



HEREDITARY

HetERogeneous sEmantic Data integration for the guT-bRain
interplaY

Deliverable 2.2

Ethical guidelines, data collection and sharing, revision

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.



**Funded by
the European Union**

EXECUTIVE SUMMARY

This deliverable presents the updated version of Deliverable 2.1 “Ethical Guidelines, Data Collection and Sharing” and provides an updated overview of the legal, ethical, regulatory, and data governance framework applicable to the HEREDITARY project. Since the submission of D2.1 in June 2024, several important developments have occurred, including the implementation of Regulation (EU) 2024/1689 (AI Act), the evolution of the European Health Data Space (EHDS), increasing cybersecurity requirements associated with the NIS2 Directive, and the operational development of semantic, FAIRification, and federated discovery infrastructures within HEREDITARY. The deliverable reviews these developments and assesses their implications for multimodal health data processing, secondary use of health data, distributed data discovery, privacy-preserving analytics, and AI governance.

In addition, the document reports updates concerning ethical approvals, data management, and governance procedures across consortium partners, including the role of the Data Management Plan, FAIRification activities, the HEREDITARY Ontology (HERO), and Hereditary Data Network (HDN). It also introduces a prospective Optical Coherence Tomography (OCT) cohort of Amyotrophic Lateral Sclerosis (ALS) patients at the University of Turin, currently in the preparatory and ethics approval phase. Once approved, the cohort will support future multimodal and multicenter analyses involving ALS, Parkinson’s disease, and Multiple Sclerosis datasets available within the consortium. Overall, D2.2 confirms the continued alignment of HEREDITARY with applicable European and national legal, ethical, and regulatory requirements while supporting privacy-preserving, trustworthy, and responsible health data research.

DOCUMENT INFORMATION

Deliverable ID	D2.2
Deliverable Title	Ethical guidelines, data collection and sharing, revision
Work Package	WP2
Lead Partner	UNITO
Due date	30/06/2026
Date of submission	26/06/2026
Type of deliverable	R
Dissemination level	PU

AUTHORS

Name	Organization
Umberto Manera	UNITO
Adriano Chiò	UNITO
Carla D'Agostino	UNITO
Marcos Casado Barbero (Internal reviewer)	EMBL
Mallory Freeberg (Internal reviewer)	EMBL
Maja Nisevic	KUL
Anna Romanovych (Internal reviewer)	UNIPD

REVISION HISTORY

Version	Date	Author	Document history/approvals
V0.1	05.06.2026	UNITO	First draft for internal review submission
V0.2	09.06.2026	EMBL	Partner's review
V0.3	10.06.2026	KU Leuven	Partner's review
V1.0	21.06.2026	UNITO	Final formatting edition
V1.1	24.06.2026	UNITO	Final formatting with minor correction
V1.2	26.06.2026	UNIPD	Finalization of the layout

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Contents

1	INTRODUCTION	8
1.1	Relationship with D2.1 and D7.1	8
2	BACKGROUND AND REGULATORY EVOLUTION	9
2.1	Evolution from D2.1	9
2.2	Ethics Summary Report (ESR) considerations	9
2.3	Evolution of the European legal and regulatory framework since June 2024	10
2.3.1	AI Act	10
2.3.2	European Health Data Space	11
2.3.3	NIS2 Directive and cybersecurity	11
2.3.4	Emerging EHDS metadata and interoperability standards	12
3	ETHICS AND LEGAL FRAMEWORKS UPDATES	13
3.1	GDPR and federated processing updates	13
3.2	AI governance and AI Act considerations	14
3.2.1	Evolution from D2.1	14
3.2.2	AI Systems Developed within HEREDITARY	15
3.2.3	Potential AI Act Relevance and High-Risk Considerations	16
3.2.4	Future Translational Scenarios	18
3.3	EHDS, semantic interoperability and secondary use of health data	18
3.4	Cybersecurity and infrastructure protection	20
3.5	Governance and ethical oversight updates	22
3.5.1	Evolution of the Governance Framework	22
3.5.2	Governance-by-Design in the HEREDITARY Infrastructure	23
3.5.3	AI Governance and Continuous Ethical Monitoring	23
4	PROSPECTIVE DATA COLLECTION	24
4.1	Rationale and objectives	24
4.2	Prospective ALS-OCT Cohort in Turin	25
4.3	Multimodal and Multicenter OCT Integration	26
5	ETHICAL APPROVALS UPDATES	26
6	CHALLENGES AND LIMITATIONS	30
7	CONCLUSIONS	30
	REFERENCES	33

List of Tables

Table 1 Consolidated overview of ethical approvals, waivers and authorizations supporting HEREDITARY activities. CET, Comitato Etico Territoriale. 27

List of Acronyms

Acronym	Full word
AAI	Anschutz Accelerated Initiative
AI	Artificial Intelligence
ALS	Amyotrophic Lateral Sclerosis
ALSFRS-R	Amyotrophic Lateral Sclerosis Functional Rating Scale – Revised
CET	Comitato Etico Territoriale (Local Ethical Committee)
CMO	Commissie Mensgebonden Onderzoek (Committee on Research Involving Human Subjects)
COMIRB	Colorado Multiple Institutional Review Board
CORIS	Clinical Ophthalmology Retrospective Imaging System
DCAT-AP	Data Catalogue Vocabulary Application Profile
DGA	Data Governance Act
EHDS	European Health Data Space
EHR	Electronic Health Record
ESR	Ethic summary report
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
fMRI	Functional Magnetic Resonance Imaging
GDI	Genomic Data Infrastructure
GDPR	General Data Protection Regulation
HDN	Hereditary Data Network
HealthDCAT-AP	Health Data Catalogue Application Profile

HERO	Hereditary Ontology
HLEG	High-Level Expert Group on Artificial Intelligence
MS	Multiple Sclerosis
NIS 2 Directive	Network Information Security 2 Directive
OBDA	Ontology-Based Data Access
OBDF	Ontology-Based Data Federation
OBO	Open Biological and Biomedical Ontologies
OCT	Optical Coherence Tomography
OECD	Organisation for Economic Co-operation and Development
OMOP-CDM	Observational Medical Outcomes Partnership - Common Data Model
PARALS	Piemonte and Valle d'Aosta Amyotrophic Lateral Sclerosis Register
RDF	Resource Description Framework
RNFL	Retinal Nerve Fibre Layer
RUMC	Radboud University Medical Centre
SaMD	Software as a Medical Device
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
UCD	University of Colorado Denver
UNIPD	University of Padova
UNITO	University of Turin
WMO	Wet Medisch-Wetenschappelijk Onderzoek (Medical Research Involving Human Subjects Act)
WP	Work Package

1 INTRODUCTION

Deliverable D2.2 constitutes the revision of Deliverable D2.1¹ “Ethical Guidelines, Data Collection and Sharing” developed within Task 2.1 of the HEREDITARY project. The objective of this deliverable is to provide updated guidance on the legal, ethical, and regulatory aspects relevant to the collection, processing, sharing, and analysis of multimodal health data within the HEREDITARY framework.

The document reflects the operational evolution of the project since the submission of D2.1 in June 2024 and incorporates updates stemming from:

- the operational implementation of federated learning infrastructures;
- the activities carried out within WP7;
- evolving European legislation and governance frameworks;
- the implementation of prospective data collection activities;
- and the ethical and regulatory requirements applicable to AI-driven health research.

1.1 Relationship with D2.1 and D7.1

D2.1 established the foundational ethical and legal framework applicable to HEREDITARY activities, focusing primarily on data protection, AI ethics, federated learning principles, and the legal context surrounding multimodal health data processing.

During the first phase of the project implementation (M0-M6), additional operational, regulatory, and governance considerations were raised, particularly concerning:

- the implementation of federated learning architectures;
- governance of AI systems and model sharing;
- secondary use of health data;
- cybersecurity obligations;
- prospective data collection for model validation.

In parallel, Deliverable D7.1² provided a detailed inventory of the legal, ethical, and regulatory frameworks applicable to the project, including the General Data Protection Regulation (GDPR), AI Act, EHDS, Data Governance Act, and cybersecurity legislation.

This revision deliverable focuses on operational updates and implementation-oriented aspects starting from the general legal and ethical overview already provided in D2.1 and D7.1.

¹ <https://doi.org/10.5281/zenodo.14033771>

² <https://zenodo.org/records/14628224>

2 BACKGROUND AND REGULATORY EVOLUTION

2.1 Evolution from D2.1

Deliverable D2.1³ provided the initial ethical and legal framework applicable to the HEREDITARY project and introduced the core principles governing the collection, processing, sharing, and analysis of multimodal health data.

Since its submission, the HEREDITARY consortium has progressed from a conceptual and governance-definition phase toward the operational implementation of federated infrastructures, multimodal analytics pipelines, and clinical use-case activities.

This operational transition highlighted the need to further refine several aspects related to:

- governance of federated learning environments;
- cybersecurity and infrastructure protection;
- management, stewardship, and long-term sustainability of health data resources;
- governance of secondary use and controlled access to health data;
- AI governance and transparency;
- prospective data collection for model validation;
- and ethical oversight procedures.

In addition, the operational deployment of HEREDITARY highlighted the importance of data stewardship and long-term sustainability considerations. Beyond enabling access to existing datasets during the project lifetime, the consortium increasingly considers how multimodal resources, metadata, semantic models, and federated infrastructures can remain discoverable, reusable, and maintainable after the end of the project. These aspects are closely linked to FAIRification activities, emerging European initiatives such as the EHDS and HealthData@EU⁴, and the sustainability and exploitation activities carried out within WP8.

Furthermore, the activities carried out within WP7 enabled a more detailed understanding of the evolving European legal and regulatory framework applicable to HEREDITARY.

2.2 Ethics Summary Report (ESR) considerations

The Ethics Summary Report (ESR)⁵ accompanying the HEREDITARY project identified several areas requiring continuous monitoring and assessment during the project lifecycle.

In particular, the ESR highlighted the importance of:

- preventing potential misuse of AI algorithms and open-source outputs;
- ensuring appropriate governance of federated learning infrastructures;

³ <https://zenodo.org/records/14033771>

⁴ European Commission. (2025). HealthData@EU: European infrastructure for the secondary use of health data. European Health Information Portal. <https://www.healthinformationportal.eu/european-health-data-space>

⁵ See ESR in D2.2 <https://zenodo.org/records/14033771>

- monitoring possible implications related to international collaboration and data processing;
- verifying compliance of prospective data collection activities with ethical and data protection requirements;
- evaluating potential implications of AI systems within clinical validation scenarios.

In response to these considerations, the HEREDITARY consortium implemented additional governance measures including:

- strengthened ethical oversight procedures;
- continuous monitoring of legal and regulatory developments;
- controlled access procedures for federated infrastructures;
- enhanced documentation and transparency measures;
- and integration of privacy-preserving and cybersecurity safeguards within the technical architecture.

2.3 Evolution of the European legal and regulatory framework since June 2024

Since the submission of D2.1, several relevant European legislative and regulatory developments have taken place.

2.3.1 AI Act

A major development concerns the adoption and entry into force of Regulation (EU) 2024/1689⁶ laying down harmonized rules on artificial intelligence (AI Act). While the Regulation entered into force in 2024, its provisions will apply gradually over a transitional period. In parallel, ongoing EU regulatory simplification initiatives, including the Digital Omnibus package⁷, may further influence the implementation and practical application of the EU digital regulatory framework.

The AI Act establishes a *risk-based regulatory framework* for artificial intelligence systems within the European Union, introducing obligations that vary according to the level of risk associated with a specific AI application. The Regulation places particular emphasis on the implementation of risk management processes, transparency requirements, human oversight mechanisms, data governance practices, technical robustness, and cybersecurity safeguards. These principles aim to ensure that AI systems are developed and deployed in a manner that is trustworthy, safe, and respectful of fundamental rights, while fostering innovation and legal certainty across the European digital ecosystem. The D7.2 provides a detailed assessment of whether HEREDITARY

⁶ European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union, L, 2024/1689. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32024R1689>

⁷ European Commission. (2025). Commission communication on the Digital Omnibus Package: Simplification and coherence of the EU digital regulatory framework. European Commission. <https://digital-strategy.ec.europa.eu>

qualifies as a high-risk AI system or not, as well as an explanation of the meaning and scope of the research-purpose exception under the AI Act.

2.3.2 European Health Data Space

A major development since the submission of D2.1 concerns the adoption and entry into force of Regulation (EU) 2025/327 on the European Health Data Space (EHDS). The EHDS establishes a common European framework for the use and exchange of electronic health data, supporting both primary use of health data for healthcare delivery and secondary use for research, innovation, policy-making, regulatory activities, and public health purposes. The EHDS aims to improve the accessibility, quality, interoperability, and reuse of health data across Member States through the establishment of common governance mechanisms, harmonized standards, and dedicated infrastructures supporting secure cross-border access to health information.

The regulation introduces provisions concerning electronic health record interoperability, health data access bodies, trusted environments for secondary use of health data, and the development of the HealthData@EU infrastructure to facilitate research and innovation activities at the European level.

The EHDS is particularly relevant for research projects such as HEREDITARY because it reinforces the importance of interoperability, metadata standardization, secure secondary use of health data, and FAIRification, namely the process of making data and metadata Findable, Accessible, Interoperable, and Reusable according to the FAIR Guiding Principles⁸. Furthermore, the regulation promotes the development of governance frameworks and technical infrastructures capable of enabling large-scale health data analyses while preserving privacy, security, and institutional responsibilities.

The implementation of the EHDS is expected to significantly influence the future European health data ecosystem, providing new opportunities and requirements for projects relying on multimodal health data integration, federated analytics, and collaborative biomedical research.

2.3.3 NIS2 Directive and cybersecurity

Another important regulatory development since the submission of D2.1 concerns the progressive implementation of Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (the Network and Information System, NIS2 Directive)⁹. The NIS2 Directive establishes a comprehensive European framework aimed at strengthening cybersecurity, resilience, risk management, and incident reporting across a wide range of critical and important sectors, including healthcare and digital infrastructures. Compared with its predecessor, NIS2 expands the scope of organizations subject to cybersecurity obligations and introduces more stringent requirements concerning governance, risk management, supply chain security, business

⁸ Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Erratum in: *Sci Data*. 2019 Mar 19;6(1):6. doi: 10.1038/s41597-019-0009-6.

⁹ European Parliament and Council of the European Union. (2022). *Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)*. Official Journal of the European Union, L333, 80–152. <https://eur-lex.europa.eu/eli/dir/2022/2555/oj>

continuity, incident handling, access control, and security monitoring. The Directive also reinforces the responsibilities of organizations in ensuring the confidentiality, integrity, availability, and resilience of network and information systems supporting essential services.

The implementation of NIS2 is particularly relevant for research and innovation projects operating in the health domain, where increasing volumes of sensitive health and genomic data are processed through distributed digital infrastructures, as we also pointed out in D7.1 and D7.2. The Directive highlights the importance of adopting robust cybersecurity governance frameworks capable of supporting secure data sharing, interoperability, federated analytics, and cross-institutional collaboration while mitigating cybersecurity risks and ensuring operational resilience. As European health data ecosystems continue to evolve through initiatives such as the European Health Data Space (EHDS) and HealthData@EU, NIS2 provides an important regulatory foundation for the secure operation of interconnected digital infrastructures supporting health research and innovation.

2.3.4 Emerging EHDS metadata and interoperability standards

Another important development since the submission of D2.1 concerns the emergence of operational tools and standards supporting the implementation of the EHDS. In particular, the publication of HealthDCAT-AP¹⁰ represents a significant step towards the harmonization of metadata descriptions for health datasets across Europe. HealthDCAT-AP (v7¹¹) extends the Data Catalogue Vocabulary Application Profile (DCAT-AP) and provides a common framework for describing health datasets intended for secondary use, improving discoverability, interoperability, and reuse within European health data ecosystems.

The HealthDCAT-AP framework is particularly relevant for HEREDITARY because it promotes standardized metadata descriptions, FAIRification practices, and alignment with future EHDS requirements for health data access and secondary use. The adoption of common metadata specifications facilitates the integration of multimodal health resources and improves interoperability between research infrastructures, data repositories, and federated analytics platforms.

In parallel, the European Commission has released validation tools supporting compliance with emerging EHDS metadata specifications. Notably, the EHDS RDF Validator¹² enables the validation of metadata records against HealthDCAT-AP SHACL-based semantic constraints¹³, supporting data providers in assessing the consistency and quality of metadata descriptions prior to publication and exchange. Such initiatives contribute to the development of trustworthy and interoperable health data infrastructures and are aligned with HEREDITARY activities related to FAIRification, federated access, and secure secondary use of multimodal health data. Additional developments concerning the Data Governance Act (DGA) and broader European data

¹⁰ European Commission. (2025). *HealthDCAT-AP: Health Data Catalogue Application Profile*. European Health Information Portal. <https://www.healthinformationportal.eu/healthdcat-ap>

¹¹ <https://healthdataeu.pages.code.europa.eu/healthdcat-ap/releases/release-7>

¹² European Commission. (2025). EHDS RDF Validator. European Commission. <https://health-data-itb-rdf-validator.acceptance.data.health.europa.eu/shacl/ehds/upload>

¹³ <https://healthdataeu.pages.code.europa.eu/healthdcat-ap/releases/release-7/#validation-of-healthdcat-ap>

governance initiatives further emphasized the relevance of trusted data sharing ecosystems, interoperability, and secure reuse of sensitive data for research purposes.

Further developments include the gradual deployment of HealthData@EU, the infrastructure supporting cross-border secondary use of health data under the European Health Data Space. Together with ongoing work on implementing acts, common specifications, and interoperability frameworks, these initiatives are expected to provide greater legal and technical certainty regarding access to and reuse of health data for research purposes. They further reinforce the importance of standardized metadata, semantic interoperability, trusted governance mechanisms, and privacy-preserving approaches, all of which are central to the HEREDITARY federated data ecosystem.

These developments support the overall HEREDITARY strategy based on federated analytics, controlled data access, and privacy-preserving infrastructures.

3 ETHICS AND LEGAL FRAMEWORKS UPDATES

3.1 GDPR and federated processing updates

The processing of personal data and special categories of personal data within the HEREDITARY project continues to be governed by the GDPR¹⁴ and applicable national legislation. The legal and ethical foundations of personal data processing, including the concepts of personal data, pseudonymization, anonymization, lawful bases for processing, data subject rights, purpose limitation, data minimization, accountability, and processing for scientific research purposes, were extensively analyzed in Deliverable D2.1, particularly in Sections 2.1–2.3, Section 4.3.1 (General Data Protection Regulation), and Section 5.2 (Ethical Guidelines and Procedures), in D7.1 and D7.2 and its annexes.

The conclusions reached in D2.1 remain fully applicable to the project. In particular, HEREDITARY continues to process health, genomic, imaging, clinical, and environmental data under a privacy-by-design and privacy-by-default approach, combining organizational and technical safeguards with continuous legal and ethical oversight. The principles of lawfulness, fairness, transparency, purpose limitation, data minimization, storage limitation, integrity, confidentiality, and accountability remain the cornerstone of all data processing activities performed within the project. For example, these principles are reflected in the use of pseudonymization, restricted access to datasets and computing environments, local institutional control over data resources, and federated approaches that reduce unnecessary transfers of patient-level information.

Since June 2024, no substantial amendments to the GDPR framework have entered into force that would significantly alter the legal basis of HEREDITARY data processing activities. Nevertheless, several developments have reinforced the importance of robust governance mechanisms for distributed and federated data processing, particularly in

¹⁴ European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. Official Journal of the European Union, L119, 1–88. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

relation to the implementation of the European Health Data EHDS¹⁵, the increasing operationalization of federated learning infrastructures, and the broader European policy framework concerning secondary use of health data.

The federated learning and analytics infrastructures developed within HEREDITARY continue to be designed according to the principles of data minimization, privacy by design, and institutional governance. In accordance with the principles described in D2.1, sensitive data remain under the control of the originating institutions, which retain responsibility for the governance and processing of their datasets. Pseudonymization measures are applied before processing activities, access to datasets and computational infrastructures is restricted to authorized personnel, and appropriate technical and organizational safeguards are implemented throughout the data lifecycle. Within this framework, federated learning and distributed analytics provide an additional privacy-preserving approach that can reduce the need for transfers of patient-level information while supporting collaborative analyses across institutions.

The consortium recognizes that federated learning architectures do not eliminate GDPR obligations. Federated analytics and federated learning represent one possible approach for enabling collaborative analyses across institutions while limiting the exchange of patient-level information. However, they do not replace the need for appropriate governance frameworks, lawful processing bases, access control mechanisms, accountability procedures, and data protection safeguards. Consequently, HEREDITARY continues to evaluate the legal and ethical implications of federated processing models in collaboration with WP7 and in light of evolving European initiatives concerning health data sharing, interoperability, secondary use of health data, and distributed research infrastructures.

Furthermore, the emergence of metadata and interoperability frameworks supporting the future implementation of the EHDS, including HealthDCAT-AP and associated validation tools, reinforces the importance of standardized descriptions of health datasets and FAIRification practices. These developments complement the GDPR framework by facilitating discoverability and interoperability while maintaining appropriate safeguards for personal data processing within distributed research environments.

3.2 AI governance and AI Act considerations

3.2.1 Evolution from D2.1

Since the submission of D2.1, the European AI governance landscape has evolved substantially following the adoption and entry into force of Regulation (EU) 2024/1689 (AI Act)¹⁶. While D2.1 primarily discussed AI ethics through the High-Level Expert Group on Artificial Intelligence (AI HLEG) Guidelines for Trustworthy AI¹⁷ and the Organization

¹⁵ ELI: <http://data.europa.eu/eli/reg/2025/327/oj>

¹⁶ European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

¹⁷ European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. <https://digital-strategy.ec.europa.eu>

for Economic Co-operation and Development (OECD) AI Principles¹⁸, the AI Act now provides a legally binding framework based on a risk-oriented approach to the development and deployment of AI systems.

Consequently, HEREDITARY complements its ethical framework with a continuous assessment of the requirements introduced by the AI Act and their potential relevance to project activities. Although the AI methods developed within HEREDITARY are intended for research and validation purposes and are not designed to autonomously perform clinical decision-making, the consortium continuously monitors regulatory developments related to AI governance, transparency, human oversight, data governance, robustness, and accountability.

In this context, several HEREDITARY developments contribute to the implementation of trustworthy AI principles. The Hereditary Ontology (HERO), introduced in D3.1 and further extended through future semantic modelling activities, supports semantic interoperability, transparency, and consistent representation of multimodal health data. The Hereditary Data Network (HDN), described in D3.2 as the first release of the federated workflow execution engine, provides mechanisms for controlled federated analytics, provenance tracking, auditability, and governance of distributed data processing. These components will be further complemented by the M30 privacy-preserving analytics deliverable D3.8 and the M36 semantic multimodal ontology deliverable D3.10, which will respectively address privacy-preserving algorithms and broader ontological modelling across the considered diseases.

3.2.2 AI Systems Developed within HEREDITARY

HEREDITARY employs AI and machine learning approaches for multimodal analytics, knowledge extraction, disease prediction, patient stratification, federated learning, ontology-supported inference, and risk factor identification.

HEREDITARY employs AI and machine learning approaches for multimodal analytics, knowledge extraction, disease prediction, patient stratification, federated learning, ontology-supported inference, and risk factor identification. Concrete examples include the ALS-focused activities in Use Cases 1 and 2, which integrate clinical, genomic, neuroimaging, and disease progression data to support patient stratification and the identification of disease-related patterns (included in D2.17); the OCT-based Parkinson's disease analyses described in D2.8¹⁹, where deep learning models were explored for the identification of retinal imaging biomarkers; and the gut–brain axis activities in Use Cases 4 and 5, which combine histopathology, microbiome, imaging, questionnaire, and clinical data to investigate multimodal associations relevant to neurological and gastrointestinal conditions (described in D2.3²⁰ and D2.4²¹). These activities are supported by the HERO ontology and related semantic modelling work described in D3.1²², which facilitates the harmonization and interpretation of heterogeneous data sources, and by the federated workflow and Hereditary Data Network (HDN) architecture described in D3.2²³, which

¹⁸ Organization for Economic Co-operation and Development. (2019). *OECD principles on artificial intelligence*. OECD Publishing. <https://oecd.ai/en/ai-principles>

¹⁹ <https://zenodo.org/records/15847669>

²⁰ <https://zenodo.org/records/14627764>

²¹ <https://zenodo.org/records/18183663>

²² <https://zenodo.org/records/14627940>

²³ <https://zenodo.org/records/18183937>

supports distributed analytics across institutional data sources. Together, these developments support exploratory biomedical research and the creation of advanced multimodal analytical frameworks within HEREDITARY.

3.2.3 Potential AI Act Relevance and High-Risk Considerations

Although HEREDITARY does not currently deploy AI systems in clinical practice, some project activities involve functionalities that share characteristics with AI systems that may become subject to stricter regulatory scrutiny under the AI Act, particularly in future translational scenarios.

Article 2(6) of the AI Act excludes AI systems or AI models specifically developed and put into service for the sole purpose of scientific research and development. However, HEREDITARY outputs should be reassessed if they are later translated into clinical.

At the current stage, HEREDITARY activities are likely to fall within the research-oriented context recognized by the AI Act, as the AI and machine learning methods developed in the project are designed for scientific exploration, validation, and methodological development rather than for being placed on the market, put into service, or used as autonomous clinical decision-support tools. The current project outputs are not intended to make or replace clinical decisions, determine access to healthcare, or directly influence patient diagnosis, treatment, or care pathways. Instead, they support exploratory biomedical research, multimodal data integration, biomarker discovery, and model development under human scientific and clinical oversight.

For the same reasons, the AI systems currently developed within HEREDITARY are not considered, at this stage, to qualify as high-risk AI systems under the AI Act. Their use remains internal to research and validation activities, without deployment as medical device software, AI-enabled clinical decision support, or operational healthcare tools. However, the consortium recognizes that this assessment may change if project results are later exploited, deployed, commercialized, or integrated into clinical workflows. In such cases, further assessment would be required, including potential evaluation under the AI Act, the MDR/IVDR framework for medical device software, and applicable clinical, ethical, and data protection requirements. This interpretation is aligned with the legal and ethical analysis carried out within WP7 and should be read in conjunction with D7.2, where the applicability of the AI Act, including the research exception and the distinction between research activities and high-risk AI deployment, is addressed in greater detail.

Several aspects of the project justify continuous monitoring of the AI Act requirements:

Multimodal disease prediction and patient stratification

HEREDITARY develops AI models capable of identifying patterns across multimodal health datasets and generating research outputs concerning disease progression, phenotypic stratification, and treatment response. Examples include Use Case 1, where multimodal ALS data are used to support phenotyping, prognosis evaluation, and the identification of patient subgroups based on clinical, imaging, genetic, laboratory, and environmental variables, and Use Case 3, where OCT-based deep learning analyses have been explored for Parkinson's disease biomarker identification, as reported in D2.8. In D2.8, the project also identified age as a relevant confounding factor in PD prediction from retinal images, illustrating the importance of systematic performance evaluation,

bias assessment, and interpretability in AI-enabled biomedical research. While these outputs are currently used exclusively within research environments, similar functionalities may fall within the scope of AI-enabled clinical decision support systems if deployed in healthcare settings.

Potential risks include:

- incorrect predictions or misclassification of patients;
- biased model performance across different patient subgroups;
- overreliance on algorithmic outputs by end users;
- reduced interpretability of complex multimodal models.

For this reason, HEREDITARY maintains human oversight throughout model development and validation and systematically evaluates model performance, uncertainty, and fairness.

Processing of sensitive health and genomic data

The project relies on large-scale processing of health, genomic, imaging, and clinical data, all of which constitute sensitive categories of information. Although the AI Act does not directly regulate data protection matters, the combination of AI methods with highly sensitive health data requires particular attention to data governance, traceability, robustness, and privacy preservation.

Potential risks include:

- re-identification risks associated with multimodal datasets;
- privacy leakage through model outputs or federated learning processes;
- unintended disclosure of sensitive health information;
- misuse of analytical results outside their intended research context.

These risks are mitigated through GDPR-compliant governance procedures, pseudonymization, controlled access policies, federated processing approaches, and privacy-preserving analytical techniques.

Federated learning and distributed analytics

The federated learning architecture developed within HEREDITARY reduces the need for centralized storage of sensitive health data. Nevertheless, federated approaches introduce specific challenges that require ongoing monitoring.

Potential risks include:

- information leakage through model parameters;
- model inversion attacks;
- membership inference attacks;
- unauthorized access to distributed infrastructures.

Consequently, federated workflows are continuously evaluated from both a cybersecurity and AI governance perspective, integrating safeguards related to access control, secure communication, auditability, and privacy-preserving analytics.

3.2.4 Future Translational Scenarios

Although HEREDITARY remains a research-oriented project, the consortium acknowledges that future exploitation or clinical translation of project outcomes may require additional regulatory assessments, particularly regarding:

- software potentially qualifying as medical device software under the MDR/IVDR framework, including relevant MDCG guidance on software qualification and classification²⁴;
- AI-enabled clinical decision support systems;
- conformity assessment requirements;
- AI Act obligations applicable to high-risk systems;
- post-market monitoring and human oversight requirements.

The consortium therefore continues to monitor the evolving implementation timeline of the AI Act and relevant guidance issued by the European Commission and competent authorities to ensure that future developments remain aligned with European legal, ethical, and regulatory requirements.

3.3 EHDS, semantic interoperability and secondary use of health data

The HEREDITARY project continues to align its activities with the objectives and principles of the EHDS, particularly with regard to interoperability, FAIRification, secure secondary use of health data, and trusted cross-border access to distributed health data resources. The work carried out in WP3 provides an important operational basis for this alignment, as it translates the general principles of the EHDS into concrete technical and organizational mechanisms for semantic integration and federated analytics.

A central component of this alignment is the Hereditary Ontology²⁵ (HERO), described in D3.1²⁶. HERO provides a shared semantic model for representing clinical and genomic information relevant to neurological diseases, initially focusing on ALS and MS. It is structured into HERO-Clinical and HERO-Genomics and is designed to support the integration of heterogeneous data originating from different medical centers. The ontology was developed through an iterative co-design process involving clinicians, genomic data experts and technical partners, with the objective of aligning terminology, identifying common real-world entities, and establishing relationships between them. This approach directly supports the EHDS objective of improving the interoperability and reuse of health data for research purposes.

HERO also contributes to FAIRification and semantic discoverability as described in D2.17. D3.1 specifies that the ontology is openly available, documented, versioned, based on persistent identifiers, and designed in accordance with OBO and FAIR

²⁴ Medical Device Coordination Group. (2019, Rev. 1 June 2025). MDCG 2019-11 Rev.1: Guidance on qualification and classification of software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. European Commission. https://health.ec.europa.eu/document/download/b45335c5-1679-4c71-a91c-fc7a4d37f12b_en?filename=mdcg_2019_11_en.pdf

²⁵ <https://hereditary.dei.unipd.it/ontology>

²⁶ <https://zenodo.org/records/14627940>

principles. It also incorporates standard semantic resources and modelling approaches, including SKOS for controlled vocabularies, Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) alignment for clinical data interoperability, and Beacon v2-related modelling for genomic data representation. These characteristics are relevant to the EHDS framework because they support the structured description, interpretation and reuse of health data across institutional and national boundaries.

From the perspective of secondary use, HEREDITARY does not rely on data centralization. Instead, it implements an ontology-mediated and federated access model. D3.1 introduced the initial architecture based on Ontology-Based Data Access (OBDA) and Ontology-Based Data Federation (OBDF), allowing heterogeneous data sources to be queried through a common semantic layer. D3.2²⁷ further operationalizes this strategy through the Hereditary Data Network (HDN), which provides a unified semantic view over consortium data while allowing each participating institution to retain control over its local storage technologies, schemas and disclosure policies. Researchers can formulate information needs using SPARQL queries over the stable HERO conceptual schema, while the participating endpoints translate these ontology-level queries into local operations and return only admissible results.

This architecture is highly relevant to the EHDS vision because it supports secondary use of sensitive health data without requiring transfers of patient-level information. The HDN hub-and-spoke model enables the central component to coordinate vetted query templates, validate requests, dispatch queries and aggregate results, while institutional endpoints remain responsible for local execution, disclosure control and policy enforcement. This model reflects the EHDS emphasis on secure, governed and interoperable reuse of electronic health data for research, while preserving institutional autonomy and reducing the risks associated with centralized repositories.

The evolution from the D3.1 OBDF architecture to the D3.2 HDN architecture is particularly important. D3.2 shows that the project moved towards a more operational federation model in which HDN Central coordinates the query catalogue, dispatch of approved query templates and aggregation of results, while semantic translation, local policy checks and disclosure-level enforcement are performed at the endpoint level. This model strengthens alignment with privacy-by-design and data minimization principles, because institutional data remain governed locally and each node can enforce its own access, disclosure and execution policies while maintaining compliance with the shared HERO semantic contract.

At the same time, the consortium recognizes that this architecture also introduces technical and organizational challenges. These include the need to maintain mappings between local schemas and the HERO ontology at each participating node, the effort required to keep query templates, endpoint translations and disclosure policies aligned over time, and the current limitation that users can only execute predefined queries included in the approved catalogue rather than arbitrary exploratory queries. These limitations are intrinsic to a governance-oriented federation model and are therefore

²⁷ <https://zenodo.org/records/18183937>

managed as part of the project's ongoing technical development, documentation and validation activities.

HEREDITARY therefore operationalizes EHDS-related principles through several complementary mechanisms: first, semantic interoperability is supported by HERO and its alignment with recognized standards and vocabularies; second, FAIRification is supported not only by public ontology documentation, persistent identifiers, versioning and structured metadata practices, but also by the dedicated activities reported in D3.6²⁸, which address FAIRification workflows, 1+MG-aligned practices, dataset discoverability, metadata exposure, and controlled-access strategies for participating data resources; third, secondary use is supported by ontology-mediated federated querying and controlled data access models rather than relying exclusively on centralized data pooling; and fourth, governance of sensitive health data is supported by local institutional control, admissible query templates, disclosure levels, provenance-aware aggregation and auditability.

3.4 Cybersecurity and infrastructure protection

Cybersecurity and infrastructure protection constitute essential components of the HEREDITARY framework, particularly given the distributed processing of sensitive clinical, genomic, imaging, and multimodal health data across multiple institutions. Since the submission of D2.1, the technical implementation of the HDN has provided a concrete operational framework for embedding privacy, security, and governance requirements directly into the federated analytics infrastructure.

A key design principle of the HDN is the adoption of a privacy-by-design architecture, in which sensitive patient-level data remain under the governance of the originating institutions and are not transferred to a central repository by default. Participating centers expose controlled HDN endpoints that communicate through a common semantic layer based on the HERO ontology. In its current design, the HDN primarily supports harmonized discovery, controlled querying, cohort identification, and provenance-aware aggregation across distributed data sources. For example, a researcher may use the HDN to identify whether specific types of data or cohorts are available at one or more endpoints, without directly accessing patient-level records.

The transition from discovery to downstream analysis remains subject to the applicable governance, ethical, legal, and technical conditions of each participating institution. Depending on the research question, data type, approvals, and available infrastructure, subsequent analysis may follow different routes, including federated analysis workflows, execution of approved analytical pipelines at local endpoints, or controlled data access and secure data distribution under appropriate agreements such as Data Transfer Agreements, Data Access Committee approvals, or equivalent institutional procedures. Therefore, the HDN should be understood as a governance-oriented enabling layer for distributed discovery and controlled access, rather than as a replacement for all downstream data access, analysis, or legal authorization processes.

The HDN architecture follows a hub-and-spoke model composed of a central orchestration component (HDN Central) and multiple institutional endpoints. HDN

²⁸ <https://zenodo.org/records/18184066>

Central coordinates the execution of approved analytical workflows, validates requests, dispatches parameterized queries, aggregates responses, and provides provenance information, while local endpoints remain responsible for executing queries on institutional data sources according to local governance and disclosure policies. Importantly, HDN Central does not access or store raw patient-level data and operates exclusively on admissible outputs returned by participating sites.

A particularly relevant cybersecurity and governance feature of the HDN is the use of a catalogue of vetted query templates. Users do not submit arbitrary queries directly to participating institutions. Instead, requests are generated from predefined templates that are validated in advance and associated with explicit disclosure levels. Each query template is identified through a cryptographic hash and verified locally by participating endpoints before execution. This mechanism reduces the risk of unauthorized or non-compliant query execution, supports auditability, and provides an additional governance layer for controlled discovery and distributed querying activities. At the same time, the consortium recognizes that this design creates operational and usability constraints, including the need to maintain and update the query catalogue over time and the reduced flexibility for researchers who may wish to perform exploratory or highly customized queries across endpoints. These limitations are managed as part of the project's governance-by-design approach, in which flexibility is balanced against privacy, security, reproducibility, and institutional control.

The architecture also incorporates explicit disclosure control mechanisms. HDN defines multiple disclosure levels ranging from highly aggregated information to increasingly detailed data views. Before any access to local data occurs, each endpoint evaluates whether the requested disclosure level is compatible with local policies. Requests exceeding locally permitted thresholds are rejected before execution and result in an empty response. This ensures that privacy protection is enforced directly at the institutional boundary rather than after data processing has occurred.

Another important aspect concerns resilience and operational robustness. The HDN architecture follows a conservative execution strategy, meaning that the federation does not assume that all participating endpoints will always be available. Queries can be executed even when individual nodes are temporarily unavailable, fail to respond within predefined time windows, or decide not to contribute due to local disclosure restrictions. Provenance metadata accompanying the results provide transparency regarding the contributing institutions and the coverage of the federated query execution. This design reduces single points of failure and improves the resilience of the distributed infrastructure.

The endpoint architecture further strengthens infrastructure protection through layered security controls. Each endpoint includes dedicated service layers for request validation, authentication, semantic query processing, disclosure-level enforcement, and local execution. Queries are translated from the shared ontology into local database operations through OBDA mappings, allowing institutions to preserve their internal schemas and storage technologies while exposing a stable semantic interface to the federation. This separation of concerns enhances maintainability, traceability, and institutional autonomy.

Finally, HDN incorporates extensive provenance and auditability mechanisms. Query execution events, participating endpoints, response coverage, and execution metadata are recorded throughout the workflow. These capabilities support accountability, reproducibility, and governance of distributed analytics activities, while facilitating compliance with data protection, cybersecurity, and emerging European health data governance requirements.

Overall, the HDN architecture demonstrates how cybersecurity, privacy protection, institutional autonomy, auditability, and semantic interoperability can be integrated into a common federated infrastructure. These mechanisms contribute directly to the objectives of HEREDITARY and provide a practical implementation of privacy-by-design, security-by-design, and trustworthy data-sharing principles for distributed health research environments.

3.5 Governance and ethical oversight updates

3.5.1 Evolution of the Governance Framework

Deliverable D2.1 established the ethical and legal principles guiding the collection, processing, and sharing of multimodal health data within the HEREDITARY project. In this section, the term “governance framework” refers to the set of project-level roles, procedures, technical safeguards, documentation, and oversight mechanisms through which HEREDITARY manages ethical compliance, data protection, data access, data stewardship, technical implementation, and regulatory monitoring across participating institutions and work packages. Since the submission of D2.1, this governance framework has evolved from a primarily conceptual and regulatory perspective toward a more operational model integrating legal, ethical, technical, organizational, and data management safeguards across project activities. This evolution has been driven by several factors, including the implementation of federated and distributed data infrastructures, the development of the HEREDITARY Ontology (HERO) and the Hereditary Data Network (HDN), the FAIRification of participating data resources, the activities carried out within WP7, and the emergence of new European regulatory frameworks such as the AI Act and the EHDS.

The project Data Management Plan (DMP)²⁹ is also a central component of this governance framework, as it defines the principles and procedures for data management, FAIRification, access, storage, sharing, preservation, and responsibilities across the consortium. The DMP complements the ethical and legal framework described in D2.1 and its revision by translating data governance principles into operational practices for the management of HEREDITARY data resources throughout the project lifecycle.

Governance activities are therefore not limited to compliance monitoring. They increasingly focus on embedding ethical, legal, cybersecurity, FAIR data, and data protection requirements directly within the project’s technical architecture, operational procedures, and data management practices. WP1 contributes to this framework through overall project coordination, administrative governance, reporting, and monitoring of implementation across the consortium. WP2 contributes through the ethical and clinical

²⁹ <https://hereditary-project.eu/deliverables/>

data governance perspective, including oversight of ethical approvals, protocol amendments, and prospective data collection activities. WP7 supports the legal, regulatory, and compliance dimension, while technical work packages contribute by implementing governance-by-design mechanisms within the HEREDITARY infrastructure.

The consortium, through coordination among WP1, WP2, WP7, and the technical work packages, continues to monitor ethical approvals, protocol amendments, prospective data collection activities, data management procedures, and relevant regulatory developments throughout the project lifecycle. These activities are coordinated across work packages to ensure consistency between technical developments, data management obligations, legal requirements, and ethical responsibilities applicable to the HEREDITARY framework.

3.5.2 Governance-by-Design in the HEREDITARY Infrastructure

The governance-by-design approach described in HEREDITARY does not introduce a separate governance layer, but builds on the technical mechanisms already discussed in relation to the HERO ontology, the HDN architecture, FAIRification activities, and cybersecurity safeguards. These mechanisms translate project-level governance principles into operational controls, including semantic harmonization, predefined query templates, local disclosure enforcement, provenance metadata, access restrictions, and auditability.

In this sense, the HDN is relevant not because it replaces ethical, legal, or institutional governance procedures, but because it provides technical means to support their implementation in distributed settings. The main governance value of the infrastructure lies in linking data use to predefined, documented, and locally enforceable conditions, while preserving institutional responsibility over datasets and execution environments. The remaining limitations, including catalogue maintenance, restricted query flexibility, and the need for continuous alignment between local mappings, templates, and disclosure policies, are recognized as part of the project's ongoing governance and technical development activities.

3.5.3 AI Governance and Continuous Ethical Monitoring

The increasing maturity of AI methods within HEREDITARY, together with the evolving European regulatory framework, further reinforces the importance of continuous governance, ethical oversight, and regulatory monitoring. While the AI methods developed within the project are currently intended for research and validation purposes, the consortium monitors their potential legal, ethical, and societal implications through the governance activities established in D2.1, the WP7 legal and ethical framework, and the ongoing coordination between WP2, WP7, and the technical work packages. This monitoring should be understood as an internal governance and alignment process, rather than as a formal external audit.

Particular attention is devoted to ensuring human oversight throughout all stages of model development, validation, testing, and interpretation. AI-generated outputs are intended to support researchers and clinicians and are not designed to autonomously make clinical decisions. Furthermore, the consortium remains attentive to potential risks associated with multimodal AI systems, including model bias, lack of explainability,

inappropriate use of predictions, and privacy risks associated with sensitive health and genomic data.

The HEREDITARY governance framework therefore incorporates continuous monitoring of developments related to AI regulation, secondary use of health data, cybersecurity, and federated analytics. This monitoring is supported by the structure of WP7, which includes the in-depth legal and ethical study carried out under Task 7.2 and the subsequent evaluation and recommendations activities under Task 7.3. It is also linked to future technical and methodological deliverables, including D3.3 on the final release of federated workflow execution methods, D3.7 on the revision of FAIRification of participating data resources, D3.9 on privacy-preserving analytics, D4.6 on the final release of multimodal learning models, and D4.7 on self-supervised learning for fairness and uncertainty estimation. Through these activities, the consortium will continue to track the implementation phases of the AI Act, the deployment of the EHDS framework, the evolution of European interoperability standards, and the progress of initiatives such as HealthData@EU and the Genomic Data Infrastructure (GDI)³⁰. This approach helps ensure that project activities remain aligned with evolving European legal, ethical, and regulatory requirements while supporting trustworthy, transparent, and responsible use of AI and health data in biomedical research.

4 PROSPECTIVE DATA COLLECTION

4.1 Rationale and objectives

The primary strategy adopted within HEREDITARY for the development and validation of multimodal analytical methods is based on the extensive retrospective datasets already available through the consortium partners and described in previous deliverables as part of the project grant and DMP. These datasets include clinical, genomic, neuroimaging, ophthalmological, microbiome, and other multimodal resources collected through previous research projects and clinical activities, providing a unique opportunity to investigate neurological and neurodegenerative diseases through privacy-preserving and federated analytical approaches.

As discussed in Deliverable D2.1, Task 2.1 included the possibility of defining and collecting additional prospective data where necessary to support model validation and address specific scientific or methodological needs emerging during the project lifecycle. During the first half of the project, the consortium assessed the availability of existing datasets and identified ophthalmological imaging, and specifically Optical Coherence Tomography (OCT), as a modality of particular interest for multimodal analyses across several HEREDITARY use cases (see D2.21³¹ and D2.22³²).

Recent evidence suggests that retinal structures may reflect pathological processes occurring within the central nervous system. Since retinal neurons share a common embryological origin with brain structures, retinal imaging has increasingly attracted attention as a potential source of biomarkers for neurodegenerative diseases. OCT

³⁰ European Genomic Data Infrastructure Consortium. (2025). Genomic Data Infrastructure (GDI) Project. <https://gdi.onemilliongenomes.eu>

³¹ <https://zenodo.org/records/15848262>

³² <https://zenodo.org/records/18183882>

provides a non-invasive, rapid, and highly reproducible method for evaluating retinal morphology and has already been investigated in conditions such as Parkinson's disease, multiple sclerosis, Alzheimer's disease, and amyotrophic lateral sclerosis (ALS).

For these reasons, the consortium identified the opportunity to establish a new prospective cohort of ALS patients undergoing OCT examinations at the University of Turin³³. The objective is twofold. First, the prospective cohort will provide an independent dataset for validating multimodal analytical methods developed within HEREDITARY. Second, it will contribute to creating a unique multicenter OCT resource spanning multiple neurodegenerative diseases and institutions, supporting the development of federated ophthalmological analyses across the consortium.

4.2 Prospective ALS-OCT Cohort in Turin

The HEREDITARY consortium plans to establish a prospective cohort of ALS patients at the University of Turin and the Città della Salute e della Scienza Hospital³⁴. The study aims to investigate retinal structural alterations in ALS through OCT and to evaluate their relationship with disease severity, progression rate, clinical phenotype, and survival.

The scientific rationale for this activity is based on the growing body of evidence suggesting that retinal alterations may reflect neurodegenerative processes occurring within the central nervous system. Several studies have investigated retinal nerve fiber layer (RNFL) thickness and other OCT-derived parameters as potential biomarkers of ALS, although current evidence remains heterogeneous and additional prospective studies are needed to clarify their diagnostic and prognostic value.

Patients diagnosed with definite, probable, or probable laboratory-supported ALS will be prospectively recruited through the Piemonte and Valle d'Aosta ALS Register (PARALS). All participants will undergo a dedicated ophthalmological assessment, including fundus examination and OCT imaging. Patients with ocular conditions potentially affecting retinal measurements will be excluded from the study. Age- and sex-matched healthy controls will be recruited using the same protocol and acquisition procedures.

The study design envisioned the enrolment of approximately 100 ALS patients and a corresponding control population. OCT acquisitions will be performed using standardized protocols based on Heidelberg Spectralis® instrumentation, including macular and peripapillary scans centered on the fovea and optic nerve head. Clinical variables collected alongside OCT data will include demographic information, ALSFRS-R scores, progression rates, clinical phenotype classification, and survival data where available.

The resulting dataset will represent one of the first prospectively collected multimodal OCT cohorts specifically designed to support privacy-preserving and federated analytical workflows within HEREDITARY.

³³ <https://ror.org/048tbm396>

³⁴ <https://ror.org/001f7a930>

4.3 Multimodal and Multicenter OCT Integration

Beyond its direct relevance for ALS research, the prospective OCT cohort represents an important strategic asset for the broader HEREDITARY framework. OCT is one of the imaging modalities available across multiple HEREDITARY use cases and institutions, creating a unique opportunity for multicenter analyses based on a common acquisition technology.

The new ALS-OCT cohort will complement ophthalmological datasets already available within the consortium, including Parkinson's disease OCT data collected at the University of Colorado Denver³⁵ (UCD) and Multiple Sclerosis cohorts available at the University of Padua (UNIPD). This will enable the development of multimodal and multicenter analyses aimed at identifying both disease-specific and shared retinal biomarkers across different neurodegenerative conditions.

In particular, the prospective ALS cohort will support the extension and validation of the analyses initially explored in D2.8³⁶ "Parkinson's Disease in the Eye, First Approach". In D2.8, OCT images from Parkinson's disease patients were analyzed using deep learning approaches to investigate the presence of retinal biomarkers associated with neurodegeneration. The study demonstrated both the potential of OCT as a biomarker source and the importance of carefully addressing confounding factors, particularly age-related effects.

The availability of prospectively acquired ALS-OCT data will enable the consortium to investigate whether retinal biomarkers identified in one neurodegenerative disease can be generalized across different conditions or whether distinct disease-specific retinal signatures emerge. Furthermore, the multicenter availability of OCT data creates a natural use case for the federated learning infrastructure developed within HEREDITARY. Future analyses will therefore explore the feasibility of federated multimodal models integrating OCT imaging with clinical, genomic, and other biological data while preserving local data control and minimizing unnecessary transfers of sensitive information.

Through this approach, HEREDITARY aims to establish a unique federated ophthalmological research framework spanning ALS, Parkinson's disease, and multiple sclerosis cohorts, providing new opportunities for biomarker discovery and multimodal disease characterization.

5 ETHICAL APPROVALS UPDATES

Since the submission of D2.1, additional documentation concerning clinical studies, ethical approvals, and data governance procedures has been produced by consortium partners through Deliverables D2.19³⁷ (UNITO), D2.20³⁸ (RUMC), D2.21³⁹ (UNIPD), and D2.22⁴⁰ (UCD). These deliverables document the ethical, legal, and regulatory

³⁵ <https://ror.org/02hh7en24>

³⁶ <https://doi.org/10.5281/zenodo.15847669>

³⁷ <https://zenodo.org/records/14034480>

³⁸ <https://zenodo.org/records/15848061>

³⁹ <https://zenodo.org/records/15848262>

⁴⁰ <https://zenodo.org/records/18183882>

frameworks governing the datasets and clinical activities contributing to the HEREDITARY use cases.

The studies contributing data to HEREDITARY (with the only exception of OCT-ALS, which is under submission) have received the necessary ethical approvals, waivers, or institutional authorizations according to the applicable national and international regulatory frameworks. These approvals cover a broad range of multimodal data types, including clinical, genomic, neuroimaging, ophthalmological, microbiome, histopathological, and behavioral data, supporting the scientific objectives of the project. The complete overview of approvals, protocol numbers, and competent authorities is reported in Table 1.

Table 1 Consolidated overview of ethical approvals, waivers and authorizations supporting HEREDITARY activities. CET, Comitato Etico Territoriale.

Partner	Project Study	Ethics Committee / Authority	Protocol Number	Approval Date	Data Types
UNITO	BRAINTEASER	CET A.O.U. CDSS Torino	79511	26/07/2021	Clinical, Genomic
UNITO	EUROMOTOR	CET A.O.U. CDSS Torino	36344	18/05/2011	Clinical, Genomic
UNITO	INITIALS	CET A.O.U. CDSS Torino	38876	12/04/2018	Clinical, Genomic
UNITO	PERMEALS	CET A.O.U. CDSS Torino	38490	24/03/2022	Clinical, Neuroimaging
UNITO	INSPIRED	CET A.O.U. CDSS Torino	11669	02/02/2017	Clinical
UNITO	EXTRALS	CET A.O.U. CDSS Torino	64510	05/12/2012	Clinical, Neuroimaging
UNITO	OCT-ALS	CET A.O.U. CDSS Torino	<i>Under submission</i>	-	Clinical, OCT
RUMC	EXAMODE	CMO RUMC (non-WMO)	CMO 2018-4764	02/04/2019	Histopathology images
RUMC	BaCo-study	CMO Arnhem/Nijmegen (WMO)	CMO 2016-2616	12/09/2016	Histopathology, Gut & Tissue Microbiome, Clinical Data

Partner	Project Study	Ethics Committee / Authority	Protocol Number	Approval Date	Data Types
RUMC	MIND-set	CMO Arnhem/Nijmegen (WMO)	CMO 2015-1972	10/02/2016	fMRI, Gut Microbiome, Questionnaire Data
RUMC	HBS	CMO RUMC (non-WMO)	CMO 2018-4894	09/04/2019	Multimodal healthy cohort data
RUMC	HEREDITARY	CMO RUMC	CMO 2024-17669	20/11/2024	Re-use of data and biosamples for Use Cases 4–5
UCD	CORIS (Clinical Ophthalmology Retrospective Imaging System)	COMIRB	23-0284	22/03/2023	Ophthalmic imaging and EHR data
UCD	Relationship between clinical metrics for neurodegenerative diseases and ophthalmic imaging biomarkers	COMIRB	25-2326	28/10/2025	Clinical metrics and ophthalmic biomarkers
UCD	Identification of biomarkers for neurodegenerative diseases from ophthalmic images and clinical data using Federated Learning	COMIRB	25-2342	28/10/2025	Clinical metrics and ophthalmic biomarkers
UCD	AI and Oculomics for Systemic Disease Detection (AAI)	COMIRB	24-2014	12/11/2024	Prospective OCT and fundus imaging collection

Partner	Project Study	Ethics Committee / Authority	Protocol Number	Approval Date	Data Types
UNIPD	PREDICT-NEURODEGEN	Local Ethics Committee (Azienda Ospedale Università Padova)	0012082	16/02/2024	Clinical, neuroimaging, neurophysiological, biomarker, wearable sensor, and cognitive-behavioral data
UNIPD	HEREDITARY Ethics Approval Request	Local Ethics Committee (Azienda Ospedale Università Padova)	0033681	08/05/2026	Clinical, demographic, genetic, neuroimaging, biosignal, laboratory, neuropsychological, and digital health data

An important aspect that emerged during the project implementation concerns the role of the federated architecture in supporting ethical and legal compliance. Within HEREDITARY, the HEREDITARY Data Network (HDN) and the federated analytics and learning approaches developed in WP3 address different but complementary objectives. The HDN primarily supports harmonized discovery, controlled querying, and identification of relevant cohorts and data resources across participating institutions, while federated analytics and learning methods enable distributed analysis without routinely requiring the transfer of patient-level data between sites.

Rather than relying exclusively on the creation of new centralized repositories, HEREDITARY explores a distributed approach in which participating institutions retain responsibility for their local datasets and execution environments. This approach complements, rather than replaces, existing European infrastructures for sensitive health data sharing, including both centralized and federated repositories. Depending on the specific use case, data type, governance requirements, and available infrastructure, different models of data access and analysis may be appropriate, including federated analysis, local execution of approved analytical workflows, or controlled data sharing under established legal and ethical procedures.

The federated analytics framework contributes to reducing the need for routine secondary transfers of patient-level information by enabling analyses to be performed closer to the data source whenever feasible. At the same time, the consortium recognizes that distributed approaches also introduce operational challenges, including the need to maintain local infrastructures, negotiate data access arrangements where required, manage institutional governance requirements, and coordinate technical and legal procedures across multiple organizations. These considerations represent

important trade-offs between local control, privacy protection, operational complexity, and research flexibility.

Data used within HEREDITARY remain subject to the ethical approvals, governance procedures, and legal responsibilities of the originating institutions. Through the combination of controlled discovery mechanisms, federated analytics methods, and governance-by-design safeguards, the project seeks to support data minimization and privacy-by-design principles while enabling collaborative research across geographically distributed data resources.

6 CHALLENGES AND LIMITATIONS

The progress made in HEREDITARY demonstrates the value of combining semantic interoperability, FAIRification, federated discovery, and privacy-preserving analytics for sensitive health data. At the same time, the consortium recognizes that this approach presents several technical, legal, organizational, and regulatory challenges that require continued monitoring.

A first challenge concerns the distinction between data discovery, access, and downstream analysis. The HDN supports harmonized discovery, controlled querying, cohort identification, and provenance-aware aggregation across distributed data resources. However, identifying relevant data through the HDN does not automatically provide access to patient-level data or enable all forms of downstream analysis. Depending on the research question, data type, institutional governance, and ethical or legal requirements, subsequent analysis may require federated analytical workflows, local execution of approved pipelines, or controlled data sharing under appropriate access procedures.

A second challenge concerns the operational trade-offs of controlled federation. Predefined query templates, local disclosure controls, and endpoint-level semantic mappings support privacy, auditability, and institutional governance, but they also require continuous maintenance and may reduce flexibility for exploratory research. Participating datasets also differ in structure, terminology, granularity, metadata quality, and local governance conditions, making semantic harmonization and FAIRification ongoing rather than one-off activities.

A third challenge concerns the evolving regulatory and methodological landscape. The AI Act, EHDS, NIS2, MDR/IVDR framework, and related European initiatives are still moving through implementation phases that may affect future exploitation or clinical translation of project outputs. In parallel, multimodal AI models require continued attention to bias, confounding, uncertainty, explainability, cybersecurity, and human oversight. HEREDITARY therefore addresses these challenges through the DMP, WP1 project governance, WP2 ethical and clinical coordination, WP7 legal and regulatory monitoring, and future technical deliverables on federated workflows, FAIRification, privacy-preserving analytics, multimodal learning, fairness, and uncertainty estimation.

7 CONCLUSIONS

This deliverable presented the update of Deliverable D2.1 “Ethical Guidelines, Data Collection and Sharing”, providing an overview of the legal, ethical, and regulatory

developments relevant to the HEREDITARY project since June 2024. The document reflects the transition of the project from an initial conceptual and governance-definition phase towards the operational implementation of federated infrastructures, multimodal analytics workflows, and privacy-preserving collaborative research activities.

A first objective of D2.2 was to assess the evolution of the regulatory landscape applicable to HEREDITARY. Since the submission of D2.1, several important developments have occurred at the European level, including the entry into force of the Artificial Intelligence Act, the adoption and implementation of the European Health Data Space framework, the increasing relevance of cybersecurity obligations associated with the NIS2 Directive, and the emergence of new interoperability initiatives supporting secondary use of health data. These developments further strengthen the importance of trustworthy AI, privacy-preserving analytics, interoperability, and secure governance of health and genomic data. The consortium continues to monitor these regulatory developments and to align project activities with evolving European requirements.

This deliverable also highlighted the operational progress achieved within HEREDITARY regarding data governance, semantic interoperability, FAIRification, and distributed data discovery. The development of the HEREDITARY Ontology (HERO) and the Hereditary Data Network (HDN) has transformed many of the principles discussed in D2.1 into concrete technical implementations. Semantic interoperability, FAIRification, controlled discovery and querying of distributed health data resources, provenance tracking, disclosure-level governance, and privacy-preserving analytics are no longer solely conceptual objectives but are progressively embedded within the project infrastructure. In this respect, the federated architecture plays a central role in reducing unnecessary transfers of sensitive data while maintaining institutional autonomy and supporting collaborative research across multiple centers and jurisdictions.

These developments are complemented by the broader project governance and data management framework. In particular, D1.1 and D1.2, The Data Management Plan and its revision, provide the project-level governance and management context within which ethical, legal, technical, and administrative responsibilities are coordinated. These further define the operational principles for data management, FAIRification, access conditions, storage, sharing, preservation, and stewardship across the consortium. In addition, the FAIRification activities carried out under Task 3.6 and reported in D3.6 provide a dedicated framework for improving the discoverability, interoperability, and reuse potential of participating data resources, complementing the semantic and federated infrastructure described in D3.1 and D3.2.

Particular attention has been devoted to the governance of artificial intelligence systems. While the methods developed within HEREDITARY remain primarily research-oriented and are not intended for autonomous clinical decision-making, the project increasingly considers the implications of the AI Act and the broader European framework for trustworthy AI. Potential risks related to multimodal prediction, bias, explainability, privacy, and future translational applications have been identified and are continuously monitored. Human oversight, transparency, accountability, and scientific validation remain fundamental principles guiding all AI-related activities within the project.

Another important aspect of this revision concerns the prospective data collection activities foreseen under Task 2.1. While the project continues to rely primarily on

extensive retrospective datasets already available through consortium partners, the possibility of collecting prospective data for validation purposes was explicitly foreseen in D2.1. Following an assessment of scientific opportunities and available infrastructures, the consortium identified ophthalmological imaging, and specifically Optical Coherence Tomography (OCT), as a modality of strategic relevance across several HEREDITARY use cases. Consequently, a prospective cohort of ALS patients undergoing OCT examinations is being prepared at the University of Turin. At the time of submission of this deliverable, this activity is in the preparatory and ethical approval phase: the study protocol, patient information materials, consent documentation, and data management procedures have been developed or are being finalized for submission to the relevant ethics and institutional review bodies. Participant recruitment and prospective data collection will commence only after all required ethical and regulatory approvals have been obtained. Once approved, this initiative will provide an independent dataset for validation purposes and will support the development of multimodal and multicenter analyses involving ALS, Parkinson's disease, and multiple sclerosis cohorts available within the consortium.

The deliverable further reviewed the ethical approvals and governance procedures supporting HEREDITARY activities. Additional documentation produced by consortium partners demonstrates that the retrospective datasets currently employed within the project originate from studies that have received the necessary ethical approvals, waivers, or institutional authorizations according to the applicable national and international regulatory frameworks. This statement refers to the existing retrospective datasets already available to the consortium. The new prospective ALS-OCT cohort is instead subject to a separate ethics and institutional approval process, and no participant recruitment or prospective data collection will begin before the required approvals are granted. At the same time, the distributed and federated approach explored within HEREDITARY reduces the need for routine secondary transfers of patient-level information by supporting controlled discovery, local governance, and, where appropriate, privacy-preserving analytical workflows. This approach complements existing ethical approvals and reinforces compliance with GDPR principles, including data minimization, privacy by design, and accountability.

Overall, the findings presented in this deliverable confirm that HEREDITARY continues to operate within a robust ethical, legal, and regulatory framework. The project has progressed from defining principles and governance requirements toward implementing practical solutions that operationalize these principles through federated infrastructures, semantic interoperability, privacy-preserving analytics, and responsible AI methodologies. The consortium will continue monitoring regulatory developments, ethical requirements, and emerging governance frameworks throughout the remaining duration of the project, ensuring that HEREDITARY remains aligned with European standards for trustworthy, secure, and ethically responsible health data research.

REFERENCES

The bibliographic entries are arranged in lexicographical order based on the key, following the APA style. This enables us to place the entries and citations in the table and text in any sequence, allowing for later sorting while ensuring consistency.

Key	Reference
AIACT2024	European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L, 2024/1689. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32024R1689
EC2019-AI	European Commission. (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu
EC2024-1MG	European Commission. (2024). 1+ Million Genomes Initiative. Directorate-General for Communications Networks, Content and Technology. https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes
EC2025-EHDSRDF	European Commission. (2025). EHDS RDF Validator. https://health-data-itb-rdf-validator.acceptance.data.health.europa.eu/shacl/ehds/upload
EC2025-HealthDataEU	European Commission. (2025). HealthData@EU: Infrastructure for the secondary use of health data. European Health Information Portal. https://healthinformationportal.eu
EC2025-HealthDCAT	European Commission. (2025). HealthDCAT-AP: Health Data Catalogue Application Profile. European Health Information Portal. https://www.healthinformationportal.eu/healthdcat-ap
EHDS2025	European Parliament and Council of the European Union. (2025). Regulation (EU) 2025/327 on the European Health Data Space. Official Journal of the European Union.
FLI2025	Future of Life Institute. (2025). AI Act implementation timeline. https://artificialintelligenceact.eu/implementation-timeline/
GDI2025	European Genomic Data Infrastructure Consortium. (2025). Genomic Data Infrastructure (GDI) Project. https://gdi.onemilliongenomes.eu
GDPR2016	European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88.
Hereditary Project D2.1	HEREDITARY Consortium. (2024). Deliverable D2.1 – Ethical Guidelines, Data Collection and Sharing. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/14033771

Key	Reference
Hereditary Project D2.19	HEREDITARY Consortium. (2024). Deliverable D2.19 – UNITO Clinical Studies Documentation. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/14034480
Hereditary Project D2.20	HEREDITARY Consortium. (2025). Deliverable D2.20 – RUMC Clinical Studies Documentation. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/15848061
Hereditary Project D2.21	HEREDITARY Consortium. (2025). Deliverable D2.21 – UNIPD Clinical Studies Documentation. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/15848262
Hereditary Project D2.22	HEREDITARY Consortium. (2025). Deliverable D2.22 – UCD Clinical Studies Documentation. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/18183882
Hereditary Project D2.8	HEREDITARY Consortium. (2025). Deliverable D2.8 – Parkinson’s Disease in the Eye, First Approach. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/15847669
Hereditary Project D3.1	HEREDITARY Consortium. (2024). Deliverable D3.1 – Ontology and Federated Execution Methods. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/14627940
Hereditary Project D3.2	HEREDITARY Consortium. (2025). Deliverable D3.2 – Hereditary Data Network. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/18183937
Hereditary Project D7.1	HEREDITARY Consortium. (2024). Deliverable D7.1 – Legal, Ethical and Regulatory Inventory. https://hereditary-project.eu/deliverables/ https://zenodo.org/records/14628224
Kairouz2021	Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/22000000083
Mothukuri2021	Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619–640. https://doi.org/10.1016/j.future.2020.10.007
NIS22022	European Parliament and Council of the European Union. (2022). Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive). Official Journal of the European Union, L333, 80–152.
OECD2019	Organization for Economic Co-operation and Development. (2019). OECD principles on artificial intelligence. OECD Publishing. https://oecd.ai/en/ai-principles