parties (GDPR, art.44).
 - o Be aware that the GDPR does not explicitly foreseen the case of secondary use of data, but Recital 50 provides for the possibility of further use of data for a different purpose, provided that the purposes of processing are compatible. However, according to Article 5(1)b GDPR, further processing of personal data for scientific research “shall not be considered to be incompatible with initial purposes”, which means that no “compatibility test” will have to be conducted for secondary use of data. Relying on this presumption of compatibility nevertheless requires additional safeguards to be put in place as set out in Article 89(1) of the GDPR.
- Increase transparency through public information mechanisms which do not compromise health data privacy and security protections or organisations’ commercial or other legitimate interests.

Tools:

- [CINECA BBMRI Code of conduct Webinar](#)³³
- [Beacon](#)

³³ <https://www.cineca-project.eu/news-events-all/code-of-conduct>



- [FairPlus Cookbook](#)
- [Regulatory and ethics toolkit GA4GH](#)
 - o [Data Privacy and Security Policy](#)
 - o [Framework for Responsible Sharing of Genomic and Health-Related Data](#)
 - o [Ethics Review Recognition Policy](#)

7 Outlook

The considerations and recommendations can be implemented in CINECA through an update in the DMP, but they are particularly intended to guide forthcoming research projects. This outlook highlights upcoming developments for data reuse, especially with artificial intelligence applications on the horizon. Future ELSI issues regarding the reuse of health/genetic data for international data sharing will be embedded in the discussion of two main legal instruments that are rooted in ethical values:

First, in the near future a new EU Regulation will provide for an operational framework to share health/genetic data for clinical care and research. **The proposal made on May 3rd 2022 by the European Commission for a [European Health Data Space](#)³⁴ is ambitious and aims at balancing the protection of individuals' rights and at making data available for the benefit of improving the healthcare of individuals and improving research.** This future European Health Data Space aims to develop a specific ecosystem for health allowing the implementation of rules, practices, common standards, infrastructures, and a governance framework. This space aims to facilitate within Europe the use of health data to improve patient care on the European territory by allowing European citizens to control their own health data in their country but also across borders. Through this objective, the effectiveness of individuals' rights over their data will be strengthened. Moreover, this space aims to promote the re-use of health data for research, innovation and policy-making³⁵.

Second, **artificial intelligence (AI)** will soon be regulated by the [future European Regulation on Artificial Intelligence \(AI Act\)](#)³⁶ which was proposed on April 21, 2021. This future regulation aims to provide harmonized regulation of AI. The approach adopted is that of a risk-based classification because of the existing risks of violation of rights (data quality bias potentially leading to discriminatory bias), and even more so when dealing with sensitive data such as health and genetic

³⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197>

³⁵ For more information: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_fr

³⁶ <https://data.consilium.europa.eu/doc/document/ST-11124-2022-INIT/en/pdf>



data. This future regulation will also establish harmonised rules for the marketing, commissioning, and use of AI systems in order to ensure a high level of protection within the EU and guarantee the respect of EU values and fundamental rights and principles. The EU aims to promote the adoption of ethical and trustworthy AI. The European Commission has already established guidelines on ethical and trustworthy AI systems³⁷ and has published guidelines on the concept of *ethics by design*³⁸ for the design, development and use of ethical and responsible artificial intelligence solutions.

Furthermore, UNESCO has also set out principles and values to be taken into account as a basis for developing AI systems that are ethical, i.e., at the service of individuals, societies, the environment, and ecosystems³⁹. The European Commission has also adopted [two proposals to adapt liability rules to the digital age](#)⁴⁰. One of these proposals aims to harmonise national rules on liability applicable to artificial intelligence in order to make it easier for victims of damage related to artificial intelligence to obtain compensation.

The future regulations on artificial intelligence and on the European Health Data Space will offer an important place to ethical considerations related to the use of artificial intelligence and the use and reuse of health data. Beside regulations, considerations in the field of Ethics of AI highlight key ethical and societal issues of medial AI and AI-assisted clinical applications such as trustworthiness, medial “black boxes”, and bias (Balthazar et al. 2018, Geis et al. 2019, Morley et al. 2020, Ryan and Stahl 2020, Goisauf and Cano Abadía 2022).

To complement these proposals that are not fully dedicated to health and genetic data, the United Nations issued a draft recommendation for tackling the specific issues for the use of health-related data in 2018 (Mandate of the United Nations Special Rapporteur on the Right to Privacy – Task Force on Privacy and the Protection of Health-Related Data 2019). Prepared under the umbrella of the United Nations Special Rapporteur on the Right to Privacy by the Task Force on Privacy and the Protection of Health-Related Data, the document was discussed during a meeting in Strasbourg in 2019 and publicly opened to comments. The new version was adopted in 2019 by the task force group. The aim of this Recommendation is to adopt “guiding principles concerning data processing of health-related data and to emphasise the importance of a legitimate basis of data processing of health-related data by all sectors of society including public authorities and commercial organisations.”. The document recalls the main conditions for health data processing that are mostly relying on the GDPR requirements and make some additional proposals that will be of interest in the future for federated health databases:

First, a full section is dedicated to Genetic Data (7) where it is recommended to be particularly cautious with the collection and use of this data and respectful of the consent of the individuals. It

³⁷ For more information: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

³⁸

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf

³⁹ For more information: <https://en.unesco.org/artificial-intelligence/ethics>

⁴⁰ https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5807



puts an emphasis on the links between the data collection in the medical research sector and the conditions that should be met when this data can be used for the purposes of judicial proceedings which is absolutely new.

Second, the Recommendation insists on the " Indigenous Data Sovereignty" (Chapter IV) that must be scrupulously respected. The main purpose of this Chapter is to enforce the control of indigenous populations over their health data and their usage. It also underlines the right for the representatives of these people to have access to sustainable self-governance mechanisms. These requirements are fully relevant to be considered in future international research activities implying health and genetic data in order to ensure the empowerment of vulnerable populations. Besides, future data use should also be evaluated regarding FAIRER principles.



8 Abbreviations

AI	Artificial Intelligence
CHILD	Canadian Healthy Infant Longitudinal Development
CINECA	Common Infrastructure for National Cohorts in Europe, Canada, and Africa
CNIL	Commission Nationale de l'Informatique et des Libertés
DPIA	Data Protection Impact Assessment
DUO	Data Use Ontology
ECOUTER	Employing Conceptual schema for policy and translation engagement in research
EHDS	European Health Data Space
ELSI	Ethical, legal, and social/societal issues
FAIR	Findable, Accessible, Interoperable, Reusable
FAIRER	Findable, Accessible, Interoperable, Reusable, Ethical, Reproducible
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
GECKO	Genomics Cohorts Knowledge Ontology
H3Africa	Human Heredity and Health in Africa
OECD	Organisation for Economic Co-operation and Development (OECD)
PRS	Polygenic Risk Score

9 Delivery and schedule

D7.3 “First recommendations needed for implementation implemented in IT Framework (incl. a Data Management Plan)” was postponed in order to realize further co-creative activities with the technical WPs and to ensure an appropriate implementation, as well as practicality and feasibility of the deliverable to produce the best possible result.



10 References

- Akyüz, Kaya, Gauthier Chassang, Melanie Goisauf, Łukasz Kozera, Signe Mezinska, Olga Tzortzatou, and Michaela Th Mayrhofer. 2021. "Biobanking and risk assessment: a comprehensive typology of risks for an adaptive risk governance." *Life Sciences, Society and Policy* 17 (1):10. doi: 10.1186/s40504-021-00117-7.
- Austin, Claire C. 2020. "The open science ecosystem: A systematic framework anchored in values, ethics, and FAIRER data." Available at SSRN: <https://ssrn.com/abstract=3654298> or <http://dx.doi.org/10.2139/ssrn.3654298>.
- Balthazar, Patricia, Peter Harri, Adam Prater, and Nabile M. Safdar. 2018. "Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics." *Journal of the American College of Radiology* 15 (3, Part B):580-586. doi: <https://doi.org/10.1016/j.jacr.2017.11.035>.
- Bhanot, Karan, Miao Qi, John S. Erickson, Isabelle Guyon, and Kristin P. Bennett. 2021. "The Problem of Fairness in Synthetic Healthcare Data." *Entropy* 23 (9):1165.
- Clarke, Victoria, Virginia Braun, and Nikki Hayfield. 2015. "Thematic analysis." In *Qualitative psychology: A practical guide to research methods*, edited by Jonathan A. Smith, 222-248. Los Angeles, London, New Delhi, Singapore, Washington DC, Boston: Sage.
- Dankar, Fida K, Andrey Ptitsyn, and Samar K Dankar. 2018. "The development of large-scale de-identified biomedical databases in the age of genomics—principles and challenges." *Human genomics* 12 (1):19. doi: <https://doi.org/10.1186/s40246-018-0147-5>.
- Fontanillo López, César Augusto, and Abdullah Elbi. 2022. "On synthetic data: a brief introduction for data protection law dummies." *European Law Blog*, 11.10.2022. <https://europeanlawblog.eu/2022/09/22/on-synthetic-data-a-brief-introduction-for-data-protection-law-dummies/>.
- Garrison, Nanibaa' A. 2013. "Genomic Justice for Native Americans: Impact of the Havasupai Case on Genetic Research." *Science, Technology, & Human Values* 38 (2):201-223. doi: 10.1177/0162243912470009.
- Geis, J. Raymond, Adrian P. Brady, Carol C. Wu, Jack Spencer, Erik Ranschaert, Jacob L. Jaremko, Steve G. Langer, Andrea Borondy Kitts, Judy Birch, William F. Shields, Robert van den Hoven van Genderen, Elmar Kotter, Judy Wawira Gichoya, Tessa S. Cook, Matthew B. Morgan, An Tang, Nabile M. Safdar, and Marc Kohli. 2019. "Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement." *Canadian Association of Radiologists Journal* 70 (4):329-334. doi: 10.1016/j.carj.2019.08.010.
- Global Alliance for Genomics and Health. 2019. Framework for Responsible Sharing of Genomic and Health-Related Data.
- Goisauf, Melanie, Kaya Akyüz, and Gillian M. Martin. 2020. "Moving back to the future of big data-driven research: reflecting on the social in genomics." *Humanities and Social Sciences Communications* 7 (1):55. doi: 10.1057/s41599-020-00544-5.
- Goisauf, Melanie, and Mónica Cano Abadía. 2022. "Ethics of AI in Radiology: A Review of Ethical and Societal Implications." *Frontiers in Big Data* 5 (850383). doi: 10.3389/fdata.2022.850383.
- Goisauf, Melanie, and Anna Durnová. 2018. "From Engaging Publics to Engaging Knowledges: Enacting "Appropriateness" in the Austrian Biobank Infrastructure." *Public Understanding of Science* 28 (3):275-289. doi: 10.1177/0963662518806451.
- Goisauf, Melanie, Gillian Martin, Heidi Beate Bentzen, Isabelle Budin-Ljøsne, Lars Ursin, Anna Durnová, Liis Leitsalu, Katharine Smith, Sara Casati, Marialuisa Lavitrano, Deborah



- Mascalzoni, Martin Boeckhout, and Michaela Th. Mayrhofer. 2019. "Data in Question: A Survey of European Biobank Professionals on Ethical, Legal and Societal Challenges of Biobank Research." *PLOS ONE* 14 (9):e0221496. doi: 10.1371/journal.pone.0221496.
- Gottweis, Herbert, and Jane Kaye. 2012. Biobanks for Europe. A challenge for governance. Report of the Expert Group on Dealing with Ethical and Regulatory Challenges of International Biobank Research. Brussels: European Commission Directorate-General for Research and Innovation.
- H3Africa Working Group on Ethics and Regulatory Issues for the Human Heredity and Health in Africa (H3Africa) Consortium. 2017. "H3Africa Guideline for Informed Consent." https://h3africa.org/wp-content/uploads/2018/05/H3A%202017%20Revised%20IC%20guideline%20for%20SC%2020_10_2017.pdf.
- Helgesson, Gert. 2012. "In Defense of Broad Consent." *Cambridge Quarterly of Healthcare Ethics* 21 (1):40-50. doi: 10.1017/S096318011100048X.
- Hemminki, Elina, Aaro Tupasela, Piia Jallinoja, Arja R Aro, Karoliina Snell, and Sinikka Sihvo. 2009. "Finnish people's attitudes towards biomedical research and its sponsorship." *Genomics, Society and Policy* 5 (2):67-79. doi: <https://doi.org/10.1186/1746-5354-5-2-67>.
- Holm, Søren, and Thomas Ploug. 2019. "Genome studies reveal flaws in broad consent." *Science* 366 (6472):1460-1461. doi: 10.1126/science.aaz3797.
- Holub, Petr, Florian Kohlmayer, Fabian Prasser, Michaela Th. Mayrhofer, Irene Schlünder, Gillian M. Martin, Sara Casati, Lefteris Koumakis, Andrea Wutte, Łukasz Kozera, Dominik Strapagiel, Gabriele Anton, Gianluigi Zanetti, Osman Ugur Sezer, Maimuna Mendy, Dalibor Valík, Marialuisa Lavitrano, Georges Dagher, Kurt Zatloukal, GertJan B. van Ommen, and Jan-Eric Litton. 2018. "Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health." *Biopreservation and Biobanking* 16 (2):97-105. doi: 10.1089/bio.2017.0110.
- Joly, Yann, Ida Ngueng Feze, Lingqiao Song, and Bartha M. Knoppers. 2017. "Comparative Approaches to Genetic Discrimination: Chasing Shadows?" *Trends in Genetics* 33 (5):299-302. doi: <https://doi.org/10.1016/j.tig.2017.02.002>.
- Kasperbauer, TJ, Mickey Gjerris, Gunhild Waldemar, and Peter Sandøe. 2018. "Communicating Identifiability risks to Biobank donors." *Cambridge Quarterly of Healthcare Ethics* 27:123-136. doi: 10.1017/S0963180117000457.
- Knoppers, Bartha M., Alexander Bernier, Palmira Granados Moreno, and Nora Pashayan. 2021. "Of Screening, Stratification, and Scores." *Journal of Personalized Medicine* 11 (8):736.
- Lewis, Anna C. F., and Robert C. Green. 2021. "Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues." *Genome Medicine* 13 (1):14. doi: 10.1186/s13073-021-00829-7.
- Mandate of the United Nations Special Rapporteur on the Right to Privacy – Task Force on Privacy and the Protection of Health-Related Data. 2019. "Recommendation on the protection and use of health-related data." https://www.ohchr.org/sites/default/files/Documents/Issues/Privacy/SR_Privacy/UNSRPhealthrelateddataRecCLEAN.pdf.
- Mester, Jessica L., MaryBeth Mercer, Aaron Goldenberg, Rebekah A. Moore, Charis Eng, and Richard R. Sharp. 2015. "Communicating with Biobank Participants: Preferences for Receiving and Providing Updates to Researchers." *Cancer Epidemiology Biomarkers & Prevention* 24 (4):708-712. doi: 10.1158/1055-9965.Epi-13-1375.
- Middleton, Anna, Richard Milne, Mohamed A Almarri, Shamim Anwer, Jerome Atutornu, Elena E Baranova, Paul Bevan, Maria Cerezo, Yali Cong, and Christine Critchley. 2020. "Global Public



- Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data?" *The American Journal of Human Genetics* 107 (4):743-752. doi: <https://doi.org/10.1016/j.ajhg.2020.08.023>.
- Mittelstadt, Brent Daniel, and Luciano Floridi. 2016. "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts." *Science and Engineering Ethics* 22 (2):303-341. doi: 10.1007/s11948-015-9652-2.
- Moodley, Keymanthri, and Anita Kleinsmidt. 2021. "Allegations of misuse of African DNA in the UK: Will data protection legislation in South Africa be sufficient to prevent a recurrence?" *Developing World Bioethics* 21 (3):125-130. doi: 10.1111/dewb.12277.
- Morley, Jessica, Caio C. V. Machado, Christopher Burr, Josh Cows, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. 2020. "The ethics of AI in health care: A mapping review." *Social Science & Medicine* 260:113172. doi: <https://doi.org/10.1016/j.socscimed.2020.113172>.
- Murtagh, Madeleine J., Joel T. Minion, Andrew Turner, Rebecca C. Wilson, Mwenza Blell, Cynthia Ochieng, Barnaby Murtagh, Stephanie Roberts, Oliver W. Butters, and Paul R. Burton. 2017. "The ECOUTER methodology for stakeholder engagement in translational research." *BMC Medical Ethics* 18 (1):24. doi: 10.1186/s12910-017-0167-z.
- Musen, Mark A. 2022. "Without appropriate metadata, data-sharing mandates are pointless." *Nature* 609 (222). doi: <https://doi.org/10.1038/d41586-022-02820-7>.
- O'Doherty, Kieran C., Mahsa Shabani, Edward S. Dove, Heidi Beate Bentzen, Pascal Borry, Michael M. Burgess, Don Chalmers, Jantina De Vries, Lisa Eckstein, Stephanie M. Fullerton, Eric Juengst, Kazuto Kato, Jane Kaye, Bartha Maria Knoppers, Barbara A. Koenig, Spero M. Manson, Kimberly M. McGrail, Amy L. McGuire, Eric M. Meslin, Dianne Nicol, Barbara Prainsack, Sharon F. Terry, Adrian Thorogood, and Wylie Burke. 2021. "Toward better governance of human genomic data." *Nature Genetics* 53 (1):2-8. doi: 10.1038/s41588-020-00742-6.
- OECD. 2019. Recommendation of the Council on Health Data Governance, OECD/LEGAL/0433.
- Ryan, Mark, and Bernd Carsten Stahl. 2020. "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications." *Journal of Information, Communication and Ethics in Society* 19 (1):61-86. doi: 10.1108/JICES-12-2019-0138.
- Schlegel, Daniel R, and Gregoire Ficheur. 2017. "Secondary use of patient data: review of the literature published in 2016." *Yearbook of medical informatics* 26 (01):68-71. doi: 10.15265/IY-2017-032.
- Shabani, Mahsa, and Pascal Borry. 2018. "Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation." *European journal of human genetics : EJHG* 26 (2):149-156. doi: 10.1038/s41431-017-0045-7.
- Slunecka, John L., Matthijs D. van der Zee, Jeffrey J. Beck, Brandon N. Johnson, Casey T. Finnicum, René Pool, Jouke-Jan Hottenga, Eco J. C. de Geus, and Erik A. Ehli. 2021. "Implementation and implications for polygenic risk scores in healthcare." *Human Genomics* 15 (1):46. doi: 10.1186/s40246-021-00339-y.
- Tindana, Paulina, and Jantina de Vries. 2016. "Broad Consent for Genomic Research and Biobanking: Perspectives from Low- and Middle-Income Countries." *Annual Review of Genomics and Human Genetics* 17 (1):375-393. doi: 10.1146/annurev-genom-083115-022456.
- Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, and Philip E Bourne. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3. doi: 10.1038/sdata.2016.18.



11 Appendices

11.1 Appendix 1: Poster “Genetic data sharing for research in Europe: consent not only a legal basis under GDPR”, European Society of Human Genetics Conference 2022, Lisa Feriol, Gauthier Chassang, Emmanuelle Rial-Sebbag

Genetic data sharing for research in Europe: consent not only a legal basis under GDPR

