This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement Nº 101137423

D1.2 – Data Management Plan

Deliverable No.	D1.2	**Due Date**	31/12/2024
Description	This document contains information on the collection, storage, sharing procedures and strategies of Data. It serves as a roadmap for ensuring that data generated or used in the PROTECT-CHILD project is managed effectively, ethically, and in accordance with best practices and regulatory requirements.		
Type	Report	**Dissemination Level**	PU
Work Package No.	WP1	**Work Package Title**	Coordination
Version	1.0	**Status**	Final

Authors

Name and surname	Partner name	e-mail
Adrian Quesada Rodriguez	UDGA	aquesada@udgalliance.org
Vasiliki Tsiompanidou	UDGA	vtsiompanidou@udgalliance.org
Renata Radocz	UDGA	rradocz@udgalliance.org
Ana Maria Pacheco Huamani	UDGA	admin@udgalliance.org
Eugenio Gaeta	UPM	eugenio.gaeta@lst.tfo.upm.es
Franco Mercalli	MME	f.mercalli@multimedengineers.com
Alessio Fioravanti	Sapienza	alessio.fioravanti@uniroma1.it
Ioanna Drympeta	CERTH	idrympeta@iti.gr
Irene Sánchez Frías	UGR	irenesanchezfrias98@gmail.com
José Antonio Castillo Parrilla	UGR	castillop@ugr.es
Matthew Salanitro	CUB	matthew.salanitro@charite.de

Document History

Version	Date	Changes	Authors
0.1	10/10/24	Initial draft and TOE	Adrian Quesada Rodriguez
0.5	1/12/24	Updated content	Adrian Quesada Rodriguez
0.7	12/12/24	DMP Questionnaire inputs added	Adrian Quesada Rodriguez, Vasiliki Tsiompanidou
0.8	13/12/2024	Updated content	Vasiliki Tsiompanidou
1.0	26/12/2024	Final version ready for peer review	Adrian Quesada Rodriguez, Renata Radocz
2.0	3/12/2024	Final version integrating changes recommended by peer-reviewers and approved Quality Manager	Adrian Quesada Rodriguez, Renata Radocz

Key data

Keywords	Data; Data Protection; Data Management; Data Storage; FAIR Data; IPR; Ethics; Privacy
Lead Editor	Adrian Quesada Rodriguez, Vasiliki Tsiompanidou
Internal Reviewer(s)	José Antonio Castillo Parrilla, Irene Sánchez Frías - UGR

Abstract

This document provides the initial Data Management Plan (DMP) describing the strategy for data management for the PROTECT-CHILD project. More specifically, it outlines the methodologies for the efficient management, sharing, and preservation of research data generated throughout the project's lifecycle. This first version is intended to provide an overview of the envisioned partners' strategies in regard to the collection, generation, and handling of such research data. Furthermore, it provides a description of best practices and strategies for data management, including as prescribed in the European Commission's guidelines. As a living document, the DMP will be continuously updated throughout the runtime of the project, taking into account any developments in regard to the partner's data collection, storage or processing activities.

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 About this deliverable

This deliverable constitutes the first iteration of the Data Management Plan (DMP), providing a first look into the (expected) data collection, generation, storage, sharing procedures and management strategies envisaged by the partners of PROTECT-CHILD. As such, provides an initial overview of the steps towards ensuring compliance with FAIR (Findable, Accessible, Interoperable and Reusable) data and open science principles adopted within the project. Finally, it takes into account from the beginning existing and upcoming Intellectual Property Rights (IPR) requirements and future exploitation plans, so as to cover the entire project's lifecycle.

This deliverable also designs the foundational framework for the management of data throughout this project, offering insight into the conduct of partners while also serving as an informative guide for the correct handling of data, whether personal or non-personal.

What is more, this first iteration of the DMP focuses on defining and outlining basic principles that are relevant to the management of research data within the project. In order to achieve this, it leverages normative requirements, such as data-related regulations (eg. General Data Protection Regulation, Data Governance Act etc), and, relevant guidelines, with the purpose of identifying best practices that could assist partners in further refining their data-related activities.

The DMP is a living document that evolves along with the project and will, thus, be updated throughout its lifetime in order to reflect the partners' ongoing activities regarding data. Said updates will be reported in future iterations.

1.1 Deliverable context

Table 1. Deliverable context

PROJECT ITEM IN THE DoA	RELATIONSHIP
Project Objectives	Project Coordination is devoted to efficiently managing all aspects of the project, including recurring activities, monitoring and orienting technical and innovation activities, supervising and assessing the scientific and medical research, ensuring quality of results and timeliness of implementation through strict risk monitoring.
Exploitable Results	The deliverable presents a model for data management and will serve as a reference for the consortium by outlining best practices and requirements in regard to good Data Management, compliance with ethical and legal requirements and securing Intellectual Property Rights.
Workplan	The deliverable will be continuously updated according to the project's evolution to best reflect the developments. Partners will provide updates to their internal data management via DMP Questionnaires.
Milestones	D1.2 contributes to the development of the data architecture (Milestone1) and the design of the PROTECT-CHILD legal framework (Milestone2), as it provides the baseline for good data management practices and compliance requirements to be used as a reference for consortium partners.
Deliverables	D1.2, D1.3
Risks	The project's activities evolve over time, as do the related data management practices and strategies. Continuous monitoring needs to be conducted.

2 Introduction

2.1 The PROTECT-CHILD Project

The PROTECT-CHILD Project aims to research, design and implement a privacy-protecting European Platform for Child Transplants in order to improve the outcomes of rare paediatric transplant patients. In order to achieve this, the project intends to integrate multiple sources of data from registries, hospital-based and public repositories into a European Reference Network.

Taking the above into consideration, the project focuses on the co-design of a secure and privacy-preserving infrastructure, where data will be made available in harmonised standards, in alignment with relevant regulations, most notably the European Health Data Space (EHDS) Regulation and General Data Protection Regulation (GDPR) principles.

As such, the project focuses on the following central objectives:

1. Co-designing children's transplant **data-sharing frameworks based on state-of-art data integration standards**, data governance policies, and a secure, trustable and privacy-preserving IT architecture;
2. Implementing an advanced **EHDS-compliant and federated infrastructure for distributed data sharing**, designed to support the TEHDAS user journey;
3. Ensuring **data interoperability and privacy-preserving data processing** leveraging on state-of-art standards (e.g., OMOP, FHIR, OpenEHR etc.) and data models by implementing a set of services for data discovery, data governance, authentication and permit, data preparation and use compliant with the EHDS user journey;
4. Developing and assessing **intelligent Data Discovery Systems** that integrate techniques, tools and standards including NLP and metadata models for the integration and linkage of real-world (e.g. EHRs, clinical referrals, etc.), transplant registries and research data from clinical studies, also exploring the potential benefits of using quantum computing-based MPC (Multi-Party Computation) algorithms in managing private genomic data;
5. Developing **tools and services for collaborative data reuse** that comply with the EHDS, DGA and GDPR regulations for the definition of abstract platform requirements for security, privacy and audit by design approaches based on mutually recognised data governance frameworks;
6. Designing and implementing a **user-friendly platform for federated data discovery, data analysis and results presentation,** as foreseen by the EHDS user journey;
7. Deploying the PROTECT-CHILD **federated data platform and common tools and services** among paediatric units of the participating hospitals in the context of the pilot study envisioned to involve transplanted children with liver and kidney diseases;
8. Assessing the **expansion potential of the pilot experiences** beyond the involved centres and the addressed rare transplants;
9. Building sound communication and dissemination channels for **sharing of information, knowledge, scientific and technological results,** including clustering actions with related European projects and initiatives, citizens and patients' families, GPs and the medical community of the PROTECT-CHILD activities;
10. Ensuring a **robust ethical framework** for the whole PROTECT-CHILD system and infrastructure, that protects children's health data use for research and clinical decision support, by continuous monitoring by legal and ethical advisors.

3 Data Summary

Given the above-described project objectives, data is intrinsically linked to the PROTECT-CHILD objectives, as it forms the primary point of consideration when designing and implementing the project's solutions. Similarly, making data available for reuse in order to improve healthcare lies at the heart of the project, in line with open data requirements and the latest EU regulations promoting data sharing for such purposes.

In order to facilitate that, the project will develop adequate tools and services in line with harmonised standards to ensure interoperability, while also promoting transparency and audits to validate privacy and security practices.

In fact, as evidenced, the project places a heavy focus on privacy and security, aiming at ensuring secure and compliant processing, analysis, and sharing of personal data, taking into account particularly their sensitive character as health data and the involvement of minors.

The above is further reflected in ongoing discussions and partners' upcoming activities, as they have been collected through a Data Management Questionnaire that aimed to collect partners' views and practices on data management, ethics and IPR. The questions attempted to gain insights into the partners' methods of data collection, what categories of data they are collecting, how they are storing the data etc.

The present section will provide a summary of said responses collected from partners as they are envisioned and designed at this stage of the project, while the complete responses are available upon request. As this is the first iteration of the DMP, the partners were expected to provide a high-level view of their activities and are expected to provide further updates while the project progresses in order to better reflect the evolution of their data-related activities, as will be reported in the following version of this deliverable.

The present section delves deeper into the categories, nature and type of data that is being collected and/or processed by PROTECT-CHILD partners in the context of their project-related activities. It also provides more insight into the following aspects related to the PROTECT-CHILD data:

- The methodology for collecting the data;
- The data sources;
- The purposes behind data collection and/or data processing;
- The introduction and reuse of existing datasets in the context of the project;
- The requirements at a partner level regarding data storage; and
- The measures in place or envisioned to ensure data security.

Some of these elements are disclosed only at a high level due to the nature of this deliverable, additional information is available upon request.

3.1 Data Collection and Processing

3.1.1 Nature of Data Collected and Processed

Partners within the PROTECT-CHILD project are expected to collect and/or process the following two main categories of data:

a) **Non-personal data**, including information, knowledge and technical data that do not involve personally identifiable information. Such data may be collected and/or

processed in the context of the design, development and implementation of the project's solutions, as well as through anonymous questionnaires with relevant experts and stakeholders to collect input during the co-creation phase or feedback on the developed solutions and approaches.

b) **Personal data**, including any information that can lead to the identification of individuals, including, for instance, names, email or IP addresses. Personal data is primarily related to the performance of the pilots' activities, particularly implementing the PROTECT-CHILD solutions and methodologies.

Personal data may further be collected in the context of interviews and questionnaires to assess the project's activity, as it is envisioned that the views of patients and of their families will be collected to assess their needs and the results of the pilot's activities.

- **Special categories of personal data**, involving personal data revealing sensitive information according to the GDPR, such as sexual orientation, racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, as well as any health, genetic or biometric data related to the data subjects. In the context of PROTECT-CHILD, health data will be collected to perform the necessary actions, as will be described below.

  Particularly in the context of the work performed in the pilot sites (SERMAS, ISMETT, UPMC, UKE, UNIPD), health data will be utilised, with a particular focus on genomic data of patients up to 18 years old.

The following table summarises the planned and envisioned categories of data to be collected and/or processed by the project's partners, starting with the pilot sites and moving to the rest of the partners. The table focuses on the dichotomy between personal and non-personal data and it reflects the current stage of the project. As its activities and needs evolve, the table will be further completed and updated accordingly.

Table 2. Categories of data collected and/or processed by PROTECT-CHILD partners.

Partner's Name	Personal Data	Non-Personal Data
SERMAS	☑	
ISMETT	☑	
UPMC	☑	
UKE	☑	
UNIPD	☑	
BIOMERIS	TBD	TBD
BELIT	TBD	TBD
CERTH	☑	
GTRC	TBD	TBD
HL7	TBD	TBD
INTM	TBD	TBD
MME	☑	☑

Partner's Name	Personal Data	Non-Personal Data
UDGA		☑
UDEUSTO	☑	
UGR		☑
UNIROMA1	☑	☑
UPM	☑	☑
UTW	☑	☑

3.1.2 Purpose of Data Collection and Processing

In the context of the project, the use of data will serve the purpose of fulfilling the envisioned project activities in order to meet the above-described objectives, in accordance with the Description of Action.

Data is to be processed in order to implement harmonised data standards (eg. in accordance with the OMOP Data Model), so as to enable data analyses in a common format and structure, aligning it with standardised vocabulary.

Regarding, in particular, the pilots' activity, health data will be utilised in order to conduct research on two types of genetically based rare paediatric diseases that need transplant, namely liver and kidney disorders. All the clinical, lab, prescription and any other data used to characterise the patients will be obtained from the clinical records (real-world data), following a pragmatic approach.

Specifically, the pilot use of retroactive and prospectively selected data during the project aims to

a) contribute to the information of the Genome-Wide Association Studies (GWAS) studies and the inclusion of this information in European repositories;

b) identify new genes/genomic regions associated with those diseases.

c) identify new genetic biomarkers associated with pharmacogenetics response, adverse drug reactions and drug toxicity in liver and kidneys transplanted patients and estimate related risk;

d) identify and discover susceptibility to genomic markers (to drug reaction, viral illness, environmental factors, etc.) in the genomic background of the patients;

e) detect specific clusters of patients with similar disease markers, so as to discover and validate new specific methylDNA signatures for diseases/genes/variants that are not known to date, or for potential new pathologies associated with transplanted children;

f) incorporate the genomic data into the Beyond 1 M Genomes project (B1MG).

Finally, data will be collected and processed in order to validate the PROTECT-CHILD approach and solutions, including relevant feedback, so as to assess its results and design adequate strategies to move forward.

3.1.3 Data format and size

The precise data formats for each type of data and expected size are currently being defined at the project level, taking into account the work performed within T2.3 as well, aiming at establishing the data and metadata models.

That being said, the following data is anticipated within the project:

a) **Project documentation:**
   a. Format: docx, xcsl, ppt, pdf
   b. Size: MBs

b) **Dissemination and communication materials:**
   a. Format: html, ppt, pdf, jpg, psd
   b. Size: MBs

c) **Surveys:**
   a. Format: txt, CSV, JSON
   b. Size: MBs

d) **Questionnaires:**
   a. Format: txt, CSV, JSON
   b. Size: MBs

e) **Interviews, workshops and focus groups:**
   a. Format: Mp4, jpg, txt, wav, mp3,csv
   b. Size: MBs

f) **Publications:**
   a. Format: docx, web links, Digital Object Identifiers (DOI)
   b. Size: MBs

g) **PROTECT-CHILD patients' data:**
   a. Format: To be defined
   b. Size: GBs

The above list will be updated as the project progresses to reflect the ongoing work performed in the respective Work Packages (WPs).

3.1.4 Method of Data Collection and Data Sources

Even though the detailed methodology for data collection and/or processing, as well as the data sources, are currently being decided, partners have identified the primary means of data collection as follows:

- The data is collected directly from data subjects belonging to partners' research teams (eg. relevant for the performance of surveys and interviews etc);
- The data is collected directly from data subjects outside of partners' research teams (eg. patients, early adopters, beta testers, etc);

- The data is collected indirectly through other partners of the project (eg.; in order to provide support in the performance of their tasks or to harmonise data vocabularies etc);
- The data is collected indirectly through other organisations external to the project (eg. reusing survey results conducted by another organisation etc).

3.2 Data Storage and Security

Data collected in the context of the project will be stored and processed locally. Thanks to the PROTECT-CHILD federated infrastructure, each data holder/owner is able to store its data locally according to their own data sovereignty prerogatives, with no request to physically deposit data at a central repository that would be managed by the project.

Similarly, the federated architecture allows for data analyses of patients' data to be performed locally, without sharing the data with third parties, ensuring that any data analysis is performed in a decentralised collaborative learning setting. As such, no data used to produce the outputs ever leaves the local sites, while only aggregated data is shared with other parties in the context of clearly defined legal agreements.

The project's architecture ensures personal data protection and security encapsulating relevant data sources in an EHDS capsule for data sovereignty enforcement and consent management. These principles enable an interoperable, FAIR-compliant, secure federated genomic linkage infrastructure for sustainable data access across Europe.

In addition to the above, where personal data is collected and/or processed within the context of the project, an adequate lawful basis will be recognised prior to any collection/processing. Where consent is required, partners will make sure that they acquire the prior, informed, freely given, specific and explicit consent of the data subjects, while for patients below the age of 18, the consent of their legal guardians will be obtained. The project will further develop and provide consent tools for data holders to enhance the procedure for collecting granular consent from data subjects.

Where appropriate, PROTECT-CHILD will prioritise anonymisation of the data, using techniques such as K-anonymity, generalisation and statistical methods. Further technical and organisational measures implemented at the data source will include but are not limited to:

- Pseudonymisation/Anonymisation techniques;
- Encryption (including encryption at rest);
- The establishment of network access protocols (e.g. firewalls, VPN access);
- The adoption of adequate access control policies;
- The adoption of data backup policies;
- System Auditing;
- Staff cybersecurity training;
- The implementation of physical security (e.g. locked vault, CCTV controls);
- The adoption of risk assessment and contingency plans;
- The use of quantum computing techniques to enhance security.

Any data collected or processed will be retained no longer than necessary to perform the required project-related actions and in accordance with partners' internal procedures and applicable regulations.

4 Applicable Legal and Ethical Framework of Project Child

This section will provide an initial overview of the legal requirements that should be taken into consideration throughout the entire project's lifecycle, focusing on a high-level description of the primary obligations that partners need to keep in mind when designing, developing and implementing the project's data-related solutions and activities. Said initial analysis will serve as the baseline for the project's legal compliance framework which will be further analysed and expanded in D2.2.

The work performed by WP11 has further expanded on these actions through the performance of dedicated webinars which introduce project partners to compliance considerations. For this reason this section will remain at a high level with the main goal of complimenting the consortium agreement's dispositions on compliance.

4.1 Legal Requirements

The following sections present brief introductions to the top regulatory frameworks to be considered by project partners in their project-related activities. Several other legal frameworks are of relevance to this project, including the Cyber Resilience Act (CRA), the Network and Information Security Directive 2 (NIS2), the Artificial Intelligence Act (AI Act), the Digital Services Act (DSA), and the ePrivacy Regulation (ePrivacy). Additional clarity and guidance will be provided in future versions of this (and associated) deliverable as needed to address both the project goals and the evolving needs of the project partners.

4.1.1 General Data Protection Regulation

The General Data Protection Regulation (GDPR) is a key document in regard to data protection in the European Union, and, as such lays down strict guidelines for the collection, processing, storage, and sharing of personal data. Adhering to these regulations is essential for safeguarding participants' rights and maintaining the integrity of the overall project data management.

Article 5 of the GDPR lays down the following principles for data processing:

- Lawfulness, Fairness and Transparency of Processing
- Purpose Limitation
- Data Minimisation
- Accuracy
- Storage limitation
- Integrity and Confidentiality
- Accountability

Furthermore, the GDPR codifies the following rights of the Data Subjects, which have to be taken into account:

- Right to be Informed (Article 13 and 14 GDPR)
- Right of Access (Article 15 GDPR)
- Right to Rectification (Article 16 GDPR)
- Right to Erasure (Right to Be Forgotten) (Article 17 GDPR)

- Right to Restriction of Processing (Article 18 GDPR)
- Right to Data Portability (Article 20 GDPR)
- Right to Object (Article 21 GDPR)
- Rights Related to Automated Decision-Making and Profiling (Article 22 GDPR)
- Right to be notified (Articles 16, 17, and 19 GDPR)

The GDPR also separates research data into various different categories. As such, sensitive data, as defined in Article 9 GDPR, includes information that requires stronger protection due to its potential impact on the data subjects' privacy, especially if it is being mishandled. In the context of research, medical data (either concerning adults or minors) are also seen as categories of sensitive data, due to the same issues. Thus, medical data encompasses information about the physical or mental health of an individual, also including genetic and biometric data, all of which are integral to advancing medical research but must be handled with strict safeguards to ensure privacy and consent. When research involves minors, additional protections apply, as children are considered vulnerable subjects with limited legal capacity to provide informed consent.

The rights and principles laid down in the GDPR must be an integral part of the handing of research data conducted by the consortium partners. Key elements to GDPR compliance is the performance of clear and comprehensive records of data processing activities (ROPAs) and Data Protection Impact Assessments (DPIAs). To ensure that these actions are dully performed in a coordinated and transparent manner, UDGA has made available a dedicated solution[1]: DPiD (available at https://dp-id.com/). All project partners shall register their relevant data processing activities in this database to ease compliance cooperation amongst the consortium members.

[1] Data Processing ID is a global registry of public information on data processing activities. It enables companies, public administrations and research infrastructures to enhance transparency and comply with some of their legal obligations by creating a unique identifier and a public record for their data processing activities (Data Processing ID or DP-ID) It enables to:

- Inform data subjects on the processing of their personal data and their rights;
- Comply with the legal obligation to inform data subjects (art. 13 and 14 GDPR);
- Facilitate the management of data processing internally and with third parties;
- Map and monitor interdependent data processing involving multiple companies;
- Use QR-Codes, hyperlinks (URL), and widgets to share data processing required information.

The registry provides free information. The accuracy of the information remains under the exclusive responsibility of the user who has provided the information and who can be directly contacted in case of error or question.

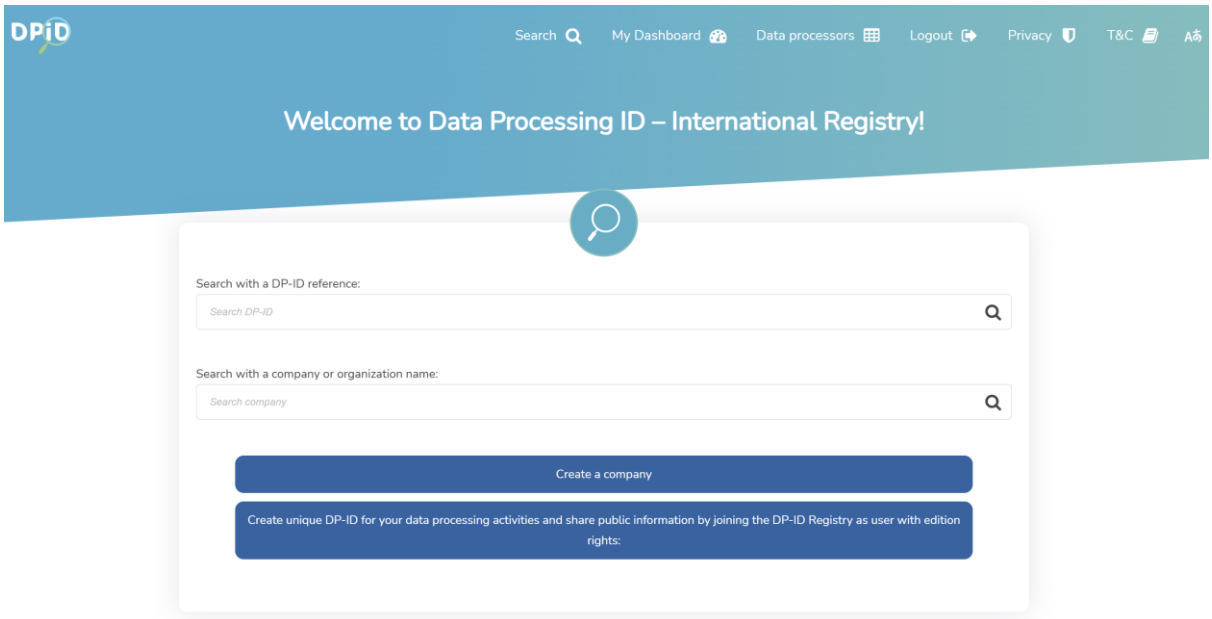

Figure 1 DPiD solution

4.1.2 European Health Data Space Regulation

The Proposal for a European Health Data Space (EHDS) has the aim of fostering the secure and efficient use of health data across EU Member States. It was officially proposed by the European Commission in May 2022, and on March 22 of 2024, the members of the European Parliament reached a provisional agreement. The Proposal still needs to be approved by the Council, once published in the EU's Official Journal, it will enter into force twenty days later. The EHDS will be applied two years after, with certain exceptions, including primary and secondary use of data categories, which will apply four to six years later, depending on the category.

The EHDS seeks to create a unified framework that enables individuals, healthcare providers, researchers, and regulators to access and share health data easily, while on the other hand still safeguarding privacy and security. Thus, this regulation specifically focuses on health data. The goals of this regulation are twofold: Empowering individuals to have greater control over their health data and, especially, promoting secondary uses of health data for research, innovation, and policymaking. Here, citizens should be able to access and share their health data more easily and without overly bureaucratic administrative barriers. In regard to the secondary use of medical data, the EHDS creates frameworks for researchers and public health authorities to utilize anonymized or pseudonymized data to drive medical advancements and improve healthcare delivery, all while ensuring strict data protection measures.

Although the legislative process for the EHDS has not yet been concluded, it is essential to consider the EHDS Proposal throughout the project, as the secondary use of health data is deeply embedded in the project's goals.

4.1.3 Data Governance Act

The Data Governance Act (DGA) is another important regulatory piece in the EU, aimed at establishing a reliable and effective framework for data sharing across various sectors and Member States. Officially adopted in May 2022, the DGA aspires to harness the potential of data as a vital catalyst for innovation and economic development, while simultaneously safeguarding fundamental rights. It acts as a foundational element of the EU's overarching European Data Strategy, which seeks to develop a secure, interoperable, and accessible single market for data, benefiting businesses, researchers, and public authorities.

The DGA introduces mechanisms to promote voluntary data sharing (without distinguish between personal and non personal data), ensuring strong protections for privacy, security, and transparency. Among its key features are the establishment of Data Intermediaries—neutral entities that facilitate data sharing between data providers and users without exploiting the data—and the encouragement of Data Altruism. This concept advocates for individuals and organizations to willingly share their data for the public good, such as for research or policy development, under well-defined and transparent conditions. Furthermore, the DGA sets the stage for the creation of European Data Spaces—sector-specific frameworks designed to improve data sharing in areas such as health, agriculture, and energy.

The DGA is particularly pertinent to data management practices, as it establishes a benchmark for ethical and sustainable data governance in the digital era. By standardizing the processes of data sharing, processing, and reuse, it addresses challenges such as data silos, trust deficits, and legal ambiguities. The Act incorporates measures to ensure interoperability and transparency, which are essential for facilitating cross-border and cross-sectoral data exchanges. Additionally, the DGA enhances confidence among individuals and organizations by emphasizing privacy and consent, especially through secure frameworks for anonymized or pseudonymized data. For both businesses and public institutions, the DGA offers a definitive guide for effective data management.

4.1.4 Data Act

The EU's Data Act is aimed at establishing equitable guidelines for data access and utilization among Member States. It seeks to guarantee fair access to data, empower both individuals and businesses, and stimulate innovation within the EU's data economy. It serves as a fundamental component of the EU's overarching European Data Strategy, complementing other significant initiatives such as the DGA and the GDPR.

The Data Act emphasizes enhancing control and utilization of data, particularly focusing on non-personal data generated by various devices and services. It aims to define explicit rights for users and organizations to access the data they generate, addressing existing disparities where manufacturers or service providers typically dominate access. The Data Act signifies a transformative change in the sharing and governance of data. By establishing clearer rights and responsibilities regarding data access and usage, it lowers barriers to innovation while promoting transparency and fairness. The Act also facilitates data interoperability, an essential characteristic for enabling seamless data exchange across various industries and borders. Its emphasis on fairness and equitable value distribution is particularly significant for businesses, ensuring that smaller entities can compete on an equal basis within the data economy. For policymakers, researchers, and the general public, the Data Act guarantees that data is managed responsibly and equitably.

4.2 Ethical Considerations

The objectives of the PROTECT-CHILD project, previously described, give rise to several ethical issues that must be addressed. These foreseen challenges primarily converge on five key areas that require careful consideration to ensure the protection of participants' rights and well-being.

1. Informed Consent

Obtaining consent in pediatric populations poses unique challenges, particularly for minors. Ensuring that consent is informed, meaningful, and age-appropriate is critical. Dynamic consent introduces complexities, such as updating minors' preferences when they reach the age of majority. Balancing patients' right to opt-out and social interest is a major concern,

2. Data Protection, privacy and data security

Handling special categories of personal data (sensitive data) demands robust safeguards to ensure compliance with data protection laws, such as the GDPR. Particular risks arise in rare disease populations, where data anonymization is more challenging. Secondary data use and post-mortem data require clear ethical frameworks to prevent misuse and respect participant confidentiality.

3. Incidental Findings

Genomic analysis may uncover incidental findings unrelated to the study's objectives. Ethical considerations include determining whether to disclose such findings, respecting participants' preferences, and providing appropriate genetic counseling to mitigate potential psychological impacts.

4. Artificial Intelligence (AI)

The use of AI systems introduces risks of algorithmic bias and lack of transparency. Ensuring that AI systems are explainable (XAI) and equitable is essential to prevent harm and promote trust in AI while benefiting from it. When deploying Artificial AI in the project, partners commit need to adhere to ethical AI principles, including transparency, explainability, human oversight, and fairness, while ensuring the protection of privacy and security.

5. Equity and Accessibility

Equity in research access is paramount or the objectives of the project, ensuring no discrimination against any pediatric participant. The project's tools and platforms must be inclusive and accessible to all users, including those from diverse socioeconomic backgrounds.

5 Data Ethics in the Context of the Project

Ethical compliance is at the forefront of the PROTECT-CHILD project encompassing the design and implementation of all of its activities, including data management. Dedicated tasks in WP2 and WP6, as well as the entirety of WP11 are focused on ensuring that both the legal and ethical requirements are considered throughout the project's lifecycle and are enshrined within each part of its activities.

The following sub-sections will present the steps adopted and envisioned to ensure ethical compliance and data ethics both at a project and at a partner level, while the complete ethical strategy will be discussed in WP2 and WP11 deliverables.

5.1 Ethical Coordination at the project level

Considering the multitude of partners involved in the development, implementation and validation of the PROTECT-CHILD solutions, it is anticipated that a degree of divergence will be present among their ethical and legal practices. This has already been taken into account from the beginning of the project and is duly reflected in all aspects of the project coordination decisions.

As such, the Consortium includes a series of experts in the field of ethics and compliance, to ensure that all relevant matters are taken into account by design and by default, further promoting discussions between partners on all of the relevant matters concerning ethics.

To promote a homogenous approach at a project level, the coordination team has developed a governance structure that encompasses partners' legal and ethics experts and Data Protection Officers (DPOs), so as to facilitate communication and decision-making regarding the ethical compliance of the project.

The internal ethics structure is finally complemented by an external independent ethical advisory board that is tasked with further supporting the project's ethical compliance, providing external ethical advice and guidance. In order to best utilise the board's contribution to the project, all of the key decision-making procedures are meant to be further validated with the ethics advisors, so as to not only ensure a solid ethics compliance framework, but also to receive recommendations for improvement or future work.

WP11 introduces dedicated ethics compliance activities, which are of required consideration to all project partners, particularly as they relate to project generated, processed, compiled, or managed (personal and non-personal) data. The above-described structure is depicted in the figure below, as this was reported in D.1.1.

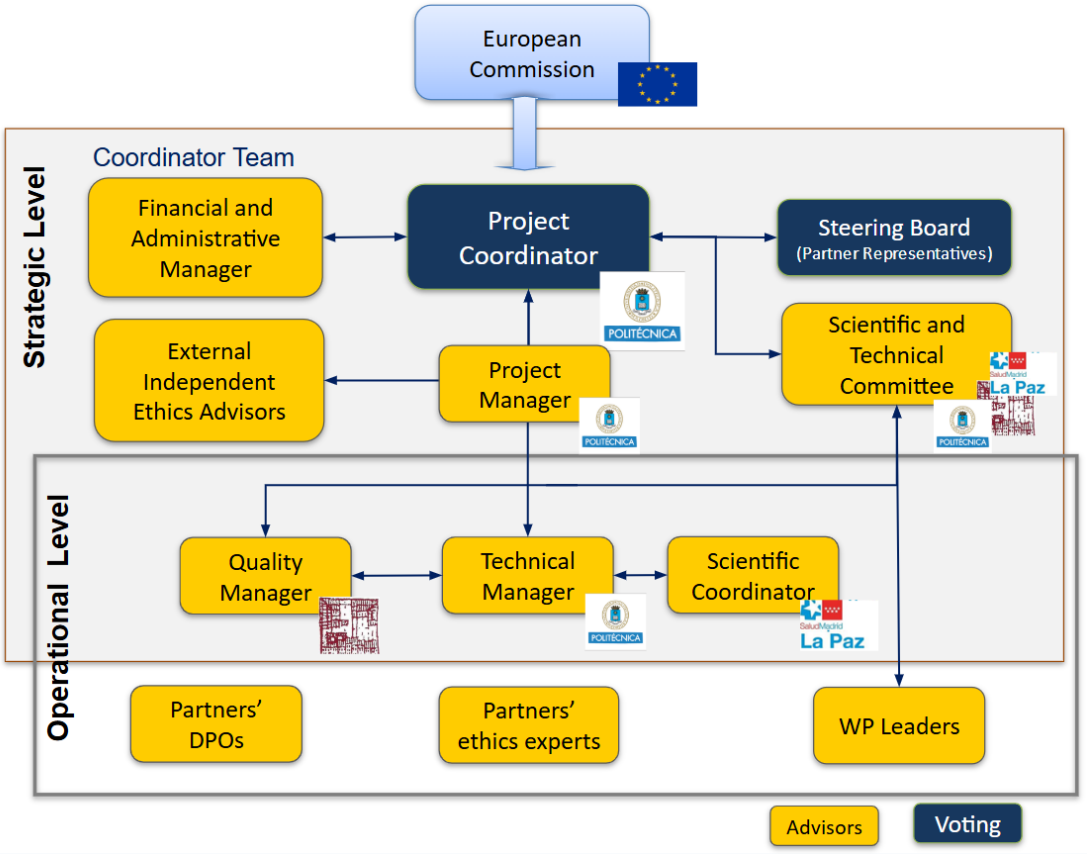

Figure 2. PROTECT-CHILD Governance Structure.

5.2 Data Ethics in PROTECT-CHILD

Having established the overall approach towards ethics and compliance within the project, each partner was called to provide more insight on their individual data ethics approach within the respective organisations.

In this regard, partners have taken great consideration of ethical matters related to privacy, abiding by the principle of data minimisation and ensuring adequate technical and organisational measures are in place to protect the security and confidentiality of personal data, including anonymisation, pseudonymisation and encryption. Where applicable, a data protection impact assessment will be performed.

Where consent is required, partners ensure that it is freely given, informed, specific and unambiguous, provided in a written form prior to any processing of the data. Where minors' data is meant to be processed, their legal guardians' consent is acquired, ensuring that the nature, purposes and objectives of the project and the envisioned processing are described in simple and clear language that is easily understood by them.

Where required, prior ethical approvals will be duly conducted by partners, according to the procedures applicable to each organisation. The project's coordination team, as well as the legal and ethical experts will provide guidance and support in these activities so as to ensure they are performed in an efficient and coordinated manner.

Where Artificial Intelligence (AI) is deployed, partners are committed to ethical AI principles, including, but not limited to, transparency, human oversight, and fairness, while safeguarding privacy and security. All of the data that will be used to train or validate AI models, as well as any data processed using AI techniques will be handled with care, ensuring that no identification of

data subjects can take place. Similarly, data subjects will not be subject to automated decision-making processes that significantly affect without human validation of any results.

Finally, partners already plan to put adequate procedures in place to assess and validate that any data used is of high quality in order to, among others, ensure that no biased or discriminatory outputs are produced.

Dedicated data and AI governance frameworks and guidelines will be generated jointly by the project partners involved in WP1-11 in direct alignment with the needs of project stakeholders. This being said, it is highly recommended that all project partner follow dedicated standards for data management and governance frameworks, and dully inform their practices to other project partners to ensure and maximize operational alignment.

6 FAIR Data Management

The FAIR principles – representing Findability, Accessibility, Interoperability, and Reusability - constitute a fundamental framework for effective data management in research initiatives. These principles were established to facilitate the discovery and utilisation of digital assets, with the objective of ensuring that research data is organised and disseminated in ways that optimise its potential value. By following the FAIR principles, researchers can improve the visibility, dependability, and influence of their findings, thereby encouraging collaboration and innovation across various fields.

This section provides further input on the FAIR practices adopted and envisioned within the project at its current state. It is noted that partners are committed to ensuring compliance with the FAIR principles and are in the design phase of the precise measures and methodologies to be adopted with the aim of ensuring that PROTECT-CHILD is in line with the relevant requirements.

6.1 Findability

Findability highlights the necessity of making data easily discoverable through well-defined and persistent metadata, indexing, and unique identifiers.

As such, partners are already in the process of identifying and ensuring data and metadata adhere to adequate standards and are followed by persistent identifiers. Standards regarding data and privacy, such as ISO 27001 and NEN 7510 are already envisioned, while DOIs are planned to be assigned where applicable.

Additionally, the work performed in T2.3 is focusing precisely on the identification and definition of the most appropriate common data model relevant for the PROTECT-CHILD data, as well as the methodologies for metadata standardisation, including aspects as well such as metadata standardisation, utility and quality labelling (e.g. data provenance, data timeliness, bias examination, etc.).

In accordance with the federated infrastructure of PROTECT-CHILD, the data will be stored locally at the pilot sites in capsules enabling multi-standard data transformation hubs that include HL7 FHIR, OMOP, among other standards. The multi-standard data transformation hubs will enable optimisation promoting the usage of standards that is better for each specific use case (for instance FHIR for data exchange, OMOP for data persistence).

Findability is further enhanced by the development of the multimodal data navigation, that allows end-users to easily “meta-query” the Data Ecosystem to find the datasets they need, respecting specific requirements in terms of accessibility, quality, variables, size, etc.

6.2 Accessibility

Accessibility guarantees that data is available to those who require it, while adhering to ethical and legal standards and maintaining transparency regarding usage conditions.

Accessibility lies at the heart of the PROTECT-CHILD project, as the primary objective is making health data, as described above, available for reuse while maintaining their security. Overall, data will be made accessible for re-use through the European infrastructures (ELIXIR, GDI) to the greater extent possible.

The accessibility planning within the project takes into account the sensitive nature of the data itself, as well as the increased need for security, in accordance with the GDPR and the EHDS

Regulation. As such, it is meant to promote access and reuse of data while adopting a privacy and security by design approach.

In addition to the secure processing environment and the federated architecture, the PROTECT-CHILD capsules, where patient data is stored, enable local analysis of private data by embedding the entire analytics software stack, including open-source distributions like TensorFlow, Apache Spark, and Apache Hadoop.

Clear and comprehensive strategies for granting data permits to data users will be defined within the project, so as to facilitate access to the data in a compliant manner.

The overall system usability, accessibility, and users satisfaction for the developed tools and their application into clinical workflows will be assessed by measuring the usage ratio of the system during the pilots, by the evaluation of learning curves for the pilot users (e.g. by having the tools used for at least one month to prepare the multidisciplinary tumour boards of participating centres) and by means of questionnaires (e.g. SUS, AttrakDiff12) administered to physicians and researchers at the participating centres after the pilot's experiments.

For non-personal data, partners are exploring the use of research repositories (eg. zenodo, OSF etc), so as to further promote open access to the project's results and knowledge generated and its reservation. HTTPS will be utilised to promote secure access to data.

Raw data will be anonymised to the greatest extent possible, in order to ensure data security, prior to its being made available in accordance with open science principles. Access to raw data will be subject to access restrictions and will only be granted where absolutely necessary.

Knowledge acquired in the context of PROTECT CHILD will be disseminated, among others, through scientific publications, in line with open science principles, benefiting from research institutions' Open Data infrastructures and tools, such as the Coordinator's Open Data Portal Configuration, the European OpenAire platform and Open Access to peer-reviewed papers in scientific journals.

6.3 Interoperability

Interoperability emphasizes the need for data to be integrated and usable across different platforms, formats, and systems, necessitating compliance with widely accepted standards and terminologies.

Given the rich content of the data envisioned within the PROTECT-CHILD project, interoperability is of utmost importance. The work conducted in T2.3, as meant to be reported in D2.3, fosters interoperability through the establishment of a common data model for the data and metadata involved.

Said plan takes into account the need to balance interoperability against the need for a secure data sharing/processing environment, so that all data is analysed in a transparent, secure, and privacy-compliant manner.

Leveraging the work performed at Georgia Tech Research Corporation to develop a highly efficient and organized infrastructure that connects their research data, stored using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), with their clinicians through a SMART-on-FHIR application, the PROTECT-CHILD project will further enhance collaboration and interoperability between the Health Level Seven (HL7) and Observational Health Data Sciences and Informatics (OHDSI) standards.

At a governance level, standardised data permits and agreements will be used to grant access to the electronic health data for secondary use solely to authorised data users.

Finally, the system's capabilities will be tested in order to assess whether compatibility with ongoing cross-country health data exchange initiatives (e.g. the PETER registry) are feasible.

6.4 Reusability

Reusability ensures that data retains its value for future research by advocating for comprehensive metadata, clear licensing, and contextual information.

In this regard, reusability of data in compliance with the EHDS is intrinsic to the project. PROTECT-CHILD data will be made available for any type of use and reuse (e.g., to inform improved healthcare processes, to support data-driven decision and policy making, to support epidemiology research, to investigate AI predictive/prognostic models, etc.) that is compliant with the upcoming EHDS Regulation.

In this regard, the project will establish a secure environment (the Capsules) that will allow for a federated data analysis to be performed. Through the PROTECT-CHILD platform, Data Users (eg. Researchers, health practitioners etc) will be able to receive the aggregated results of the data analysis in an easy and standardised manner.

The conditions for reuse, the documentation required, as well as the agreements and data permits will be standardised at a project level so as to ensure a smooth and coordinated approach to further promote reusability. In this context, the nature of the data will be taken into account, while partners Intellectual Property rights will be ensured.

As already explained, the data will be accompanied by appropriate metadata to allow the performance of the original data queries prior to the performance of the analysis, after users have been securely authenticated.

Where applicable, partners are exploring licensing options to ensure alignment with organisational requirements.

Finally, the project encompasses high-quality data ensured by the PETER registry and by the data from studies conducted. As new data is collected and generated during the project's clinical study and the data analytics performed, strict data quality procedures and appropriate data quality labelling and scores will be defined. Data quality will be assessed both for metadata and for data categories (clinical, genomic, methylomic), while data quality indicators will also be included in the metrics of performance of the predictive models, e.g. to estimate biases and confidence intervals.

While T2.3 will ensure data quality through the common data model, T5.3 will implement the agreed data quality labelling, enriching data with annotations, using semantic technologies and ontologies, by establishing a quality layer (service meta-data annotation layer) that provides relevant information about the pre-processing and cleaning process performed on the data.

7 IP Rights and Licensing

Intellectual Property Rights (IPR) is a generic term that encompasses several different issues that are covered by different laws and practices. Generally, all issues related to copyright, patents, trademarks, trade secrets and sui generis database rights are collectively indicated as IPR.

IPR management is of fundamental importance in PROTECT-CHILD, because the main software artefacts that the project will release, will be distributed with the intent that Parties, either internal or external to PROTECT-CHILD, can use it freely. In order to grant Parties these rights, it should be carefully managed about the IPR distribution terms.

In the following subsections, we will discuss three basic issues about the knowledge the project are producing:

- Access rights: who will own the basic rights?
- Licences: under which conditions is the project going to exchange it?
- Use and dissemination: how will the project exploit it?

7.1 Intellectual Property Rights

7.1.1 Copyright

Copyright is a corpus of laws that are harmonised in most nations in the world thanks to the Berne copyright convention. Copyright laws establish the rights that the authors have over their work. Copyright applies to most original and non-trivial works, be it writings, painting, music, most works of art and even software, both source and machine-readable code.

Copyright concerns the rights of copying, displaying, performing, printing, publishing, extending, modifying, translating a work. Application of copyright to software involves the rights to copy, modify or distribute the program. It does not involve the right to independently write a program performing the same actions as an original one. Generally speaking, the programmer who writes the program owns the rights. Where there is more than one programmer, the Directive (Directive 2009/24/EC) provides for co-ownership.

Software licences

A software licence is a legal document that accompanies a program. Without a software licence, according to the provisions laid down within the Berne copyright convention, a program cannot be distributed or modified without the explicit permission of the authors[2].

There are many kinds of software licences. Broadly speaking, we will divide software licences in two distinct classes: FLOSS licences and non-FLOSS licences. Licences belonging to the latter class are also termed proprietary licences. PROTECT-CHILD is mostly interested into FLOSS licences because, as stated in the Description of Work, PROTECT-CHILD execution platforms, tool components and service components shall be released as FLOSS software during the lifetime of the project. However, occasionally some programs or components may be used or written which are not released with a FLOSS licence.

FLOSS licences. FLOSS is an acronym originated in 2001. Its diffusion is mainly due to the Free/Libre and Open Source Software: Survey and Study commissioned by the EC in that same year. The term was coined to encompass all the different terms used to indicate the same class of software copyright licences. Among the FLOSS licences, two broad classes can be identified:

[2] This being true, standards beyond the Berne Convention, which are applicable in the EU, should be taken into account. For example, the aforementioned Directive 2009/24/EC, Directive 2019/790/EU, or the WIPO Copyright Treaty. All three aforementioned standards contain mandatory references for EU states on software and copyright.

copyleft and non-copyleft licences. In short, a copyleft licence allows the covered program to be redistributed only under the same licence: the licence can be said to be persistent. FLOSS licences that do not have this requirement are non-copyleft licences.

Copyleft

Copyleft software licences are a class of FLOSS licences that, like all FLOSS licences, permit unrestricted usage, copying and modification of the covered program. Like all FLOSS licences, they also permit modified or unmodified redistribution, but only using the same licence.

In practice, this means that when you obtain a program distributed with a copyleft licence, you cannot change the licence while redistributing it, whether or not you have made changes to it. Consequently, any copyleft program you distribute is granted to maintain its FLOSS state, whether it is modified or not. In other words, a copyleft licence persists through modifications and redistributions of a covered program.

The first copyleft licence, the GNU General Public License (commonly known as the GPL), is the most famous and by far the most used FLOSS licence, recently revived in its version 3. The Free Software Foundation is responsible for the GPL maintenance.

Non-copyleft licences are those licences that do not require the covered program to be redistributed under the same terms. Take the Apache licence as an example: you can lawfully get a program covered by the Apache licence and redistribute it with any other licence, even a proprietary (non-FLOSS) one, with or without modifications. Often the term permissive is used to indicate non-copyleft licences.

7.1.2 Patents

A patent is a set of exclusive rights granted by a national or international body to an inventor for a limited period of time in exchange for a public disclosure of an invention.

The procedure for granting patents, the requirements placed on the patentee, and the extent of the exclusive rights vary widely between countries according to national laws and international agreements. A patent application must include one or more claims defining the invention which must be new, non-obvious, and useful or industrially applicable. The exclusive right granted to a patentee in most countries is the right to prevent others from making, using, selling, or distributing the patented invention without permission.

Software patents are an important issue, because they can pose a real danger to FLOSS software. When a software method or algorithm is covered by a patent, the patent office has recognised the inventor's claim that the software method is original (never invented before), and non-trivial (a knowledgeable person in the field would not be able to reproduce it from state of the art). The inventors have a 20-years monopoly on the exploitation of the software method, and no one can lawfully use it without their permission.

In Europe the law disallows patents on software per se. The legal status of the European software patents is unclear, though. Application of these recommendations depending a big extent on national laws, which are not harmonized.

FLOSS communities are particularly sensible to this risk, and in fact they avoid as much as possible to use software on which patent claims are known or suspected to exist. Modern FLOSS software licences often contain provisions against the most blatant abuses of software patents. The Apache licence, for example, contains some clauses that protect the software against the use of submarine patents: if a contributor's software is covered by a patent, and that contributor makes legal attacks against users of the software, that contributor loses all rights to using the

software. The Mozilla, Eclipse and GPL licences all have some sort of protection against software patents.

7.1.3 Trademarks

A trademark is a distinctive sign, usually a word or a logo. Its usefulness is to give a brand to something and avoid that someone else takes credit for the product using the trade mark or distributes a different version of it with the same name.

7.1.4 Trade Secrets

The most generic way of protecting IPR is to just not let slip the knowledge outside of the boundaries of your organization. For its very nature, this practice is utterly incompatible with FLOSS, which is based on openness. Keeping the development secret is a risky choice, because it can easily give the impression to outsiders that the openness of the developers is just a facade, rather than a real overall policy.

In PROTECT-CHILD, trade secrets would be kept to a minimum, and development should be organised around publicly-available repositories as early as practically feasible.

7.2 PROTECT-CHILD and IPR

As of now, IPR practices have only been defined by some of the project partners. Generally speaking the project partners have noted their intentions through the Consortium Agreement, and as such it is expected that the initial results consist of reports and thematic analyses, which will be shared under open-access policies like CC BY 4.0, ensuring public availability for non-commercial purposes. Individual partners, however, may generate intellectual property, particularly software or methodologies, under specific licensing frameworks. This will be noted in further detail in upcoming versions of the DMP.

Table 3 Partner-Specific IPR Contributions and Practices (as currently defined)

Partner	Ownership & Rights	Background Licenses	Foreground Licenses	License Compatibility	Access & Dissemination
UTW	Shared consortium ownership for data/documentation; no new IP generated.	No pre-existing proprietary solutions; pre-approved tools used without restrictions.	Open-access under CC BY 4.0.	No conflicts expected; follows open-access standards.	Publicly accessible via repositories like OSF, adhering to GDPR and ethical guidelines.
UPM	Shared IPR under consortium agreements; GNU GPL ensures open-source.	Pre-existing code governed by GNU GPL; must maintain GPL compatibility.	GNU GPL license; all derivative works must also adhere to GPL.	Strict compatibility rules; conflicts resolved by re-licensing or replacing	Foreground code shared via repositories like GitHub/GitLab under GNU GPL.

Partner	Ownership & Rights	Background Licenses	Foreground Licenses	License Compatibility	Access & Dissemination
				incompatible components.	
MME	IPR likely in the form of copyrights from contributions across various work packages (WPs).	No pre-existing IPRs expected.	To be decided during project exploitation phase.	No conflicts anticipated.	Policies for public domain dedication vs confidentiality to be defined later.
UDEUSTO	Development will be open source; license yet to be decided.	Not specified.	Not defined.	Not specified.	Not specified.
UNIROMA1	TBD.	Not specified.	Not specified.	Not specified.	Not specified.
CERTH	Software components protected under copyright; details pending in future deliverables.	Access to necessary background will be under fair and reasonable conditions; exclusions apply.	Copyright protection planned for foreground software.	Compatibility details pending.	Access and dissemination details pending.

This table will be further detailed in future versions of this deliverable.

8 Allocation of resources

This section is intended to describe any additional costs incurred in order to make data available and/or maintain the PORTECT-CHILD infrastructure both within and beyond the project's duration.

It also reflects on the current allocation of data management responsibilities within the project.

8.1 Estimation of Costs

The project's partners may require additional resources in order to make the PROTECT-CHILD FAIR. Given the early stages of the project, no such resources are currently envisioned or anticipated by partners, but will be assessed as the project evolves.

Nonetheless, the Data Management Plan is a living document and will be updated as required as the project progresses in order to best reflect the project's needs, particularly with regards to the repository and the PROTECT-CHILD platform and infrastructure.

9 Conclusion

This deliverable serves as the first iteration of the Protect Child Data Management Plan, and seeks to establish the baseline elements for data management in alignment with FAIR principles and legal compliance requirements, emphasizing transparency, ethical data management and secure data handling.

Previous sections present the necessary support elements to ensure integration of project activities with ethical requirements considering the project's aims to include assessment of personal data of minors of age and genomic health information. Legal compliance with applicable requirements will be further addressed by all partners, particularly those included in WP1, 2 and 11, and will build upon the project's envisioned federated data models and research infrastructure to ensure risk minimization and privacy by design and by default compliance.

The project has setup a pathway for open-access and FLOSS management and compliance, the adoption of which by all partners will seek to support immediate project goals and long-term innovation in healthcare and data management in Europe. Furthermore, a baseline strategy and activities to ensure data sustainability, metadata standardization, data quality labelling and mechanisms for data utility beyond the project's lifespan are ongoing. The project has defined a clear strategy for responsibilities and cost allocation, and will seek to maximize a holistic approach to data management and governance in its upcoming months to match the expected activities and the goals of all partners during the project and beyond.

The next iteration of this deliverable will further expand on its contents, to provide an overview of the operationalization-oriented data management practices finally adopted by the project and a summary of its main achievements.

Appendix A Principles and Basic Notions of Data Management

This annex presents the principles and basic notions associated with data management and compliance activities, it is a non-exhaustive list of terms that will guide discussions amongst the project stakeholders, and as such it is a normative reference for future discussions which will be further tailored and extended as necessary to align it with project and partner needs.

A.1 Definitions

- **Access Control:** Mechanisms or policies restricting data access to authorized users based on roles, responsibilities, or other criteria.
- **Accountability of the controller:** In generally, data controllers are more so in the position to ensure and demonstrate compliance with data protection principles (article 5.2 GRPR). As such, the principle of accountability requires that controllers put in place internal mechanisms and control systems that ensure compliance and provide evidence of this in order to demonstrate compliance with external stakeholders, including supervisory authorities.
- **Anonymization:** The process by which personal data is altered to prevent the identification of a specific individual, either directly or indirectly. Anonymized data is no longer subject to GDPR as it does not constitute personal data.
- **Audit Trail:** A chronological record of data access, changes, and transfers, designed to provide evidence of compliance and traceability.
- **Automated individual decision:** Article 22 of the GDPR gives individuals the right to object to decisions based solely on automated means, unless certain conditions are met or appropriate safeguards are in place. Automated individual decisions are decisions that significantly affect a person based on automated data collection and processing.
- **Biometric data:** As defined by article 4(14) of the GDPR, biometric data is “personal information resulting from specific technical processing relating to a person's physical, physiological, or behavioral characteristics, which can be used to identify or confirm the natural person, for example facial images or dactyloscopic data.”.
- **Bias Mitigation:** Processes and strategies to identify and address bias in data collection, processing, and analysis.
- **Confidentiality:** Confidentiality in general refers to the obligation not to disclose information to persons who are not authorised to receive it. This refers to the confidentiality of communications provided for in Article 5 of the E-Privacy Directive 2009/136/EC.
- **Consent:** In the context of data protection, consent is defined as a voluntary, specific, and informed expression of a data subject's preferences, through which they agree to the processing of their personal data (refer to Article 4, paragraph 11 of Regulation (EU) 2016/679). Consent plays a crucial role in data protection laws, serving as one of the conditions that can validate the processing of personal data. When consent is the basis for processing, the data subject must have clearly provided their [written or verbal] consent for a particular processing activity, and they must have been adequately informed about it. The consent obtained is strictly limited to the specific processing

activity for which it was granted and can generally be revoked without affecting prior processing.

- **Controller:** Article 4(7) of the GDPR explains the concept of data controller as the following: “controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided by Union or Member State law”.
- **Data Breach Notification:** The obligation of data controllers to inform the relevant data protection authority (and, in some cases, affected individuals) within 72 hours of discovering a breach involving personal data.
- **Data Backup and Recovery:** Regular copying of data to secure locations and plans for restoring data in case of loss, corruption, or disaster.
- **Data Catalog:** A centralized repository describing available datasets, including metadata, ownership, access rights, and usage conditions.
- **Data concerning health:** Health data is defined by article 4(15) of the GDPR as the following: “data concerning health means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status”.
- **Data Ethics Board**: A governance body responsible for reviewing and approving data-related activities, ensuring alignment with ethical and legal standards.
- **Data Encryption Standards**: Specific protocols or algorithms used to encode data to protect its confidentiality and integrity during storage and transfer.
- **Data Integrity:** Ensuring the accuracy and consistency of data throughout its lifecycle, including protection against unauthorized alteration or corruption.
- **Data Quality:** Characteristics of data that determine its suitability for purpose, including accuracy, completeness, reliability, and timeliness.
- **Data minimization:** The principle of "data minimization" stipulates that a data controller must restrict the gathering of personal information to what is essential and pertinent for achieving a particular objective. Furthermore, they are required to retain the data solely for the duration necessary to fulfill that objective. In essence, data controllers should only collect the personal data that is truly required and maintain it only for the time it is needed. This principle is articulated in Article 5(1)(c) of the General Data Protection Regulation (GDPR).
- **Data mining:** Data mining refers to the analytical process of examining data from various viewpoints and distilling it into valuable insights. Data mining software serves as one of several tools designed for data interrogation. This software enables users to explore data from multiple dimensions, categorize it, and summarize the identified relationships. From a technical standpoint, data mining involves uncovering correlations or patterns across numerous fields within extensive relational databases. It is widely applied in diverse profiling activities, including marketing, surveillance, fraud detection, and scientific research. Clearly, for data mining to yield effective results, it is essential to analyze substantial volumes of previously gathered data.

- **Data Portability:** The right of individuals to receive their personal data in a structured, commonly used, and machine-readable format and to transmit it to another controller without hindrance.
- **Data protection authority:** A Data Protection Authority (DPA) is an autonomous entity responsible for the following functions:
    - overseeing the handling of personal data within its jurisdiction, which may encompass a country, region, or international organization;
    - offering guidance to relevant authorities concerning legislative and administrative actions related to personal data processing;
    - addressing grievances submitted by individuals regarding the safeguarding of their data protection rights.

  In accordance with Article 51 of the GDPR, every Member State is required to establish at least one data protection authority within its borders, which must possess investigative powers (including access to data and information gathering), corrective powers (such as the authority to mandate data deletion, impose fines, or prohibit processing), and authorization or advisory powers (including the ability to issue opinions and accredit certification bodies).

  National data protection authorities have been instituted across all European nations, as well as in numerous other countries globally.
- **Data Protection Impact Assessment (DPIA):** The data controller is required to conduct an evaluation of the potential effects of the proposed processing activities on the safeguarding of personal data when such processing is expected to pose a significant risk to the rights and freedoms of individuals. This evaluation must be performed before the commencement of the processing and, especially when employing new technologies, should take into account the nature, scope, context, and objectives of the processing.
- **Data of Minors:** The concept of "data of minors" refers to personal data belonging to individuals under the age of 18, with specific provisions for children under 16 in certain contexts. Under the GDPR, data controllers must implement additional safeguards when processing minors' data, recognizing their heightened vulnerability and need for protection. Consent for processing such data must typically be obtained from a parent or guardian unless processing is necessary for specific, lawful purposes. This concept is primarily governed by Recital 38 and Article 8 of the General Data Protection Regulation (GDPR).
- **Data protection officer (DPO):** The Data Protection Officer (DPO) possesses specialized knowledge of data protection legislation and practices and must function autonomously within the organization. It is the DPO's responsibility to guarantee the internal implementation of the Regulation and to ensure that the rights and freedoms of data subjects are not at risk of being negatively impacted by data processing activities.
- **Data Processing Notice:** (or Fair processing notice) A statement or policy provided to data subjects, outlining how their data will be collected, used, shared, and stored, as required under GDPR's transparency obligations.
- **Data retention:** Data retention encompasses the responsibilities of data controllers to maintain personal information for specified purposes. The Data Retention Directive (Directive 2006/24/EC) mandates that providers of electronic communication services retain traffic and location data related to communications, such as telephone calls and

emails. This retention is intended to facilitate the investigation, detection, and prosecution of serious criminal activities.

- **Data Sovereignty:** The concept that data is subject to the laws and governance structures of the nation where it is collected or processed.
- **Data Stewardship:** The assignment of specific roles and responsibilities to ensure the ethical and effective management of data throughout its lifecycle.
- **Data subject:** The individual whose personal data is collected, stored, and processed is referred to as the data subject.
- **Data transfer:** Transfers are governed by particular protections when the recipient resides in a nation outside the EU/European Economic Area (EEA), as outlined in Chapter V of the GDPR.
- **Data Versioning:** Tracking and managing changes to data over time by creating distinct versions.
- **Encryption:** The process of encoding data to ensure that it remains secure and inaccessible to unauthorized parties, typically through cryptographic methods.
- **Legitimate Interests:** A lawful basis for processing personal data under GDPR, where the processing is necessary for the legitimate interests of the controller or a third party, provided it does not override the rights and freedoms of the data subject (article 6.1.f GDPR).
- **Metadata**: Information that describes data, including its source, structure, format, and usage restrictions.
- **Open Data:** Data that is freely available for use, reuse, and distribution, typically accompanied by licensing terms.
- **Personal data:** Article 4(1) of the GDPR defines "personal data" as any information that pertains to an identified or identifiable natural person, referred to as the "data subject." An identifiable natural person is one who can be recognized, either directly or indirectly, particularly through identifiers such as a name, identification number, location data, online identifier, or various factors that are specific to their physical, physiological, genetic, mental, economic, cultural, or social identity.

  Examples of personal data include a person's name and social security number, which are directly linked to an individual. However, the definition is broader and also includes items such as email addresses and an employee's office phone number. Additional instances of personal data can be found in details regarding physical disabilities, medical records, and performance evaluations of employees. Advances in data analytics techniques and the possibilities of re-identification or inference are progressively increasing the scope of what can be considered an identifiable person.

  Personal data processed in connection with the data subject's work remains personal and is protected under applicable data protection laws, which aim to safeguard the privacy and integrity of individuals. Consequently, data protection legislation does not apply to legal entities, except in rare cases where information about a legal entity is also associated with a natural person.
- **Personal data breach:** Article 4(12) of the GDPR defines a personal data breach as a security incident that results in the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data that has been transmitted, stored, or otherwise processed.

- **Privacy:** Privacy refers to an individual's capacity to remain undisturbed, shielded from public scrutiny, and to maintain authority over personal information. It is possible to differentiate between the prevention of encroachment upon one's physical environment, known as "physical privacy" (such as safeguarding one's home), and the management of the gathering and dissemination of personal information, termed "informational privacy." Consequently, the notion of privacy intersects with, yet is distinct from, the notion of data protection. The right to privacy is recognized in the Universal Declaration of Human Rights (Article 12) and the European Convention on Human Rights (Article 8).
- **Privacy by design:** The concept of privacy by design focuses on integrating privacy and data protection measures from the outset within the design specifications and architecture of information and communication systems and technologies (article 25 GDPR). This approach is intended to ensure adherence to privacy and data protection principles.
- **Profiling:** According to Article 4(2) of the General Data Protection Regulation (GDPR), the term “profiling” means “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements”. .
- **Processing (of personal data):** According to Article 4(2) of the General Data Protection Regulation (GDPR), the term "processing of personal data" encompasses any operation or series of operations conducted on personal data or collections of personal data, regardless of whether these actions are performed through automated means. This includes activities such as collecting, recording, organizing, structuring, storing, modifying, retrieving, consulting, utilizing, disclosing through transmission, disseminating, or otherwise making available, aligning or combining, restricting, erasing, or destroying the data. Personal data can be processed in various activities related to the professional life of an individual, as well as in the context of health care treatments.
- **Processor:** As stated in Article 4(8) of the General Data Protection Regulation (GDPR), a processor refers to an individual or entity, including a public authority, agency, or any other organization, that handles personal data on behalf of the data controller.
- **Processor agreement:** The transfer of personal data from a data controller to a data processor requires the establishment of a data processor agreement. This agreement must adhere to specific minimum standards outlined in Article 28 of the GDPR. It is essential that the contract specifies that the data processor will operate solely based on the instructions provided by the data controller. Furthermore, the data processor is obligated to offer adequate assurances regarding the technical and organizational security measures related to the processing activities and must ensure adherence to these measures.
- **Pseudonymisation:** As stated in Article 4(5) of the GDPR, pseudonymisation refers to the processing of personal data in a way that prevents the data from being linked to a specific individual without the use of supplementary information. This additional information must be stored separately and must be protected by technical and organizational measures to ensure that the personal data cannot be associated with an identified or identifiable individual.
- **Retention periods:** Data retention encompasses the responsibilities of data controllers to maintain personal data for specific purposes and for no longer than needed. Limiting the duration for which personal data is held is also linked to data minimization. The

general guideline is to retain data "no longer than necessary," although certain legal requirements may dictate specific retention periods. By ensuring that data is not retained beyond its necessary duration, the risk of it being misused or accessed by unauthorized individuals is mitigated. Therefore, establishing and adhering to defined retention periods is crucial for safeguarding the individuals whose data is being processed.

- **Risk Assessment in Data Management:** Identifying and mitigating potential risks associated with data handling, including breaches, misuse, and obsolescence.
- **Sustainability of Data:** Long-term maintenance of datasets, ensuring accessibility, usability, and preservation after the project lifecycle.
- **Sensitive Processing:** Processing activities involving special categories of personal data that require additional safeguards and justification.
- **Special categories of personal data:** Special categories of personal data encompass information that discloses: "racial or ethnic background, political beliefs, religious or philosophical convictions, or membership in trade unions, as well as genetic data, biometric data intended for the unique identification of an individual, health-related information, and details regarding an individual's sexual life or sexual orientation" (Article 9 of the GDPR).

  The processing of such sensitive information is generally prohibited, with certain exceptions. For instance, it may be permissible to process this type of data if it is essential for medical diagnosis, if specific safeguards are in place within employment law, or if the data subject has provided explicit consent.
- **Transparency:** The obligation of data controllers to ensure that data subjects are fully informed about the processing of their personal data, including purposes, rights, and safeguards.
- **Third Countries:** Nations outside the European Union or European Economic Area, where specific conditions or safeguards (e.g., adequacy decisions or Standard Contractual Clauses) are required for data transfer.