



# HEREDITARY

HetERogeneous sEmantic Data integration for the guT-bRain interplaY

## Deliverable 2.21

### UNIPD clinical studies documentation

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 outlines the ethical, legal, and data protection framework adopted within the HEREDITARY project, with particular attention to the management of sensitive health and genetic data. The project uses retrospective, pseudonymised clinical data collected over the past two decades from over 6,000 patients affected by complex neurological and psychiatric conditions. These include neurodegenerative diseases (e.g. Parkinson's disease, ALS), cerebrovascular and autoimmune disorders, neuropathies, gliomas, and psychiatric conditions such as anorexia nervosa.

In accordance with Article 9(2)(j) of the GDPR, Article 89 of the GDPR, and Article 110 of the Italian Privacy Code, informed consent is not required for the use of such data at the University of Padova, due to the large number of subjects, the retrospective nature of the data, and the fact that many of the patients are deceased or no longer contactable. This is justified by the disproportionate effort required and the risk of compromising the scientific goals of the project. These criteria are further supported by provision no. 146/2019 of Italian Data Protection Authority (DPA), along with the related General Authorisations no. 8/2016 and 9/2016.

To ensure robust data protection, HEREDITARY implements a federated learning architecture. All raw data remain on secure servers at each clinical site; no identifying or raw data are transferred between centres. Each site performs local analysis and shares only aggregated outputs or features. This approach guarantees full compliance with data minimization and privacy-by-design principles, while enabling advanced multimodal analysis across institutions.



## DOCUMENT INFORMATION

<b>Deliverable ID</b>	D2.21
<b>Deliverable Title</b>	UNIPD clinical studies documentation
<b>Work Package</b>	WP2
<b>Lead Partner</b>	UNIPD
<b>Due date</b>	30.06.2025
<b>Date of submission</b>	30.06.2025
<b>Type of deliverable</b>	R
<b>Dissemination level</b>	PU

## AUTHORS

<b>Name</b>	<b>Organisation</b>
<b>Manfredo Atzori (Author)</b>	UNIPD
<b>Umberto Manera (Internal reviewer)</b>	UNITO
<b>Anna Romanovych (Contributor)</b>	UNIPD

## REVISION HISTORY

<b>Version</b>	<b>Date</b>	<b>Author</b>	<b>Document history/approvals</b>
<b>0.1</b>	12/06/2025	Manfredo Atzori	Introduction, Data inclusion, Ethical conduct, Data management, Conclusions
<b>0.2</b>	17/06/2025	Umberto Manera	Review, adjustments, and clarifications
<b>0.3</b>	26/06/2025	Manfredo Atzori	Final review, adaptations, formatting
<b>1.0</b>	27/06/2025	Anna Romanovych	Final formatting of the deliverable
<b>1.1</b>	29/06/2025	Manfredo Atzori	Final technical check and editing

*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

<b>List of abbreviations .....</b>	<b>7</b>
<b>1 Introduction .....</b>	<b>8</b>
<b>2 Data to be included in the study .....</b>	<b>12</b>
2.1 Summary of data provided by UNIPD .....	12
<b>3 Ethical conduct of the study .....</b>	<b>14</b>
3.1 Ethical Approvals .....	14
3.2 Ethical Approval Documentation .....	16
<b>4 Data Management .....</b>	<b>18</b>
4.1 Data Minimization and Pseudonymization .....	19
4.2 Data Transfers to Third Parties and International Transfers .....	19
<b>5 Incidental Findings Policy .....</b>	<b>20</b>
<b>6 Conclusion .....</b>	<b>21</b>
<b>Annexes .....</b>	<b>22</b>

## List of abbreviations

Abbreviation	Original version
<b>GDPR</b>	General Data Protection Regulation
<b>ALS</b>	Amyotrophic Lateral Sclerosis
<b>MS</b>	Multiple Sclerosis
<b>PD</b>	Parkinson's Disease
<b>MRI</b>	Magnetic Resonance Imaging
<b>EEG</b>	Electroencephalography
<b>PFBC</b>	Primary Familial Brain Calcifications
<b>PLS</b>	Primary Lateral Sclerosis
<b>HSP</b>	Hereditary Spastic Paraplegia
<b>NIHSS</b>	National Institutes of Health Stroke Scale
<b>ALSFRS-R</b>	Amyotrophic Lateral Sclerosis Functional Rating Scale - Revised
<b>EDSS</b>	Expanded Disability Status Scale
<b>MoCA</b>	Montreal Cognitive Assessment
<b>PET</b>	Positron Emission Tomography
<b>OCT</b>	Optical Coherence Tomography
<b>GPPs</b>	Guidelines for Good Pharmacoepidemiology Practices
<b>ISPE</b>	International Society for Pharmacoepidemiology
<b>VPN</b>	Virtual Private Network
<b>SSH</b>	Secure Shell
<b>UOC</b>	Unità Operativa Complessa (Complex Operating Unit)

## 1 Introduction

This deliverable presents the legal framework and supporting documentation for UNIPD's participation in Use Cases 1 through 5 of the HEREDITARY project. The Use Cases constitute the main pillars of the project, which is focused on improving the knowledge and management of neurologic pathologies, such as amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), Parkinson's disease (PD), and stroke, by applying techniques for privacy-preserving integrated analytics.

HEREDITARY's methods are designed to operate on multimodal biomedical data in general and they are built to be easily adaptable to additional pathologies. Accordingly, data related to other neurological and psychiatric disorders have also been included in the regulatory, ethical and legal documentation. These datasets, available through UNIPD, are characterised by heterogeneous phenotypes, suggesting strong potential for benefit from the methods developed within the project.

The HEREDITARY project seeks to address the shortcomings of conventional classifications of neurodegenerative disorders based solely on symptoms, which often overlook the biological diversity among patients. Through the integration of diverse data sources, such as genetic profiles, clinical documentation, biomarkers, neuroimaging, and physiological signals, the project aims to advance the precision of disease characterization, enable more accurate patient grouping and outcome forecasting, discover new biomarkers across and within modalities, and explore underlying disease mechanisms.

This strategy involves handling large volumes of sensitive personal information. In line with the goals outlined in the Horizon Europe call "Tools and technologies for a healthy society, HORIZON-HLTH-2023-TOOL-05", the HEREDITARY consortium is creating federated workflows to support secure biomedical data analysis and visualisation.

The core Use Cases in HEREDITARY are:

- **Use Case 1: Neurodegenerative diseases phenotyping and prognosis evaluation.** This use case focuses on understanding ALS by identifying endophenotypes through the integration of multimodal data. It aims to detect genomic variants and biological pathways associated with clinical features and patient survival.
- **Use Case 2: Next-generation diagnosing and treatment response for neurodegenerative diseases.** This use case aims to develop a biologically informed, biomarker-driven classification system for neurodegenerative diseases, enabling precision medicine approaches.
- **Use Case 3: Signs of Parkinson's disease in multimodal data.** This use case focuses on identifying Parkinson's disease via ophthalmic imaging combined with multimodal data (MRI, EEG, genetics) using deep learning, federated learning and self-supervised methods to detect, predict, and correlate PD biomarkers across sites.
- **Use Case 4: Phenotyping of the gut-brain axis in healthy individuals to understand deviations in disorders.** This use case studies the healthy gut-brain axis via microbiome, imaging, and genetic data to detect deviations linked



to disease. Linked ICA and correlation analyses explore links to brain, behavior, stress, and health factors.

- **Use Case 5: Gut-Brain linkage and disease relevance.** Gut-brain linkage is explored in disorders like PD and ALS using microbiome, imaging, and clinical data. Deep learning detects gut alterations and links them to symptoms, enabling insights into disease and personalized probiotic strategies.

At UNIPD, the following additional diseases have been included in the ethics approval requests:

- Sensory and motor neuropathies (e.g. CANVAS, anti-MAG neuropathy)
- Genetic myopathies (e.g. muscular dystrophy)
- Brain tumours (e.g. gliomas)
- Conditions involving the upper motor neuron (e.g. primary lateral sclerosis, PLS or hereditary spastic paraplegia, HSP)
- Primary familial brain calcifications (PFBC)
- Other rare genetic or hereditary neurological disorders
- Psychiatric disorders, such as anorexia nervosa

Naturally, the inclusion of additional diseases has expanded the scope and complexity of the regulatory, ethical, and legal documentation. However, this broader coverage is expected to enhance the overall impact of HEREDITARY and improve outcomes for a wider range of patients. The early phase of WP4 analytics is primarily focused on pre-training models using large, publicly available datasets, so immediate access to proprietary data is not required. Additionally, requesting a second set of approvals for the inclusion of additional pathologies would have created a significant administrative burden, likely preventing the consortium from analysing such data. For these reasons, we chose to follow a longer, but more inclusive, path during this phase of the project.

Due to the nature of the data being processed and the extensive scope of the initiative, it is essential to implement a solid legal and ethical structure that adheres to the applicable laws and safeguards the rights of data subjects. This document presents the main legal principles regulating the processing of personal information in scientific research, with a focus on the General Data Protection Regulation (GDPR) and pertinent European and national legal frameworks.

UNIPD is an Italian university and receives data via protocols which regulate the interaction between UNIPD and the institution where data are acquired, i.e. "Azienda Ospedale Università Padova". As both institutions are Italian entities, this deliverable particularly focuses on adherence to Italian and European legislation regarding data privacy and security. Data collection and processing is carried out in compliance with the Italian Legislative Decree 196/2003 (as amended by Legislative Decree 101/2018), the EU Regulation 2016/679 (GDPR), and the guidelines of the Italian Data Protection Authority, including provision no. 146/2019 and its related General Authorisations no. 8/2016 (for genetic data) and 9/2016 (for research purposes). Data processing is also carried out in accordance with Article 110 of the Italian Privacy Code, which regulates the use of health and genetic data for scientific research purposes. It adheres to the deontological rules for treatments for statistical or scientific research purposes (Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica - Garante, 19/12/2018), and to international and European regulations and ethical frameworks such

as the Oviedo Convention (Law no. 145/2001), the Council of Europe Recommendation R(97)5, and the Declaration of Helsinki.

Within this framework, personal data encompasses any detail that pertains to an individual who is either directly known or could potentially be recognized. Identification may occur through various means, such as names, identification numbers, geographical information, digital identifiers, or distinct traits related to the person's physical, biological, psychological, financial, cultural, or societal attributes.

The processing of personal data for scientific research in the biomedical domain is allowed when the data subject's informed consent is obtained. However, the law also permits such processing without consent under certain conditions. Specifically, this applies when contacting individuals is unfeasible or would entail a disproportionate effort, or if doing so would compromise or obstruct the research objectives (as outlined in Article 110 of the Italian Privacy Code). The applicable provisions, particularly Section 5.3 of the Italian Data Protection Authority's Provision no. 146/2019, describe the acceptable ethical and organizational reasons that justify this exception. These justifications must be clearly stated in the research protocol. In these cases, the data controller is required to implement suitable safeguards to uphold the rights and interests of the data subjects. Furthermore, the research must be approved by the appropriate territorial Ethics Committee, and a Data Protection Impact Assessment (DPIA) must be conducted and made accessible upon request to the "Garante per la protezione dei dati personali" (Italian Data Protection Authority), pursuant to Article 36 of the GDPR. These provisions are based on Article 110 of the Privacy Code (Legislative Decree 196/2003, as amended) and Articles 9(2)(j), 9(4), and 89 of the General Data Protection Regulation.

The data minimization principle plays a fundamental role, requiring that any data collected and processed must be appropriate, relevant, and strictly necessary for achieving the intended purposes (Article 5(1)(c) of the GDPR). Thus, when conducting scientific research in medical, biomedical, or epidemiological contexts, the use of sensitive data such as health information, details about sexual life or orientation, and data on racial or ethnic background is permitted only if such information is essential to the research objectives. This requirement is reinforced by Section 5.4 of Provision no. 146/2019 issued by the Italian Data Protection Authority.

As a general rule, the handling of personal data must respect the principle of transparency, as established in Article 5(1)(a) of the GDPR. This entails that individuals must receive clear and easily accessible information about how their data is being used. When data is gathered directly from the subject, Article 13 applies.

Personal data must be retained in a form that allows the identification of data subjects only for as long as necessary to fulfill the purposes for which they were collected. However, extended retention is permitted when the data are processed exclusively for archiving in the public interest, scientific or historical research, or statistical purposes. In such cases, appropriate technical and organizational safeguards must be applied to protect the rights and freedoms of data subjects, as specified in Article 5(1)(e) of the GDPR.

Under Article 5 of the GDPR, the accountability principle is especially important. It requires data controllers not only to comply with the Regulation but also to be able to demonstrate that compliance (Articles 5(2) and 24). This responsibility is closely tied to the obligation to incorporate data protection into processing activities from the outset and

to ensure that only the necessary data are processed, in line with the principles of privacy by design and privacy by default (Article 25).

To comply with the principle of data protection by design and by default, data controllers must proactively apply data protection principles to ensure meaningful safeguards. Article 25 of the GDPR requires controllers to embed data protection into processing workflows from the outset and maintain it throughout the data lifecycle. This includes implementing appropriate technical and organizational measures to enforce compliance and safeguard the rights and freedoms of data subjects.

The next sections present an in-depth analysis of the actions taken by UNIPD to meet the ethical and legal obligations tied to the HEREDITARY project's use cases and datasets. They detail the concrete steps adopted to guarantee alignment with the GDPR and explain how data processing and analytical practices comply with the ethical standards required for handling sensitive personal data, safeguarding participants' rights and freedom throughout the research.

## 2 Data to be included in the study

The HEREDITARY project utilizes UNIPD's retrospective observational data extracted from clinical records available at Azienda Ospedale Università Padova. Ethics approvals have been obtained for similar and related data usages, as in the context of the currently active project PREDICT- NEURODEGEN - PNRR-MAD-2022-12376415.

In HEREDITARY, data usage is governed by specific regulatory, ethical, and legal authorizations. All processing is carried out in a pseudonymized setting, maximizing privacy and security.

### 2.1 Summary of data provided by UNIPD

The study cohort will comprise individuals diagnosed with the following medical conditions who have been taken care by the Azienda Ospedale Università Padova:

- Neurological diseases, including neurodegenerative disorders (e.g., Parkinson's disease, ALS)
- Cerebrovascular diseases (e.g., stroke)
- Autoimmune diseases (e.g., multiple sclerosis)
- Sensory and motor neuropathies (e.g., CANVAS, anti-MAG neuropathy)
- Genetic myopathies (e.g., muscular dystrophy)
- Brain tumours (e.g., gliomas)
- Upper motor neuron diseases
- Primary familial brain calcifications (PFBC)
- Other rare genetic or hereditary neurological diseases
- Psychiatric disorders, such as anorexia nervosa

#### Inclusion Criteria

Only adults are eligible to take part in the study. Patients must have been followed at the Padova Hospital for one of the above-listed pathologies and/or enrolled in studies related to the above pathologies. Retrospective, observational data must be available at Azienda Ospedale Università Padova and be usable according to ethics approval. Data must be pseudonymized.

#### Exclusion Criteria

Individuals who are minors and/or legally or mentally incapacitated are excluded from the study.

#### List of data to be included in the study

The project exclusively uses retrospective, pseudonymized data already available at the participating healthcare facilities. The data include:

- Demographic data (age, sex)
- Socio-demographic data (education level, profession)
- Anthropometric data (weight, height, dominant hand)
- Health and disease-related data (diagnoses, therapies, outcomes)
- Clinical scales (e.g., NIHSS, ALSFRS-R, EDSS, Rankin, Barthel, MoCA)
- Genetic data (variants, mutations)
- Biomedical images (e.g., MRI, PET, EEG, OCT) and biosignals (EEG)

- Anamnestic and clinical context data
- Laboratory and biological data (e.g., blood count, cerebrospinal fluid, biomarkers, gut microbiome)
- Digital, instrumental, and sensor-based data
- Neuropsychological data (e.g., cognitive tests)
- Clinical classifications and codifications (e.g., SNOMED)

### 3 Ethical conduct of the study

All research analyses involving data from UNIPD will follow internationally recognized ethical and scientific standards. These include the principles outlined in the Declaration of Helsinki and its updates, and any other relevant national laws and regulatory frameworks. These measures aim to uphold both the scientific validity of the work and the ethical protection of individuals involved.

#### 3.1 Ethical Approvals

The project involves the use of retrospective clinical data concerning stroke, Parkinson's disease, multiple sclerosis, ALS, and additional conditions detailed in Section 2 of this document. Article 110 of the Italian Privacy Code (Code of Conduct for the processing of personal data for statistical or scientific research purposes under Articles 2-quater and 106, enacted on May 9, 2024, and published in the Gazzetta Ufficiale No. 130 on June 5, 2024) introduced updated provisions for managing personal data, particularly for cases involving deceased individuals or those who cannot be contacted. The amendment removes the requirement to submit the research protocol and its associated impact assessment for preliminary consultation in situations where notifying the subjects is unfeasible, excessively burdensome, or would compromise the scientific validity of the research. Instead, it is sufficient to comply with the safeguards expressly defined by the Italian Data Protection Authority.

The Italian Data Protection Authority requires that, in situations where informed consent cannot be obtained, data controllers must secure approval from the relevant territorial Ethics Committee and provide a well-founded explanation of the ethical or practical constraints that prevent recontacting patients, as well as a Data Protection Impact Assessment (DPIA). Such organizational limitations may exist when the study involves an exceptionally large number of participants or when, despite appropriate verification efforts (such as confirming patient survival, examining clinical records, utilizing existing contact details, and consulting public databases) the individuals are determined to be deceased or untraceable at the time of inclusion.

Personal and genetic data within the HEREDITARY project will be processed for medical, biomedical, and epidemiological research purposes, in accordance with the requirements set out in Sections 4.5, 4.11.2 and 5.4 of the provision no. 146/2019 of the Italian Data Protection Authority and the provisions regarding the processing of genetic data (General Authorization no. 8/2016).

The objectives of the Horizon Europe call under which HEREDITARY was funded include the development of analytical methods aimed at ensuring the full protection of patient privacy. The ultimate goal is to develop methods for the extraction of medical knowledge from retrospective, observational, and multimodal data, including genetic data.

Therefore, the HEREDITARY project and the validation of the methods developed therein require access to patients' genetic data for exploratory purposes and for validating the methods themselves. The project would be compromised without access to this data.

The personal and genetic data considered will be extracted from a cohort of over 6,000 patients, collected over more than twenty years of clinical practice, making efforts to contact patients for their informed consent disproportionate for the project. Moreover, the characteristics of the considered diseases, reported below with indicative estimates of

age of onset and 10-year mortality, make it reasonable to assume that a large part of the patients could be dead or otherwise unreachable via the available contact information.

- **Neurodegenerative diseases** - e.g., ALS, whose onset is between 60 and 70 years of age, is a rapidly progressive disease with high mortality: median survival is about 3–5 years from diagnosis, with over 80% mortality at 10 years; Parkinson's disease, usually begins at around 60 years of age, it has slower progression but with increasing motor and cognitive disability; 10-year mortality estimated between 30% and 50%, higher in individuals with early-onset dementia, postural instability, and cardiovascular comorbidities.
- **Brain tumours** - e.g., multiform glioblastoma, typical onset between 45 and 75 years of age, 10-year mortality over 95%.
- **Cerebrovascular diseases** - e.g., stroke, typical onset over age 65; highly prevalent in the elderly, a leading cause of mortality and long-term disability; 1-year mortality may exceed 30% in older and more fragile individuals, while 10-year survival is heavily influenced by severity and comorbidities.
- **Sensory and motor neuropathies** - e.g., CANVAS, typical onset around 60–65 years of age, slowly progressive and disabling condition, with potentially reduced life expectancy; anti-MAG neuropathy, typical onset over age 60, chronic course and generally not fatal.
- **Autoimmune diseases** - e.g., multiple sclerosis, typical onset between 20 and 40 years of age; clinical course is variable but generally chronic and progressive, especially in primary or secondary progressive forms. Although direct mortality is lower than in other severe neurological diseases, the condition can result in significant long-term disability. Life expectancy is reduced on average by about 6–10 years compared to the general population, with increased mortality in cases of severe disability, infections, and comorbidities.
- **Genetic myopathies** - e.g., muscular dystrophy, typical onset between 2 and 6 years of age in the most severe forms, such as Duchenne muscular dystrophy; progressive clinical course with muscular weakness, early motor disability, and multisystem involvement including respiratory and cardiovascular systems. Without treatment, life expectancy is significantly reduced, with mortality exceeding 40% by age 30. However, with current treatments and assisted ventilation, life expectancy can be significantly prolonged.
- **Conditions involving the upper motor neuron** - typical onset between 50 and 70 years of age; includes forms of hereditary spastic paraplegia and degenerative conditions with corticospinal tract impairment. The course is progressive and may lead to significant motor disability. Prognosis varies depending on aetiology: in degenerative forms associated with other motor neuron diseases such as ALS, mortality can be high; in purely spastic forms, life expectancy is often normal, though quality of life may be significantly affected.
- **Primary familial brain calcifications (PFBC)** - typical onset after age 50; a rare genetic condition characterised by calcium deposits in the basal ganglia, cerebellum, and other brain regions. Clinical presentation is variable: some patients remain asymptomatic, while others develop motor, cognitive, or



psychiatric symptoms. Progression may be slow. Although survival is often preserved, symptomatic forms can significantly impact quality of life.

- **Other rare genetic or hereditary neurological disorders.**
- **Psychiatric disorders** - e.g., anorexia nervosa, typically onset in adolescence or early adulthood, predominantly in females; complex and multifactorial disorder with high morbidity and among the highest mortality rates among psychiatric conditions, estimated around 5–10% in the long term. Anorexia nervosa severely affects physical and mental health, with potentially severe medical complications, especially in the absence of timely and effective treatment.

The data processing will be carried out according to Article 9(2)(j) of the GDPR for scientific research purposes, in compliance with Article 89 and with the adoption of appropriate safeguards, and in accordance with Sections 4.5, 4.11.2 and 5.3 of Provision No. 146/2019 of the Italian DPA concerning personal and genetic data.

Within the HEREDITARY project, the following ethically justified or organizationally impossible conditions mentioned in the provision are met:

- The data processing is carried out for scientific research purposes.
- The principle of proportionality is respected – the conditions considered are highly disabling with highly heterogeneous phenotypes.
- Data will be analysed following approval by the Ethics Committee.
- Adequate security measures are in place – pseudonymization, limited access, protected storage.
- It is impossible to inform the patients – a significant percentage of the patients are deceased or uncontactable due to the conditions mentioned in the following list.

The following co-factors contribute to the objective impossibility of contacting the patients considered in the study:

- **High mortality rate of many of the considered diseases** – a significant proportion of patients have died from the disease itself.
- **Advanced age of onset of many of the considered diseases** – a significant proportion of patients have died from other causes.
- **Disproportionate effort to obtain consent** – the efforts required to contact patients are incompatible with the execution of the project due to the number of patients (over 6,000) and the timeframe considered.
- **Long timeframe of data acquisition** – contact information originally provided by patients is often outdated or no longer valid.

For transparency and informational purposes, a non-individualised notice will be made available on the institutional website of the project, including the possibility for patients to request exclusion from the project.

### 3.2 Ethical Approval Documentation

All legal documentation for the HEREDITARY project, including Annex 1, has been completed and submitted to the local ethics committee. Verification and approval of the



documents is ensured by the local ethics committee, which supervise adherence to legal and regulatory standards in data collection, processing, and sharing.

Moreover, part of the data used in HEREDITARY originates from previous ethically approved studies, such as PREDICT-NEURODEGEN (Annex 2). Where necessary, collaboration with the original project teams will be established to align HEREDITARY's analyses with the original study's objectives.

In addition, clinical data gathered through prior EU-funded initiatives have been made openly available, reinforcing HEREDITARY's robustness with respect to data availability and accessibility. Specifically, the consortium is currently working with open-access data released through the Horizon 2020 project BRAINTEASER (Grant Agreement 101017598), which is available on Zenodo (<https://zenodo.org/records/8083181>).

## 4 Data Management

UNIPD data will be stored on secure local servers, hosted at the Neurology Unit (UOC) of Azienda Ospedale Università Padova, at the Padova Neuroscience Center and at the Department of Neuroscience of the University of Padova. The following measures are foreseen:

- Pseudonymization of data at the source with key separation, prior to analysis
- Access restricted to authorised personnel
- Server access is protected by VPN, strong authentication via Secure Shell (SSH), and access logging
- Encrypted storage for the project duration
- Encrypted offline backups
- Secure destruction of data and backups at the end of the life cycle

All analyses will be conducted using a federated single-centre approach. No raw data will be transferred between centres. The UNIPD centre will perform local analyses on its own data and will only share metadata (e.g., aggregated statistics, features extracted from local models). This infrastructure ensures patient privacy and full compliance with current data protection regulations through the following approaches:

- No sharing of identifiable or raw data: all patient data will remain on local servers, securely hosted at the Neurology Unit (UOC) of the University Hospital of Padova, at the Padova Neuroscience Center and the Department of Neuroscience of the University of Padova
- Pseudonymization of data at the source with key separation, prior to analysis
- Server access is protected by VPN, strong authentication via Secure Shell (SSH), and access logging
- Full compliance with the GDPR and Italian data protection regulations

The HEREDITARY architecture is based on an asynchronous federated topology with centralised coordination. Each clinical node will be configured as an independent federated instance and will include the following components:

- Data extraction from the local database
- Data pre-processing and harmonization modules (Python, Pandas, NumPy)
- Federated analysis wrappers (e.g., PySyft, Flower, TensorFlow Federated)
- Local logger for audit, versioning, and logging
- Cryptographic module for encrypted communication of sensitive information (e.g. TLS, OpenSSL)

The results obtained from the federated single-centre models will be integrated through:

- Statistical meta-analysis
- Sharing of high-level features

Thanks to these methods, the central server will aggregate the models and results obtained at the various centres and/or the model weights using secure strategies (e.g., secure aggregation, differential privacy).

#### **4.1 Data Minimization and Pseudonymization**

HEREDITARY applies strict data minimization, processing only information essential to its research aims, as required by the GDPR. To enhance privacy safeguards, UNIPD pseudonymizes clinical records by substituting personal details with codes, limiting identification risks while preserving data value for scientific analysis.

#### **4.2 Data Transfers to Third Parties and International Transfers**

UNIPD does not foresee any data transfers, either within or outside the EU. If such transfers become necessary, they will comply fully with GDPR requirements, ensuring appropriate protection through adequacy decisions or safeguards such as standard contractual clauses, and obtaining all relevant ethical and legal approvals.

## 5 Incidental Findings Policy

In studies involving human data, researchers must anticipate the possible occurrence of incidental findings and define how to handle them. This usually includes informing participants about such possibilities and outlining how confidentiality and communication will be managed. Although HEREDITARY relies on pseudonymized data that prevents clinical follow-up, the research team must still reflect on the ethical treatment of such findings.

Incidental findings are results that emerge unintentionally and are unrelated to the study's core aims. They may be predictable, linked to known risks of procedures, or completely unforeseen given current scientific knowledge. While HEREDITARY cannot prepare specific procedures for unknown scenarios, it is necessary to account for their potential ethical and legal relevance.

One of the central ethical questions is whether participants should be informed. In the context of HEREDITARY, this is improbable due to the nature of pseudonymization, which restricts re-identification. Nevertheless, the duty of beneficence obliges researchers to consider ways to minimize harm and enhance benefits.

Within HEREDITARY, researchers are not responsible for identifying or intervening in case of incidental discoveries. Should any such result arise, it must be communicated to the Project Coordinator, who, in consultation with ethics and privacy authorities, will determine the proper course of action. Any decisions and related actions will be confidentially recorded and securely archived.

## 6 Conclusion

The HEREDITARY project, as implemented at the University of Padova (UNIPD), is fully compliant with national and European regulatory and ethical frameworks, particularly the GDPR and the Italian Privacy Code. All activities involving clinical data at UNIPD are conducted using retrospective, pseudonymized datasets collected over more than two decades of clinical activity; in accordance with the legal requirements, informed consent is not required.

UNIPD adheres to stringent data security and privacy-preserving protocols. In addition, HEREDITARY employs a federated analysis approach in which no raw or identifying data are shared. Instead, only aggregated outputs or high-level features are exchanged, ensuring full compliance with privacy obligations. UNIPD has obtained a favourable opinion from the relevant Ethics Committee for a related previous project, and a new, dedicated ethics approval request has been submitted for HEREDITARY.

Furthermore, the presented documentation clarifies the inclusion of neurologic and psychiatric disorders beyond the four primary conditions (ALS, MS, PD, and stroke), recognising the relevance of multimodal analysis and the importance of capturing complex, heterogeneous phenotypes across brain diseases.

This deliverable substantiates HEREDITARY's commitment to protecting personal and genetic data while enabling advanced scientific analysis in line with its ethical and legal obligations.

## Annexes

Annex	Name
1	Legal documentation and ethics approval request for the HEREDITARY project
2	Legal documentation and ethics approval for the project: PREDICT-NEURODEGEN