IMI2 101034344 – EPND
European Platform for Neurodegenerative Diseases
WP2 – Legal and Ethical Regulations

D2.2 – White Paper V1: Gaps and Requirements

Lead contributor (partner organisation)	**UNILU**
Other Contributors	UCB
	SARD
	EATRIS
	BBMRI-ERIC
	GV
	AE
	ACI

WP2	
Due date	31/10/2022
Delivery date	31/10/2023
Delivery type[1]	Report
Dissemination level[2]	Public

Description of Action	
V2.0	27/10/2022

Document History

Version	**Date**	**Status**
V1.0	October 2022	Draft for WP2 review
V2.0	15/01/2023	Submitted to PMT
V3.0	22/06/2023	Resubmitted to PMT
V4.0	09/08/2023	Resubmitted to PMT
V5.0	30/10/2023	Final

The Current State of Biomedical Data and Sample Sharing for Research Purposes in Europe: Perspectives from the European Neurodegenerative Disease Community

The European Platform for Neurodegenerative Diseases (EPND)

Table of Contents

Glossary

A29WP	Article 29 Working Party
AD	Alzheimer's Disease
ADDI	Alzheimer's Disease Data Initiative
AD Workbench	Alzheimer's Disease Workbench
ADNI	Alzheimer's Disease Neuroimaging Initiative
CSF	Cerebrospinal fluid
CTR	Clinical Trials Regulation of the European Union
DAC	Data Access Committee
DLB	Dementia with Lewy bodies
DPAU	Dementias Platforms Australia
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DPUK	Dementias Platforms UK
DSAC	Data and Samples Access Committee
DZNE	German Centre for Neurodegenerative Diseases
EATRIS	European Infrastructure for Translational Medicine
EDPB	European Data Protection Board
EEA	European Economic Area
EFPIA	European Federation of Pharmaceutical Industries and Associations
EGA	European Genome-Phenome Archive
EHDS	European Health Data Space
ELSI	Ethical, Legal, Social Issues
EMA	European Medicines Agency
EPND	European Platform for Neurodegenerative Diseases
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FDSA	Federated Data Sharing Appliance
GA4GH	Global Alliance for Global Health
GAAIN	Global Alzheimer’s Association Interactive Network
GDPR	General Data Protection Regulation
IC	Informed Consent
ICF	Informed Consent Form
ICO	Information Commissioner's Office
IMI	Innovative Medicines Initiative
IRB	Institutional Review Board
ITF	Innovation Task Force
MDTA	Material and Data Transfer Agreement
MTA/DTA	Material Transfer Agreement / Data Transfer Agreement
ND	Neurodegenerative Disease
PD	Parkinson's Disease
PHT	Personal Health Train

PI	Principal Investigator
PPP	Public-Private Partnership
QA	Quality Assurance
QC	Quality Control
REC	Research Ethics Committee
SME	Small and Medium-sized Enterprise
SOPs	Standardised Operating Procedures
SPE	Secure Processing Environment
TPP	Target Product Profile
UK	United Kingdom
US	United States
WP	Work Package

Lists of Tables, Figures, and Text Boxes

Executive Summary

Neurodegenerative diseases (NDs), such as Alzheimer's Disease and Parkinson's Disease, are complex, multifactorial conditions whose progression is determined by an incompletely understood interplay of various genetic, biological, environmental, and lifestyle factors. A critical unmet need in precision medicine research into NDs is scaling up sample and data access for research projects across networks of cohorts and biobanks. These networks are essential to increase the scale and statistical power of research analyses, as well as to identify rare associations relevant to biomarker discovery.

The European Platform for Neurodegenerative Diseases (EPND) is a public-private partnership initiative whose goal is to build a scalable and sustainable platform that will facilitate discovery, access, and analysis of a wealth of high-quality clinical and biological samples and associated data held across European ND cohorts. Funded under the European Innovative Medicines Initiative (IMI) framework, the EPND aims to establish an integrated European ecosystem whereby existing ND sample and data collections can be shared and used for research purposes in a streamlined manner, thus maximising the utility of these collections.

The present document is part of the White Paper series authored by members of the EPND consortium. Prepared during year one of the 5-year IMI project, this document has two main objectives. First, to elucidate the complex European legal, governance, and policy frameworks, with emphasis on the potential barriers these frameworks pose to sharing of human biosample and data collections for the purposes of biomedical research. Understanding the nature of these barriers is crucial for the EPND consortium in formulating its governance policies and legal compliance strategies that will allow the EPND to achieve its objectives in a compliant manner. Second, to provide a broad perspective on sustainability aspects of the EPND, which is intended to help inform refinement of the consortium's strategic roadmap, both during the IMI funding period and in the years beyond.

The structure of the document is as follows. Chapter 1 of the document provides an overview of the ND research landscape in Europe, with a focus on the existing sample and data collections that could be used to identify novel biomarkers or methods aimed at predicting, diagnosing and/or treating NDs. The chapter describes some of the key challenges to the discovery, access, and research use of these European collections, while also outlining potential technical, infrastructural, and governance capabilities that could enable an initiative such as the EPND to effectively address them.

Chapter 2 focuses on European data protection laws, primarily the General Data Protection Regulation (GDPR) and its national implementations by Member States. The chapter provides clarity around the key GDPR concepts bearing on European platforms and initiatives such as the EPND. More specifically, it elucidates the scope of the GDPR with a special focus on the legal concept of anonymity/anonymisation, discusses GDPR roles (controller/processor) in the context of complex multi-party data processing environments, and provides an in-depth analysis of the GDPR legal bases within the meaning of Art. 6(1) in conjunction with Art 9(2) GDPR. The chapter also expounds on the

GDPR compliance obligations for the relevant parties (i.e., cohorts, the platform, and data users), with the obligations vis-a-vis data subjects being explored in the greatest detail.

In Chapter 3, a broader view of the legal and governance challenges to biomedical sample and data sharing in Europe is outlined. To help conceptually frame the discussion, Chapter 3 refers to three pillars of the relevant European legal and policy frameworks governing: i) (retrospective) biomedical research using existing sample and data collections; ii) biobanking and research use of human biosamples/tissues; and iii) data protection and privacy. This chapter describes the complex interrelatedness of these pillars at the local (institutional), national, and international levels, which gives rise to a variety of legal and governance interoperability barriers, complicating cross-border retrospective biomedical research in Europe. Subsequently, the chapter uses these insights to outline key features of a desirable governance framework a European initiative such as the EPND could implement to mitigate some of the legal and governance barriers unique to the European context.

Chapter 4 focuses on the sustainability of the EPND beyond the IMI funding period of five years. The discussion in Chapter 4 is informed by two lines of background research activities carried out by members of the sustainability work package of the EPND consortium: i) a stakeholder study, consisting of written surveys and in-person interviews with potential users of EPND's services; and ii) a benchmarking exercise whereby platforms and initiatives sharing similarities with the EPND were reviewed and characterised. Collectively, these efforts have led to the identification of several unmet community needs, which translate into valuable insights that could help the EPND provide a differentiated offer to successfully establish itself as a leading European research ecosystem in the ND space.

Finally, Chapter 5 draws on the analysis and insights presented in the preceding chapters to formulate actionable strategic recommendations for the EPND consortium. The recommendations are intended to help the consortium calibrate its strategic priorities over the years ahead, serving as a high-level roadmap for steering the IMI project towards its successful completion, with a view to achieving long-term sustainability well beyond the IMI funding period of 2021-2026.

Chapter 1: Background

Neurodegenerative diseases (NDs), such as Parkinson's disease (PD), Alzheimer's disease (AD), and other causes of dementia, pose a significant public health challenge. The socio-economic impact of NDs is particularly pronounced in developed countries with ageing populations. For example, in Europe, AD and PD alone are expected to be responsible for a combined economic burden of €267 billion by 2030 (Maresova et al. 2016). As the population continues to age in most countries around the world, the global socio-economic impact of NDs is looming large.

Some of the more common NDs, including AD and PD, are complex, multifactorial conditions. Their progression is thought to be determined by an incompletely understood interplay of various genetic, biological, environmental, and lifestyle factors (Gentile, La Cognata, and Cavallaro 2021). This complex aetiology of common NDs translates into variable clinical presentations in affected patients, including differences in the age of onset, rate of disease progression, as well as the nature and severity of clinical symptoms (Maresova et al. 2016).

Due to the multifactorial nature and heterogeneous clinical presentations of NDs, there are currently few possibilities to accurately predict the course of a disease, or to evaluate the effectiveness of a newly initiated medical therapy in a given patient. Consequently, the lack of reliable biomarkers for accurate prediction and monitoring of disease progression has been identified as one of the key challenges hindering the design of new therapies for NDs (Mammadova et al. 2020). However, the complexity of NDs means that the development of reliable biomarkers requires a big data-driven approach, enabled by the analysis of rich, high-quality healthcare data from a large number of patients (Gentile, La Cognata, and Cavallaro 2021; Myszczynska et al. 2020).

Encouragingly, over the past decades, numerous medical centres and research organisations have established valuable collections of biological samples and the related datasets derived from patients with NDs (Reijs et al. 2015; Giannella et al. 2021). In Europe, a recent review of ND resources found that as of 2021, there were at least 60 sample and/or data collections held by medical and research institutions across 17 European countries (Table 1). These collections represent an invaluable resource for researchers working on elucidating the pathophysiological mechanisms of ND development and progression. By aggregating rich biospecimens and health data from different patient populations, researchers can significantly enhance the statistical power of their analyses and gain novel insights into NDs. Furthermore, large aggregate datasets from multiple cohorts may facilitate meaningful patient stratification into sub-populations, thus enabling more targeted study of specific patient groups with shared characteristics (Myszczynska et al. 2020).

Table 1 - Overview of European ND cohorts associated with EPND

Institute	Cohort name	Country	Setting	Focus	Cog. nor	AD spec	PD spec	DLB spec
UMC Innsbruck	Memory clinic	AT	C	PD	350	300	330	15
KU Leuven	LEU_DEM	BE	C	AD	495	237	0	20
KU Leuven	ARCK	BE	C	AD	111	190	0	0
KU Leuven	F-PACK	BE	R	AD	233	79	0	0
UMC Saint-Luc Brussels	UCL-2016/NCHM	BE	C	AD	294	228	15	17
UMC Brussel	Memory clinic	BE	C	Mix	100	200	150	50
University of Antwerp	Memory clinic	BE	C	Mix	350	2750	260	100
Lausanne University Hospital	CLEMENS	CH	C	AD	244	620	20	40
University of Geneva	SNAP/gMAD	CH	C	AD	61	163	1	20
University of Geneva	Geneva outpatient clinic	CH	C	AD	241	307	5	16
University of Geneva	Bus Sante	CH	P	Age	700	0	0	0
IMI PharmaCog	PharmaCog	CH	R	AD	0	81	0	0
UMC Motol	CBAS	CZ	C	AD	92	324	7	25
DZNE	4 Cohorts	DE	R	Mix	1375	320	544	5
DZNE-Tuebingen	Tuepac	DE	C	PD	94	0	1466	108
DZNE-Tuebingen	TREND	DE	C	PD	1200	25	220	0
UMC Berlin Charité	Memory Clinic	DE	C	AD	176	1137	0	0
UMC Munich	ActiGliA	DE	R	AD	28	92	0	0
University of Erlangen	DCN multi center study	DE	C	AD	243	1870	0	0
UMC Goettingen	Kassel	DE	C	PD	0	0	205	20
UMC Goettingen	KAMSA	DE	R	PD	0	0	250	0
UMC Goettingen	DeNoPa	DE	R	PD	80	0	150	0
Danish Dem Biobank	DDB	DK	C	AD	230	1095	30	205
FACE Barcelona	Memory clinic	ES	C	AD	3930	11380	108	503
Hospital St Pau Barcelona	SPIN	ES	C	AD	330	484	0	135
CITA-AD Foundation	GAP/DEBA	ES	R	AD	499	165	0	0
UMC Virgen Rocio Sevilla	Movement Clinic	ES	C	PD	1666	0	712	29
Hospital Salpetrière Paris	SOCRATES	FR	R	AD	0	158	0	20
Lille Memory Clinic	COMAJ/YOD/Baltazar	FR	C	AD	70	760	0	16
Hospital Lariboisiere Paris	BioCokBank	FR	C	AD	300	670	30	160
Hospital Salpetrière Paris	Insight pre-AD	FR	R	AD	230	103	0	1
Hospital Salpetrière Paris-ICM	NS-PARK	FR	C	PD	0	0	25000	0
Hospital Salpetrière Paris-ICM	ICM NGC, Iceberg, DIG-PD	FR	C	PD	162	0	5000	0
University of Toulouse	French MSA cohort	FR	C	PD/MSA	0	0	645	0
University of Athens (NKUA)	ALBION	GR	C	AD	90	65	0	0
University of Thessaloniki	Memory Clinic	GR	C	AD	340	1700	65	125
University of Athens (NKUA)	HELIAD	GR	P	Age	3100	67	154	9
IRCCS Brescia	ARWIBO	IT	C	AD	261	217	0	0
University of Genoa	Genoa-EADC	IT	C	AD	70	600	220	80
University of Milano	CAA cohort	IT	C	AD	0	600	0	0
University of Perugia	ReGAl 2.0	IT	C	AD	255	490	0	0
University of Pavia	PD cohort	IT	C	PD	1600	0	1800	30
University of Luxembourg	NCER	LU	C	PD	1632	0	681	17
University of Maastricht	BBACL	NL	C	AD	342	349	4	0

VU Medical Center Amsterdam	ADC	NL	C	AD	1200	2449	0	260
VU Medical Center Amsterdam	EMIF PreAD/90+	NL	R	AD	210	90	0	0
VU Medical Center Amsterdam	EPAD-VUmc	NL	R	AD	150	37	0	0
University of Maastricht	Maastricht study	NL	R	Age	9000	0	0	0
VU Medical Center Amsterdam	LASA	NL	P	Age	4000	0	0	0
VU Medical Center Amsterdam	PROGRESS-PD	NL	C	PD	52	0	149	0
Multicentre Netherlands	PROPARK	NL	C	PD	265	0	1250	0
Erasmus MC	Rotterdam Study	NL	P	Age	13000	800	0	0
RIVM	Doetichem cohort	NL	P	Age	4700	0	0	0
Lifelines	Lifelines	NL	R	Age	160k	0	0	0
KNAW	DutchBrainBank	NL	R	Mix	175	776	247	82
UMC Utrecht	SMART-MR	NL	R	Mix	1269	0	0	0
UMC Oslo	PROSPOS/DDI	NO	C	AD	418	173	76	0
UMC Coimbra	Memory clinic	PT	C	AD	300	1270	140	140
University of Lisbon	BBA	PT	C	AD	0	228	0	0
University of Aveiro	iBiMED	PT	C	Age	500	9	0	0
UMC Carol Davila	ANALS	RO	C	AD	20	45	25	17
University of Gothenburg	Gothenburg MCI	SE	R	AD	362	50	0	0
Karolinska Institute	GEDOC	SE	C	AD	4700	6000	50	50
Dokuz Eylul University	Turkish MCnetwork	TR	C	Mix	97	41	60	0
University of Oxford	Discovery	UK	C	PD	322	0	1241	0
Multicentre Europe	E-DLB	UK	C	DLB	0	0	0	900
University of Cambridge	CamPaIGN	UK	C	PD	160	0	179	0
University of Glasgow	PRoBaND	UK	C	PD	0	0	2266	0
Newcastle University	ICICLE-PD	UK	C	PD	99	0	257	0
University of Cambridge	PICNICS	UK	C	PD	118	0	276	1
Imperial College London	Chariot	UK	R	Age	450	0	0	0
University of Bristol	UK Brain Banks Network	UK	R	Mix	131	1700	1140	
		Total			63272	41494	45428	3216

Adapted from the EPND Description of Action. Accurate as at June 2021. Setting: C=clinical, R=research, P=population-based; Main focus cohort: AD=Alzheimer's disease, PD=Parkinson's disease, DLB= Lewy Body dementia; Mix: mixed; Age: ageing study

However, at present, existing ND sample and data collections are not always readily accessible for external researchers to effectively leverage. Sample and data sharing is often hampered by scientific, technical, sustainability, and ethical/legal issues. Samples may vary significantly, not only in their quality and quantity, but also in the methods and protocols used to collect them. Similarly, datasets associated with samples may lack standardisation, follow a format specifically tailored to an in-house research project, or contain errors and omissions. As a consequence, these collections are not always research-ready and may require considerable pre-processing before they can be meaningfully used by external researchers. In response, efforts have been underway to establish quality assurance standards and good research practices relating to biological samples from ND patients (Reijs et al. 2015). At the same time, the field of ND research has been progressing towards adopting FAIR principles: making the data Findable, Accessible, Interoperable, and Reusable for research purposes (Frey et al. 2019; Scott et al. 2020).

Some collections, particularly those obtained opportunistically in a clinical context and in the absence of a clear framework for their future research use, may lack the necessary governance to enable sharing

with external researchers. Even where sample and data collections are intended to be shared for research purposes, access to these resources by external researchers is often cumbersome and unpredictable, due to a lack of transparent and efficient access processes. Access to a single collection typically requires approval from an access committee and sometimes also a research ethics committee. For big data-driven projects, researchers will need to access multiple data and sample collections across different institutions. Gaining access to multiple collections can quickly become administratively burdensome and time-consuming (Giannella et al. 2021). Lack of incentives to share, alongside the costs and sustainability of the associated scientific, technical and governance processes, are also major barriers to collaboration.

Various ND initiatives aim to tackle different barriers to meaningful sharing and research use of ND data and sample collections, including: the Global Alzheimer's Association Interactive Network (GAAIN), Dementias Platforms UK (DPUK) and its Australian counterpart Dementias Platforms Australia (DPAU), and Alzheimer's Disease Data Initiative (ADDI). However, to date, no comparable platforms or initiatives have been established with a focus on cross-border data and sample sharing among ND researchers in the Europe.

The European Platform for Neurodegenerative Diseases

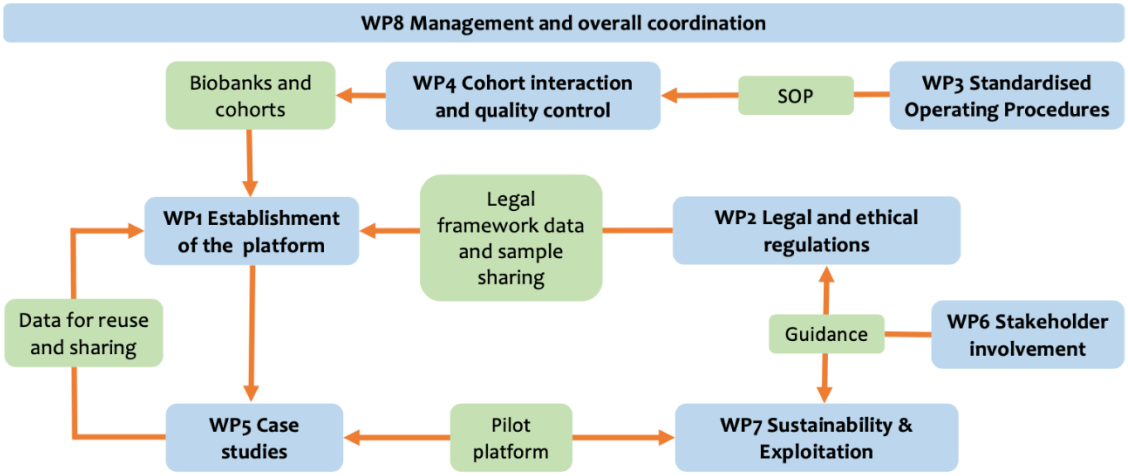
The European Platform for Neurodegenerative Diseases (EPND) is an international effort whose goal is to build "a scalable and self-sustainable platform that will facilitate discovery, access and analysis of a wealth of high-quality clinical and biological samples [and associated data] held across numerous European cohorts"[1]. The initial five-year period of the EPND, spanning from November 2021 to October 2026, is funded under the European Innovative Medicines Initiative (IMI) framework. In the context of the IMI-supported project, the EPND brings together a diverse consortium of public and private partners across the following work packages (WPs):

- WP1 - Establishment of the platform (Technical Hub)
- WP2 - Legal and ethical regulations
- WP3 - Standardised Operating Procedures (SOPs)
- WP4 – Cohort interaction
- WP5 – Case studies
- WP6 - Stakeholder involvement, external communication and dissemination
- WP7 – Sustainability
- WP8 - Management and overall coordination

A high-level overview of these WPs, including their interrelatedness, is captured in the figure below. A more detailed description of the EPND IMI consortium can be found on the EPND website (https://epnd.org/about)

[1] https://epnd.org/about

Figure 1 - Relationships between EPND work packages (WPs)



With its primary focus on medical and research institutions in Europe, the EPND aims to create a comprehensive infrastructure in three important ways. First, the EPND will include both biological samples and data across a wide range of sample and data types. Second, the EPND is being designed to play a pivotal role at every stage of data (and, potentially, sample) pre-processing, quality control, and standardisation, making the available collections research-ready. Third, the EPND will be supported by an online portal designed to simplify the process of searching for, accessing and analysing data resources held across multiple institutions. The necessary technical capabilities enabling the EPND user portal will be supported by AD Workbench, which will provide services for data management, hosting and analysis in the research environment workspaces (see details in Box 1 below). The online EPND user portal will act as a single point of contact for external researchers, reducing the administrative complexity of requesting data and sample access from multiple institutions, while also providing useful tools for data analysis across collections. A key motivation for the EPND consortium is to establish a trustworthy, secure IT platform in Europe with resource discovery, data access and data analysis capabilities (i.e., the EPND's Technical Hub) that demonstrates compliance with the European data protection law, while also enabling international connectivity and collaboration. The EPND consortium has deemed the AD Workbench the most suitable ND-centred data platform currently available for achieving this ambition. The consortium is currently deploying AD Workbench solutions through multiple EPND partner cohorts. The consortium's experiences with the AD Workbench will be reported in the final EPND White Paper, expected by October 2025.

Box 1 - Description of the AD Workbench[2]

ADDI's AD Workbench provides a global community of researchers with a cloud-based interoperability infrastructure to support the discovery, querying, and requesting of Alzheimer's and related dementias datasets. Private Workspaces are offered to users free of charge and provide a secure cloud-based environment where open-source tools, apps, and virtual machines are provided so that users can curate & harmonise data, perform analyses, work with direct collaborators, and save their work.

- Currently available data include observational trials, clinical trials, longitudinal cohort studies, synthetic cohort studies, and synthetic brain images.
- Users can search and preview curated and organized field-level metadata before requesting access to improve data discovery.
- Users can leverage off-the-shelf workflows (e.g., NGS or GWAS pipelines) or share algorithms in a community docker repository.
- The AD Workbench enables federated querying of data when centralised or distributed access to data is not an option. Additional information is available in the Knowledge Base "how-to" guides.

Data Access – Workflow, Authorisation and Security Measures

The AD Workbench enables datasets to be discoverable and available for analysis to permissioned users via interoperability: with existing data platforms; utilising distributed or federated connections to data sources (on a private cloud or on premise); or hosted centrally on the Azure cloud based in a data centre in the Netherlands. The AD Workbench supports end-to-end secure data access workflows and audit logging.

Data contributors are offered hosting services at no cost. For federated and distributed connections, ADDI provides software (federated data sharing appliance - FDSA) enabling data to be hosted locally by the data contributor. Metadata and record-level data are queried and analysed remotely, while the approved results of the analysis are transferred back to the user's workspace.

Through the dataset search functionality, users can explore datasets and field-level metadata using the FAIR Search for data discovery. Users will be able to analyse data in their personal workspace. The request access workflow within FAIR Search allows the user to access data in order to perform the intended data analysis.

The EPND aims to establish itself as an inclusive ecosystem, attracting sample and data collections across a broad spectrum of NDs, while facilitating their meaningful reuse for both commercial and academic research. Furthermore, the EPND seeks to accommodate collections of various sizes and sample/data types, irrespective of whether they were generated in the context of medical research or as part of routine clinical practice. Although the inclusive approach pursued by the EPND holds promise of creating a uniquely rich resource for ND research, the resultant heterogeneity of the prospective participating cohorts also raises important challenges.

Heterogeneity of collections, samples and data

As it can be seen in Table 1, existing cohorts vary significantly in disease areas, spanning rare NDs such as Dementia with Lewy bodies (DLB), common NDs such as AD, and other aging-related neurological conditions or general cognitive decline. Such a wide range of disease areas inevitably translates into clinically divergent patient cohorts, different types of biological samples, and disjointed datasets, many of which were originally established to support a particular research project. Even within the same disease areas, there may be heterogeneity or uncertainty over the quality of collections.

[2] Sources:
https://www.alzheimersdata.org/-/media/files/addi/addi_workbench_welcome.pdf
https://www.alzheimersdata.org/ad-workbench/how-to-use

Across different clinical and research contexts, different protocols have been used to perform sample and data collection. The EPND will need to address this heterogeneity in the context of its sample collection discovery and request workflows to simplify external researchers' access to the biosamples of interest. Furthermore, in order to make datasets from different cohorts amenable to analyses in an interoperable manner, the EPND will need to overcome significant data harmonisation and standardisation challenges.

Existing cohorts also vary significantly in their size, ranging from samples and data obtained from fewer than 100 patients or research participants to considerably larger collections including more than 10,000 individuals. Different sizes mean different incentives and support needs. Since smaller cohorts are less likely to be supported by dedicated personnel and infrastructure, assessing their eligibility for inclusion and carrying out the necessary quality control may pose considerable practical challenges. Additionally, clinical and research institutions housing smaller collections may have fewer incentives to participate in the EPND, as they may deem their collections less valuable for third-party research and decide against expending their efforts to collaborate with the EPND. The EPND will implement processes that minimise the workload for different contributing institutions, while at the same time developing appropriate incentive schemes to encourage participation of smaller cohorts.

Diverse ethical and legal access and use conditions

Sample and data collections may come with varying restrictions and use conditions. Sources of these restrictions could include preferences expressed by the data subject, as documented through informed consent forms. Such restrictions are more likely to be attached to established prospective research cohorts with well-elaborated informed consent forms and otherwise clearly defined use conditions.

As an illustrative example, consider the informed consent form (ICF) used to recruit research participants for a PD-centred study in Luxembourg, one of the ND studies committed to contribute its sample and data collections to the EPND. As shown in Box 2, the ICF presents prospective research participants with a number of options to define data/sample use conditions. For example, research participants may consent to the use of their samples for ND-focused research but opt out of future studies in other biomedical domains. Similarly, they can request that their data not be shared with entities based outside the EU, effectively excluding such entities from using their data for research purposes. Furthermore, research participants are asked to indicate whether they would like to undergo genetic analysis. Those participants who decide against genetic analysis are placing an additional restriction to the use of their samples, in this case based on the methodology employed by research projects.

Box 2 - Examples of options provided to research participants in a Luxembourg-based PD study

I give my consent:	
To donate my biosample material to the IBBL and to the LCSB to be used for biomedical research in the domain of neurodegenerative diseases.	☐ Yes ☐ No
To donate my biosample material to the IBBL and to the LCSB to be used for future biomedical research in other health related research domains.	☐ Yes ☐ No
For genetic analyses including Next Generation Sequencing techniques to be undertaken on my collected biosample materials.	☐ Yes ☐ No
To allow transfer of my collected pseudonymised data to countries outside the European Union that may not provide the same level of data protection as provided within the European Union, within the terms explained in the Subject Information Sheet.	☐ Yes ☐ No

Key: IBBL - Integrated Biobank of Luxembourg; LCSB - Luxembourg Centre for Systems Biomedicine

Other common restrictions on the use of samples/data may concern whether collections can be shared with for-profit entities and/or used for commercial research projects. For example, samples and personal data collected as part of the ENLIST-UK study, another prospective EPND cohort, cannot be shared with commercial organisations, effectively preventing private companies and other for-profit entities from accessing their samples and data. In the majority of cases, however, restrictions on commercial research use of samples and data will be less prescriptive, allowing for greater flexibility in their interpretation. For example, it may be acceptable to share collections with for-profit entities, as long as their use is limited to non-commercial research projects. Under these circumstances, it will typically be up to the cohort to specify how the restriction should be interpreted. (For a more extensive discussion of cohort-defined sample/data access and use conditions, see Chapter 3: Governance View, sub-section “Best practices for EPND Cohort sample/data access governance”)

To overcome these challenges, the EPND consortium is developing a comprehensive roadmap towards realising its ambitious mission. The overarching goal of the EPND is to establish a versatile and sustainable infrastructure enabling research use of existing data and samples from clinically divergent ND cohorts. To this end, the EPND will be equipped with a wide range of infrastructural capabilities designed to achieve the following objectives:

- Data standardisation across collections
- Determining data and sample use conditions
- Ensuring data and sample discovery, including enabling preliminary feasibility assessment of an intended research project
- Streamlining data and sample access requests
- Enabling (federated or centralised) data analysis through an online user portal and dedicated virtual workspaces (supported by AD Workbench)

Data standardisation

The purpose of data standardisation is to structure data in a more readily accessible format, using intuitive variable labelling and commonly accepted data representations. In practice, many research datasets are constructed based on a conceptual framework tailored to a particular in-house project. As a consequence, external researchers may find these dataset structures unintuitive or incompatible with other databases (Bauermeister et al. 2021). By curating the data to a defined standard, EPND can help make the data research-ready, effectively addressing issues with the usability and compatibility of data. Another important benefit of data standardisation is that it enables streamlining data structures and metadata catalogues, such as common variable names, data and sample use ontologies, and other metadata describing research-ready resources. This, in turn, can significantly improve the discoverability of data and samples by researchers who are interested in assessing the availability of samples/data for the purpose of their research project.

Standardisation is also closely linked to quality control. Technical assessment concerns the quality of biological samples, including the robustness of the protocols used to collect and store them. Similarly, the quality of the data will be assessed based on the frequency of errors, missing entries, and the dataset's overall conformity to the data standards in NDs. Quality control processes and metrics will be key to meeting regulatory standards.

Admittedly, the practical challenges associated with data standardisation can be significant. Past efforts aimed at harmonising non-interoperable datasets from ND cohorts have revealed that the process is time-consuming, entails laborious manual work, and may not always result in common data structures accurately fitting all of the disjointed datasets (Bauermeister et al. 2021; Birkenbihl et al. 2021). Therefore, one of the key strategic challenges facing the EPND consortium will be to design and implement an effective strategy for data standardisation in order to ensure interoperability of the datasets provided by different EPND cohorts.

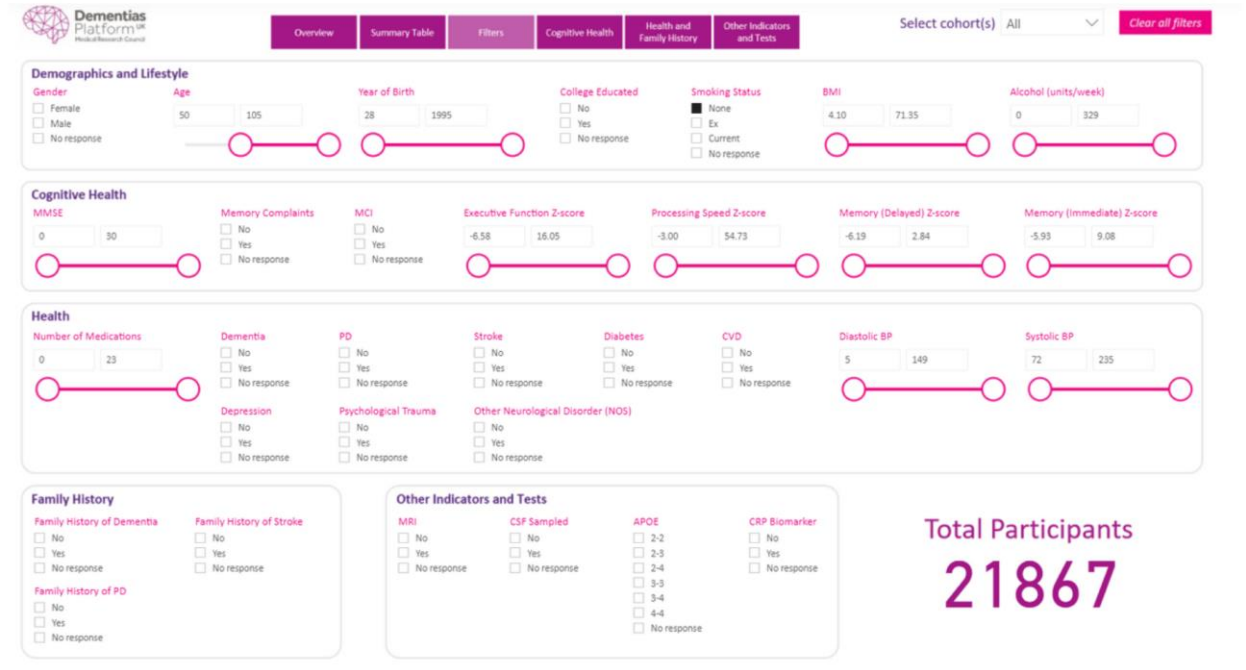
Enabling Discovery

Many research platforms incorporate tools enabling some form of data discoverability. In the simplest form, discoverability can be operationalised as a searchable metadata catalogue allowing the user to find cohorts of interest (i.e., cohort-level discoverability). However, superficial data discovery tools have been criticised for their limited utility in helping researchers understand whether data needed for carrying out a particular research project is available.

As a consequence, more advanced data discovery tools are currently being developed by ND research platforms. The goal of these tools is to enable feasibility assessment of the proposed research project based on the data/samples available at the participating institutions. An illustrative example of one such tool has been developed by the Dementias Platform UK (DPUK) and is available to registered users of DPUK's data platform (Figure 2). As it can be seen in Figure 2, the data discovery tool allows users to perform advanced search using filters based on various demographic, clinical, and other health-related variables. With the aid of this granular data discovery tool, external researchers querying the web portal can apply filters to gain an overview of the sub-set of data collections required for their research project. This, in turn, may enable researchers to obtain substantial insights into the feasibility

of their proposed research project, potentially obviating the need to inquire about data availability with data-providing institutions (Bauermeister et al. 2021).

Figure 2 - A filter-based data discovery tool from the Dementias Platform UK (DPUK).



Owing to the large number and heterogeneity of cohorts to be included in the EPND ecosystem, the EPND would greatly benefit from the implementation of an advanced sample and data discovery tool.

Streamlining data and sample access governance

Given the variety of data/sample use restrictions both across and within ND cohorts, it is essential to devise practical solutions for accurately and transparently attributing these restrictions to the corresponding individual samples and their associated data. Moreover, some of the largest ND-related sample and data collections in Europe have been obtained through routine clinical practice, which will necessitate their careful assessment for ethical and legal eligibility for research purposes.

The EPND can support generic ethico-legal assessments by cohorts to identify any special conditions or restrictions to making data and sample collections available for future research through the EPND. Some of the reasons preventing data and sample sharing could be: restrictions stemming from European data protection laws and regulations (discussed in Chapters 2 and 3); a documented expressed objection of the data subject to having their samples and associated data used in future research projects; or a national law significantly limiting export of human biological samples outside the country. Additionally, there will be cases falling under the grey areas, where an in-depth analysis will need to be performed to ascertain availability. This is expected to be the case with legacy samples and their associated data, collected opportunistically in the context of routine clinical practice or as part of a specific research project with a well-defined scope and timeframe. Among other issues seen with such collections is that they tend to lack adequate informed consent for future research use, making their ethical/legal eligibility dubious (Wallace, Kirby, and Knoppers 2020). One promising

approach is to develop and implement a dedicated data use ontology for ND research. Data use ontologies have been actively explored in several other areas of health research (Lawson et al. 2021) and may also prove valuable in the context of NDs.

One of the key aspects in which European ND cohorts differ from one another is their access governance models, which comprises institutional policies, rules and processes concerning data/sample access by external researchers. A review of the prospective EPND cohorts has revealed that the nature of the access governance model adopted by a particular institution can be influenced by various factors, including the characteristics of the ND cohort, the availability of physical and digital infrastructure, as well as whether the cohort is already part of a multicenter research consortium. In general, larger cohorts established as part of longitudinal observational studies tend to be supported by more developed access governance models, which typically include a dedicated data and/or sample access committee (D(S)AC), a standard application form for requesting access, and binding data/sample use agreements aimed at ensuring accountability after external researchers have been granted access. Examples of prospective EPND cohorts with dedicated access governance mechanisms include the Tracking Parkinson's (or PRoBaND) study in the UK, and several prospective observational studies based in the Netherlands, including the Longitudinal Aging Study Amsterdam (LASA), LifeLines, and the Maastricht Study[3]. By contrast, the majority of smaller ND cohorts, as well as sample and data collections not obtained in the context of prospective observational studies, are typically not supported by formal access governance frameworks. Furthermore, it is not always clear whom external researchers should approach with sample/data access requests, or whether access to the resources is even possible in the first place.

The lack of a consistent access governance framework poses significant administrative hurdles to biomedical researchers seeking access to sample/data collections held by multiple institutions. Researchers may be required to prepare and submit a separate access request to each institution, while encountering inconsistent requirements, variable decision-making timeframes, and incongruous decisions (Devriendt et al. 2021). It is therefore paramount that the EPND identifies ways to streamline the data and sample access process, thus reducing the administrative burden for external researchers. To eliminate inconsistencies across data/sample access decisions, and improve the efficiency of access, the EPND should systematically identify opportunities for streamlining (including, to the extent feasible, by centralising) various steps along the process. These steps include submission of an access request by an external researcher, review of the access request by the competent committee(s), communication of the access decision to the requesting researcher, providing access, and (in the case of centralised access modalities) supporting user authentication, authorisation and activity monitoring for compliance purposes.

[3] Information concerning data and sample access from these cohorts:
Tracking Parkinson's / PRoBaND: https://www.trackingparkinsons.org.uk/about-1/data/data-access/
LASA: https://lasa-vu.nl/en/request-data/
LifeLines: https://www.lifelines.nl/landingspaginas/lifelines-data-provides-many-opportunities-for-good-research
The Maastricht Study: https://www.demaastrichtstudie.nl/research/data-guidelines

Enabling data analysis through the online portal

There is a clear consensus within the biomedical research community that leveraging interoperable datasets for research purposes can significantly enhance the statistical power of data analysis, allowing for the elucidation of novel insights from the available data (Bergeron et al. 2018). However, until recently, directly linking datasets held by different institutions was hindered by numerous technical, privacy, and legal barriers. As a consequence, researchers have traditionally sought access to each dataset of interest separately, with few opportunities to integrate the data into a single composite dataset for the purpose of analysis.

In recent years, the emergence of research platforms coupled with the growing interest in data FAIRification has led to the proliferation of methods enabling integrated data analysis across datasets held by different institutions. In the space of NDs, several platforms have implemented solutions that allow users to perform analyses of varying complexity via the central research portal. One approach that has seen growing adoption by ND platforms is federated data analysis, whereby data remains stored at local institutions while being queried and analysed centrally. However, as discussed in subsequent chapters, fully federated data analysis approaches are challenging to implement and come with significant technical as well as governance limitations. Consequently, in practice, many ND data platforms continue to rely on centralised solutions, in which data collections provided by the participating institutions are stored (hosted) in a central data repository where the data can be accessed and analysed by the external researcher.

Centralised data hosting and analysis approaches offer significant operational efficiencies and, as such, are being actively explored by the EPND. However, owing to the highly diverse governance frameworks of the European ND data collections, as well as complex and granular ethico-legal constraints to their availability for research purposes, it is currently unclear to what extent the EPND can rely on centralised data hosting and analysis solutions. Striking an optimal balance between operational efficiency and ethico-legal considerations will be a key objective for the EPND consortium when refining its data hosting, access, and analysis approaches in the years ahead. These issues are explored in greater detail in Chapter 3: Governance View.

Overarching challenges: data protection laws, lack of incentives to share, and uncertain utility of ND collections

In addition to the issues stemming from the heterogeneity of European ND sample and data collections, there are three types of overarching challenges complicating access to, and research use of, biomedical samples and data in Europe. Notably, these challenges can be collectively described as uniquely European, in that they are either non-existent in other jurisdictions/regions or are particularly pronounced in Europe for a variety of legal, policy, and other contextual reasons.

First, the European data protection laws, primarily the General Data Protection Regulation (GDPR) and its national implementations, impose a significant compliance burden on parties intending to utilise personal data of European data subjects for research purposes. As will be discussed extensively in subsequent chapters (in particular, Chapter 2: Data Protection View), demonstrating compliance with the principles of the GDPR when re-using existing data collections for research purposes can be

procedurally burdensome, potentially discouraging institutions from engaging in data-sharing practices. Moreover, owing to the broad definition of "personal data" under the GDPR, these compliance requirements usually apply to their full extent even where the research project uses only a limited subset of the existing data.

Second, it is highly challenging to incentivise European biomedical research institutions to make their sample and - especially - data collections available for external researchers. Although lack of incentives to share valuable research resources is a well-documented global issue (Rowhani-Farid, Allen, and Barnett 2017; Devriendt, Shabani, and Borry 2021), European institutions are additionally deterred from data sharing by the complex and uncertain regulatory framework. Moreover, these regulatory issues also prevent European funding agencies from mandating that biomedical research institutions deposit the data resulting from grant-supported research activities into designated data repositories (Devriendt, Shabani, and Borry 2022), an approach increasingly successfully utilised outside Europe as an effective policy tool to incentivise data sharing.

Third, even where European ND collections can be made widely available for external researchers, ascertaining the utility of such data/samples is not straightforward. This is partly due to the diverging quality standards, sample collection protocols, and data formats described previously. However, another source of uncertainty stems from the rapidly evolving European regulatory framework around the use of existing sample and data collections to facilitate the discovery, qualification, and validation of novel biomarkers and methods. The European Medicines Agency (EMA) has made available certain support tools and stakeholder engagement venues whose detailed overview is provided in a supplementary document (see Annex G: Regulatory guidance – Support tools offered by the European Medicines Agency for Innovations). However, at present, there is limited regulatory guidance regarding to what extent existing data and sample collections obtained outside of clinical trials can be relied upon to generate the necessary evidence for bringing novel biomarkers and methods to the European market.

Aims of the Document

The present document focuses on the key ethical and legal challenges and barriers impacting on the main objectives of the EPND consortium. Developed during Year 1 of the 5-year (2021 – 2026) Innovative Medicines Initiative (IMI) project, it places particular emphasis on the fundamental legal, governance, and policy challenges associated with sharing and research reuse of ND sample and data collections in Europe. However, as these fundamental challenges are not limited to the EPND, much of the analysis presented in this paper has implications beyond the IMI project and is of relevance to the broader European biomedical research community.

The paper explores the principal challenges facing European biomedical research platforms from three broad perspectives: i) data protection (privacy); ii) governance; and iii) sustainability and exploitation. The three perspectives correspond, in that order, to the subsequent three chapters. The chapters summarise the key challenges under each perspective and highlight their relevance for the EPND. A more extensive analysis supporting this discussion, including a nuanced examination of the implications for the EPND, can be found in the corresponding annexes (A – G) of the document. Collectively, the considerations discussed in the three chapters, - and further elaborated in the annexes

– give rise to overarching conclusions that are subsequently translated into actionable strategic recommendations for the EPND consortium. The recommendations are intended to inform the development of the EPND's long-term roadmap, not only for the remainder of the IMI project (until November 2026), but for the years beyond.

Chapter 2: Data Protection View

The utility of EPND resources for scientific research will, to a large extent, be determined by the richness of individual-level data collected from ND cohorts. In this respect, the EPND will seek to continuously enhance its research-ready data repositories with high-quality multimodal data pertaining to the patients and/or research participants enrolled in EPND-associated cohorts. The types of data made available to external researchers through the EPND may include information extracted from patient health records, neuroimaging data, genomic and other multi-omic data, and clinical standardised test scores. Many of the cohorts will have integrated biobanks, allowing an additional sharing of biosamples for biomarker research, potentially including genetic and genomic research.

The critical importance of high-quality, multimodal data means that the EPND will enable the sharing and re-use of personal health data concerning individuals participating in EPND cohorts. As a consequence, parties participating in the EPND (e.g., cohorts, platforms, and research organisations) will be subject to various data protection laws and regulations.

This chapter is devoted to the key privacy- and data protection-related legal frameworks relevant to the EPND. The centrepiece of this analysis is the EU Regulation 2016/679, the European General Data Protection Regulation (GDPR), whose provisions have profound and multifaceted implications for the EPND, impacting on the design, governance, and day-to-day operations of the platform in numerous complex ways. Additionally, this section briefly touches upon other pertinent legal frameworks, including medical confidentiality laws, legal oversight of medical research, and the forthcoming European Health Data Space regulation.

GDPR: scope and application

The GDPR went into force in May 2018. The GDPR aims to protect the fundamental rights and freedoms of individuals in relation to processing their data, while at the same time fostering a harmonised legal framework for enabling the movement of data across the European Union (EU). General provisions of the GDPR (Articles 1-4) define the scope of the Regulation along three main axes: types of data covered by the GDPR, parties subject to the Regulation, and the territorial scope. Broadly speaking, the GDPR establishes a comprehensive legal framework pertaining to processing of *personal data* – that is, any information relating to *data subjects*, who are defined as "identified or identifiable natural persons" (Art. 4(1) GDPR). Where data subjects are on the territory of the European Union, the GDPR applies to processing of their personal data by any party, irrespective of whether the party is based in or outside the EU (Art 3(2) GDPR).

As a general rule, the data relating to patients and research participants undergoing processing through the EPND will constitute personal data, and hence its processing will be subject to the GDPR. This includes data processing by, or on the instruction of (e.g., under federated and/or distributed analysis without direct data access), an external researcher using the EPND data and sample collections.

The analysis contained in the following appendix deconstructs these general provisions concerning the scope of the GDPR, elucidating their relevance and implications for the EPND.

Annex A: GDPR Scope: Data Types and Territorial Considerations

Roles and responsibilities of the parties involved in data processing

The GDPR formally delineates the roles that can be assumed by parties involved in the processing of personal data. In the context of platform-enabled research into NDs, the following GDPR roles are of particular importance: **controllers, joint controllers,** and **processors.** It is critical to correctly allocate these roles to the parties processing personal data, as this has significant implications for the parties' obligations vis-a-vis one another, data subjects, and the regulators. The need for - alongside the challenges associated with - clearly delineating GDPR roles among parties is particularly highlighted in complex, multi-party data processing environments, such as the ecosystem to be fostered by the EPND.

Under most scenarios, the parties responsible for the infrastructural components of the EPND (including the EPND's Technical Hub supported by the AD Workbench) will be acting as processors in relation to the processing of the personal data contained in cohorts' data collections. The cohorts will typically be (sole) controllers for the processing operations required for making the data collections available through the EPND, whereas external researchers accessing and/or analysing these data collections will be acting as (sole) controllers when using the data for own research purposes. There is some uncertainty as to GDPR roles in relation to processing operations required for disclosing data to external researchers (or, under a federated scenario with no direct access, granting permission to perform federated analysis). In the short-term, while the EPND is in the process of building its compliance capabilities, and lacks the status of a legal entity, it is foreseen that cohorts making their data available via the EPND will remain sole controllers also in relation to subsequent data disclosures/permissions for each instance of data reuse by an external researcher. In the longer term, the allocation of the GDPR roles will depend on the precise nature of EPND's data access governance framework, to be developed by WP2 over the course of the IMI Project.

The analyses in the following annex seek to clarify the different GDPR roles parties can assume and discuss their relevance to the EPND.

Annex B: Definition and implications of GDPR Roles; For a more detailed discussion concerning access governance framework, see Chapter 3 (Governance View)

Lawfulness of Sharing data through the EPND

All parties within the EPND ecosystem must ensure that they process personal data in a GDPR-compliant manner. The most relevant parties in this context are the following:

- Data providers: cohorts submitting data to the EPND
- The legal entity/entities responsible for managing the infrastructural components of the platform (including the EPND's Technical Hub)
- Data recipients: external researchers accessing EPND resources for research purposes

Of these parties, data providers and data recipients act as controllers for various processing operations with respect to cohort data, and as such, are responsible for complying with the principles of the GDPR, as well as for demonstrating their compliance, in accordance with Art. 5 GDPR. Under these circumstances, the principles of the GDPR do not directly apply to the entity managing the platform, which will typically act as a processor (with respect to the cohort data) in the sense of the GDPR. However, from a strategic point of view, it is crucial to consider what services and capabilities the platform may need to offer to enable controllers meet their compliance obligations under Art. 5 GDPR (see Box 4 in Annex B for the list of Art. 5 GDPR principles uniquely applicable to controllers). This section focuses on the steps and strategies the parties involved in the EPND-facilitated data use lifecycle may be collectively required to adopt, in order to ensure that the controllers are compliant with the principles of the GDPR.

Data providers are medical or research institutions responsible for ND cohorts, including managing biological samples and data associated with the cohorts. These institutions routinely collect and/or generate personal data pertaining to ND patients or research participants. From the point of view of a (prospective) data provider, the key question is whether, and under what circumstances, inclusion of their dataset in the EPND would be GDPR-compliant. As it will be shown below, this is a complex question, the answer to which will be context-dependent. In some cases, data providers will be able to readily deposit their datasets with the EPND and make them widely available for external researchers, whereas for some ND cohorts, data sharing may not be permissible under the GDPR. Still in other cases, the provider may be able to deposit the data with the EPND and allow for limited processing, such as data cleaning, standardisation and long-term storage/hosting, while at the same time being legally required to make future data access and reuse by downstream controllers (external researchers) conditional upon additional steps such as recontacting and reconsenting the data subjects.

In the context of GDPR compliance, the "permissibility" of sharing data via the EPND is encapsulated by the principle of "lawfulness" of processing, enshrined in Article 5(1)(a) GDPR, according to which personal data "shall be processed lawfully, fairly and in a transparent manner in relation to the data subject". Art. 6(1) GDPR further elaborates on the notion of lawfulness, clarifying that in order for processing to be lawful, it must be grounded on one of the conditions listed in Box 3. Below.

Box 3 - Article 6(1) sub-para 1 of the GDPR

Processing shall be lawful only if and to the extent that at least one of the following applies:

1. the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
2. processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
3. processing is necessary for compliance with a legal obligation to which the controller is subject;

4. processing is necessary in order to protect the vital interests of the data subject or of another natural person;
5. processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
6. processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

These conditions, commonly referred to as the 'legal bases' or 'lawful bases' apply to all forms of processing of personal data. Without grounding a processing operation in (at least) one of the six legal bases, the processing cannot be lawful. This means that the party acting as the controller for a particular processing operation must select a legal basis it deems most suitable in view of the nature of processing. Where special categories of data are concerned, including data relating to health and genetic data, processing must also be based on one of the 10 conditions in Art. 9(2) GDPR, in addition to the legal bases delineated in Art. 6(1) GDPR. The practical implication for the EPND is that for each processing operation undertaken on the special categories of personal data within the EPND's Technical Hub (see Table 5 in Annex B for types of such processing operations), the party/parties acting as the controller(s) for the processing must select a valid Art. 6(1) legal basis, coupled with a suitable Art. 9(2) condition.[4]

In principle, the controller has a considerable discretion to choose among the six legal bases as the grounds for processing. However, in practice, the range of options available to the controller will be limited, with the choice being dictated by the nature of the intended processing operation, the national implementation of the GDPR in the controller's country, and other relevant constraints. For example, in the context of scientific research requiring processing of personal data, the European Data Protection Board (EDPB) states that the choice of legal bases will often be limited to the following three options: data subject's explicit consent (Art. 6(1)(a)), a task carried out in the public interest (Art. 6(1)(e)), and/or the legitimate interests of the controller (Art. 6(1)(f)).[5] The EDPB also offers guidance on which of the Art. 9(2) conditions should be relied upon, in conjunction with Art. 6(1) legal bases, when special categories of data are concerned.

Each of the three Art. 6(1) options commonly available to researchers, alongside their corresponding Art. 9(2) grounds for processing special categories of data, come with certain conditions, caveats, and limitations, as explored in Annex C: Consent, public interest and legitimate interest.

[4] Notably, there is a widespread misconception that when processing of previously collected data is carried out for scientific research purposes, the processing automatically meets this requirement of lawfulness, and the controller need not identify a valid legal basis to justify the processing. A recent publication by Becker et al. elucidates the origins of this misconception and explains why this view is inconsistent with the principle of lawfulness (Becker, Chokoshvili, et al. 2022).

[5] For certain ancillary processing operations, which are closely associated with, but not part of core scientific research activities, other legal bases may also be used. In its Opinion 3/2019, the EDPB particularly highlights that any processing carried out to comply with applicable legal obligations - such as mandatory archiving or reporting of data to regulatory bodies - should be based on Art. 6(1)(c) of the GDPR (European Data Protection Board 2019).

Apart from their essential role in enabling GDPR-compliant processing of personal data, Art 6(1) GDPR legal bases are also relevant for determining whether, and to what extent, different rights of the data subjects apply. The topic of data subjects' rights under the GDPR is explored further in Annex D: Data Subjects' rights.

Potential compliance needs

The findings of the analysis presented in Annexes C-D highlight potential needs of prospective EPND participants regarding their GDPR compliance obligations. It can be concluded that ND cohorts, who will typically act as initial controllers vis-a-vis data subjects (i.e., controllers who have collected personal data from data subjects), would particularly benefit from ongoing support by the EPND in relation to cohorts' GDPR compliance. For example, the EPND could provide operational and infrastructural support, such as digitally-enabled services aimed at streamlining cohorts' capabilities to recontact their data subjects, be it for information notification purposes under Art. 13 GDPR, or in order to obtain a new consent, where necessary. Additionally, the EPND would be well-placed to provide advisory services to the cohorts, offering valuable guidance in the areas requiring greater interpretive clarity. This may include, for example, assisting cohorts in assessing the applicability of data subjects' right to access, right to data portability, and right to rectification, among others. Although cohorts as controllers would remain responsible for demonstrating their compliance with the principles of the GDPR, they may nevertheless benefit from the availability of a service such as the EPND ELSI helpdesk that cohorts can address for specific compliance-related questions. Finally, the EPND could support cohorts in meeting certain procedural and documentation obligations in their capacity as controllers under the GDPR, such as performing a Data Protection Impact Assessment (DPIA), upon a cohort's request.

Chapter 3: Governance View

Part of the mission of the EPND is to enable a large European resource pool of high-quality sample and data collections that can be meaningfully utilised in research focused on NDs. Fundamentally, this translates into establishing an integrated cross-border ecosystem aimed at facilitating the discovery, access, and research use of ND sample and data collections held by institutions across the European countries. The success of this ambitious goal is contingent, to a large extent, upon the feasibility of developing a coherent framework that streamlines the governance of the EPND, while at the same time aligning with the numerous laws, standards, and policies under which institutions comprising the EPND ecosystem must operate.

This chapter provides an overview of the diverse legal and governance frameworks pertinent to European institutions engaged in biomedical research, including research into NDs. The chapter predominantly focuses on organisations facilitating and/or performing research that employs previously collected human biological samples and data. This overview is helpful for identifying the key governance challenges that need to be addressed by a European research platform such as the EPND, particularly with a view to enabling seamless discovery, streamlined access to, and meaningful research use of existing European ND resources. Subsequently, the chapter outlines the key foundational principles and preliminary recommendations for designing a suitable access governance framework for the EPND. The access governance framework consists of a set of overarching rules, organisational procedures, and accountability tools around the governance of ND sample and data collections in the EPND.

Research Involving Existing Biological Sample and Data Collections: the European Landscape

Human ND biological sample and data collections with potential scientific value can be generated across a wide range of contexts. The emergence of such collections can be concomitant to the routine healthcare practice, whereby patients with a known or suspected ND undergo various screening, diagnostic, or treatment procedures. The collections may also be generated as part of a specific research project in the area of NDs where, upon the project's completion, it is foreseen that the samples and data can be retained to make them reusable in other studies. Finally, ND sample and data collections can be established with the expressed primary objective of enabling future (yet-to-be-specified) research into NDs, as is the case with prospective biobanks and longitudinal observational cohorts. A shared commonality across these contexts is that they all give rise to human biological samples and/or data collections that can be of significant scientific value in the future for researchers working in the area of NDs.

In Europe, the practice of research based on existing human biological sample and data collections (i.e., retrospective biomedical research) is governed by a complex set of legal, ethical, and policy frameworks across the following three key areas:

- Research ethics (specifically, ethics of research involving human subjects)

- Biobanking (i.e., collection, storage, and treatment of human biological samples)
- Privacy and data protection

The latter area, privacy and data protection, is characterised by a comparatively high degree of legal harmonisation across Europe. This is due to the fact that the regulatory framework introduced by the GDPR applies throughout the EU and has also been embedded in the data protection laws of other European - especially (but not limited to) EEA - countries. However, the same is not true for research ethics and biobanking legal/policy frameworks in Europe, as these areas lack similar harmonisation, resulting in a more heterogeneous landscape across European countries. Furthermore, as it will be discussed below, the scope of the national laws and policies governing these three areas often overlap, creating practical governance challenges to enabling research use of existing sample and data collections.

See for a comprehensive exploration of the three pillars - Annex E: The Three Pillars

Interrelatedness and Complexity

The inherent interrelatedness of the different legislative pillars concerning retrospective biomedical research highlights the critical need for European countries to streamline their legal frameworks, ensuring complementarity of the applicable national laws. However, in view of the growing complexity and the fragmentation of the national legal landscapes in European countries, achieving this goal requires that legislators have an increasingly holistic as well as nuanced knowledge of the various laws applicable to retrospective biomedical research in their country.

It should be highlighted that the complex interplay of the relevant national laws increasingly translates into practical hurdles hindering the implementation of retrospective biomedical research projects. Apart from potentially subjecting researchers to burdensome administrative and compliance requirements, complex legal frameworks have also been responsible for creating confusion as to the roles and responsibilities of various stakeholders whose formal legal mandates are defined under different national laws. The prime example of such confusion is a potential overlap between the legal mandate and competences of Research Ethics Committees (RECs) with those of data protection authorities, DPOs, and other stakeholders overseeing privacy-related matters. In this regard, one of the early high-profile cases involved the Oslo University Hospital, Norway, where the local DPO had been given broad authority by the institution to decide on various matters related to healthcare and research uses of personal data. This resulted in a series of decisions by the DPO that the local research community saw not only as overly restrictive, but also as a clear overreach by the DPO, infringing on the legal competences of the REC. As a consequence, 32 researchers signed a petition against the DPO and submitted it to the Norwegian Board of Health in late 2018, catalysing a protracted national policy debate over the overlapping legal competences of DPOs and RECs (Befring 2021). Legislators in other European countries have also recognised this potential issue and have taken steps to address it by clearly defining legal mandates of the parties. For example, in Greece and Spain, the assessment of privacy- and data protection-related issues in the context of retrospective biomedical research is explicitly incorporated into the ethics review process under the national research ethics law and biobanking law, respectively (Tzortzatou and Siapka 2021; M. Romeo-Casabona 2021). A somewhat

different approach has been adopted by Italy, where although the initial assessment of retrospective research projects for their GDPR compliance is the controller's responsibility, a separate review by the national data protection authority (GARANTE) is mandatory under certain conditions ("Decreto Legislativo 30 giugno 2003, n. 196 'Codice in materia di protezione dei dati personali'" 2003). The aforementioned examples highlight the European national legislators' early efforts to pre-empt conflicting overlaps among the relevant laws and regulations bearing on retrospective biomedical research. However, a significant amount of additional legislative work is required to achieve this goal. To our knowledge, overlapping roles and competences of various actors and/or bodies designated under different laws continue to pose practical challenges at all levels of governance in Europe, including at the local (within-institution), regional (e.g., in federal countries), national, and international levels. Unsurprisingly, the magnitude of these challenges is particularly pronounced at the international level, where inconsistent legal frameworks or conflicting roles and competences of various national and local actors engender significant practical barriers that hinder cross-border research.

To illustrate the nature of these barriers, Annex F: Informed consent and ethics approval focuses on two pertinent overarching issues that tend to be regulated in a fragmentary manner under the different pillars of the relevant national legal frameworks: i) informed consent; and ii) ethics approvals for data/sample use.

While these two examples are non-exhaustive in illustrating legal and policy interoperability issues in the context of cross-border biological sample and data transfers, they capture the essence of the challenges that need to be tackled by the EPND community. Collectively, these challenges can be grouped under the broad theme of Access Governance, a critically important component of an international research platform such as the EPND.

Access Governance

Access governance aims to ensure effective, efficient and responsible access to scientific resources such as biospecimens and data. Access governance seeks to enable researchers to unlock the scientific value of biomedical collections while at the same time ensuring compliance with the relevant legal, ethical, policy, and (where applicable) business requirements. The sources of these requirements can be the following:

a) The three pillars of the legal and policy frameworks governing retrospective biomedical research (see Annex E: The Three Pillars)

Research ethics: aims to ensure science is conducted in a high-quality manner, and in a way that respects the rights and interests of research participants. Research ethics norms may be found in human subjects' research laws and guidelines, as well as biobanking/human tissue laws and guidelines. The main types of requirements found in these norms relate to informed consent, research ethics review and oversight, privacy protection and data security (overlapping with data protection norms), among other considerations.

Biobanking: concerns the collection, storage and use of human biological materials which may have been derived in clinical or research settings. The practice can be regulated under specific national biobanking laws or, in the absence thereof, in a more fragmentary and often incomplete manner where various aspects of the practice are addressed by different laws and policies. National biobanking laws, where they exist, may partly overlap with the other two pillars of the applicable legal and policy frameworks.

Privacy/Data protection: specifically concerns the protection of personal data relating to data subjects (namely cohort patients and research participants). The GDPR and its complementary national data protection laws apply.

b) Sustainability and incentives

Access governance may include considerations relating to publication and commercialisation rules that ensure fair and appropriate incentives are in place that encourage cohorts to share samples and data, and that encourage researchers to request and use samples and data. Well-defined rules and policies concerning incentives for sharing will ensure clarity around matters of strategic importance and help establish a level playing field where the interests of both sides - resource contributors and resource users - are taken into consideration.

c) Patient and public involvement

Access governance, particularly when implemented in well-articulated and comprehensive manner, may explicitly address patient and public involvement considerations. Increasingly, patients and publics play a partnership role in establishing strategic as well as sample/data governance in health research. Ethically sound access governance should ensure that the research use of sample and data collections are aligned with the values, interests, and reasonable expectations of the communities from which the collections have been sourced. To this end, a meaningful engagement of patients, their representatives, and other members of the community is an important goal.

Defining the Needs for the EPND Access Governance Framework

Access governance frameworks are established to implement and operationalise governance principles and requirements. Access governance frameworks comprise specific rules, procedures, processes, and contractual agreements. Through well-designed access governance frameworks, samples and data within a research ecosystem can be effectively and fairly accessed for appropriate research purposes, while also ensuring legal and ethical requirements are met, and risks are minimised.

Individual institutions, cohorts, data repositories, or biobanks may have their own access governance policies and processes. However, even if single resource access governance works effectively and efficiently, requestors may face challenges seeking to link resources at the level of consortia, or research networks. This is especially true for large ecosystems such as the network being established

by the EPND, which seeks to bring together a large number of institutions from different countries. For any complex ecosystem consisting of heterogeneous resources and research organisations, it is crucial to have an additional, overarching access governance framework at the level of the ecosystem. The overarching access governance framework will aim to improve the transparency of, and the degree of harmonisation among local governance processes. Additionally, it will seek to establish common policies and coordinated processes for requestors seeking access to multiple different data and sample resources.

As highlighted previously, there are various legal, policy, and governance challenges hindering data and sample transfers between institutions based in different European countries.

- Certain datasets may not be accessible to researchers who are required to process personal data under a particular GDPR legal basis not permitted by the cohort (see Annexes C-D, and F);
- A biobank or a cohort may require a formal (and possibly jurisdiction-specific) REC approval as a precondition for providing samples/data, which researchers in some jurisdictions may not be able to obtain (see Annex F)
- In certain jurisdictions, cross-border transfers of human biological samples for research purposes are subject to additional special approvals. This is, for example, the case in France, where both import and export of human biological samples require a formal authorisation by the French Research Ministry[6] (see Annexes E-F)
- Additionally, as discussed below in this chapter, cohorts and biobanks may impose various bespoke restrictions or conditions for the use of their collections, which will further limit the pool of the eligible researchers and/or research projects that can make use of the collections.

In the context of sample and data transfers taking place between two parties, or within a multiparty research consortium, such interoperability issues usually significantly delay the implementation of an intended research project, as also evidenced by the experience of the EPND consortium in EPND case studies (discussed below). Where the interoperability challenges cannot be resolved, they may even preclude the implementation of the research project altogether. However, the negative implications of the legal/governance interoperability issues, in principle, can be alleviated by an intermediary platform such as the EPND. Assuming the platform succeeds in attracting a sufficiently large number of sample and data contributors from different jurisdictions, the platform could overcome the interoperability issues by simply matching a requestor with the cohorts whose data/sample access and use conditions the requestor clearly meets. For example, a requestor unable to obtain a REC approval would be matched by the platform with cohorts that either don't require a REC approval as a precondition for transfers, or are willing to grant an exemption under certain conditions, and the platform can verify that the requestor meets these conditions.

Therefore, from the platform's perspective, legal or governance interoperability challenges between any cohort-requestor pair are not necessarily problematic, provided that i) there are sufficiently many

[6] More detailed information about the authorisation can be found on the following page: https://www.esrf.fr/Infrastructure/Safety/Experiments/Biology_Experiments/samples-from-humans-and-or-derivatives

institutions participating in the network served by the platform, and ii) the platform is able to match the requestor with suitable cohort(s) whose collections the requestor can access. Consequently, for the platform, the principal challenge is transparency: it must ensure that it has accurate, comprehensive, and up-to-date information concerning the availability of cohorts' resources. For this, the platform requires a detailed picture of the applicable conditions for data and sample use, as well as specific restrictions or prohibitions. Because these considerations will necessarily vary not only across cohorts but also among individual sample donors/data subjects within the same cohort, the information required by the platform is highly detailed and granular. Of note, within-cohort variability in sample/data use conditions and restrictions is particularly common where the cohort relies on a layered informed consent with embedded options when recruiting research participants. Recall, for example, the excerpt from an informed consent form in Box 2, whereby individual research participants can impose highly specific restrictions on the research use of their samples/data based on the purpose, geographical location, and methodology of the prospective research projects.[7] The experience with the initial EPND case studies has made it clear that in practice, at least a small proportion of research participants make use of the available discriminatory choices regarding the use conditions and restrictions applicable to their samples and the associated data, giving rise to the aforementioned within-cohort variability. At the same time, certain types of use conditions or restrictions may apply to an entire cohort in a uniform manner. These include, for example, the requirement that a requestor provide a valid REC approval for the intended study as a precondition to data/sample access, or the (un)availability of the data depending on the requestor's GDPR legal basis for processing personal data.

Given the wide range of data/sample use conditions and restrictions, both across and within cohorts, the value of transparent communication around these matters cannot be overemphasised. Arguably, one of the most valuable functions the EPND consortium could perform for the European ND research community consists of accurately and comprehensively mapping sample/data use conditions and restrictions to each collection within the ecosystem of cohorts, with a view to leveraging this information to facilitate meaningful resource search and discovery by external researchers. Essentially, the EPND will need to compare the data/sample use conditions and restrictions against the characteristics of the intended research project of a requestor, enabling the EPND to accurately determine whether the EPND cohorts, collectively, hold the collections the requestor can lawfully use to carry out the intended research project.

These considerations clearly indicate that ensuring transparency is the key enabling requirement for the EPND to function effectively as a research-facilitating intermediary. Transparency is crucial for the establishment of an access governance framework enabling the EPND to deliver its meaningful resource discovery, streamlined data access, and centralised or federated analysis services in an

[7] As discussed in Annexes C and F, a granular/layered informed consent with embedded options would be required under the GDPR if the data collection is taking place based on Art. 6(1)(a) in conjunction with Art. 9(2)(a) GDPR. In the case of another GDPR legal basis, such as Art. 6(1)(e) coupled with Art. 9(2)(j) GDPR, informed consent forms for the research project need not explicitly ask research participants to opt in for specific uses of their data. For a detailed comparison of the "GDPR consent" and the "research ethics (informed) consent", see Annex F, section "Informed Consent in retrospective biomedical research".

efficient and compliant manner. What follows are preliminary best-practice recommendations to inform the EPND Access Governance design choices.

Best practices for EPND Cohort sample/data access governance

Define Purposes / Access and Use Conditions

The starting point for designing any access governance framework is to clearly define the purpose(s) for which a cohort intends to provide external researchers with access to a resource, and then to identify the conditions that apply to access and use. This step is key for the cohort to comply with the applicable laws and policies. It is also valuable to make this information transparently available (as part of a published access policy), enabling potential requesters to determine not only whether the resources are scientifically relevant for their research project, but also whether the samples and data can legally and ethically be made available to them for the project.

It is worth highlighting, however, that the purposes(s) for which EPND collections may or may not be made available, alongside other detailed access and use conditions pertaining to specific collections will, to a large extent, be shaped by the scope of the services provided by the EPND. The EPND needs to be clear about the scope of uses it aims to facilitate determination of access and use conditions. Presumably, the scope of EPND activities will naturally be focused on health and/or biomedical research aiming to improve the understanding, prevention, and treatment of human disease - in particular, neurodegenerative diseases. This is because of the nature of the actors and resources involved. However, it remains to be determined by the EPND consortium whether there are types of foreseeable scientific research that would be considered out of scope, such as health and wellness research focused on cognitive enhancement. Furthermore, there is a question as to whether there are potential other types of purposes beyond scientific research that would be permitted. The most prominent examples include digital innovation activities, such as the development of novel artificial intelligence/machine learning-based methods, as well as the use of EPND sample and data collections for quality assurance and/or research equipment calibration purposes.

There are also questions whether the research activities enabled by the EPND are limited to providing access to existing samples/data for data-intensive research projects and biomarker studies, or would also extend to other activities such as recontacting of participants for recruitment into clinical trials based on their health and/or genetic profile. If the EPND decides to provide this additional function of acting as a registry of potential clinical trial participants (including patients with NDs) who agree to be contacted for recruitment, this could be considered an additional purpose, requiring its own governance rules and procedures to ensure that recontacting of patients is pursued in an efficient and responsible manner. The EPND's choice whether to provide such additional services will likely depend on its precise role in supporting actors within the medical industry towards the discovery, qualification, and validation of novel biomarkers and methods (for more information on regulatory support, see Annex G: Regulatory guidance – Support tools offered by the European Medicines Agency for Innovations)

Determining the scope of EPND purposes is key from a data protection and ethics perspective, as well as from an access governance perspective, as a multi-purpose processing ecosystem will typically require more complex governance frameworks.

Collection-specific Access and Use Conditions

Notwithstanding the uncertainty regarding the purposes to be supported by the EPND, it is crucial to be mindful that the sample and data collections included in the EPND will typically have associated access and use conditions specific to each collection. These conditions will determine whether, and under what circumstances, samples and/or data from a particular EPND collection can be used in a proposed research project. From a practical point of view, it might be useful to differentiate among the following four layers of data/sample access and use conditions.

- *Purposes (What types of uses are permitted / prohibited?)*

For example, some collections can only be used in research projects focused on a specific disease, or a specific disease area. The availability of some collections may also be restricted to non-commercial research projects.

- *Users (Who can access and use the collections?)*

There may be restrictions to the types of entities that can access the collections, based on the entities' legal status (e.g., for-profit vs publicly funded research institutions), geographical location, and the jurisdiction in which the entity operates, among other factors.

- *Prerequisites for access (What are the substantive and procedural requirements the requestor must meet prior to being granted access?)*

These may include, to the extent applicable, obtaining a favourable opinion from a competent REC, an approval by the cohort's D(S)AC, and any other review or approval mandated by the cohort. In relation to the processing of personal data, the cohort may also require that the requestor carry out a data protection impact assessment (DPIA) for the intended research project.

- *Scientific, publication and commercial access and use conditions*

This is a broad set of conditions encompassing the obligations of the requestor vis-a-vis the cohort owner institutions. They only become relevant after the use of the sample and/or data collection has been deemed ethically and legally permissible, following a preliminary assessment of the proposed research project. These conditions may include, to the extent applicable, the user's obligation to:

- Acknowledge the source of the samples/data in resultant scientific publications (in certain cases, it may also be required to include researchers affiliated with the cohort's owner institution as co-authors of the paper)
- Return any newly generated or enriched data to the cohort owner

- Report any medically actionable incidental findings pertaining to individual patients/research participants
- Financially compensate the cohort owner for any direct costs (e.g., costs associated with recontacting/reconsenting research participants, transportation of samples, procedural and compliance-related costs, where applicable)

It should be noted that the first two sets of conditions - the permitted / prohibited purposes and users - are typically "fundamental" access and use conditions in the sense that they stem from the legislation, policies, approvals and consent agreements that are binding on the cohort/biobank. They cannot be easily modified or negotiated for existing sample and data collections (e.g., without reconsenting the individuals, or seeking renewed approvals), and thus they are essentially non-negotiable.

The third set of access and use conditions - prerequisites for access, such as various forms of approvals - may include both fundamental and negotiable conditions, depending on the applicable laws and institutional policies. For example, in the case studies involving obtaining samples and data from DZNE (discussed more extensively in Annex F, section "Ethics approvals for data/sample use"), the EPND consortium learned that the requirement to submit a REC approval was a fundamental, that is, non-negotiable access and use condition. In contrast, the authors of this paper are aware of various European research institutions that, despite having the same nominal requirement as part of their access governance rules, will commonly exempt access-requesting researchers from a REC approval. The same observation holds generally true for other approvals, research project reviews, and risk assessments (including the DPIA): whether, and the extent to which these prerequisites are negotiable varies across the jurisdictions and institutional policies.

Finally, the access and use conditions relating to scientific, publication, and commercial matters tend to be, with a few exceptions, negotiable, as they rarely stem from prescriptive legal obligations or rigid institutional policies. Although cohort owners may have preferences regarding these conditions, they will typically have greater discretion to engage in a negotiation with the prospective sample and/or data users.

The aforementioned considerations are of relevance for the EPND access governance framework in two important ways. First, understanding cohort-specific access and use conditions are crucial for enabling meaningful resource discovery and a preliminary study feasibility assessment by researchers interested in using EPND collections. By comprehensively mapping out cohorts' access and use conditions, as well as indicating whether these conditions are negotiable, the EPND can enable researchers to make an informed decision whether it is worthwhile to initiate a formal access request. To facilitate meaningful discovery, the EPND could integrate one of the existing research resource discovery tools, or components thereof, into the EPND user web portal. One such tool is the Digital Use Conditions (DUC), developed by the University of Leicester, the UK[8]. The DUC is a layered resource discovery solution with an embedded search logic that reflects resource-specific access and use conditions. The EPND can also seek insights from various decision-making tools used by research-

[8] Available at: https://duc.le.ac.uk

oriented data-sharing platforms. For example, ADDI has developed a decision tree[9] to help determine whether informed consent forms permit sharing data with third parties for secondary research on Alzheimer's disease. The decision tree covers providing access to (certain) external research organisations, restrictions on the subject matter (e.g., disease-specific research), and special safeguards (e.g., removal of particular identifiers).

Second, surveying cohort-specific access and use conditions in a comprehensive manner is crucial for the EPND to identify key commonalities and differences across cohorts. In the areas of greater convergence, there will be opportunities for the EPND to help streamline access and use conditions across the network of the participating institutions. The experience of other ND research platforms and consortia indicates that some access and use conditions - for example, those pertaining to scientific publications based on the use of community resources - are particularly amenable to harmonisation.[10] A detailed understanding of the EPND cohorts' access and use conditions may allow the EPND to discern multiple elements that can be brought in alignment across the participating institutions and streamlined under the overarching EPND access governance framework.

Barriers to Establishing Transparent Access and Use Policies

Interpretation. One challenge is accurately interpreting sample/data access and use conditions, especially on a generic level (i.e., for unspecified future research projects). This requires a certain level of sophistication on the part of the cohort – i.e., being able to determine what requirements apply to samples and data.

As part of the ongoing ethical and legal support work within the EPND consortium, members of WP2 reviewed various legal documents, including but not limited to informed consent forms, pertaining to prospective EPND cohorts. During this review, WP2 identified numerous statements regarding data/sample access and use conditions that were deemed ambiguous by the WP2 members. Several examples of such ambiguous statements are listed below.

[9] Available at: https://www.alzheimersdata.org/resources/tools-for-researchers
[10] E.g., See common data sharing and publication policies from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

Table 2 - Ambiguities in relation to interpreting data/sample access or use conditions

Data/sample access or use condition	Ambiguity
Samples/data from the cohort cannot be used for commercial purposes	It is unclear whether the condition specifically refers to the development of commercial products, or is meant to be interpreted in a more restrictive manner (e.g., for-profit entities are not allowed to access and use the collection, even as part of an academic research collaboration led by public institutions)
An ethics approval for the proposed research project is required, but an exemption can be granted under certain circumstances	Conditions for the exemption have not been specified.
Access to data is contingent upon the researchers' willingness to provide additional assurances demonstrating GDPR compliance. This may include carrying out a data protection impact assessment (DPIA).	It is unclear what criteria the cohort owner uses when deciding whether a DPIA must be performed.
A research participant has opted out of "genome sequencing", "genomic analysis", or "next-generation sequencing"	Depending on the interpretation, these terms can refer to any form of genomic analysis, from highly targeted genotyping for a set of predefined variants, to whole-genome and whole-exome sequencing (WES/WGS). It is possible that the research participant meant to opt out of WES/WGS, for example, out of concern that such broad analysis often results in incidental findings, without necessarily objecting to targeted genotyping for genomic biomarkers specific to a neurodegenerative disease (e.g., GBA and APOE genotypes). On the other hand, it is also possible that the research participant is opposed to any use of their genomic data based on unspecified moral grounds or beliefs.
A research participant has consented to be enrolled in studies focused on neurodegenerative diseases, but has opted out of other categories of research	Several types of research fall in the grey area (e.g., healthy ageing studies with a neurological component). More importantly, while it appears that the research participant was concerned about potential misuse of their samples or data, the specific nature of these concerns is unclear. This makes it difficult to deduce the research participant's reasonable expectations and ensure that they are respected in future research projects.

When contacting the organisations responsible for ND cohorts for additional clarifications, WP2 members observed that they were often reluctant to frame these and other generic conditions in an unambiguous manner. Furthermore, based on WP2 members' preliminary interactions with different

stakeholders responsible for managing cohorts (including PIs, RECs, D(S)ACs, and DPOs), it became clear that these stakeholders generally do not consider ambiguous access and use conditions as inherently problematic. This is due largely to the fact that currently, the prevailing practice at research organisations is to evaluate data and/or sample access requests on a case-by-case basis. The competent D(S)AC carrying out this evaluation will take into consideration various contextual factors relating to the proposed research project. Based on these factors, the D(S)AC may decide whether to require additional assurances (e.g., in the form of a REC approval or a DPIA) or how restrictive it must be in interpreting particular data/sample use conditions. However, this approach conflicts with the goal of the EPND to ensure transparency and clarity around the availability of data and sample collections. As a consequence, the EPND may need to expend considerable effort towards engaging stakeholders responsible for managing ND cohorts around these matters. The EPND could better educate such stakeholders on the benefits of clearly defined data/sample access and use conditions in the context of meaningful resource discovery, as well as seek to negotiate with them to redefine some of the more ambiguous access and use conditions. The EPND should particularly emphasise that by clarifying access and use conditions, the cohort owners will not be relinquishing their authority in relation to making access decisions; they will only help the EPND ensure that access requests are submitted by eligible researchers and for eligible research projects. If communicated effectively, this message may well be regarded as a welcome development by the stakeholders responsible for data and sample collections.

Another challenge is negotiating conditions relating to scientific, commercial, and, to the extent applicable, intellectual property aspects. There are often concerns that the conditions of access would be unfair to the interests of sample and data providers. Concerns over these aspects may be higher for newly generated resources that have not yet been analysed or published, as there will be worries as to losing out on scientific and commercial opportunities. The involvement of commercial requestors may also lead to uncertainty over the appropriate and fair conditions of access. This also ties to incentives and sustainability issues: if a cohort is not motivated to provide access in the first place, it may be unwilling to invest resources to do this assessment.

Strategies to address these concerns include establishing a set of overarching agreements and policies that negotiate these aspects at a general level. (Scientific and commercial ownership of findings generated using EPND data/sample collections, including intellectual property rights, will be explored in the final EPND White Paper, expected to be published in late 2025). As a key guiding principle, cohorts joining the EPND community should agree to make their collections available for research purposes on a non-discriminatory basis. Cohorts should also agree to follow the EPND community standards and policies that establish fair, standard rules relating to these non-fundamental access and use conditions. The EPND, primarily through WP2 and with input from WP7 (Sustainability), will develop these standards and policies as part of establishing the access governance framework throughout Phase II (2022-2026) of the IMI project.

Another general strategic recommendation is that the EPND should engage cohorts towards developing or refining, as relevant, cohort-specific governance frameworks around sample/data sharing and future reuse. The goal of these engagements should be to maximally align the rules, conditions, and required procedures across the cohorts, thus effectively bringing local (cohort-specific)

access governance frameworks in alignment with the streamlined EPND access governance framework. When (re)defining components of their access governance frameworks, the cohorts may choose to seek institutional or research ethics committee approval, if applicable. The approval will confirm that the newly adopted (or modified) rules, conditions and procedural requirements accurately reflect the institution's policies as well as the consent agreements (where applicable) interpreted in the context of applicable regulatory frameworks. A cohort's willingness and ability to align with the EPND's access governance framework will be an important factor in determining the extent to which the cohort will be participating in the EPND.

Define Access Processes and Data/Sample Access Committees (D(S)ACs)

Currently, WP2 is exploring various ways to help streamline the data access processes within the EPND ecosystem. This includes assessing the feasibility of establishing a central Data and Samples Access Committee (DSAC) to be tasked with facilitating the evaluation of access requests in a centralised manner.

During the first year of the EPND IMI project, ongoing interactions with partner cohorts has made it clear that the majority of European institutions contributing sample and data collection to the EPND cannot fully delegate their authority over access decisions to an external body. This is due to a variety of legal (including GDPR-related) reasons and procedural requirements the institutions must follow prior to granting access. However, this observation does not necessarily preclude a central EPND DSAC. It could be feasible to establish the central DSAC in a manner that leaves the decision-making *authority* with the cohorts, while centralising the decision-making *process* through the DSAC. More specifically, this can be achieved where the cohort provides the EPND with sufficiently detailed access and use conditions in a generic manner and tasks the EPND with assessing future data/sample requests based on these conditions. In this manner, the EPND, through its DSAC, essentially implements the cohort's detailed instructions, ensuring that the cohort remains in full control over access decisions. To operationally guarantee the cohort's role as the access decision-maker, the EPND could additionally implement a two-step process. In this approach, an access request would be initially assessed by the EPND's central DSAC, based on the access and use conditions predefined by the cohort. Where the conditions are not met, the central DSAC could promptly inform the requestor about the missing requirements, avoiding unnecessary delays and allowing the requestor to decide whether to resubmit a modified request. Where, on the other hand, the central DSAC deems that the proposed research use meets the predefined conditions, the central DSAC can notify the cohort about the eligible request. The final decisions concerning whether to grant access would depend on the approval by the cohort. In this two-step approach, although the access decision-making process is not fully centralised, it is nevertheless streamlined, enabling the EPND to reduce inefficiencies and long delays traditionally associated with the review of access requests.

The feasibility of the different modalities for streamlined access decisions will be explored more systematically by WP2 during 2023. Even if the evaluation of access requests cannot be streamlined due to practical constraints and/or unwillingness on the part of cohorts, the EPND may still be able to improve the overall access process by enhancing transparency. For example, the EPND, via its website

and the user portal, could explicitly and transparently explain to prospective data/sample requestors the following steps:

1. How a requestor discovers available resources
2. How a requestor makes an access request (e.g., through a web portal)
3. What information must the requestor provide in an access request (e.g., identity of the requestor, description of the project aims and methods, sample/data types required, justification for requesting pseudonymised data, requested/justified access period, technical and organisational measures in place to protect the rights and interests of data subjects and data providers, applicable legal basis under the GDPR, and information on any ethical assessments, as applicable).
4. How access requests are reviewed by cohorts, including indicative timelines, types of possible decisions (e.g., approval, requests for clarification/modification, refusal with justifications), and the assessment criteria typically employed by cohorts.
5. In case of a positive decision, how access is provided (i.e., via a dedicated secure data processing environment vs a direct data transfer) and what material and data access/use agreement must be signed.
6. Any conditions in relation to the monitoring, periodic reporting, auditing, and termination of access (during or after completion of the project).

As a general principle, the access process should be proportionate to the level of risk to ensure a proper balance is struck between the aims of access and the aims of protection. The access process should be fair, efficient, and, to the extent possible, non-discriminatory vis-a-vis requestors. Building an access policy that incorporates the aforementioned transparency requirements will be an important task of WP2 from year 2 of the IMI project. Encouragingly, for the past few years, various leading biomedical research consortia and ecosystems have been working towards developing effective and efficient data access policies. The EPND WP2 members can draw valuable insights from these existing frameworks, which include, among other sources, the GA4GH Data Access Committee Review Standards (DACReS), and the work by the Expert Advisory Group on Data Access (EAGDA) of the Wellcome Trust.

The issue arising when the access policy and process is not explicit and transparent is that it may not be clear which individuals or units within an organisation will need to review an access request, and approve the associated access/use agreement, until after the request is launched. Subsequently, the agreement may move from the Principal Investigator or the research team, to consultations with the Data Protection Officer, to review by other organisational units such as legal counsel and technology transfer, review by the REC and possibly also go through several layers of the institutional hierarchy to obtain the appropriate approvals and signatures. At European research institutions without well-defined access governance frameworks, this process usually takes several months, as also reported by other authors (Devriendt et al. 2021).

Establish Appropriate Safeguards

Access to EPND data collections can only be granted with appropriate safeguards in place, such as data use agreements and appropriate technical and organisational measures that would be deemed adequate under the GDPR. Some measures such as pseudonymisation may be applied by the cohort before a data collection is even included in the EPND, whereas other measures such as cybersecurity safeguards will need to be provided by the platform utilised for the data analysis.

Data Access/Use Agreements

Contractual agreements with users are a common and generic safeguard to ensure the responsible use of samples and data. These are typically material and data transfer/access/use agreements. The clauses of these agreements typically include:

- General terms applicable to all projects (often found in a standard cohort, consortia, or community template)
    - Use for authorised purpose(s) only.
    - Prohibition of data subject re-identification or establishing individual-level linkages.
    - Data breach reporting obligations.
- Details specific to the project/user (project purpose, resources, duration, fees, technical/organisational measures in place). These elements are also commonly referenced within the general terms and conditions of the online research portal through which data are accessed.
- Reiteration of dataset-specific access and use conditions, reflecting the outcome of the access decision in the previous step.

There are various potential challenges complicating the adoption of standardised data access/use agreements by a research platform such as the EPND. Many of these challenges may arise due to the fact that the main stakeholders responsible for legal compliance at the prospective signatory institutions, including the DPOs, privacy officers, and other legal personnel, may disagree on how to interpret the key GDPR concepts in the envisaged data use scenario. As part of the initial EPND case studies, WP2 members encountered such disagreements around the definition of personal data, as well as the assignment of the controller/processor roles and the associated GDPR obligations to the participating institutions. This experience has been immensely helpful in understanding the risk-based approach to contractual agreements adopted by legal compliance support personnel responsible for overseeing ND sample and data collections. In the coming years, it will be critically important for the EPND consortium to develop contractual frameworks and legal accountability tools that reflect the views, concerns, and preferences of the legal compliance support personnel acting as veto players at the research institutions.

Another emerging requirement, particularly among data-providing institutions is that the data user is only authorised to access data in a secure processing environment and may not export data from that environment. Only the summary results of the analysis can be exported or downloaded by the user. The nature of the secure processing environment may differ; it may be on-premise (i.e., incorporated into the infrastructure of the cohort owner institution), an external platform provider designated and

vetted by the cohort or the EPND, or may refer more generically to any environment selected by the data user that meets certain standards or has obtained certain certifications.

Secure Processing Environments (SPEs)

Traditionally, cohorts sharing data, or data hubs storing data to support such sharing, have provided users access to the data by means of direct data transfers. This means that data may be processed by a large number of parties in many different environments, exposing the cohort to significant data security risks. Consequently, there is a growing interest in data access and analysis solutions that do not necessitate physical transfers of the data to the end-users (and their local processing environments).

One approach that has been considered as highly desirable in the context of biomedical research is federated data access and analysis, where the data remains stored with the institution that generated it, while being remotely queried for analysis by external users. However, in practice, fully federated approaches to biomedical data analysis (i.e., where multiple datasets are remotely quired simultaneously, in a distributed manner) are rare, owing to their technical limitations, as well as due to the fact that they do not necessarily resolve the fundamental legal challenges of data sharing.

In Europe, a notable example of a federated data analysis infrastructure is being implemented by the Personal Health Train (PHT) project, a digital research platform developed in the Netherlands. The PHT provides a suite of pre-defined data analysis tools and functionalities that researchers can choose from. Each data analysis tool is supported by a bespoke federated learning algorithm which allows querying all of the participating datasets simultaneously, without revealing data subjects' personal information to the user[11]. This approach to federated data analysis may offer increased speed and efficiency while at the same time minimising the risks of unintended disclosure of personal data. However, the approach is limited by the availability of predefined data analysis options researchers can choose from. Moreover, from the point of view of the GDPR, it makes no fundamental difference whether the party pursuing data analysis directly accesses the data, or instructs the data holder to perform the analysis (Thorogood et al. 2021). From the standpoint of the GDPR, in both cases, the external party is the controller for the analysis, and must be able to demonstrate its compliance with the principles of the GDPR.

Owing to these limitations to federated data *analysis* approaches, platforms are increasingly utilising "hybrid" approaches which combine elements of federated infrastructures with those of centralised solutions (i.e., where cohorts deposit their data in a central database). The principal example of such a hybrid infrastructure in the NDs space is the AD Workbench, which provides capabilities for centralised as well as federated or distributed querying/analysis[12]. There are multiple ways in which such infrastructures can be "hybrid". For example, they can enable centralised *discovery* and submission of *access requests*, while providing data *access* and/or *analysis* in a distributed manner,

[11] A more extensive description of the federated data analysis approach utilised by PHT can be found in the following document: https://distributedlearning.ai/documents/6/2020-11-12_DPIA.pdf

[12] An overview of the AD Workbench data access and analysis workflow can be found on the following page: https://www.alzheimersdata.org/-/media/files/addi/addi_workbench_welcome.pdf

via multiple instances of a Secure Processing Environment (SPE) implemented on-premise by the cohorts. Alternatively, access requests may be reviewed and approved by DACs of the cohorts individually, whereas access to the approved datasets be provided via a central SPE which also allows for centralised data analysis.

In principle, cohorts may decide to either implement instances of SPEs on premise, or they may outsource hosting to an external service provider, subject to technical constraints and the availability of resources. (For example, ADDI's AD Workbench can be adopted by cohorts under both implementation modalities). However, irrespective of the modality, the important consideration is that data access/analysis is provided within the secure and monitored environment. A key advantage of SPEs is increased trust among the cohorts, as well as other stakeholders, in the level of data security across the data lifecycle. In turn, this reduces the need for data users to report detailed information about security measures in their requests, or for access committees to review those measures. Although the adoption of SPEs does not completely obviate the need for Data Access/Use Agreements (e.g., user-side security (remote access), and prohibitions on unauthorised access and data exports will still need to be covered in agreements), it significantly reduces parties' reliance on contractual agreements as primary accountability tools.

A key challenge for SPEs is linking different data that live in different SPEs. Some of the interoperability issues can be addressed where a single platform provides SPE services to multiple cohorts, who each control their own secure cloud locker. Throughout the remaining funding period of the EPND IMI project, WP1 and WP2 will be working closely to address these challenges and implement ADDI's AD Workbench-supported SPE solutions in a technically interoperable and legally compliant manner.

Foster patient/research participant and public engagement

Increasingly, participatory models of governance are emerging as best practices in retrospective biomedical research. Participatory approaches to the governance of human sample and data collections aim to ensure that the research using these resources is aligned with the values and the reasonable expectations of the communities impacted by the research. These include patients and research participants who have contributed their samples/data, their family members, as well as the broader ND patient community and other stakeholders with a legitimate interest in advancing responsible scientific research into NDs.

Participatory models of governance vary in the extent and the nature of participation they afford to the relevant stakeholder groups. At the one end of the spectrum are approaches whereby stakeholders are meaningfully informed about how the community resources are utilised. These information-centric approaches prioritise transparency by ensuring that accurate and up-to-date information is provided to the community members (Ada Lovelace Institute 2021). At a minimum, the communication should describe the nature of the completed, ongoing, and planned research projects that make use of the community resources, with the level of detail and the means of communication being tailored to the needs of the target audience. However, information-centric models may not necessarily provide

community members with tools and mechanisms to directly influence day-to-day governance of the research resources.

At the other end of the spectrum are participatory models of governance aimed at empowering the key stakeholders by enabling them to decide how community resources are used (Ada Lovelace Institute 2021). In the case of data/sample donors and their family members, this can be accomplished by allowing them to decide whether their samples and data can be used in a particular research project. In recent years, various digitally enabled solutions, often described as "dynamic consent" tools, have been developed that allow research participants to not only determine the conditions under which their samples/data can be used, but also grant or restrict data and sample access to specific studies (DNV GL, Group Research and Development, Precision Medicine Program 2021; Mascalzoni et al. 2022).[13] Moreover, stakeholder empowerment in the governance context can also be achieved at a broader community level, whereby representatives of the key stakeholder groups are invited to participate in the governance process and are given a formal mandate to influence decision-making, including by vetoing decisions (Ada Lovelace Institute 2021).

One promising form of participatory approaches to the governance of community research resources is the consultative governance model. In this model, representatives of the key stakeholder groups such as patients, research participants and their family members are given an advisory role that is formally incorporated into the governance framework. These representatives are tasked with supporting the research organisation towards bringing the governance process into a closer alignment with the broader community values, while ensuring that the research projects using community resources respect the rights, interests, and reasonable expectations of the donors and their family members. In certain cases, the formal advisory role of the community representatives may explicitly incorporate the mandate to engage the organisation's REC and/or D(S)AC, as applicable. In this manner, the organisation can ensure that the voices of the key stakeholder groups are well-represented in the deliberations around whether community sample and data collections can be used in specific research projects (Thiel et al. 2014; Milne, Sorbie, and Dixon-Woods 2021).

Inclusion of community representatives in the governance process may be especially suitable in the context of ND research, where many members of the key stakeholder groups - such as ND patients and research participants - may not be able to directly represent their own interests. In this respect, the EPND should carefully identify candidates who are qualified to represent the communities most directly affected by the research facilitated through the EPND. Certain European patient advocacy groups with a well-documented history of advancing voices of their communities would be an ideal place for the EPND to look for qualified candidates. One such organisation is Alzheimer Europe,

[13] While the concept of digitally-enabled dynamic consent in biomedical research was introduced over a decade ago, initially it failed to garner widespread support within the medical research community. However, with the improving digital literacy of the general population, coupled with the ubiquity of digital applications, dynamic consent solutions are gaining greater acceptance in the biomedical research context. The potential utility of such dynamic consent tools is further highlighted by the controller's obligations vis-a-vis data subjects under the GDPR. When processing personal data based on consent (Art. 6(1)(a) and Art. 9(2)(a) GDPR), requesting study-specific consent from research participants through such tools may allow controllers to demonstrate their compliance with the consent requirements under the GDPR (see Annexes C-D for more information regarding controllers' transparency, information and consenting obligations vis-à-vis data subjects under different scenarios and GDPR legal bases).

whose European Working Group of People with Dementia has extensive experience representing the collective interests of Alzheimer's and other dementia patients.[14] However, as the EPND establishes its access governance framework, the EPND should seek to broaden its dialogue with the ND patient community, potentially engaging other organisations and advocacy groups to this end.

Conclusion: key challenges and considerations towards the EPND access governance framework

A critical unmet need in precision medicine research in NDs is scaling up sample and data access for research projects across networks of cohorts and biobanks. These networks are essential to increase the scale and statistical power of research analysis, as well as to identify rare associations relevant to biomarker discovery. However, access to samples and data across such networks can be complicated by different applicable access and use conditions, as well as the need to go through a different access process for each resource, including signing and respecting different access/use agreements.

On its face, efficient access to networks of databases and biobanks requires harmonisation of access policies, processes, agreements (and increasingly also the SPEs in which data are accessed) to increase predictability and consistency of outcomes. However, it is far from straightforward how the EPND could harmonise access policies and centralise access processes, especially when dealing with diverse resources from different countries, with disparate existing governance structures.

IT platforms can provide services to facilitate the discovery and access to samples and data. These platforms can also improve transparency over access and use conditions. Increasingly, it is being acknowledged that some form of "data intermediary" - beyond an IT platform - may be necessary to connect networks of multiple cohorts/biobanks to networks of potential data users. Not only may these research-facilitating intermediaries assist with harmonisation of sample/data governance across resources, but they may additionally centralise all or some aspects of responsibility for the access process, such as review by a central Data and Sample Access Committee (if legally feasible) and signing of a single Data Access/Use agreement. If successful, the EPND could become such an intermediary for the European ND research community, with AD Workbench serving as the EPND's IT platform (i.e., the Technical Hub).

First step towards building the EPND access governance framework would be a proposal aimed at establishing common principles to govern sample/data access. As a general principle, the EPND should seek to adopt a maximally harmonised access governance framework whereby the same rules, procedural steps, and access and use conditions apply to the different EPND collections. However, in order to accommodate the special needs of the cohorts - particularly in relation to fundamental access and use conditions that are legally or contractually binding on the cohort - limited exceptions should be permitted. The EPND will need to perform a comprehensive survey of the European ND cohorts' access governance frameworks in order to determine to what extent their core elements can be harmonised under an overarching, EPND access governance framework.

[14] For more information, see https://www.alzheimer-europe.org/membership (Accessed 14 January 2023)

Chapter 4: Sustainability and Exploitation perspective

Context and Background

Building an interoperable European platform for data and sample access to facilitate sharing and broader global reach provides an opportunity to answer a high unmet need, delivering enduring value to the research community and ultimately patients. Sustainability has been identified as a priority for IMI funded projects (Aartsen et al. 2018). A handful of white papers and publications in specialist journals tackle the issue of sustainability of Public Private Partnerships (PPPs), infrastructures, biobanks, and comparable initiatives (McQueen et al. 2014; Watson et al. 2014; Hellström, Wikström, and Eriksson 2021; Abdaljaleel, Singer, and Yong 2019; Sargsyan et al. 2015; Henderson, Simeon-Dubach, and Albert 2015). However, further research is required to understand the current landscape and specific challenges to be tackled to deliver a successful strategy to ensure EPND's persistence beyond the IMI funded period.

The objective of WP7 (sustainability) during year 1 was to deliver the first draft of the sustainability and exploitation strategy plan, a roadmap to guide the development of viable and scalable self-sustainable operating model for the EPND beyond year 5. The approach adopted to achieve this goal is based on the business plan methodology, tailored to the specific context of the EPND. As the first step, an initial set of attributes were gathered from a preliminary stakeholder survey and benchmarking exercise, with the aim to subsequently inform defining of the EPND as a differentiated offer. The conceptual framework for delineating and prioritising the relevant attributes was grounded in the target product profile (TPP) methodology commonly utilised in the pharmaceutical industry.

Year one research and findings

Year one efforts of WP7 were largely driven by fact-finding to identify the needs of key stakeholders (using a questionnaire and interviews) and mapping of the current competitive landscape (desktop benchmarking), with the results of both exercises reported and prioritised in a TPP framework. A detailed description of these activities and their outcomes have been reported in internal EPND documents. What follows below is a high-level overview of the preliminary insights generated by the consortium's sustainability work package during Year 1 of the IMI Project.

- **Key Stakeholders**

An initial series of workshops with members of WP7 and representatives of other WPs were held to identify EPND stakeholders. The key stakeholders identified were grouped in the following categories:

  - Users: community that will access resources (samples, data, services);
  - Depositors: cohort community that will contribute resources (samples and data);

- Advocates: community that will collaborate and champion the platform's mission (policy makers, regulators, and funding bodies) also acting as potential sponsors, where appropriate.

- **Users and depositor survey**

To initiate the mapping of the relevant unmet needs of the ND research community, the list of stakeholders was narrowed down to the segment that would have the closest and most immediate experience of the platform, namely the users and depositors. A set of users and depositors from the pharmaceutical industry and various academic institutions were asked to answer a short questionnaire online or to participate in a 30-minute interview to gather insights into how these groups accessed or shared data and samples as well as their frustrations, and expectations. The full outcome of this exercise, described in more detail in the first draft of the sustainability and exploitation plan (an internal project deliverable), is beyond the scope of this review. However, it is noteworthy to highlight that legal and ethical barriers to obtain samples/data in a timely fashion as well as frustrations around the scarcity of certain samples (e.g., CSF, or healthy controls) and lack of completeness, quality or outright absence of associated data were the most frequently cited hurdles. When it came to expectations, unsurprisingly, a single-point access of high-quality samples with rich associated information and pre-negotiated legal agreements to transfer materials and samples in a reasonable time frame scored high. It ought to be noted that this information was collected from a small sample of individuals and answers often varied depending on the role of the respondent (e.g., data analysist, cohort manager, wet lab researcher, etc). A follow-up survey encompassing a wider range of stakeholders and roles would be needed to minimise bias.

- **Desktop benchmarking**

The main objective of the initial desktop benchmarking research was to identify existing organisations, platforms, and initiatives with objectives analogous to the EPND, to gain insights on common/best practices as well as potential gaps in the existing offers. This landscape mapping activity was not limited to Europe, as the aim was to identify a comprehensive set of features, attributes, and/or services offered by relevant data and sample-sharing initiatives worldwide. The feasibility of adopting a particular feature/attribute/service by the EPND (including, based on legal and regulatory constraints uniquely applicable to the European context) will be explored at a later stage by WP7 in collaboration with WPs 1 and 2, with findings to be reported in the final EPND White Paper, (forthcoming in October 2025).

Preliminary search for the initiatives with objectives comparable to that of the EPND resulted in the identification of 21 initiatives deemed relevant for the purposes of this review, and subsequently analysed in greater detail. The main finding of this initial research on 21 initiatives was the fragmentation of the landscape and the lack of clear "success formula". This led to the conclusion that EPND has the potential to fill a gap in the current offering by proposing a single point of access to a comprehensive well organised catalogue of existing and future resources in the ND space. However, competition for users, depositors and advocates is likely to be strong and poses a risk to the EPND's

aims. The key elements to mitigate this risk and ensure sustainability will be to develop a differentiated offering and adopt an effective communication strategy that increases EPND's visibility (see for example (Klinger et al. 2021)). In year two, as part of the more in-depth benchmarking exercise, WP7 intends to develop objective metrics for assessing and comparing relevant initiatives across several key criteria defining "what success looks like". Aspects where no clear conclusions could be drawn from the desktop exercise, including, for example, the types of legal entities and defining characteristics of governance models supporting the relevant initiatives, will also be closely re-examined during the year 2 benchmarking research.

- **TPP**

To consolidate the insights gathered from the fact-finding process and derive a unique value proposition for the EPND, a strategic decision was made to adopt a TPP framework to inform the subsequent development of EPND's differentiated offering. TPPs are tools used in the pharmaceutical industry to communicate the differentiating features and target characteristics of new marketable assets to different stakeholders (Singh 2018; Breder, Du, and Tundall 2017; Tyndall, Du, and Breder 2017). The first draft of EPND's TPP focussed on attributes and sub-attributes relating to the data/sample discovery, access, and analysis functionalities as well as their associated processes and workflows to be implemented by the EPND consortium during the IMI funding period. This initial TPP will be a living document, a tool to guide the definition of a target product with key technical specifications, and will guide subsequent years' efforts to deliver a draft of a broader set of critical characteristics for the EPND's differentiated offering.

Preliminary conclusions and future perspectives

The initial assessment of the competitive landscape, coupled with a preliminary exploration of the users'/depositors' needs and expectations has led to the following conclusions:

- Continuous and adaptative assessment of users' and depositors' needs, alongside stakeholder prioritisation ought to be at the core of strategic decisions for the EPND consortium. The methods to consult the different groups will evolve and become more targeted as the nature of the offering evolves. Stakeholders' needs should be examined in light of, and balanced against, the existing constraints under the applicable legal frameworks (such as the GDPR and its national implementations), as well as technical feasibility and cost considerations.
- The large number of platforms, initiatives and organisations with overarching objectives analogous to those of the EPND clearly signals an opportunity for the EPND consortium to consolidate the fragmentary landscape by providing a differentiated offer that better serves the needs of the target community. Understanding and addressing the core challenges behind the unmet needs requires a close cooperation among EPND partners, ideally complemented by coordination and alignment with external parties with valuable experience in the context of sample and data sharing.
- Central to developing services that are attractive and contribute to the sustainability of the platform is to ensure that evidence generated using EPND resources is fit for purpose to support the efforts of the pharmaceutical industry to bring new drugs and diagnostic tools to the market.

This will be a key priority for EPND's Regulatory group in the years ahead. (Some of the most important regulatory support tools available in Europe in the context of biomarker/method discovery, qualification and validation are described in Annex G). The EPND consortium will need to strategically position itself as a leading network of research experts in the ND space, acting as a bridge between the ND translational science community and European regulators. Advisory services aimed at pharma players in relation to biomarker and/or method discovery, qualification and validation, will also be explored during the remaining years of the IMI project.

- Once the blueprint for the technical offering and an initial portfolio of desirable services is established, the evidence-based decision-making process will be implemented to derive suitable operational and governance models as well as funding strategies. A close collaboration among EPND consortium partners across different WPs is key to success.

Chapter 5: General Conclusions and Recommendations

In this paper, we have explored various strategic challenges facing European biomedical research platforms and initiatives, including the EPND. As the document was prepared during Year One of a 5-year (November 2021 – October 2026) IMI Project, it places a particularly heavy emphasis on the most immediate challenges associated with bringing existing European biomedical sample and data collections into the EPND, making them discoverable, accessible, and reusable for retrospective research purposes. More specifically, the document provides an in-depth analysis of the fundamental legal, policy, and governance barriers to enabling cross-border retrospective biomedical research in Europe. These barriers, outlined in Chapters 2 (Data Protection View) and 3 (Governance View) – and examined in greater detail throughout Annexes A-F -, are not unique to the EPND; they apply to most European biomedical research platforms, large consortia, and other European initiatives aimed at facilitating retrospective biomedical research involving the use of existing sample and/or data collections. As such, the analysis presented in this document has implications beyond the EPND, being of relevance to the broader European biomedical research community.

However, it is worth highlighting that the legal, policy, and governance barriers to sample and data sharing constitute only part of the strategic challenges to be overcome by the EPND. . Additionally, as the project matures, the EPND consortium will need to develop strategies to address regulatory aspects of bringing novel biomarkers and methods to the market. The utility of research-ready sample and data collections is determined, in considerable measure, by the extent to which they can be used to discover, qualify, and validate promising biomarkers and methods aimed at predicting, diagnosing, or treating NDs. To this end, the EPND consortium will make use of the available support tools to engage European regulators such as the European Medicines Agency to better understand the EPND's potential role in the context of biomarker/method development based on EPND sample and data collections (see Annex G: Regulatory guidance – Support tools offered by the European Medicines Agency for Innovations for an overview of the relevant EMA support tools).

What follows in the concluding sections below is a set of strategic recommendations for the EPND consortium based on the analysis presented in the previous chapters of the document. These recommendations are intended to inform the development and refinement of the EPND's strategic roadmap during Phase II of the IMI project (Years 2 – 5, ending in October 2026), as well for the years after the project's completion. The recommendations are grouped under three themes, corresponding to the three core chapters of the document: i) GDPR/Data Protection, ii) Governance; and iii) Long-term sustainability aspects.

GDPR / Data Protection

1. The roles under the GDPR (Controller/Joint Controller/Processor) should be defined in a granular manner and reflected in the EPND's contractual agreements among the relevant consortium members, data providers (cohorts) and data users.

Where personal data are processed via the EPND's Technical Hub – i.e., as part of data contribution (by a cohort), or data discovery, access, and analysis (by a data user) – the parties supporting the operations of the EPND's Technical Hub will be acting as processors under the GDPR. In their capacity as operators of the EPND's Technical Hub, these parties will be executing specified instructions of the controllers. This allocation of the GDPR roles is in line with the analysis presented in Chapter 2 and Annex B of this document, and summarised in Table 5 of Annex B.

At present, as the EPND lacks the status of a legal entity, the controllers for the processing operations carried out by the EPND's Technical Hub will primarily be cohorts and data users. Accordingly, in the short term, the EPND consortium's internal GDPR expertise should be directed primarily at supporting cohorts and data users towards demonstrating their GDPR compliance as controllers (see also the next recommendation). During the IMI funding period (2021 – 2026), the EPND consortium has adequate internal resources and expertise to provide ad-hoc compliance guidance and support to both cohorts and data users. However, to ensure the sustainability of compliance efforts beyond 2026, the EPND consortium will be exploring ways to internalise the GDPR compliance burden associated with the role of a controller in a more permanent manner. The ambition of the consortium is to do so by establishing the EPND as a legal entity by 2026 and ensuring that this entity has a clear legal mandate to become a controller in relation to the processing of patient and research participant data provided by European ND cohorts.

During the IMI funding period, should situations arise where a particular EPND consortium partner becomes a controller for one or more processing operations, this partner will need to enter in a suitable agreement (e.g., a joint controller agreement) with the relevant party, such as the cohort or the data user. The same is true where one of the entities operating the EPND's Technical Hub seeks to undertake further processing (in the sense of Art. 5(1)(b) GDPR) of the data contributed by the EPND cohorts, provided that such further processing is for a compatible purpose and meets other criteria for lawful processing under the GDPR.

Additionally, it is worth highlighting that the aforementioned considerations are limited to the processing of personal data relating to patients and research participants enrolled in ND clinical and research cohorts. With respect to other categories of data subjects, the EPND consortium partners operating components of the Technical Hub will be routinely acting as controllers. These additional categories of data subjects include person(s) responsible for managing cohort-related information in the EPND system (e.g., Principal Investigators of the cohorts) and researchers accessing cohorts' data collections via the EPND user portal. Insofar as the EPND (or the legal entity/entities operating the EPND's Technical Hub, as applicable) acts as a controller under the GDPR, it must ensure it complies with Art. 5(1) principles of the GDPR, as well as takes necessary steps to be able to demonstrate this compliance in accordance with the Accountability Principle.

2. The EPND should provide ad-hoc GDPR compliance support services to cohorts and data/sample users. The scope of such services should reflect the allocation of GDPR roles (controller/processor) among the cohort, the platform, and the data user.

Cohorts and data users utilising EPND's services could benefit from the availability of GDPR compliance advisory support from the platform. The EPND has the internal expertise (WP2 – Legal

and Ethical Regulations) to provide such support and intends to establish a dedicated "ELSI helpdesk" as one of the project deliverables.

It must be highlighted, however, that the scope of the GDPR compliance support should reflect the parties' roles under the GDPR. For example, the EPND, where acting as a processor with respect to cohorts' data collections, cannot offer extensive guidance on compliance with the principles of the GDPR, which is the sole responsibility of controllers (i.e., cohorts and data users). As an illustrative example highlighting the limitations of EPND's advisory role in relation to compliance with the GDPR principles, consider the principle of lawfulness. In order for processing to be lawful under the GDPR, it should be grounded in a valid Art. 6(1) GDPR legal basis and further legitimised under a suitable Art. 9(2) exemption. However, the EPND may have no means to assess whether the legal basis selected by the controller is valid. For example, a controller may state that it is processing personal data based on Art. 6(1)(e) in conjunction with Art. 9(2)(j) GDPR. The validity of this legal basis depends on whether the controller is subject to a (national) law giving it a mandate to process special categories of personal data for research purposes. Consequently, verifying the validity of the legal basis (and, by extension, ascertaining the controller's compliance with the GDPR principle of lawfulness) requires intimate familiarity with the applicable national legal framework in the controller's country, which is beyond the remit of the EPND's ELSI helpdesk.

More generally, by helping controllers to comply with the principles of the GDPR, the EPND would be effectively advising controllers on various crucial aspects of defining the nature of processing. This would expose the EPND to a significant risk of becoming a joint controller for various relevant processing operation(s), contravening the spirit of the previous recommendation.

In view of these considerations, the EPND's ELSI helpdesk should generally refrain from providing advisory services in relation to the GDPR compliance obligations that uniquely apply to controllers. Instead, the helpdesk should support controllers by identifying implementation solutions that meet controllers' GDPR compliance requirements, as interpreted by controllers themselves. For example, cohorts interpreting their obligations under the transparency principle in a manner that requires them to periodically recontact their data subjects, could be offered additional privacy notification services by the EPND aimed at supporting the controller's transparency obligations. Similarly, the helpdesk could act as a point of liaison between controllers and the EPND personnel operating the platform's technical infrastructure to help the controller meet its obligations vis-à-vis its data subjects, including exercising the data subjects' rights under Chapter III GDPR. Finally, the EPND ELSI Helpdesk should be prepared to assist the controller in meeting certain procedural and documentation requirements, such as performing a Data Protection Impact Assessment, upon the controller's request.

3. The EPND should explore ways to strategically position itself in view of the evolving regulatory landscape, especially in relation to the forthcoming European Health Data Space (EHDS) Regulation.

As the European legislators are becoming increasingly aware of the GDPR-related legal challenges to the sharing of personal data for research purposes, they can be reasonably expected to seek legislative changes aimed at remedying these issues. Revisions to the existing European data protection legal framework in Europe are expected during the EPND's IMI funding period (2021 - 2026), particularly in view of the upcoming comprehensive evaluation report of the GDPR due in May 2024, in accordance with Art. 97 GDPR. It is crucial for the EPND consortium to stay abreast of the relevant legislative developments and promptly assess the expected impact of the forthcoming legislative measures on the governance and sustainability of the EPND.

To that end, the EPND consortium will be closely following the progress of the EHDS Regulation proposal, while exploring ways to engage the EU legislators to ensure that the views of the European ND research community are heard in the legislative discussions. In parallel, the EPND consortium will be monitoring the implementation of other newly adopted relevant European laws and legal frameworks such as the Data Governance Act, and the legal instruments established under the European Digital Policy Programme 2030. The implications of these legislative development for the European ND research community will be extensively evaluated by the EPND consortium through its dedicated ELSI work package, WP2.

Governance

1. To enable meaningful, record-level discovery of the data/sample collections of EPND cohorts, the EPND's discovery tool must accurately capture access and use conditions in a comprehensive manner.

Our analysis and practical experience with ND cohorts suggests that accurately capturing data/sample access and use conditions is a significant challenge (see Chapter 3 and Annexes E-F). Cohorts with highly complex access and use conditions may be unwilling - or, in certain cases, unable - to communicate these conditions in a generic manner. Comprehensive (ELSI) Metadata structures/ontologies, and digital solutions implementing them are valuable tools for capturing this requisite information. However, representatives of cohorts (e.g., PIs) inputting this information into digital forms could be prone to errors and may incorrectly report crucial sample/data access and use conditions. Complementary approaches to supporting cohorts in this regard (potentially via the EPND ELSI Helpdesk) will need to be explored.

In addition to capturing detailed access and use conditions per cohort, the EPND should also accurately differentiate between fundamental and negotiable conditions. As a general rule, access and use conditions concerning permissible data/sample uses (purposes) and the types of entities that are allowed to access the resource, tend to be fundamental, as they are often dictated by the applicable laws, institutional policies, and consent forms that are binding on the cohort. By contrast, certain prerequisites to access (e.g., obtaining a REC approval or conducting a DPIA by the data recipient) may be either fundamental or negotiable, depending on the cohort and the context of intended data/sample use. Finally, publication co-authorship, financial compensation in return to access, as well

as other sustainability and business considerations generally tend to be negotiable, as cohorts usually have greater freedom in deciding on these matters.

The EPND consortium needs to overcome cohorts' resistance to defining its access and use conditions in a generic manner. To achieve this goal, the consortium must educate cohort owners on the importance of having clear-cut access and use conditions for an efficient resource discovery functionality and a transparent EPND access policy. Some cohorts may need to be additionally reassured that by providing detailed access and use conditions, they will not relinquish their authority over the future access decisions.

2. The EPND's Access Governance Framework, to be developed by WP2 during Phase II of the IMI project (years 2-5, ending in October 2026), should maximally streamline and standardise the process of data/sample access

(Subject to constraints such as compliance with national laws, as well as bespoke requirements from individual institutions managing neurodegenerative disease cohorts).

2.1. To the extent feasible, WP2 must standardise agreements, contractual tools, and policies. Where a complete standardisation is not possible, agreements should use modular clauses to accommodate specific needs and requirements of the intended signatories.

These agreements/tools will include: a standard data/sample transfer agreement; a data/sample use agreement; EPND publication policy, and IP rights policy (the latter to be developed in collaboration with WP7 – Sustainability). The complete set of the essential agreements and policies under the EPND Access Governance Framework will be identified by WP2, alongside defining a roadmap for developing these agreements/policies.

2.2. To the extent feasible, the EPND should streamline the process whereby external researchers request access to data and/or samples from multiple cohorts.

This task will require a close collaboration between WPs 1 (Establishment of the Platform) and 2. Governance-related aspects (e.g., preparing a standard data/sample access request form) of this work will be driven by WP2, whereas technical implementation (e.g., submission of the online form via the EPND website or through the user portal) will be decided by WP1.

2.3. During early stages of Phase II of the IMI project (M13 - M18), WP2 should further assess the feasibility of a streamlined access request review process, including by the EPND's central Data and Sample Access Committee (DSAC).

Our preliminary interactions with cohorts suggest that a fully centralised access request review process will be challenging to implement. Cohorts often prefer (some are required) to ensure that each access request is reviewed and approved by the cohort's own institutional Data (and Samples) Access Committee. However, given the significant potential efficiency gains associated with streamlining this process, it is warranted to further explore the feasibility of a central DSAC, alongside other modalities

of centralising or streamlining the access request review process. The outcomes of this feasibility assessment will be described in an internal EPND report (due in October 2023), and subsequently communicated with the wider audience by the EPND consortium.

As discussed in Chapter 3, even in the absence of fully centralised access *decisions*, there will be opportunities for the EPND to streamline certain aspects of the access *process*, which spans the following steps: submission of an access request (by an external researcher); review of the request (by the competent access committee(s)); communication of the decision(s) to the requestor; and the act of providing access. The optimal approaches for streamlining specific technical and operational aspects of these steps will be further explored by WP2 in the coming months.

3. The design and oversight of the EPND Governance Framework should be a participatory process involving representatives of the key stakeholder groups. At a minimum, representatives of patients with neurodegenerative diseases must be meaningfully involved.

While much of the participatory governance-related work around the EPND will naturally fall under the remit of EPND's external advisory boards (forthcoming during the IMI Project), the consortium must ensure that at any given time, patients and/or their representatives are meaningfully involved. Participatory and inclusive approaches to the governance of the EPND are essential for fostering trust vis-a-vis patient communities and the general public alike. While at this early stage in the project the present document does not prescribe a particular participatory governance model, Chapter 3 of the document provides an overview of the participatory governance approaches commonly employed by community-oriented biobanks, data-sharing initiatives, and platforms.

To further enhance the public's trust in the governance of the EPND, the consortium should make its governance structure and processes maximally transparent via its website and other means of public communication.

4. The EPND may decide to establish a Central Research Ethics Committee that will be available to review prospective data users' research proposals.

A central REC will be helpful in those cases where the institution responsible for the cohort's collection requires a pre-existing ethics approval as a non-negotiable access and use condition, whereas the prospective user is unable to obtain such an approval due to differences in the applicable national laws and/or institutional policies. However, the feasibility of such a central Research Ethics Committee is currently unclear and should be further assessed by the consortium members. For example, organisationally, in order for this committee to qualify as a legally competent REC, it will likely need to be embedded within an external research institution. Alternatively, an existing REC within one of the EPND partner institutions can be designated as the EPND's Central REC, subject to legal and financial feasibility assessment.

Long-Term Sustainability Aspects

1.The EPND consortium needs to devise a concrete strategy for attracting members of the neurodegenerative diseases research community through appropriate incentives. The strategy should particularly prioritise cohorts (data and sample depositors).

Apart from well-documented traditional factors discouraging cohorts from data and sample sharing (including administrative burden, need for ethics approvals, lack of resources, motivation to retain exclusive access to own data), European cohorts are now also subject to strict GDPR compliance obligations, further discouraging them from engaging in data sharing with external partners. As such, European ND cohorts will likely require strong incentives to participate in the EPND and actively share their collections through the platform.

In the coming years, WP7 will seek to systematically identify, characterise, and prioritise such incentives. This work will be partly informed by the outcomes of a forthcoming stakeholder study that will focus on prospective data/sample depositors.

2. Further benchmarking of the existing platforms and initiatives is needed in order to identify opportunities for the EPND to provide a differentiated offering.

As evidenced by the preliminary review of the platforms and initiatives comparable to the EPND, there are a large number of public as well as private initiatives aimed at facilitating access to ND sample and data collections. The conclusions stemming from this finding, as also highlighted in Chapter 4, are two-fold. First, there is a clear interest among the ND research community in the solutions that enhance the ease of access to existing collections. Second, the current landscape is highly fragmented, with no single platform or initiative holding a dominant position. Both considerations suggest that there are opportunities for a new entrant such as the EPND to establish itself in this field, provided that the EPND succeeds at differentiating its services from those of its competitors. In order to identify concrete strategies for providing a differentiated offer, WP7 will perform a more extensive analysis of the existing sample/data sharing platforms and initiatives, combining the resultant insights with the findings of stakeholder studies.

3. The EPND should maximise the transparency in relation to the availability of samples and data.

Consistent with the analysis by WP2 described in Chapter 3, the preliminary stakeholder study conducted by WP7 found that prospective data/sample users are often frustrated by the lack of adequate information regarding the availability of samples and data. This includes the quality as well as the quantity of existing samples/data, alongside the access and use conditions the users are expected to meet. In practice, potential users seeking access to ND collections often learn about the unavailability of samples/data for the intended research use only after they have submitted a detailed access request, resulting in significant operational inefficiencies and frustrating experiences. The EPND should therefore prioritise maximising transparency in relation to the availability of samples and data (see also recommendation 1 under Governance).

4. The Regulatory group of the EPND consortium, in collaboration with other partners, should establish the framework for assessing the utility of prospective EPND sample and data collections in the context of discovery, qualification, and validation of novel biomarkers

Over the remaining 4 years of the IMI project, the EPND consortium will need to develop internal quality standards and metrics to better understand whether and to what extent the evidence generated based on EPND collections can be used to help bring novel biomarkers/methods to the market. To this end, the EPND should take advantage of the available support tools to engage European regulators to formulate a clear regulatory strategy (See Annex G: Regulatory guidance – Support tools offered by the European Medicines Agency for Innovations).

Concluding Remarks

This document was prepared by the EPND consortium during Year 1 of the 5-year IMI-funded project (November 2021 – October 2026). As such, it focuses on the key strategic challenges (particularly legal, governance, regulatory, and sustainability) facing the consortium. However, owing to the generic nature of these challenges, many issues discussed herein are of considerable relevance to other European research initiatives and platforms aimed at facilitating the exchange of biomedical samples and data for research purposes. The conclusions and recommendations discussed in this document will be revised and updated at the end of year two of the IMI project (October 2023), in view of the latest information. The updates to be described (initially in an internal report) will cover, among other content elements, the following: additional information regarding the feasibility of the EPND DSAC, a more detailed roadmap for building EPND's governance framework, as well as more nuanced insights into the legal and legislative roadmap bearing on the EPND's objectives.

The final EPND White Paper is planned for Month 48 of the IMI project (October 2025). This forthcoming White Paper will focus on the implementation aspects specific to the EPND, building on the experiences of the consortium over the initial four years of the project. More specifically, it will detail the implementation of the ADDI's AD Workbench in the European context, report on the progress concerning the development of EPND's governance framework and provide an up-to-date analysis of the changing European legal landscape, with a focus on the implications for the EPND. The final EPND White Paper will also incorporate more extended chapters addressing the sustainability and regulatory strategies for years beyond the IMI funding period.

References

Aartsen, Wendy, Paul Peeters, Scott Wagers, and Bryn Williams-Jones. 2018. “Getting Digital Assets from Public–Private Partnership Research Projects through ‘The Valley of Death,’ and Making Them Sustainable.” *Frontiers in Medicine* 5. https://www.frontiersin.org/articles/10.3389/fmed.2018.00065.

Abdaljaleel, Maram, Elyse J. Singer, and William H. Yong. 2019. “Sustainability in Biobanking.” In *Biobanking*, edited by William H. Yong, 1897:1–6. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-8935-5_1.

Ada Lovelace Institute. 2021. “Participatory Data Stewardship: A Framework for Involving People in the Use of Data.” https://www.adalovelaceinstitute.org/wp-content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf.

Amur, Shashi, L. LaVange, I. Zineh, S Buckman-Garner, J Woodcock. 2015. "Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization" *Clinical Pharmacology & Therapeutics*. 98: 34-46. https://doi.org/10.1002/cpt.136

Artene, Stefan-Alexandru, Marius Eugen Ciurea, Stefana Oana Purcaru, Daniela Elise Tache, Ligia Gabriela Tataranu, Mihaela Lupu, and Anica Dricu. 2013. “Biobanking in a Constantly Developing Medical World.” *The Scientific World Journal* 2013 (September): e343275. https://doi.org/10.1155/2013/343275.

Article 29 Data Protection Working Party. 2014. “Opinion 05/2014 on Anonymisation Techniques.” Opinion 05/2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

———. 2017. “Guidelines on the Right to Data Portability.” WP260rev.01. https://ec.europa.eu/newsroom/article29/items/611233/en.

———. 2018a. “Guidelines on Consent under Regulation 2016/679.” WP259 rev.01. https://ec.europa.eu/newsroom/article29/items/623051.

———. 2018b. “Guidelines on Transparency under Regulation 2016/679.” WP260rev.01. https://ec.europa.eu/newsroom/article29/items/622227/en.

Bak, Marieke A. R., M. Corrette Ploem, Hakan Ateşyürek, Marieke T. Blom, Hanno L. Tan, and Dick L. Willems. 2020. “Stakeholders’ Perspectives on the Post-Mortem Use of Genetic and Health-Related Data for Research: A Systematic Review.” *European Journal of Human Genetics* 28 (4): 403–16. https://doi.org/10.1038/s41431-019-0503-5.

Bakker, Elisabeth, Natalie M. Hendrikse, Falk Ehmann, Daniëlla S. van der Meer, Jordi Llinares Garcia, Thorsten Vetter, Viktoriia Starokozhko, Peter G.M. Mol. 2022. "Biomarker Qualification at the European Medicines Agency: A Review of Biomarker Qualification Procedures From 2008 to 2020". *Clinical Pharmacology & Therapeutics*, 112: 28-30. https://doi.org/10.1002/cpt.2554

Barbosa, Carla, and Andreia Andrade da Costa. 2021. “Biobanks and GDPR: A Look at the Portuguese Panorama.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:345–62. Law, Governance and Technology Series. Springer.

Bauermeister, Sarah, Joshua R Bauermeister, Ruth Bridgman, Caterina Felici, Mark Newbury, Laura North, Christopher Orton, et al. 2021. “Research-Ready Data for Multi-Cohort Analyses: The Dementias Platform UK (DPUK) C-Surv Data Model.” Preprint. In Review. https://doi.org/10.21203/rs.3.rs-937113/v1.

Becker, Regina, Davit Chokoshvili, Giovanni Comandé, Edward S. Dove, Alison Hall, Colin Mitchell, Fruzsina Molnár-Gábor, Pilar Nicolàs, Sini Tervo, and Adrian Thorogood. 2022. “Secondary Use of Personal Health Data: When Is It ‘Further Processing’ Under the GDPR, and What Are the Implications for Data Controllers?” *European Journal of Health Law*, August, 1–29. https://doi.org/10.1163/15718093-bja10094.

Becker, Regina, Adrian Thorogood, Jasper Bovenberg, Colin Mitchell, and Alison Hall. 2022. “Applying GDPR Roles and Responsibilities to Scientific Data Sharing.” *International Data Privacy Law* 12 (3): 207–19. https://doi.org/10.1093/idpl/ipac011.

Befring, Anne Kjersti. 2021. “Norwegian Biobanks: Increased Complexity with GDPR and National Law.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, edited by Santa Slokenberga, Olga Tzortzatou, and Jane Reichel, 323–44. Law, Governance and Technology Series. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-49388-2_18.

Bergeron, Julie, Dany Doiron, Yannick Marcon, Vincent Ferretti, and Isabel Fortier. 2018. “Fostering Population-Based Cohort Data Discovery: The Maelstrom Research Cataloguing Toolkit.” *PLOS ONE* 13 (7): e0200926. https://doi.org/10.1371/journal.pone.0200926.

Birkenbihl, Colin, Sarah Westwood, Liu Shi, Alejo Nevado-Holgado, Eric Westman, Simon Lovestone, and Martin Hofmann-Apitius. 2021. “ANMerge: A Comprehensive and Accessible Alzheimer’s Disease Patient-Level Dataset.” *Journal of Alzheimer’s Disease* 79 (1): 423–31. https://doi.org/10.3233/JAD-200948.

Breder, Christopher, Wenny Du, and Adria Tyndall. 2017. "What’s the Regulatory Value of a Target Product Profile?" *Trends in Biotechnology* 35 (7): 576-579. https://doi.org/10.1016/j.tibtech.2017.02.011

CCMO - Central Committee on Research Involving Human Subjects. n.d. “Non-WMO Research.” Accessed January 14, 2023. https://english.ccmo.nl/investigators/additional-requirements-for-certain-types-of-research/non-wmo-research/file-research.

Chen, Jiahong, Edward S. Dove, and Himani Bhakuni. 2021. “Explicit Consent and Alternative Data Protection Processing Grounds for Health Research.” Preprint. SocArXiv. https://doi.org/10.31235/osf.io/4fdsj.

Christofidou, Maria, Nathan Lea, and Pascal Coorevits. 2021. “A Literature Review on the GDPR, COVID-19 and the Ethical Considerations of Data Protection During a Time of Crisis.” *Yearbook of Medical Informatics* 30 (01): 226–32. https://doi.org/10.1055/s-0041-1726512.

Cool, Alison. 2019. “Impossible, Unknowable, Accountable: Dramas and Dilemmas of Data Law.” *Social Studies of Science* 49 (4): 503–30. https://doi.org/10.1177/0306312719846557.

Dankar, Fida K., Marton Gergely, and Samar K. Dankar. 2019. “Informed Consent in Biomedical Research.” *Computational and Structural Biotechnology Journal* 17 (January): 463–74. https://doi.org/10.1016/j.csbj.2019.03.010.

“Decreto Legislativo 30 giugno 2003, n. 196 ‘Codice in materia di protezione dei dati personali.’” 2003. https://archivio.pubblica.istruzione.it/amministrazione/allegati/dlg300603.pdf.

Devriendt, Thijs, Clemens Ammann, Folkert W. Asselbergs, Alexander Bernier, Rodrigo Costas, Matthias G. Friedrich, Josep L. Gelpi, et al. 2021. “An Agenda-Setting Paper on Data Sharing Platforms: EuCanSHare Workshop.” *Open Research Europe* 1 (July): 80. https://doi.org/10.12688/openreseurope.13860.1.

Devriendt, Thijs, Mahsa Shabani, and Pascal Borry. 2021. “Data Sharing in Biomedical Sciences: A Systematic Review of Incentives.” *Biopreservation and Biobanking* 19 (3): 219–27. https://doi.org/10.1089/bio.2020.0037.

———. 2022. “Policies to Regulate Data Sharing of Cohorts via Data Infrastructures: An Interview Study with Funding Agencies.” *International Journal of Medical Informatics* 168 (December): 104900. https://doi.org/10.1016/j.ijmedinf.2022.104900.

Dimitrova, Diana. 2021. “The Rise of the Personal Data Quality Principle. Is It Legal and Does It Have an Impact on the Right to Rectification?” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3790602.

DNV GL, Group Research and Development, Precision Medicine Program. 2021. “Dynamic Consent in Clinical Genetics: Implementation Barriers.” https://www.dnv.com/research/healthcare-programme/dynamic-consent-whitepaper.html.

Erdos, David. 2021. “Identification in EU Data Protection Law.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3761068.

European Commission Directorate-General for Health and Food Safety. 2022. “Proposal for a Regulation - The European Health Data Space.” https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en.

European Data Protection Board. 2019. “Opinion 3/2019 Concerning the Questions and Answers on the Interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR).” 3/2019. https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers_en.

———. 2020. “Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0.” 4/2019. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf.

———. 2021a. “Guidelines 01/2022 on Data Subject Rights - Right of Access.” 01/2022. https://edpb.europa.eu/system/files/2022-01/edpb_guidelines_012022_right-of-access_0.pdf.

———. 2021b. “EDPB Document on Response to the Request from the European Commission for Clarifications on the Consistent Application of the GDPR, Focusing on Health Research.” 07/2020. https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en.

———. 2021c. “Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR.” 07/2020. https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-072020-concepts-controller-and_en.

Finck, Michèle, and Frank Pallas. 2020. “They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR.” *International Data Privacy Law* 10 (1): 11–36. https://doi.org/10.1093/idpl/ipz026.

Freeberg, Mallory Ann, Lauren A Fromont, Teresa D’Altri, Anna Foix Romero, Jorge Izquierdo Ciges, Aina Jene, Giselle Kerry, et al. 2022. “The European Genome-Phenome Archive in 2021.” *Nucleic Acids Research* 50 (D1): D980–87. https://doi.org/10.1093/nar/gkab1059.

Frey, Anna-Lena, Michael Karran, Rafael C. Jimenez, James Baxter, Michael Adeogun, Niranjan Bose, Dennis Chan, et al. 2019. “Harnessing the Potential of Digital Technologies for the Early Detection of Neurodegenerative Diseases (EDoN).” Preprint. Open Science Framework. https://doi.org/10.31219/osf.io/u49z5.

Gefenas, Eugenijus, J. Lekstutiene, V. Lukaseviciene, M. Hartlev, M. Mourby, and K.Ó Cathaoir. 2022. “Controversies between Regulations of Research Ethics and Protection of Personal Data: Informed Consent at a Cross-Road.” *Medicine, Health Care and Philosophy* 25 (1): 23–30. https://doi.org/10.1007/s11019-021-10060-1.

Gentile, Giulia, Valentina La Cognata, and Sebastiano Cavallaro. 2021. “The Contribution of CNVs to the Most Common Aging-Related Neurodegenerative Diseases.” *Aging Clinical and Experimental Research* 33 (5): 1187–95. https://doi.org/10.1007/s40520-020-01485-4.

Georgiou, Dimitra, and Costas Lambrinoudakis. 2021. “Data Protection Impact Assessment (DPIA) for Cloud-Based Health Organizations.” *Future Internet* 13 (3): 66. https://doi.org/10.3390/fi13030066.

Giannella, Emilia, Valentino Notarangelo, Caterina Motta, and Giulia Sancesario. 2021. “Biobanking for Neurodegenerative Diseases: Challenge for Translational Research and Data Privacy.” *The Neuroscientist*, August, 107385842110366. https://doi.org/10.1177/10738584211036693.

Hallinan, Dara. 2020. “Broad Consent under the GDPR: An Optimistic Perspective on a Bright Future.” *Life Sciences, Society and Policy* 16 (1): 1. https://doi.org/10.1186/s40504-019-0096-3.

Hartlev, Mette. 2021. “Balancing of Individual Rights and Research Interests in Danish Biobank Regulation.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:215–26. Law, Governance and Technology Series. Springer.

Haslacher, Helmuth, Michaela Bayer, Heidelinde Fiegl, Marlene Gerner, Philipp Hofer, Melanie Korb, Sabrina Neururer, et al. 2019. “Quality Management at the National Biobanking Level – Establishing a Culture of Mutual Trust and Support: The BBMRI.at Example.” *Clinical Chemistry and Laboratory Medicine (CCLM)* 57 (12): e301–5. https://doi.org/10.1515/cclm-2019-0491.

Hellström, Magnus, Kim Wikström, and Kent Eriksson. 2021. “Sustainable Infrastructure Projects: Systemic versus Traditional Delivery Models.” *Sustainability* 13 (11): 6273. https://doi.org/10.3390/su13116273.

Henderson, Marianne, Daniel Simeon-Dubach, and Monique Albert. 2015. “Finding the Path to Biobank Sustainability Through Sound Business Planning.” *Biopreservation and Biobanking* 13 (6): 385–86. https://doi.org/10.1089/bio.2015.29039.mh.

Information Commissioner’s Office (ICO). n.d. “Legitimate Interests.” Guide to the General Data Protection Regulation (GDPR). Accessed January 13, 2023. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/.

Kamara, Irene, and Paul De Hert. 2018. “Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3228369.

Kindylidi, Iakovina, and Inês Antas de Barros. 2021. “AI Training Datasets & Article 14 GDPR: A Risk Assessment for the Proportionality Exemption of the Obligation to Provide Information.” *Law, State and Telecommunications Review* 13 (2): 1–27. https://doi.org/10.26512/lstr.v13i2.36253.

Klingler, Corinna, Magdaléna von Jagwitz-Biegnitz, Ronny Baber, Karl-Friedrich Becker, Edgar Dahl, Cornelius Eibner, Jörg Fuchs, et al. 2022. “Stakeholder Engagement to Ensure the Sustainability of Biobanks: A Survey of Potential Users of Biobank Services.” *European Journal of Human Genetics* 30 (12): 1344–54. https://doi.org/10.1038/s41431-021-00905-x.

Lawson, Jonathan, Moran N. Cabili, Giselle Kerry, Tiffany Boughtwood, Adrian Thorogood, Pinar Alper, Sarion R. Bowers, et al. 2021. “The Data Use Ontology to Streamline Responsible Access to Human Biomedical Datasets.” *Cell Genomics* 1 (2): 100028. https://doi.org/10.1016/j.xgen.2021.100028.

Lindén, Krister, Aleksei Kelli, and Alexandros Nousias. 2019. “A CLARIN Contractual Framework for Sharing Personal Data for Scientific Research.” In *Selected Papers from the CLARIN Annual Conference 2019*. Linköping, Sweden. https://doi.org/10.3384/ecp2020172.

“Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter.” 2011. https://www.retsinformation.dk/eli/lta/2011/593.

M. Romeo-Casabona, Carlos. 2021. “The New European Legal Framework on Personal Data Protection and the Legal Status of Biological Samples and Biobanks for Biomedical Research Purposes in Spanish Law.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:363–78. Law, Governance and Technology Series. Springer.

Mammadova, Najiba, Tricia K. Neppl, Natalie L. Denburg, and M. Heather West Greenlee. 2020. “Reduced Retinal Thickness Predicts Age-Related Changes in Cognitive Function.” *Frontiers in Aging Neuroscience* 12. https://www.frontiersin.org/articles/10.3389/fnagi.2020.00081.

Maresova, Petra, Blanka Klimova, Michal Novotny, and Kamil Kuca. 2016. “Alzheimer’s and Parkinson’s Diseases: Expected Economic Impact on Europe—A Call for a Uniform European Strategy.” *Journal of Alzheimer’s Disease* 54 (3): 1123–33. https://doi.org/10.3233/JAD-160484.

Mascalzoni, Deborah, Roberto Melotti, Cristian Pattaro, Peter Paul Pramstaller, Martin Gögele, Alessandro De Grandi, and Roberta Biasiotto. 2022. “Ten Years of Dynamic Consent in the CHRIS Study: Informed Consent as a Dynamic Process.” *European Journal of Human Genetics* 30 (12): 1391–97. https://doi.org/10.1038/s41431-022-01160-4.

McGuire, Amy L., and Laura M. Beskow. 2010. “Informed Consent in Genomics and Genetic Research.” *Annual Review of Genomics and Human Genetics* 11 (1): 361–81. https://doi.org/10.1146/annurev-genom-082509-141711.

McQueen, M.J., J.L. Keys, K. Bamford, and K. Hall. 2014. “The Challenge of Establishing, Growing and Sustaining a Large Biobank.” *Clinical Biochemistry* 47 (4–5): 239–44. https://doi.org/10.1016/j.clinbiochem.2013.11.017.

Milne, Richard, Annie Sorbie, and Mary Dixon-Woods. 2021. “What Can Data Trusts for Health Research Learn from Participatory Governance in Biobanks?” *Journal of Medical Ethics*, March, medethics-2020-107020. https://doi.org/10.1136/medethics-2020-107020.

Mitchell, Colin, Corrette Ploem, Valesca Retèl, Sjef Gevers, and Raoul Hennekam. 2020. “Experts Reflecting on the Duty to Recontact Patients and Research Participants; Why Professionals Should Take the Lead in Developing Guidelines.” *European Journal of Medical Genetics* 63 (2): 103642. https://doi.org/10.1016/j.ejmg.2019.03.006.

Mourby, Miranda, Heather Gowans, Stergios Aidinlis, Hannah Smith, and Jane Kaye. 2019. “Governance of Academic Research Data under the GDPR—Lessons from the UK.” *International Data Privacy Law* 9 (3): 192–206. https://doi.org/10.1093/idpl/ipz010.

Myszczynska, Monika A., Poojitha N. Ojamies, Alix M. B. Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M. Hautbergue, Joanna D. Holbrook, and Laura Ferraiuolo. 2020. “Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases.” *Nature Reviews Neurology* 16 (8): 440–56. https://doi.org/10.1038/s41582-020-0377-8.

Peloquin, David, Michael DiMaio, Barbara Bierer, and Mark Barnes. 2020. “Disruptive and Avoidable: GDPR Challenges to Secondary Research Uses of Data.” *European Journal of Human Genetics* 28 (6): 697–705. https://doi.org/10.1038/s41431-020-0596-x.

Pormeister, Kärt. 2021. “Regulatory Environment for Biobanking in Estonia.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:227–42. Law, Governance and Technology Series. Springer.

Rambla, Jordi, Michael Baudis, Roberto Ariosa, Tim Beck, Lauren A. Fromont, Arcadi Navarro, Rahel Paloots, et al. 2022. “Beacon v2 and Beacon Networks: A ‘Lingua Franca’ for Federated Data Discovery in Biomedical Genomics, and Beyond.” *Human Mutation*, April, humu.24369. https://doi.org/10.1002/humu.24369.

Reijs, Babette L. R., Charlotte E. Teunissen, Nikolai Goncharenko, Fay Betsou, Kaj Blennow, Inês Baldeiras, Frederic Brosseron, et al. 2015. “The Central Biobank and Virtual Biobank of BIOMARKAPD: A Resource for Studies on Neurodegenerative Diseases.” *Frontiers in Neurology* 6. https://www.frontiersin.org/articles/10.3389/fneur.2015.00216.

Rowhani-Farid, Anisa, Michelle Allen, and Adrian G. Barnett. 2017. “What Incentives Increase Data Sharing in Health and Medical Research? A Systematic Review.” *Research Integrity and Peer Review* 2 (1): 4. https://doi.org/10.1186/s41073-017-0028-9.

Salokannel, Marjut, Heta Tarkkala, and Karoliina Snell. 2019. “Legacy Samples in Finnish Biobanks: Social and Legal Issues Related to the Transfer of Old Sample Collections into Biobanks.” *Human Genetics* 138 (11–12): 1287–99. https://doi.org/10.1007/s00439-019-02070-0.

Sargsyan, Karine, Tanja Macheiner, Petra Story, Manuela Strahlhofer-Augsten, Katharina Plattner, Skaiste Riegler, Gabriele Granitz, Michaela Bayer, and Berthold Huppertz. 2015. “Sustainability in Biobanking: Model of Biobank Graz.” *Biopreservation and Biobanking* 13 (6): 410–20. https://doi.org/10.1089/bio.2015.0087.

Scott, Christopher J. M., Stephen R. Arnott, Aditi Chemparathy, Fan Dong, Igor Solovey, Tom Gee, Tanya Schmah, et al. 2020. “An Overview of the Quality Assurance and Quality Control of Magnetic Resonance Imaging Data for the Ontario Neurodegenerative Disease Research Initiative (ONDRI): Pipeline Development and Neuroinformatics.” bioRxiv. https://doi.org/10.1101/2020.01.10.896415.

Shabani, Mahsa, and Pascal Borry. 2018. “Rules for Processing Genetic Data for Research Purposes in View of the New EU General Data Protection Regulation.” *European Journal of Human Genetics* 26 (2): 149–56. https://doi.org/10.1038/s41431-017-0045-7.

Shabani, Mahsa, and Luca Marelli. 2019. “Re-Identifiability of Genomic Data and the GDPR: Assessing the Re-identifiability of Genomic Data in Light of the General Data Protection Regulation.” *EMBO Reports* 20 (6). https://doi.org/10.15252/embr.201948316.

Simell, Birgit A., Outi M. Törnwall, Iiro Hämäläinen, H.-Erich Wichmann, Gabriele Anton, Paul Brennan, Laurene Bouvard, et al. 2019. “Transnational Access to Large Prospective Cohorts in Europe: Current Trends and Unmet Needs.” *New Biotechnology* 49 (March): 98–103. https://doi.org/10.1016/j.nbt.2018.10.001.

Singh, Gursharan. 2018. “Target Product Profile and Clinical Development Plan.” In *Pharmaceutical Medicine and Translational Clinical Research*, 65–80. Elsevier. https://doi.org/10.1016/B978-0-12-802103-3.00005-5.

Slokenberga, Santa, Olga Tzortzatou, and Jane Reichel, eds. 2021. *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*. Vol. 43. Law, Governance and Technology Series. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-49388-2.

Souche, R., S. Mas, O. Scatton, J.-M. Fabre, L. Gimeno, A. Herrero, and S. Gaujoux. 2022. “French Legislation on Retrospective Clinical Research: What to Know and What to Do.” *Journal of Visceral Surgery* 159 (3): 222–28. https://doi.org/10.1016/j.jviscsurg.2021.12.002.

Southerington, Tom. 2021. “Access to Biomedical Research Material and the Right to Data Protection in Finland.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:243–56. Law, Governance and Technology Series. Springer.

Staunton, Ciara. 2021. “Individual Rights in Biobank Research Under the GDPR.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:91–104. Law, Governance and Technology Series. Springer.

Stenbeck, Magnus, Sonja Eaker Fält, and Reichel Jane. 2021. “Swedish Law on Personal Data in Biobank Research: Permissible But Complex.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:379–96. Law, Governance and Technology Series. Springer.

Thiel, Daniel B., Tevah Platt, Jodyn Platt, Susan B. King, and Sharon L. R. Kardia. 2014. “Community Perspectives on Public Health Biobanking: An Analysis of Community Meetings on the Michigan BioTrust for Health.” *Journal of Community Genetics* 5 (2): 125–38. https://doi.org/10.1007/s12687-013-0162-0.

Thorogood, Adrian, Heidi L. Rehm, Peter Goodhand, Angela J.H. Page, Yann Joly, Michael Baudis, Jordi Rambla, et al. 2021. “International Federation of Genomic Medicine Databases Using GA4GH Standards.” *Cell Genomics* 1 (2): 100032. https://doi.org/10.1016/j.xgen.2021.100032.

Tyndall, Adria, Wenny Du, and Christopher Breder. 2017. " The target product profile as a tool for regulatory communication: advantageous but underused". *Nature Reviews Drug Discovery*. 156 (16). https://doi.org/10.1038/nrd.2016.264

Tzortzatou, Olga, and Anastasia Siapka. 2021. “Mapping the Biobank Landscape in Greece.” In *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*, 43:291–308. Law, Governance and Technology Series. Springer.

Verhenneman, G., K. Claes, J. J. Derèze, P. Herijgers, C. Mathieu, F. E. Rademakers, R. Reyda, and M. Vanautgaerden. 2020. “How GDPR Enhances Transparency and Fosters Pseudonymisation in Academic Medical Research.” *European Journal of Health Law* 27 (1): 35–57. https://doi.org/10.1163/15718093-12251009.

Wallace, Susan E., Emily Kirby, and Bartha Maria Knoppers. 2020. “How Can We Not Waste Legacy Genomic Research Data?” *Frontiers in Genetics* 11: 446. https://doi.org/10.3389/fgene.2020.00446.

Watson, Peter H., Sara Y. Nussbeck, Candace Carter, Sheila O’Donoghue, Stefanie Cheah, Lise A.M. Matzke, Rebecca O. Barnes, et al. 2014. “A Framework for Biobank Sustainability.” *Biopreservation and Biobanking* 12 (1): 60–68. https://doi.org/10.1089/bio.2013.0064.

Weitzenboeck, Emily M, Pierre Lison, Malgorzata Cyndecka, and Malcolm Langford. 2022. “The GDPR and Unstructured Data: Is Anonymization Possible?” *International Data Privacy Law* 12 (3): 184–206. https://doi.org/10.1093/idpl/ipac008.

Annex A: GDPR Scope: Data Types and Territorial Considerations

Types of data within GDPR scope

The GDPR applies to the processing of personal data of living persons. As such, personal data concerning deceased persons is beyond the scope of the Regulation. Member States have a prerogative to adopt national data protection laws that effectively extend some of the provisions of the GDPR to the personally identifiable information relating to deceased individuals, a prerogative affirmed by Recital 27 GDPR. In practice, however, there is a considerable variation in the extent to which Member States' national data protection laws address post-mortem data protection (Bak et al. 2020).

The legal framework applicable to the personal data of deceased individuals is relevant for an ND research platform such as the EPND. Due to the nature of ND cohorts, most participants enrolled in them tend to be elderly persons, some with an advanced form of a ND. Therefore, when existing sample and data collections from ND cohorts are included in the EPND, it should be expected that some of the biological samples and their associated data will be from deceased donors. Where the institution responsible for an ND cohort can confirm the death of a donor, there may be a possibility to process the data pertaining to the deceased donor under different conditions, potentially making the data more widely available to external researchers. However, it is crucial that all research institutions involved in processing consult the applicable national data protection laws which may impose certain restrictions and/or compliance obligations in relation to processing data from deceased persons. Moreover, the principles of ethical data governance command that any future research using the data from deceased donors must respect the expressed wishes (if known) and reasonable expectations of the deceased donors regarding the continued use of their data.

Another important factor in determining whether the GDPR applies to data processing is the question of what constitutes *personal data* under the GDPR. As noted above, the GDPR defines personal data in a broad manner, namely as any information concerning an identified or identifiable natural person (i.e., the data subject) (Art. 4(1) GDPR). Recital 26 GDPR additionally clarifies that even where a particular piece of information, in and of itself, is insufficient to allow for identifying a data subject, it should nevertheless be considered personal data as long as it "could be attributed to a natural person by the use of additional information". The principal example of such information is the data that has undergone pseudonymisation, a process whereby personal identifiers relating to a natural person are replaced with different data, and the information required for re-identifying the person is kept separately (Art. 4(5) GDPR). The resultant pseudonymous data is considered personal data, and hence within the scope of the GDPR, provided that re-identification remains feasible. The feasibility of re-identification should be assessed contextually, taking into account the "means reasonably likely to be used … to identify the natural person directly or indirectly" (Recital 26 GDPR). Importantly, Recital 26 explicitly states that this condition is not limited to the party who is provided with the pseudonymised data and is also applicable to the cases where re-identification of data subjects can be carried out by "another person", i.e., a third party.

In the context of the EPND, this means that even though data-recipient researchers themselves may not have sufficient information to infer the identities of the data subjects, from the point of view of the

GDPR, they may nevertheless be processing personal data. This is due to the fact that another party, in particular, the medical or research institution that provided data, may retain information required for re-identifying data subjects, such as conversion tables or the code for decrypting pseudonymised data. Importantly, even if any direct link to data subjects were to be removed (e.g., through the deletion of conversion tables), it may still be possible to infer the identities of data subjects by gleaning insights from multiple sources of data, including publicly available information. The feasibility of identity inference in this manner has been extensively discussed in the context of genomics, where it is often possible to deduce the identities of purportedly anonymous individuals by combining their genomic data with the publicly available information in online ancestry and genealogy databases (Shabani and Marelli 2019). Consequently, irreversible anonymisation is difficult to achieve and in any case may not even be desirable from a research perspective, as this will significantly limit the utility of the data, precluding drawing meaningful associations across datasets that could help elucidate novel insights into NDs.

Of note, there have been considerable academic and policy discussions around the concept of data anonymisation within the meaning of the GDPR. Authors have cautioned that data anonymisation, as currently defined in European guidelines,[15] makes it exceedingly difficult to fully anonymise personal data. Consequently, it has been recommended that European regulators adopt a contextual and risk-based view on data anonymity, whereby data could be deemed effectively anonymous in context even if it is theoretically possible to re-identify data subjects (Finck and Pallas 2020; Weitzenboeck et al. 2022). The Information Commissioner's Office (ICO) of the United Kingdom has also signalled its support for risk-based approaches to interpreting anonymisation. In a May 2021 draft guidance, the ICO stressed that identifiability of data subjects can be viewed as a spectrum, where "identifiability depends on the specific circumstances and risks posed. Essentially, information may 'move' along the spectrum of identifiability to the point that data protection law starts to apply to it (or, conversely, stops applying to it)."[16]

A contextual definition of data anonymity would be a welcome development for the scientific research community in the EU, as this would allow for a meaningful use of data while also reducing compliance obligations for researchers. However, to date (October 2022), EU regulators and national data protection authorities in the Member States have not formally communicated an intention to endorse a less restrictive definition of data anonymity for research purposes. As a consequence, currently, a prudent approach for a European initiative such as the EPND would be to generally treat the data provided by ND cohorts, as well as any biomarker data derived by analysing biological samples, as personal data whose processing must comply with the GDPR. In parallel, stakeholders operating EU-based research platforms may choose to explore promising strategies with a potential to help them demonstrate the anonymity of the data undergoing processing. Such strategies may include, among others, leveraging synthetic data approaches to enable research using sensitive personal data. As additional guidance from the EU regulators and national data protection authorities emerges over the next few years, more clarity can be expected as to viable strategies for demonstrating data anonymity.

[15] For a commonly cited guidance by an authoritative body, see Article 29 Data Protection Working Party Opinion 05/2014 on Anonymisation Techniques (Article 29 Data Protection Working Party 2014)

[16] ICO call for views: Anonymisation, pseudonymisation and privacy enhancing technologies guidance. Chapter 2. Published online on 28 May 2021

One of the more salient challenges for the EPND stemming from the prevailing interpretation of data anonymity is associated with the discoverability of EPND data and sample resources. As discussed in Chapter 1 of this paper, one of the main objectives of the EPND's user portal is to enable meaningful search and discovery of EPND collections by external researchers, prior to initiating a formal data access request. This may be achieved by publishing metadata describing the collections, or by permitting requestors to submit structured queries to cohorts for ascertaining the availability of the personal data of interest. However, given the high threshold for demonstrating data anonymity under the GDPR, there is a clear tension between the goals of maximising the usefulness of EPND data/sample discovery tools and ensuring that such tools reveal no personal information pertaining to data subjects. On the one hand, certain types of information of highly generalised nature, such as the disease area, sample/data types, and the number of individuals enrolled in a given ND cohort, can be deemed anonymous within the meaning of the GDPR, and can be published in an open-access data catalogue. However, such generalised information is, arguably, of limited value for external researchers seeking to assess the feasibility of their intended research project. On the other hand, providing more granular information about the cohort, such as aggregate descriptive statistics by demographic and clinical variables, increases the risk of inadvertently disclosing personal data. In some cases, third parties may be able to combine this data with other information accessible to them, including publicly available information, and successfully deduce the identities of individuals enrolled in an ND cohort.

This tension between the informational value of EPND data and sample discovery tools on the one hand, and the risk of revealing personal data on the other hand, calls for careful considerations when implementing the EPND resource discovery functionality. The EPND could potentially draw on the experience of other research platforms that have sought to address this challenge. For example, the European Genome-Phenome Archive (EGA), until recently, had relied on a two-tier metadata structure to enable data discovery and preliminary assessment by external researchers. The so-called non-sensitive metadata, primarily consisting of technical information pertaining to an aggregate dataset (such as overall dataset format and data quality) could be accessed by any third party visiting the website of the EGA. The second-tier, or sensitive metadata, contained aggregate descriptive statistical information concerning the data subjects within a dataset. Access to the sensitive metadata was restricted to registered users only, i.e., users whose identities had been authenticated and who had accepted terms of use (Freeberg et al. 2022). However, the EGA subsequently abolished the two-tier metadata structure, placing the onus on the data-submitting researchers to ensure - and confirm - that the metadata they share via the EGA contains no personal data relating to their data subjects.

Another notable approach to striking an optimal balance between enabling meaningful data discovery and minimising the risk of personal data disclosure has been pioneered by the Global Alliance for Global Health (GA4GH) in their Beacon protocol. Beacon is an interactive data discovery tool that allows an external user to ascertain whether a dataset contains a specified observation of interest, such as the presence of a particular genetic variant or a clinical trait. The latest version of the Beacon protocol allows users to submit queries and receive responses in a manner that stratifies data subjects into medically and demographically meaningful categories. Although in some cases it may be theoretically possible for external users to obtain personal data by submitting a series of targeted

queries, such risks are addressed by: (i) implementing solutions that ensure the outputs of the search are “safe” (e.g., sufficiently aggregated); (ii) monitoring the activity of the users, and (iii) limiting the number of queries they can submit (Rambla et al. 2022).

These considerations will help inform the design and implementation of EPND resource discovery tools in the coming years. The key principle to be reflected in the resource discovery functionality is a risk-based approach to publishing metadata or permitting search queries. Risks of accidental disclosure of personal data through the EPND resource discovery tool will need to be addressed by implementing proportionate technical safeguards, which may include user monitoring systems and tiered search whereby the granularity of the search results corresponds to the user’s authorisation level.

Territorial Scope

The final important consideration regarding the relevance of the GDPR to the EPND concerns the territorial scope of the Regulation. As highlighted above, the GDPR applies to the processing of personal data of data subjects who are in the EU/EEA (“the Union”). This condition holds true irrespective of whether the party undertaking such processing is based in Europe or in a third country. Considering the fact that the majority of cohorts currently envisaged to be part of the EPND are based in EU/EEA countries, it can be assumed that the access and use of EPND resources by researchers located outside of the Union will be subject to the GDPR. More specifically, Chapter V of the Regulation, which concerns international data transfers, will apply. Consequently, when transferring personal data outside the EU/EEA, one of the safeguards or conditions required for international transfers under Chapter V GDPR must be in place. Additionally, processing of personal data concerning natural persons outside the Union is also within the scope of the GDPR where the party undertaking such processing is based in a Member State. This consideration is relevant to scenarios where non-European cohorts choose to collaborate with the EPND, making their data and sample collections accessible for research through the platform. The EPND should ensure that the patients and research participants of such cohorts are afforded the same fundamental rights and privacy protections as the data subjects enrolled in ND cohorts based within the Union.

Annex B: Definition and implications of GDPR Roles

Definition of the GDPR roles

Article 4(7) GDPR defines the controller as "a 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". Implied in this definition is the possibility of more than one party acting as a controller under the GDPR. This is reflected in Article 26(1) GDPR, which introduces the concept of joint controllership: "Where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers". As regards to processorship, the processor is defined in Article 4(8) GDPR as "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller".

Additional valuable guidance and clarifications regarding GDPR roles are provided by the European Data Protection Board (EDPB) in the Guidelines 07/2020 ("On the Concepts of Controller and Processor in the GDPR"). In this interpretive guidance, the EDPB offers a more nuanced analysis of the overarching condition for controllership, namely that the controller determines "the purposes and means" of processing (i.e., "the *why* and the *how* of processing" (European Data Protection Board 2021c)). With respect to determining purposes, the EDPB notes that the interpretation is relatively straightforward: only the controller can choose a purpose for which data is processed. A processor cannot pursue an independent purpose for processing. If it does so, it no longer qualifies as a processor and should be considered as a controller for the processing operation for which it determined the purpose.

In regard to determining means of processing, the EDPB recognises that in practice, processors acting at the request of controllers may exercise certain discretion as to how the data will be processed in order to achieve the purpose specified by the controller. Consequently, both the controller and the processor may participate in defining the means of processing. In order to address this potential confusion, the EDPB draws a distinction between "essential" and "non-essential means" of processing. The EDPB provides the following definition of essential means, accompanied by useful indicative questions:

> *" "Essential means" are means that are closely linked to the purpose and the scope of the processing, such as the type of personal data which are processed ("which data shall be processed?"), the duration of the processing ("for how long shall they be processed?"), the categories of recipients ("who shall have access to them?") and the categories of data subjects ("whose personal data are being processed?"). (European Data Protection Board 2021c)*

According to the EDPB, determination of the essential means is inherently part of controllership. By contrast, processors may determine other, "non-essential" means relating to processing, which concern "practical aspects of implementation, such as the choice for a particular type of hard- or software or the detailed security measures" (European Data Protection Board 2021c).

Despite extensive guidance provided by the EDPB, alongside multiple illustrative examples, the EDPB recognises that in certain data processing environments, the allocation of GDPR roles may still be a challenging task. This is certainly the case with complex research-oriented ecosystems such as the network of institutions to be supported by the EPND infrastructure. Thus, the allocation of GDPR roles within the EPND will be a crucial aspect for ensuring EPND's GDPR compliance, with the outcome of such allocation to be dictated by contextual factors associated with a particular processing operation.

In general, when defining GDPR roles in a complex, multi-party data-processing environment, it is crucial to be mindful of the following two considerations. First, as emphasised by the EDPB, the concepts of controller, joint controller, and processor are functional concepts, meaning that they allocate responsibilities based on the actual role parties play in data processing. As such, the decision needs to be informed by concrete contextual factors pertaining to a given processing operation and should not be shaped by factors extraneous to the processing operation, such as formal descriptions in contracts of the relationship among the organisations involved. For example, if two parties enter into a business relationship whereby they act as a client and a service provider, this does not necessarily mean that the service provider will be a processor within the meaning of the GDPR for all processing operations. The roles and the responsibilities of the parties under the GDPR will be dictated by the extent to which the parties determine the purpose and means of a particular processing operation. Emphasising the need for contextual assessment of a party's role in relation to each processing operation highlights the second important consideration: it is possible for the same party to act as a controller for certain processing operations, while being a processor for other processing operations with respect to the same data subject, or indeed, the same data. A relevant example of this dual role of the same entity, also discussed by the EDPB, is that of scientific research collaborations between institutions. A data-holding research institution engaged in a collaborative project may need to carry out certain processing operations (e.g., a particular form of data analysis), at the request of its external partner, in which case it would be acting as a processor for the external partner (controller). At the same time, the data-holding institution remains the controller with respect to using the same data for its own research purposes (European Data Protection Board 2021c). A more nuanced discussion of this example, alongside its variations, will be provided in the sections below.

Compliance obligations of controllers and processors

Although both the controller and the processor are subject to the GDPR, they have different responsibilities and legal obligations under the Regulation.

The controller is the party responsible for ensuring that processing of personal data takes place in accordance with the principles of the GDPR laid out in Art. 5(1) of the Regulation. This obligation is formalised by the Accountability Principle, introduced under Art. 5(2) GDPR, which states that "the controller shall be responsible for, and be able to demonstrate compliance with" the principles of the Regulation (see box 4 below).

Box 4 - Principles relating to processing of personal data (Article 5 GDPR)

1. Personal data shall be:

a. processed lawfully, fairly and in a transparent manner in relation to the data subject ('**lawfulness, fairness and transparency**');
b. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('**purpose limitation**');
c. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('**data minimisation**');
d. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('**accuracy**');
e. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('**storage limitation**');
f. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('**integrity and confidentiality**').

2. The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('**accountability**').

The Accountability Principle is notable in two ways. First, the Accountability Principle means that the party ultimately responsible for the compliance with Art. 5(1) GDPR principles is the controller, irrespective of who performs the processing. That is, even where the controller delegates processing to another party (who will be acting as a processor), ensuring that processing is carried out in accordance with the principles of the Regulation remains the sole responsibility of the controller. Second, the accountability principle places emphasis on the requirement of demonstrating compliance with the principles laid out in Art. 5(1) of the Regulation. Importantly, demonstrating compliance with the principles of the GDPR is an overarching obligation that should not be interpreted narrowly, as in complying with specific procedural, technical, and organisational requirements laid down elsewhere in the GDPR.[17] However, the GDPR is non-prescriptive as to the exact manner in which the controller may be able to "demonstrate compliance" in relation to the Art. 5(1) principles. In its 2019 guidelines on Data Protection by Design and by Default, the EDPB reaffirmed controllers' obligation to be able demonstrate their compliance with the principles of the GDPR:

[17] Examples of such specific requirements include, among others: an obligation to carry out a Data Protection Impact Assessment, where applicable (Art. 35 GDPR), adopt one of the safeguards described in Chapter V GDPR when engaging in international data transfers, and implement technical and organisational measures referenced in Art. 89(1) GDPR when processing personal data for scientific research purposes.

> *"The controller needs to be able to demonstrate compliance with the principles. In doing so, the controller may demonstrate the effects of the measures taken to protect the data subjects' rights, and why the measures are considered to be appropriate and effective. For example, demonstrating why a measure is appropriate to ensure the principle of storage limitation in an effective manner". (European Data Protection Board 2020)*

Although helpful for indicative purposes, the non-exhaustive nature of the guidance above highlights that demonstrating compliance with the Art. 5(1) principles is to be interpreted in a contextual manner. Essentially, the controller will need to: i) assess, by taking into consideration all the relevant contextual factors, whether the processing complies with the Art. 5(1) principles of the GDPR; and ii) decide on the appropriate way to demonstrate this compliance.

These two aspects embedded in the Accountability Principle – ensuring and demonstrating compliance with the GDPR principles – clearly promote a deliberative approach to the design of processing operations. Controllers have a strong incentive to ascertain that the intended processing complies with the Art. 5(1) GDPR principles, while carefully assessing whether the proposed measures are sufficient to demonstrate this compliance. However, the lack of clarity around the precise threshold of demonstrating compliance with the principles may also encourage controllers to adopt a risk-averse approach, potentially refraining from processing altogether. Arguably, this is particularly true in the context of data sharing with other parties (controllers or processors), where the controller sharing the data is naturally incentivised to require additional assurances from the recipient, in order to ascertain that the processing by the recipient also takes place in accordance with the principles of the Regulation. Overall, it has been argued that the GDPR has, on balance, made controllers less willing to share data with other parties, an observation that has also been widely reported in the context of biomedical research (Cool 2019; Christofidou, Lea, and Coorevits 2021). These systemic disincentives to data-sharing, introduced by the GDPR, constitute a critical risk factor bearing on the viability of research platforms such as the EPND. Exploring strategies to mitigate this risk is expected to be one of the key strategic priorities of the EPND consortium in the years ahead.

Apart from the overarching general obligation to comply with the Art. 5 principles of the GDPR, there are numerous additional concrete requirements (e.g., procedural, contractual, organisational, and technical) that uniquely apply to the controller. Most of these requirements do not directly concern the entities operating the EPND's Technical Hub, whose role is currently envisaged to be that of a processor (see the next section), and hence will not be discussed in detail in the present document. However, as the EPND seeks to explore ways in which it can support controllers' GDPR compliance efforts, the EPND may choose to map out a comprehensive set of GDPR compliance requirements to which controllers (i.e., cohorts and data users) are subject, potentially identifying areas where the EPND can provide meaningful support. One possible approach in this respect is to perform a generic GDPR Gap analysis from the point of view of a hypothetical controller, an exercise that could be considered during Phase II of the IMI project (years 2-5 of the 5-year project, ending in October 2026). (For an illustrative example of what such a gap analysis might entail, see Georgiou and Lambrinoudakis 2021). As preparatory work towards simplifying controllers' GDPR compliance in the future, WP2 of the IMI consortium carried out a generic Data Protection Impact Assessment

(DPIA) for the EPND in Year 1 of the project. The EPND DPIA captures multiple potential scenarios of data processing through the technical platform to be established by the consortium. The DPIA is publicly available[18] and provides an extensive analysis of the privacy risks associated with the processing of personal data in research-oriented biomedical data platforms such as the EPND. Following the identification of appropriate risk-mitigating measures and controls, the DPIA concluded that the residual risks were low, indicating that a prior consultation with a supervisory authority, in the sense of Art. 36(1) GDPR, is not required for the EPND. The DPIA will be updated during Phase II of the project, with the scope of the DPIA centred on data processing through the ADDI's AD Workbench.

The obligations associated with controllership remain the same where two or more parties act as joint controllers for data processing. Under these circumstances, Article 26(1) GDPR mandates that the joint controllers agree on their respective responsibilities for meeting controllership obligations under the GDPR in a transparent manner. The responsibilities of the joint controllers should be clearly delineated in a contract or a similar legal act and must correspond to each party's precise role in data processing. It is recommended to explicitly define GDPR roles and responsibilities in a granular manner, that is, across all envisaged processing operations and the phases of processing. Doing so will not only enable greater operational clarity, but also ensure that the controllers limit their potential liability to the processing operations (and temporal phases thereof) for which they are actually responsible. In the absence of such a contractual legal agreement, all parties acting as controllers vis-a-vis the data subject *could* be held jointly and severally liable for any damage resulting from an infringement of the GDPR (Becker, Thorogood, et al. 2022).

With regards to processors, the GDPR defines a different set of compliance obligations. The processor must implement technical and organisational measures appropriate for the processing operations it performs (Art. 32 GDPR), and, under certain circumstances, designate a data protection officer (Art. 37 GDPR). The processor is additionally required to establish and maintain a comprehensive record of the categories of data processing activities, in accordance with Article 30(2) GDPR. Other obligations of the processor are primarily vis-à-vis the controller on whose behalf it processes personal data. This includes, for example, assisting the controller to ensure and demonstrate its GDPR compliance, duly inform the controller in the event of detecting a personal data breach, and to fulfil contractual obligations such as those laid out in Art. 28(3) GDPR.

The EDPB recognises that due to the differences in the obligations between the controller and the processor, parties involved in data processing may have an incentive to allocate GDPR roles based on the ease of compliance and other convenience considerations. In this respect, the EDPB reiterates that the responsibilities of the parties must be defined based on an objective assessment of each party's role in a given processing operation. The EDPB particularly discourages any attempt to circumvent compliance obligations associated with a GDPR role by misrepresenting the function parties play in processing. The EDPB cautions that "[it] is not possible either to become a controller or to escape controller obligations simply by shaping the contract in a certain way where the factual circumstances

[18] Available at: https://epnd.org/resources/deliverable-2-1-data-protection-impact-assessment

say something else." This consideration further highlights the importance of correctly assessing the function each party plays in data processing, and allocating GDPR roles accordingly.

The section below discusses the application of the GDPR roles framework to the cases of data sharing among scientific research institutions, including scenarios where sharing is intermediated by a platform such as the EPND.

GDPR roles and sharing of research data

The practice of data sharing/transfers among scientific research institutions has been profoundly impacted by the GDPR in various ways, not least of which is the requirement to allocate GDPR roles (controller, joint controller, and processor) among the parties, and determine their corresponding compliance obligations.

Direct transfers between two research organisations

The simplest case of data transfers occurs when two research organisations agree on a direct transfer of data, without the involvement of an intermediary. The roles and the responsibilities of the two parties under the GDPR can be determined by assessing to what extent they influence the design of the research project for which the data are intended. The analysis below is modelled on the examples discussed in EDPB guidelines and considers the following two scenarios:

- The data-providing and the data-recipient institutions collaborate to design the study together, including the objectives and the research methodology of the study (collaborative research project)
- The data-recipient institution is the sole entity pursuing the study.

In the former scenario, i.e., where the parties design the study in a collaborative manner, the institutions are joint controllers, since they both play a critical role in determining the purposes and essential means of processing. Consequently, for all processing operations directly associated with performing the study, the overarching condition for establishing joint controllership is fulfilled. The roles of the two institutions under the assumption of collaborative research are summarised in Table 3 below.

Table 3 - GDPR roles under a bilateral data transfer: research collaboration

	Processing operation(s)	Data Provider	Data Recipient
1	Data transfer	Joint Controller	Joint Controller
2	Data storage by recipient	Joint Controller	Joint Controller
3	Data use for research	Joint Controller	Joint Controller

The roles of the parties change if the research project is pursued by the data-recipient institution independently of the data provider. In other words, where the data recipient is the sole party designing the study, including its objectives and methodology. In this case, the role of the data provider is limited

to reviewing and accepting the data-recipient's study design, followed by a data transfer to enable the proposed study. Merely accepting a study proposal already elaborated by another party and ensuring the transfer complies with data protection principles (e.g., data minimisation) would be insufficient for the data provider to qualify as a controller for the processing taking place following the data transfer. Hence, in this scenario, the data recipient, which independently determines the purposes and essential means of the processing related to the study, would be the sole controller for the processing it undertakes following the receipt of the data. However, the data-providing party should be deemed the sole controller for data transfer specifically, given that the party decides whether or not to approve the use of its data by the intended recipient. By making this decision, the data-providing party determines the purposes and means of this processing operation (transfer data to enable the recipient's research project). Table 4 describes modified GDPR roles of the parties under the assumption of independently pursued research by the data recipient.

Table 4 - GDPR roles under a bilateral data transfer: research pursued by the data recipient

	Processing operation(s)	Data Provider	Data Recipient
1	Data transfer	Controller	Third party
2	Data storage by recipient	Third party	Controller
3	Data use for research	Third party	Controller

One scenario which is implicitly captured by the two cases above, but merits separate consideration is a situation where the data-recipient institution is the sole party pursuing a study it has independently designed, but the data provider, after reviewing the study protocol, requests changes to the protocol as a condition for transferring data. Under these circumstances, the GDPR roles of the parties are less clear-cut, and the outcome will depend on the substantiveness of the changes to the study protocol requested by the data provider. Should the changes be deemed significant enough so as to alter the purposes and essential means of processing, the data provider will have to be considered a joint controller for the study, and hence the allocation of roles under Table 3 will apply. On the other hand, in the case of minor changes to the study protocol that have no substantive impact on the purposes and essential means of processing, the data provider will likely remain the controller for data transfer and won't have a role in the study proposed by the data recipient, resulting in the allocation of roles under Table 4.

Data transfers via a third party (research platform)

Understanding the implications of bilateral transfers on the GDPR roles of parties is an important stepping stone towards accurately allocating GDPR roles in more complex environments where data transfers among research institutions are mediated by a platform such as the EPND's Technical Hub. Table 5 shows the allocation of GDPR roles among the three principal actors in this scenario: the data provider (i.e., clinical or research institution responsible for an ND cohort), the platform, and the data recipient (external researcher seeking data access).

Table 5 - GDPR roles under a platform-mediated data transfer

	Processing operation(s)	Data Provider	Platform	Data Recipient
1	Data transfer to the platform	Controller	Processor	Third party
2	Data storage	Controller	Processor	Third party
3	Data standardisation*	Controller	Processor	Third party
4	Data disclosure (granting access) to the Data Recipient	Controller	Processor	Third party
5	Data analysis for research**	Third party or Joint Controller	Processor	Controller or Joint Controller

* Depending on the circumstances, may include: generation of new data from samples, data pre-processing, cleaning, harmonisation, FAIRifying

** Applies to all modalities of platform-enabled analysis (e.g., centralised, federated, hybrid)

Importantly, the table assumes that the data provider's participation in the EPND entails provision of personal data. If the data provider only shares anonymous metadata (such as the cohort size and types and format of available datasets) to be displayed in the EPND cohort catalogue, but no personal data is made available for access and/or (federated) analysis via the EPND's Technical Hub, Table 5 is irrelevant.

Another assumption underpinning Table 5 is that the EPND offers standardised tools and services to all data providers. Although data providers will have access to certain customisation options, such as the possibility to specify data and sample use conditions, the intention of the EPND consortium is to make these tools available to all participating institutions in a uniform manner and allow the institutions to select modalities that best fit their compliance needs. Under these circumstances, the guidance from the EDPB clearly indicates that the data platform providing standardised tools and services would be a processor,[19] carrying out processing at the request of the data provider and/or the data recipient, depending on the processing operation. Importantly, in step 4 (data disclosure, granting permission to analyse data), it is assumed that the cohort providing the data retains full authority over the access decisions. As discussed in Chapter 3 of this document (Governance View), the EPND is currently exploring the feasibility of a central Data and Sample Access Committee (DSAC). Depending on the (legal and organisational) feasibility of such a central EPND DSAC and (if deemed feasible), its governance blueprint, the assumptions made in relation to parties' GDPR roles under step 4 may need to be re-evaluated.

It is worth highlighting that the processing operations 4 (Data disclosure/granting access to the Data Recipient) and 5 (Data analysis for research) in Table 5 mirror the logic of allocating GDPR roles under bilateral transfers, as discussed previously. Namely, the role of the data provider will depend on

[19] For indicative cases discussed within the EDPB guidance, see "Example: standardised cloud storage service" and "Example: Cloud Service Provider" (European Data Protection Board 2021c).

the extent to which the data provider influences the intended research use of the data by the recipient. Where the data provider, either through its local/institutional DSAC or via the EPND central DSAC, merely accepts a data access request, the role of the data provider will be limited: it will act as a controller for data disclosure to the data recipient and won't be involved in the data analysis carried out by the recipient. Accordingly, the data recipient will be acting as the sole/independent controller for the research use of the data. However, the data provider may request substantive changes to the proposed research project (for example, adding or removing a research question, or agreeing to only grant access to a sub-set of the requested data), in which case it would be effectively redefining the purposes and essential means of processing. In this case, the data provider and the data recipient would act as joint controllers with regards to processing operations 4 and 5 (data disclosure and data analysis, respectively).

Finally, as a caveat, it should emphasised that the allocation of the GDPR roles discussed in this section is limited to the personal data of patients and research participants enrolled in ND cohorts. The parties operating the EPND's Technical Hub, at least in the short term, will be acting as a processor with respect to the personal data of these data subjects. However, the EPND consortium, as part of its activities, may need to process personal data pertaining to other categories of data subjects. These additional data subjects may include, for example: person(s) responsible for managing cohort-related information in the EPND cohort catalogue (e.g., Principal Investigators/PIs of the cohorts), individuals requesting data access through the EPND, and authorised researchers accessing cohorts' data collections via the EPND user portal. With respect to the personal data concerning these data subjects, the EPND - more specifically, the entity/entities responsible for the infrastructural components of the platform – will be acting as the controller(s) under the GDPR. Given the relatively limited and (as of yet) incompletely defined nature of processing operations in relation to these data subjects, the processing where the EPND acts as the controller is not addressed in greater detail in this paper. It is however worth stressing that as part of its routine activities, the EPND will be the controller vis-a-vis data subjects other than research participants/patients whose data are included in ND cohorts.

Annex C: Consent, public interest and legitimate interest

Data subject's consent (Art. 6(1)(a) GDPR)

Data subject's consent constitutes one of the six legal bases for processing personal data and is grounded in Art. 6(1)(a) GDPR. Where special categories of data are processed, this legal basis will typically be coupled with the condition under Art. 9(2)(a) GDPR whereby the data subject provides an "explicit consent to the processing of those personal data for one or more specified purposes". The EDPB also considers it appropriate to rely on the Art. 6(1)(a) legal basis and the Art. 9(2)(a) condition in conjunction (European Data Protection Board 2021b).

Importantly, consent as a GDPR legal basis for processing personal data should not be confused with the informed consent provided by patients and other prospective research participants of a medical study. The two types of consent share many similarities, for example, they must both convey key informational elements in a transparent and understandable manner, and be provided by the consenting individual voluntarily, that is, in the absence of any undue influence. However, the two types of consent also differ in several important ways. On the one hand, obtaining informed consent for research participation is a standard practice and often a legal (non-GDPR) requirement for certain types of biomedical research. On the other hand, the GDPR Art. 6(1) legal basis and Art 9(2) legitimation for processing personal data in the context of biomedical research projects need not be consent; in fact, as will be discussed later, consent within the meaning of the GDPR is increasingly considered a sub-optimal legal basis for processing personal data in biomedical research. For a more extensive overview of the differences between consent as a GDPR legal basis and informed consent for biomedical research, see Annex F, and in particular Table 8 therein.

According to the GDPR, in order for a data subject's consent to be valid, it must be "freely given, specific, informed and unambiguous" (Art. 4(11) GDPR). This definition, and in particular, the requirement that consent be specific, is typically interpreted as limiting the purposes for which the data can be lawfully processed under Art. 6(1)(a) GDPR. However, there is considerable debate as to how narrowly the purpose(s) of data processing must be defined, particularly in the context of scientific research. This discussion stems from Recital 33 of the GDPR, which allows for consent "for certain areas of research" where it is not possible "to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection" (Verhenneman et al. 2020; Gefenas et al. 2022). One interpretation of this condition is that through Recital 33, the legislator intended to enable the so-called "broad consent", whereby data subjects, at the time of initial data collection, consent to having their data used in various future research projects falling under the predefined areas of research. Consequently, this initial broad consent can be used to legitimise processing of the data by all such subsequent research projects, even if their precise nature (e.g., specific research question, methodology, and identity of the controller performing research) was unknown at the time of the initial data collection (Hallinan 2020; Gefenas et al. 2022). However, this interpretation was expressly rejected by the EDPB in their Guidelines on Consent under the GDPR (Guidelines 05/2020)[20],

[20] These guidelines were originally elaborated by the EDPB's predecessor, the Article 29 Working Party (A29WP): "Guidelines on Consent under Regulation 2016/679; WP259". The document was subsequently endorsed, in a minimally altered form, by the EDPB in 2020 Guidelines 05/2020).

particularly in relation to processing special categories of data. According to the EDPB, "When regarded as a whole, the GDPR cannot be interpreted to allow for a controller to navigate around the key principle of specifying purposes for which consent of the data subject is asked." The EDPB further elaborates on the limits of broad consent in the context of research using special categories of data:

> *"When research purposes cannot be fully specified, a controller must seek other ways to ensure the essence of the consent requirements are served best, for example, to allow data subjects to consent for a research purpose in more general terms and for specific stages of a research project that are already known to take place at the outset. As the research advances, consent for subsequent steps in the project can be obtained before that next stage begins".* (The European Data Protection Board)

In making this statement, the EDPB has effectively argued that, where consent (Art. 6(1)(a) in conjunction with Art. 9(2)(a) GDPR) is the legal basis for processing special categories of data, each subsequent use of the data for a new research study requires an explicit consent by the data subject.

An analysis by several co-authors of this document of what constitutes a valid consent under the GDPR also suggests that the notion of broad consent, as a one-off consent at the time of data collection, cannot be relied upon to legitimise future yet-to-be-specified reuses of the data where consent is the legal basis for processing personal data (*publication forthcoming*). The conclusions of the analysis agree with the EDPB's interpretation that the use of broad consent under these circumstances will likely give rise to unlawful processing of personal data. The case against the validity of a broad consent is particularly strong where data are shared with external controllers for their research purposes, as opposed to being re-used by the organisation that initially collected the data. This is due to a provision in Recital 42 of the GDPR, according to which when data processing is based on consent, the data subject should be made aware, among other things, of the identity of the data controller (Recital 42 GDPR). Because the data recipient is a controller for the downstream research project (see Table 5 above) whose identity was not communicated to the data subject as part of the initial consent, Recital 42 clearly suggests that the initial consent cannot be valid for this downstream research use.

The aforementioned considerations suggest that cohorts that rely on consent as the legal basis to process personal data are not well-placed to share the data with external researchers who seek data access for their own purposes and hence are acting as controllers in relation to the reuse of data. This conclusion reinforces the emerging consensus in the data protection community that consent - that is, Art. 6(1)(a) GDPR, in conjunction with Art. 9(2)(a) GDPR where special categories of data are processed - is **not** the most suitable legal basis to enable lawful data processing for scientific research (European Data Protection Board 2019; 2021b). By contrast, under alternative GDPR legal bases for processing personal data in the research context, no explicit consent is required from data subjects prior to data reuse in a new research project or by a new controller. This makes it considerably easier for controllers to lawfully re-use previously collected data within their organisation, as well as to make the data widely available to external researchers in a GDPR-compliant manner.

Nevertheless, the reality is that as of 2022, consent remains a widely utilised legal basis by research organisations in Europe, a generally held perception that was also confirmed by our experience when interfacing with consortium partners as part of EPND case studies. This is due largely to the fact that other GDPR legal bases, even if in theory more suitable for scientific research purposes, are not uniformly available across European countries and types of legal entities (see the next section). Consequently, it is important that the EPND find ways to avoid systematically excluding the cohorts that rely on consent as the GDPR legal basis to process personal data. In this regard, three considerations are worth highlighting.

First, even where a cohort relies on consent as the GDPR legal basis and cannot readily share its data with an external researcher for the latter's research project, the cohort may still be able to lawfully deposit the data with a repository and make limited use of the repository's services. For example, it may well be permissible under the existing consent for the cohort to utilise an external data hosting service, alongside other value-added services around data quality improvement, such as data cleaning or standardisation. Insofar as these processing operations are enabling the cohort to pursue the purpose(s) for which the cohort is already lawfully processing the data – for example, conducting a longitudinal study focused on a specific ND -, the GDPR consent obtained by the cohort remains valid. Although this consent does not enable lawful sharing of the same data with previously unspecified controllers, the EPND may nevertheless consider offering these limited services to cohorts relying on consent as the legal basis to process personal data. Arguably, processing operations such as data hosting, cleaning and/or standardisation constitute important value-added services in their own right, and their viability as stand-alone services may be worth exploring in the coming years. Perhaps more importantly, the fact that data are already stored on the EPND-managed hosting environment may incentivise cohorts to take additional steps towards enabling lawful data sharing via the EPND (e.g., through one of the paths described in the subsequent paragraph).

Next, although data processed under consent as the GDPR legal basis cannot be readily reused by previously unspecified controllers, additional steps can be taken to enable lawful data sharing and reuse via the EPND. For example, the initial controller (cohort) may be able – and willing – to recontact its data subjects to obtain a new consent for the intended data-recipient controller's research project. Alternatively, if the cohort can confirm that a data subject is deceased, it may be possible to lawfully share and reuse the data (subject to any constraints to post-mortem data processing under the national data protection laws, as applicable). Moreover, depending on the nature and richness of the data requested by the external researcher, the cohort may be able to confirm that the data are anonymous within the meaning of the GDPR, rendering the Regulation inapplicable to the intended sharing. Finally, under certain circumstances – for example, following a legislative change – the cohort hitherto relying on consent as the GDPR legal basis may gain access to another legal basis to process personal data for scientific research purposes.[21] Although the cohort may still need to meet certain transparency

[21] Such legislative changes may take place at the national as well as the European level. Regarding the latter, the forthcoming European Health Data Space (EHDS) Regulation is of particular relevance. First draft of the EHDS Regulation was published by the European Commission in May 2022. Among other key objectives of the Regulation is to provide a harmonised framework for uniform legal bases when processing health and biomedical data for research purposes. However, the proposed EHDS Regulation is currently in an early stage of the Ordinary Legislative Procedure, and its implementation timeline is yet to be determined.

obligations vis-à-vis its data subjects upon changing the legal basis (see Annex D: *Data Subjects' Rights*), the cohort should be able to switch to a different GDPR legal basis to share the data with external researchers, obviating the need for a new consent from its data subjects.

Lastly, it is worth reiterating the different GDPR compliance obligations applicable to controllers and processors. Compliance with the principles of the GDPR, including the principle of lawfulness, is the sole responsibility of the controller. Because the parties operating the EPND's Technical Hub, at least in the short term, will be acting as processors intermediating between controllers, the EPND consortium will be under no obligation to demonstrate that processing of personal data takes place in accordance with the principles of the GDPR. As such, the simplest solution for the EPND would be not to inquire about the cohort's legal basis for processing data. The relevant EPND consortium partners could simply enter in a data processing agreement with the cohort – for example, structured as a data deposit or data transfer agreement – whereby the cohort declares that it deems the transfer of its data to the EPND lawful under the GDPR. Although this approach seems appealing from the EPND's point of view due to its contractual simplicity, it comes with two important disadvantages. First, because the datasets included in the EPND's Technical Hub on the basis of data subjects' consent are not readily reusable to external researchers, accepting such datasets without taking due care to appropriately label their use restrictions may give rise to problematic situations. External researchers who discover the dataset on EPND's cohort catalogue need to be made aware whether the data can be lawfully accessed and used for the intended research project. In the absence of this information, researchers may initiate a formal data access request, only to discover much later that the data are not available due to the restrictions stemming from the GDPR legal basis. EPND's failure to proactively disclose this information will likely expose the EPND to reputational risks. Second, the decision to distance itself from the controllers' compliance obligations may run counter to the EPND's aspiration to provide the participants of its ecosystem – most notably cohorts – with guidance and ongoing support. By choosing not to engage cohorts in the context of cohorts' legal compliance, the EPND may miss out on an opportunity of providing value-added services that would differentiate the EPND from other data-sharing initiatives in the space of NDs. Determining how to navigate these complex interactions with cohorts in relation to cohorts' GDPR compliance will be one of the key priorities for Work Package 2 (Legal and Ethical Regulations) of the consortium during Phase II (years 2022-2026) of the IMI project.

A task carried out in the public interest (Art. 6(1)(e) GDPR)

According to the EDPB, a task carried out in the public interest (Art. 6(1)(e) GDPR) is a more suitable legal basis than the data subject's consent (Art. 6(1)(a) GDPR) for data processing operations in the context of scientific research (European Data Protection Board 2021b). Additionally, Art. 6(1)(e) GDPR legal basis offers advantages also over Art. 6(1)(f) GDPR, *legitimate interest*. While both Art. 6(1)(e) and Art 6(1)(f) legal bases obviate the need for reconsenting data subjects prior to future instances of data reuse, the Art. 6(1)(e) legal basis comes with additional characteristics that can potentially further reduce the burden associated with the controller's GDPR compliance. Most notably, as it will be discussed in *Annex F Data Subjects' Rights*, the controller relying on Art. 6(1)(e) GDPR

to process personal data can usually benefit from certain exemptions and derogations from its obligations vis-à-vis its data subjects.

When processing biomedical data for research purposes, Art. 6(1)(e) is to be coupled with Art. 9(2)(j) GDPR, which concerns conditions under which special categories of data can be processed for scientific research (Chen, Dove, and Bhakuni 2021). However, it should be highlighted that there are limitations as to the nature of controllers as well as contexts where the two Articles can be relied upon to legitimise the processing. More specifically, in order for a controller to rely on Art. 6(1)(e) in conjunction with Art. 9(2)(j), additional applicable law(s) must be in place giving the controller a mandate to process special categories of personal data for research purposes in the public interest. The law(s) giving the controller this mandate may be either European (i.e., EU regulations) or national law(s) of a Member State where the controller is based (Arts. 6(1)(e), 6(3), and 9(2)(j) GDPR). However, at present, there is no European regulation applicable to the context of the EPND (i.e., retrospective research using existing data collections, as opposed to prospective clinical trials, which is regulated under the EU Clinical Trials Regulation) conferring such a mandate on designated controllers. The forthcoming EHDS Regulation is intended to address this regulatory gap at the European level, creating a pathway for eligible controllers to rely on GDPR Art. 9(2)(j) to lawfully process special categories of personal data for the purposes of biomedical research. However, until the EHDS Regulation comes into effect, the availability of this exemption (i.e., Art. 9(2) GDPR) to the participants of the EPND ecosystem is poised to remain contingent upon the national laws of the country where each participant is based.

To date, not all Member States (and other countries where the GDPR applies to full extent) have implemented national laws that would enable controllers from these countries to rely on GDPR Arts. 6(1)(e) and 9(2)(j). Moreover, even where a suitable national law is in place, its scope may be limited to certain categories of research institutions, such as public universities (Lindén, Kelli, and Nousias 2019). As a consequence, other categories of controllers from the same country, notably, private entities, will not be able to rely on the Art. 6(1)(e) legal basis in conjunction with the Art. 9(2)(j) when pursuing scientific research.

Due to the above-mentioned considerations, the landscape of the national legal frameworks regarding the applicability of GDPR Arts. 6(1)(e) and Art 9(2)(j) has grown highly heterogenous in recent years. Given that the landscape is rapidly evolving, coupled with the European legislator's clear intention to eliminate this heterogeneity altogether through the proposed EHDS Regulation, the present document does not discuss the current situation in specific European countries (For a comprehensive, albeit less up-to-date overview of the availability of GDPR legal bases, see the 2021 report by European Commission Directorate-General for Health and Food Safety).[22] As a general conclusion, it can be stated that GDPR Art. 6(1)(e) coupled with Art. 9(2)(j) is theoretically the most suitable legal basis for the controllers participating in the EPND ecosystem. However, whether and to what extent the controllers can rely on this legal basis in practice, currently varies across countries and the types of legal entities.

[22] Available at: https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf

Legitimate interest (Art. 6(1)(f) GDPR)

Where a controller cannot rely on a law to process personal data for scientific research purposes under Art. 6(1)(e) GDPR, the controller may consider using legitimate interest (Art. 6(1)(f) GDPR) as an alternative legal basis. To rely on the Art. 6(1)(f) GDPR legal basis, the controller must identify a legitimate interest for which processing is necessary. Subsequently, the controller must balance this legitimate interest against the interests, fundamental rights, and freedoms of the data subject, taking into account, among other factors, reasonable expectations of the data subject (Recital 47 GDPR). This assessment, also known as the balancing test (Kamara and De Hert 2018; Information Commissioner's Office (ICO)) is necessary for ascertaining that no overriding data subjects' interests exist that would prevent the controller from relying on the Art. 6(1)(f) GDPR legal basis. As a general rule, for research projects that meet Art. 89(1) GDPR requirements (technical and organisational measures), and comply with Art. 5 principles of the GDPR, the balancing test will typically yield a positive outcome, enabling the controller to base the intended processing on Art. 6(1)(f) GDPR.

When processing special categories of data, the EDPB has identified Art. 9(2)(j) GDPR as a suitable additional condition to be used in conjunction with Art. 6(1)(f) GDPR. According to Art. 9(2)(j) GDPR, the controller will be able to invoke this condition when "*processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued ...*". As both Art. 6(1)(e) GDPR and Art. 6(1)(f) legal bases are commonly complemented by Art. 9(2)(j) GDPR to legitimise scientific research using special categories of data, they both need to be grounded in a "Union or Member State law". However, it should be emphasised that the relevant "Union or Member State law" may be different, depending on whether a controller processes the data based on the Art. 6(1)(e) or the Art. 6(1)(f) legal basis. For example, as discussed previously, in a given Member State, the national implementation of the GDPR may limit the applicability of the Art. 6(1)(e) legal basis to publicly funded research institutions. This means that a private research organisation from the same country cannot rely on the same national law, and in the current absence of an equivalent European regulation, will not be able to use the Art. 6(1)(e) legal basis. Notably, the proposed EHDS Regulation explicitly states that while it provides for suitable Art. 9(2) GDPR exemptions to enable processing of special categories of data (including Art. 9(2)(j) GDPR), it does not provide for the legal basis in the sense of Art. 6(1) GDPR (Recital 37 EHDS Regulation proposal). Consequently, the private research organisation may choose to explore the Art. 6(1)(f) legal basis, leading to a situation where even though both organisations are relying on the same Art. 9(2) GDPR exemption – and indeed the same law providing for this Art. 9(2) GDPR exemption – they differ in their choice of the Art. 6(1) GDPR legal bases. As a consequence, the two controllers processing special categories of personal data for research purposes will likely be subject to different data protection obligations. For example, the private organisation relying on the Art. 6(1)(f) GDPR legal basis would need to take additional steps to comply with its GDPR obligations, including performing a balancing test, and communicating its legitimate interest to the data subject.

Another potential limitation is that the Art. 6(1)(f) legal basis is not available to public authorities, where the intended processing is pursued as part of the performance of their tasks (GDPR Art. 6(1), *sub-para 2*). In some countries, research organisations affiliated with government agencies such as

ministries of health, research, or science and education, will also qualify as public authorities for legal purposes. Depending on the specific provisions in the relevant national law, such organisations may include public research universities. Where this is the case, the research university cannot rely on the Art. 6(1)(f) legal basis to process personal data for research purposes.

Annex D: Data Subjects' rights

Chapter III of the GDPR (Arts. 12 - 23), *Rights of the data subject,* lays down a set of specific rights, alongside the conditions for their applicability and potential exemptions or derogations, afforded by the Regulation to data subjects. Grounded in the principles of the Regulation, and reflecting controllers' obligations towards data subjects, these rights are instrumental for ensuring the protection of the data subjects.

In the context of data processing for research purposes, the following rights of the data subject are of particular relevance:

- Right to be informed (Arts. 13 and 14 GDPR)
- Right of access (Art. 15 GDPR)
- Right to rectification (Art. 16 GDPR)
- Right to erasure ('right to be forgotten') (Art. 17 GDPR)
- Right to restriction of processing (Art. 18 GDPR)
- Right to data portability (Art. 20 GDPR)
- Right to object (Art. 21 GDPR)

It should be highlighted that whether, and to what extent, the aforementioned rights apply in a given context will depend on various factors, including the selection of the Art. 6(1) legal basis in conjunction with the Art. 9(2) legitimation, and the national implementation of the GDPR in the country where the controller is based.

This annex (Annex D) explores in detail the GDPR rights of data subjects, with a focus on their applicability and relevance to data reuse scenarios facilitated by research intermediaries such as the EPND. A summary-level overview of the potential restrictions to, or derogations from, the data subjects' rights is captured in Table 7.

The right to be informed

Data subjects' right to be informed about processing is grounded in the principle of transparency and is embedded in multiple articles of the GDPR, most extensively in Arts. 13-14 of the Regulation.

The GDPR emphasises the crucial importance of providing data subjects with clear, concise, and meaningful information about data processing in a timely manner. For example, Recital 39 of the Regulation stresses that "[natural] persons should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing". It should be noted that the right to be informed is critical for effectuating other rights of the data subjects, since in the absence of meaningful information referred to in Recital 39, data subjects will not be fully aware of the processing and will therefore be unable to exercise their rights. This makes the right to be informed an important factor enabling compliance with other GDPR principles, primarily the principles of fairness and lawfulness of processing (Kindylidi and Antas de Barros 2021; Staunton 2021).

Article 13 GDPR concerns data subject's right to be informed where personal data are collected directly from the data subject, whereas Article 14 GDPR applies where personal data are not obtained from the data subject (e.g., obtained from a third party that already holds the personal data relating to the data subject). Articles 13 and 14 GDPR lay down mandatory minimum disclosure requirements, listing the types of information the controller must provide to the data subject prior to engaging in the intended processing. Although the disclosure requirements under the two articles overlap to a significant extent, they also contain several differences that reflect the circumstances specific to each data collection context. In particular, where data are not collected from the data subject, the controller is required to inform the data subject about the source(s) and categories of the personal data it intends to process. This information is not explicitly required where the data are collected directly from the data subject. The table below (adapted from Kindylidi & de Barros 2021) summarises the types of information disclosure requirements under Articles 13 and 14 GDPR.

Table 6 - Information disclosure requirements for controllers vis-à-vis their Data Subjects

Type of information to be provided	Article 13 *	Article 14 **
Name and Contact details of Controllers	Required	Required
Name and Contact Details of Representative	If Applicable	If Applicable
Name and Contact Details of DPO	If Applicable	If Applicable
Purpose of Processing	Required	Required
Lawful Basis for Processing	Required	Required
Legitimate Interests for Processing	If Applicable	If Applicable
Categories of Personal Data Obtained	**Not Required**	Required
Recipients or Categories of Recipients of Data	If Applicable	If Applicable
Transfer of Data to Third Countries	If Applicable	If Applicable
Retention Period	Required	Required
Right(s) Available to Data Subjects	Required	Required
Right to Withdraw Consent	If Applicable	If Applicable
Right to Lodge a Complaint with a Supervisory Authority	Required	Required
Source of the Personal Data	**Not Required**	Required
Information Regarding Automated Decision-Making Profiling	If Applicable	If Applicable

Adapted from Kindylidi & de Barros, 2021

*Article 13 requirements do not apply if the relevant controller can demonstrate that the data subject already has the information

** Article 14 requirements do not apply if the relevant controller can demonstrate that one of the conditions under Art. 14(5) GDPR is met

Importantly, the GDPR prescribes neither the means nor the format for providing the requisite information to data subjects. For example, Recital 60 of the Regulation mentions the possibility of utilising visual aids such as standardised machine-readable icons to enhance the effectiveness of communication with data subjects, without expressly mandating the use of such tools by controllers. Additionally, under Article 14, the GDPR affords some discretion to controllers in regard to the timing for informing data subjects about the intended processing. Accordingly, the transparency guidelines issued by the Article 29 Working Party (A29WP), and subsequently endorsed by the EDPB, emphasise that controllers will need to decide on the practical implementation aspects of Articles 13 and 14 GDPR based on factors such as the nature of the intended processing and the implications for the data subjects (Article 29 Data Protection Working Party 2018b).

Arguably, the most consequential difference between Articles 13 and 14, in relation to data processing for scientific research purposes involving data sharing among controllers, concerns the conditions under which a controller is exempt from disclosure requirements. Under Art. 13 GDPR, disclosure requirements do not apply when the data subject already has the information. By contrast, Art. 14 GDPR envisages additional exemptions to disclosure requirements, including where "the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes ..." (Art. 14(5)(b) GDPR). The absence of an equivalent provision in Art. 13 GDPR has been interpreted to mean that the limitation to disclosure obligations vis-a-vis data subjects on the grounds of impossibility or disproportionate effort uniquely applies to controllers that obtain personal data from a source other than the data subject (i.e., 'downstream controllers') (Erdos 2021). This distinction is also clearly emphasised in the A29WP Guidance on Transparency, which states that, in the context of Art. 14(5)(b) GDPR, "impossibility or disproportionate effort only arise by virtue of circumstances which do not apply if the personal data is collected from the data subject. In other words, the impossibility or disproportionate effort must be directly connected to the fact that the personal data was obtained other than from the data subject." (Article 29 Data Protection Working Party 2018b)

Importantly, Art. 14(5)(b) GDPR should not be interpreted as affording a blanket exemption from disclosure obligations to downstream controllers. Any downstream controller intending to rely on this exemption to process personal data for scientific research purposes must be able to demonstrate that informing data subjects about the intended processing would indeed prove impossible or require disproportionate effort. A29WP particularly emphasises that to justify its reliance on the "disproportionate effort" exemption, the controller must carry out a balancing exercise and document its outcomes in accordance with the controller's accountability obligations. The balancing exercise must "assess the effort involved for the data controller to provide the information to the data subject against the impact and effects on the data subject if he or she was not provided with the information." (Article 29 Data Protection Working Party 2018b). In carrying out this assessment, the controller must take into consideration factors such as "the number of data subjects, the age of the data and any appropriate safeguards adopted [by the controller]". (Recital 62. GDPR)(Article 29 Data Protection Working Party 2018b). Where the balancing exercise leads the downstream controller to conclude that proactively informing individual data subjects about the intended processing constitutes disproportionate effort, the controller can rely on the exemption under Art. 14(5)(b) GDPR, provided the processing takes place for scientific research purposes and employs appropriate safeguards in line with GDPR Art. 89(1). However, the controller remains responsible for making the information about the processing publicly available (Art 14(5)(b) GDPR). This can be accomplished by, for example, publishing the information on the controller's website, which is one way the controller may demonstrate its compliance with the transparency principle (Erdos 2021).

Of note, there is considerable uncertainty regarding the circumstances that would trigger a controller's obligation to proactively inform the data subject about intended and ongoing processing of previously collected data. One notable exception to this is the case of further processing: when a controller intends to use existing personal data for a purpose other than that for which the controller initially collected the data, the controller is required to notify the data subject (Arts. 13(3) and 14(4) GDPR). However, the GDPR is less prescriptive on other circumstances that would trigger an obligation to inform data

subjects (Article 29 Data Protection Working Party 2018b; Erdos 2021). A29WP adopted the view that any intended changes to data processing that impacts on data subjects in a material way – for example, by affecting data subjects' ability to exercise their rights under Articles 15-22 GDPR - should be communicated to data subjects in a manner that is transparent and respects the principle of fairness (e.g., "well in advance of the change actually taking effect", allowing the data subjects sufficient time to opt out of the intended future processing). A reasonable interpretation of this guidance is that material changes such as the transfer of data to a new controller, or identification of a new Art. 6(1) GDPR legal basis would typically trigger an obligation to inform data subjects even where the data is being processed strictly for the purpose(s) that motivated the initial data collection (i.e., no further processing takes place). However, it is unclear whether, and to what extent, more technical or operational changes to the processing must also be communicated to the data subjects under the principles of fairness and transparency. Such changes could include, for example, the implementation of a new data encryption solution that alters the risk profile of the ongoing processing. Moreover, A29WP also stresses that even in the absence of material changes to how processing is conducted, the controller should be mindful that data subjects may not be able to retain information over extended periods of time. Consequently, where personal data are processed on an ongoing basis, the principle of fairness means that controllers may have to "re-acquaint data subjects with the scope of the data processing, for example by way of reminder of the privacy statement/ notice notified at appropriate intervals" (Article 29 Data Protection Working Party 2018b). The approach espoused by the A29WP emphasises the controller's ongoing responsibility to assess the need for notifying data subjects about current and planned processing activities. However, apart from the clear-cut cases (e.g., undertaking further processing, transferring data to a new controller, and/or changing the Art. 6(1) GDPR legal basis), the decision whether a disclosure obligation applies under the transparency and fairness principles, will require the controller's judgement.

Notwithstanding the legal uncertainty outlined in the preceding paragraph, the following generalised conclusions can be made regarding controllers' disclosure obligations towards data subjects under Articles 13 and 14 GDPR. First, downstream controllers obtaining existing data from other controllers are, in principle, afforded more exemptions from mandatory disclosure obligations, compared to initial controllers who collect the data directly from data subjects (Article 29 Data Protection Working Party 2018b; Erdos 2021). Second, the initial controller is required to inform data subjects when the controller: i) intends to further process the data for a purpose other than the purpose(s) for which the personal data were obtained; ii) plans to introduce a change to the processing that materially impacts on data subjects' rights, including what rights continue to apply and how they can be exercised by the data subject; iii) deems that it is warranted to re-acquaint the data subjects with the nature of the ongoing processing to ensure that data subjects remain adequately informed.

These conclusions have important implications for the EPND. It is significant that the mandatory disclosure obligations appear to disproportionately affect EPND cohorts, who are the parties responsible for sample and data collection from patients and research participants, making them data controllers subject to Art. 13 GDPR. By contrast, external researchers who intend to use existing sample and data collections for a new research project would qualify as downstream controllers subject to Art. 14 GDPR, including its wider range of possible exemptions under Art.14(5)(b). Consequently, cohorts willing to share their data through the EPND are required to notify their data subjects under

certain circumstances and cannot rely on any in-built provisions within the GDPR to seek exemption from the obligation. Importantly, cohorts, as controllers, have the sole responsibility to decide whether a particular event triggers the obligation to notify data subjects under the principles of transparency and fairness. It is therefore conceivable that some cohorts may choose to allow research re-use of their data through the EPND's Technical Hub without deeming it necessary to notify their data subjects. It is important for the EPND to accommodate cohorts with divergent requirements and perceived needs in relation to their Art. 13 GDPR transparency obligations vis-a-vis their data subjects. For example, cohorts requiring frequent transparency notifications to inform their data subjects about current and planned data processing could benefit from the availability of additional EPND services, such as periodic and/or event-driven updates from the EPND, as relevant, capturing the content elements required under Art. 13 GDPR in an exhaustive manner. It may also be worthwhile to explore potential digitally-enabled solutions that the EPND could employ to support cohorts in their communication efforts vis-a-vis data subjects. Such ancillary services would be particularly valuable where a cohort's GDPR legal basis for the initial data collection was data subjects' consent (Art. 6(1)(a) GDPR), and the cohort deems the existing consent inadequate for enabling data reuse through the EPND, thus necessitating a new consent or a change in the legal basis accompanied by appropriate transparency disclosures.

The right of access and the right to data portability

Apart from mandatory disclosure obligations to which controllers are subject under Articles 13 and 14 GDPR, the Regulation also affords data subjects other informational rights. In this respect, Art. 15 GDPR (right of access) and Art. 20 (right to data portability) are of particular relevance.

Art. 15 GDPR entitles the data subject to obtain, upon request, information from the controller concerning processing of their personal data by the controller. This information includes the purposes of processing, the categories of the personal data being processed, and the identities of additional data recipients with whom the controller has shared the data.

According to Art. 15 GDPR, the data subject is also entitled to access the entirety of the data concerning him or her being processed by the controller. Upon the data subject's request, the controller is additionally required to provide the data subject with a copy of this data. The Article further specifies that where requests to access data are made electronically, the data should be provided "in a commonly used electronic form" (Art. 15(3) GDPR). However, in an interpretive guidance concerning the right of access, the EDPB emphasises that the format and medium of communication will depend on the request. In some cases, particularly where large amounts of data are processed, the EDPB recommends providing the data in a structured, layered format, in order to aid comprehensibility (European Data Protection Board 2021a).

The EDPB further stresses that the controller is required to comply with access requests in a manner that does not adversely affect the rights of others. For example, where appropriate, the controller will need to assess whether the data to be provided to a data subject in response to an access request contains personal data concerning another data subject. However, the controller cannot deny the requestor data subject the right of access solely on the grounds that the data under consideration includes personal

data of another natural person. Rather, the controller is expected to consider ways to comply with the request in a manner that balances the rights of the parties concerned. For example, the controller may choose to provide an edited copy of the data where the data pertaining to another data subject has been deleted or redacted (European Data Protection Board 2021a).

The EDPB recognises that the controller may incur costs or other administrative burden when complying with data subjects' access requests. However, the EDPB notes that the controller should be prepared to comply with access requests, including providing a complete copy of the data subject's personal data undergoing processing, free of charge for the data subject. While, under certain circumstances, the controller may charge the data subject a reasonable fee for access, this should be limited to the cases where the controller can demonstrate that the access request(s) from the data subject have been "manifestly unfounded or excessive", or where the data subject has requested multiple copies of the data. By contrast, the EDPB explicitly states that the "controller must ensure that the first copy is free of charge, even where it considers the cost of reproduction to be high" (European Data Protection Board 2021a).

The right to obtain a copy of personal data under Art. 15 GDPR shares some similarities with the right to data portability (Art. 20 GDPR). Data portability is defined in the GDPR as the data subject's right to "... receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format …" (Art. 20(1) GDPR). However, the two rights are not identical and differ in several significant ways.

First, the right to data portability only applies where the legal basis for data processing is either the data subject's consent (Art. 6(1)(a) GDPR, in conjunction with Art. 9(2)(a) GDPR when processing special categories of data) or performance of a contract (Art. 6(1)(b) GDPR). This means that in the context of the EPND, the right to data portability will not be applicable where personal data are processed based on public interest or legitimate interest, that is, Arts. 6(1)(e) and 6(1)(f), respectively.

Second, there is a general consensus that while the right of access concerns the entirety of the personal data undergoing processing, the right to data portability only applies to: i) the data that the data subject has directly provided to the controller; and ii) derived or inferred data, insofar as generation of this data is inextricably linked to and is a direct consequence of the service the data subject is using (e.g., generation of physical activity tracking data resulting from the use of a wearable device by the data subject) (Article 29 Data Protection Working Party 2017). This limits the scope of the data to be provided to a data subject following the data subject's request to obtain a copy under Art. 20 GDPR. For example, in the context of the EPND, the categories of data within the scope may include demographic, medical history, or lifestyle data provided directly by the data subject through, for example, medical questionnaires. It may also include raw data generated following a medical evaluation procedure, such as pre-processed imaging data or (uninterpreted) genomic data. However, the output of any subsequent transformation or analysis of this data would fall beyond the scope of Art. 20 GDPR. This is different from the scope of Art. 15 GDPR, which concerns all types of personal data undergoing processing.

Third, and perhaps most importantly, Art. 20 GDPR refers to providing a copy of data in a "structured, commonly used and machine-readable format". There is a general consensus that the intention of the legislator behind this requirement was to enable data subjects to freely transfer their personal data across controllers, thus improving competitive practices within the European Union. However, in the healthcare sector, the implementation of the right to data portability has been hindered by the lack of interoperability standards. As a consequence, even where Art. 20 GDPR applies in principle, it may not be clear what concrete steps a controller must take to effectuate data subjects' right to data portability, limiting the extent to which the right can be exercised in practice. One of the stated objectives of the forthcoming European Health Data Space (EHDS) regulation, whose draft proposal was published in May 2022, is to enable data subjects to exercise their data portability right, by fostering common European standards and interoperability of health data (European Commission Directorate-General for Health and Food Safety 2022). However, at present, compliance implications of the right to data portability for controllers such as EPND cohorts and biomedical research organisations remains unclear.

Right to rectification, erasure, restriction of processing, and the right to object

The GDPR lays down additional rights aimed at enabling the data subject to exercise a degree of control over how their data are processed following initial data collection.

The right to rectification (Art. 16 GDPR) refers to the right of the data subject to have "inaccurate personal data concerning him or her" amended by the controller. However, there is considerable uncertainty as to the scope of the right. In particular, it has been noted (e.g., Dimitrova 2021) that there is a tendency to conflate "data accuracy" with the broader notion of "data quality". One scenario in which this conflation is relevant for the EPND may arise where a data subject, after accessing their data under Art. 15 GDPR, requests rectification of health data, such as medical diagnosis. The data subject may justify this request on the grounds that based on a recent, more reliable diagnostic procedure, the medical information currently held by the EPND cohort is no longer up to date.

However, Art. 16. GDPR also notes the relevance of taking into account the purposes of the processing in interpreting the right to rectification. Therefore, in the context of research, it may be more appropriate to conceptualise the right more narrowly, allowing the controller to limit the applicability of the right, if deemed appropriate in view of the purposes of the processing. For example, the data subject may have the right to have contact details or other operational data rectified, but not the biomedical data derived through research analysis. This may be particularly relevant where the controller needs to retain unmodified data for the purposes of enabling study reproducibility, or where for the consistency of data analysis it is important to ensure that future analyses use the same data.

The right to erasure (Art 17. GDPR) is commonly conceptualised in the biomedical research discourse as a research participant's right to have their personal data removed through, for example, withdrawing their consent to study participation. Importantly, the right to erasure is not limited to the situations where personal data are processed based on the Art. 6(1)(a) GDPR legal basis. It may also apply to processing under Articles 6(1)(f) and 6(1)(e) GDPR, albeit here, the right to erasure is contingent on

the applicability of the right to object (Art. 21 GDPR), and must be preceded by the data subject's expressed objection to processing in accordance with Art. 21(1) GDPR.

It is generally recognised that there are inherent practical limitations to the right to erasure in the context of scientific research. For example, it may not be feasible to retroactively remove the data from the data analyses already completed. Furthermore, retention of personal data may be necessary to ensure the reproducibility of research (e.g., Peloquin et al. 2020). These considerations are also reflected within the GDPR, with Art. 17(3)(d) providing an in-built restriction of the right where processing is carried out for scientific research purposes in accordance with Art. 89(1) GDPR.

The right to restriction of processing (Art. 18 GDPR) complements Art. 17 GDPR, by allowing data subjects to obtain restriction of processing without data deletion. Article 18 refers to two cases in which the right to restriction of processing serves as an alternative to data deletion: i) where processing is unlawful but the data subject opposes data deletion; and ii) where the data are no longer needed, but the data subject requires the data for the purpose of the establishment, exercise or defence of legal claims. Of note, unlike the right to erasure, there are no in-built restrictions to this right where data processing takes place for scientific research purposes. Consequently, the right to restriction of processing would normally apply unless scientific research is grounded in a national law that derogates from the right to restriction of processing based on Art. 89(2) GDPR.

The right to object (Art. 21 GDPR) applies specifically where data processing is grounded in the legal bases under Art. 6(1)(e) or 6(1)(f) GDPR. Following the notification of an objection to processing from a data subject under Art. 21 GDPR, the controller has three options: i) to comply with the request and no longer process the data; ii) to demonstrate that there is a compelling legitimate ground for processing which override the interests, rights and freedoms of the data subject; iii) to demonstrate that continued processing is necessary for the establishment, exercise or defence of (the controller's) legal claims. Importantly, the GDPR makes it explicit that the right to object applies to processing for scientific research purposes (Art. 21(6) GDPR), "unless the processing is necessary for the performance of a task carried out for reasons of public interest" (Art. 21(6) GDPR). An implication of this provision is that the right to object, similarly to the right to restriction of processing, can only be restricted based on a law that makes use of Art. 89(2) GDPR derogations.

Derogations from data subjects' rights based on a Member State laws

Art. 89(2) GDPR provides grounds for Member States[23] to adopt national laws derogating from certain data subject rights where processing takes place for scientific research purposes.

[23] NB: Art. 89(2) refers to a "Union or Member State law", meaning that in principle, these derogations can also be allowed under a European regulation. However, as of 2022, there is no European regulation relevant to the context of the EPND that implements Art. 89(2) GDPR derogations from certain data subjects' rights. Although this reality will likely change in the future, particularly in view of the forthcoming EHDS Regulation, the present paper focuses exclusively on the national laws, as many countries have already incorporated Art. 89(2) provisions into their national laws.

Notably, Article 89(2) explicitly mentions derogations from only a subset of rights. Namely, it lists the following Articles that can be derogated from based on national laws: Art. 15 (the right of access), Art. 16 (the right to rectification, Art. 18 (the right to restriction of processing), and Art. 21 (the right to object). Article 89(2) does not refer to other rights of data subjects, which can be interpreted as the legislator's intention to exclude some of the rights from the scope of derogations. However, this appears incongruous with Recital 156 GDPR, which provides a more comprehensive list of data subject rights that can be derogated by national laws governing scientific research (Shabani and Borry 2018).

Furthermore, some uncertainty is introduced by Art. 23 GDPR, which states that data subjects' rights can be restricted based on a Union or national law, while Art. 23 concerns all of the rights of data subjects under Chapter III GDPR. According to Art. 23(1), one of the aims of a Union or national law restricting data subjects' rights may be to safeguard "important objectives of general public interest" (Art. 23(1)(e) GDPR). However, the article does not explicitly include scientific research as part of "important objectives of general public interest". This has created uncertainties around the Legislator's intention to differentiate processing falling under this article from processing carried out in public interest within the meaning of Art. 6(1)(e) GDPR, where the latter may include scientific research. Should "important objectives of general public interest" be interpreted in a manner that includes scientific research, it might be possible to adopt laws governing research that derogate from the data subjects' rights beyond those referred to in Article 89(2) GDPR. On the other hand, a narrower interpretation would mean that only the rights explicitly listed in Article 89(2) can be derogated from (Shabani and Borry 2018; Mourby et al. 2019).

To summarise, the table below provides a comprehensive overview of the potential restrictions of data subjects' rights in the context of scientific research. There are two forms of such restrictions: i) in-built exemptions embedded within the GDPR itself; and ii) derogations based on Member States' laws. The former, in-built exemptions can in principle be relied upon by any controller processing personal data for scientific research purposes, provided that the controller is compliant with the general organisational and technical requirements under Art. 89(1) GDPR. By contrast, derogations from data subjects' rights are not available to all controllers; the derogations only apply if, and to the extent that, the Member State in which a controller is based has implemented a law derogating from data subjects' rights, as permitted under Art. 89(2) GDPR.

Table 7 - Restrictions to Data Subjects' Rights under the GDPR

Right	GDPR Article	In-built restriction for research	Derogation based on a Member State law*
Right to be Informed *(data collected from the data subject)*	Art. 13	-	-
Right to be Informed *(data not collected from the data subject)*	Art. 14	Art. 14(5)(b): the right can be restricted where "disproportionate effort" is required	-
Right of Access	Art. 15	-	Foreseen under Art. 89(2)
Right to Rectification	Art. 16	-	Foreseen under Art. 89(2)
Right to Erasure (Right to be Forgotten)	Art. 17	Art. 17(3)(d): the right doesn't apply where processing is for scientific research	-
Right to Restriction of Processing	Art. 18	-	Foreseen under Art. 89(2)
Right to Data Portability	Art. 20	-	-
Right to Object	Art. 21	-	Foreseen under Art. 89(2)

* The availability of the derogations varies across Member States. It depends on whether, and to what extent, the data protection laws in a given country make use of the Art. 89(2) prerogative to explicitly derogate from these rights in the context of scientific research.

Of note, even among the countries whose national laws make use of the Art. 89(2) GDPR prerogative, the extent to which the relevant data subjects' rights are derogated from may be different. One example of this is Norway, where the Art. 89(2) GDPR prerogative has been implemented in a manner that derogates from some, but not all of the rights mentioned in Art. 89(2) (Befring 2021). Such differences across countries once again highlight the importance of understanding subtle, yet consequential ways in which the diverging national implementations of the GDPR may affect various aspects of data processing in a cross-border context.

Implications for controllers' obligations vis-a-vis data subjects

Our analysis suggests that where personal data are processed in complex environments involving multiple controllers, there are important differences in the GDPR compliance obligations of the different controllers vis-a-vis data subjects. In particular, these obligations appear to disproportionately fall on the initial controller, that is, the party that collected personal data directly from the data subject. There are generally fewer exemptions or derogations from data subjects' rights available to the initial controller, compared to downstream controllers who do not collect personal data directly from the data subjects. For example, mandatory informational disclosure requirements under Art. 13 GDPR appear to always apply in principle, which implies that the initial controller will generally have the obligation to recontact data subjects prior to introducing material changes to data processing, and/or provide periodic notifications concerning ongoing processing. This could be problematic in the context of scientific research, given that recontacting research participants is widely considered as administratively burdensome and is often perceived as a significant practical barrier to continued

research use of personal data (Mitchell et al. 2020). Furthermore, as discussed previously, it is unclear to what extent Union or national laws can derogate from Art. 13 obligations. On the other hand, downstream controllers who process personal data for research purposes can rely on several potential exemptions from disclosure obligations, as per Art. 14(5)(b) GDPR.

In complex data processing environments, the initial controller is also responsible for supporting data subjects in exercising their rights with respect to other controllers (Art. 19 GDPR). When a data subject contacts the initial controller with a request to exercise their rights, particularly the rights under Articles 16, 17, and 18 GDPR, the initial controller is required to notify all recipients of the personal data, including downstream controllers, of the data subject's data. The obligation to notify data recipients applies unless compliance is "impossible or involves disproportionate effort" (Art. 19 GDPR). A common interpretation of the Article 19 obligation, at least in the context of biobanking, is that the controller is required to notify those recipients with whom the controller has directly disclosed the data, whereas notification of any subsequent recipients, if possible at all, would typically amount to disproportionate effort (Staunton 2021). It is important to further emphasise that in practice, support from the initial controller is necessary for exercising even those data subject rights that are not explicitly mentioned in Art. 19 GDPR. Downstream controllers may have no means to authenticate the identity of the data subject; Art. 11 GDPR expressly states that controllers are not required to store information needed for re-identifying data subjects for the sole purpose of complying with data subjects' potential future requests under Articles 15-20 GDPR. Because scientific research commonly relies on robustly pseudonymised data, downstream controllers are unlikely to collect information required for the authentication of data subjects' identities. As a consequence, it is probable that the involvement of the initial controller will be necessary for enabling the exercise of all rights of the data subject.

Annex E: The Three Pillars

The three pillars of legal and policy frameworks governing research use of existing sample and data collections in biomedical research concern the following areas:

- Research ethics (specifically, ethics of research involving human subjects)
- Biobanking (i.e., collection, storage, and treatment of human biological samples); and
- Privacy and data protection

Research Ethics Laws and Policy Frameworks

The majority of European countries have national laws around the conduct of biomedical research in humans. These laws are designed to provide a clear legal framework for the implementation of the fundamental ethical principles centred on ensuring the protection of, and respect for, research participants. Broadly speaking, such laws define the norms, practices, and substantive and procedural requirements for the conduct of legally compliant research. They also commonly lay down the process whereby research proposals are reviewed, approved, and monitored for their ethical and legal compliance. This involves establishing a formal mandate for competent ethics review bodies, such as local, regional, and/or national Research Ethics Committees (RECs)[24] tasked with evaluating the soundness and ethical-legal compliance of research proposals. Additionally, national research ethics laws typically prescribe the conditions and requirements for a valid informed consent for research participation. However, informed consent requirements are less straightforward in the context of retrospective research, and may vary considerably across countries, governance frameworks, and the nature of the intended retrospective research project (see Annex F: Informed Consent and Ethics Approval).

Although European research ethics laws share the same foundational principles and core objectives, they differ in significant ways relevant to research platforms such as the EPND. In particular, retrospective research using existing sample and data collections is not explicitly covered in the national research ethics laws in some European countries. This is, for example, the case in France, where the national legislation on scientific research involving human subjects, known as the Jardé law, defines specific categories of biomedical research that it regulates. Notably, retrospective research based on existing sample and data collections does not fall under any of these categories, effectively leaving it beyond the scope of the Jardé law (Souche et al. 2022). Similarly, the corresponding national law in the Netherlands, the Medical Research Involving Human Subjects Act (WMO), does not cover retrospective research where no new samples and/or data are obtained from research participants

[24] In this paper, we use the term REC to refer to any committee, board, or a distinct legal entity with a formal mandate of assessing the ethical-legal compliance aspects of a research proposal. This includes institutional review boards (IRBs) embedded within research organisations, as well as dedicated regional and/or national ethics review bodies/authorities designated by the law. However, RECs should be generally distinguished from data (and sample) access committees, or D(S)ACs: whilst RECs are mandated to ascertain the overall ethical-legal compliance of proposed research studies, D(S)ACs decide on the permissibility of using a particular data and/or sample collection in a given study, alongside determining any collection-specific data or sample use restrictions the study must observe, if applicable.

(CCMO - Central Committee on Research Involving Human Subjects). Importantly, even though retrospective biomedical research involving existing sample and data collections is not explicitly regulated by the national research ethics laws in all European countries, some of the key aspects of the practice will necessarily fall under other European or national laws. In particular, in Europe, the processing of personal data as part of retrospective research is within the scope of the GDPR, alongside other data protection laws complementing the GDPR, if available in a given country. At the same time, research use of human biological sample collections can be addressed by national biobanking and human tissue laws. As discussed in the next section of this annex, large number of European countries have such laws, providing a dedicated legal framework for the retrospective research use of human biological samples (and, in many cases, of the associated data).

However, comprehensive national biobanking/human tissue laws do not exist in all European countries, whereas the national implementations of the GDPR, even where it is available and addresses the use of personal data for scientific research, may not cover more specific organisational and procedural matters, such as establishing a detailed blueprint for the functioning of RECs and/or D(S)ACs. These considerations are true for France and the Netherlands. Since neither country has a national research ethics law explicitly regulating retrospective biomedical research, several key procedural aspects relating to this form of research, such as the ethics review and approval process, are not explicitly addressed by any applicable law in these countries. This results in diverging institutional policies and practices across research organisations within the two countries. For example, as part of our ethical-legal support work during the EPND case studies in 2022, in WP2 we learned that the RECs of the EPND partner institutions based in the Netherlands did not provide a formal approval for retrospective research projects, effectively exempting them from the ethics review process. At the same time, we are aware of several other research institutions in the Netherlands where this type of research is subject to the conventional ethics review process. Similarly, Souche et al. report that in France, although a growing number of biomedical research institutions mandate that retrospective biomedical research proposals undergo a formal review by the institution's REC, considerable within-country heterogeneity persists (Souche et al. 2022).

It should be emphasised that from an ethical point of view, compelling arguments can be made in favour of generally exempting retrospective biomedical research from a traditional REC review process. Retrospective research using existing samples and data is, by definition, of non-invasive nature, posing no physical risks to research participants. While privacy, confidentiality, and data misuse risks exist, these are, to a substantial extent, addressed by the current European data protection regulation. By requiring researchers to demonstrate their compliance with data protection principles, including by adopting appropriate technical and organisational measures, the GDPR and its national implementations arguably provide for effective legal safeguards for mitigating these risks. Therefore, the decision of some European countries - and research institutions therein - to exempt retrospective biomedical research from an ethics review should not be seen as inherently problematic. An alternative approach, whereby retrospective research involving existing biological samples and associated data is effectively subject to the same substantive and procedural requirements as other forms of biomedical research, also exists in some European countries. For example, in Denmark, the Danish Act on Research Ethics Review of Health Research Projects is a broad-scope law whose provisions apply to a wide range of biomedical research projects, including retrospective research on human biological

samples. Although the law envisages certain exemptions in relation to retrospective research, such as the possibility to waive informed consent from donors when existing samples and data are used, these exemptions can only be granted by a relevant REC on a case-by-case basis. Therefore, in principle, all retrospective biomedical research proposals involving existing sample collections are subject to the same ethics review and approval process as higher-risk, invasive research projects (Hartlev 2021; "Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter" 2011). This requirement has been considered as unduly burdensome by members of the Danish biomedical research community. In recognition of these challenges, Danish legislators have initiated the process of drafting a new national biobank law to establish a more favourable legal framework for the conduct of retrospective biomedical research *(Tzortzatou et al. 2023, forthcoming; authors' personal correspondence).*

The absence of a harmonised European legal framework concerning retrospective biomedical research poses considerable practical challenges, particularly where the research is of international nature and entails data/sample sharing across multiple European countries. Researchers may be confronted with vexing, sometimes intractable, interoperability issues where their institutions operate under legal or policy frameworks that are in conflict with those of the partner research institutions based in other countries.

Legal and policy frameworks governing research use of human biological samples

Biobanking can be broadly defined as the set of practices encompassing the collection, storage and use of biological materials such as tissues, cells, blood, and serum (Artene et al. 2013). In the context of NDs specifically, the types of human biological materials used in biobanking also commonly includes cerebrospinal fluid (CSF) and, with the growing interest in the role of human microbiome in NDs, stool and other microbiome-rich samples (Giannella et al. 2021).

In Europe, there is some degree of legal harmonisation of biobanking in healthcare. In particular, human biological materials intended for human applications are subject to the EU Tissue and Cells Directive of 2004 which has been transposed into Member States' legislation. However, governance of human biological materials intended for (non-clinical) research purposes, i.e., where the use of biological materials does not entail their application to a human body, is beyond the scope of the directive. As a consequence, non-clinical biobanking-based research is largely regulated at the national level in Europe. Broadly speaking, there are two ways in which the practice can be regulated in a given country: i) the country may have a specific biobanking law addressing various aspects of the practice within a dedicated legal act. This is currently (as of 2022) the case in Sweden, Finland, Norway, Belgium, Estonia, Spain, and Portugal, among others; ii) the country may not have a dedicated biobanking law, in which case different aspects of biobanking-based research will be addressed by a cluster of European and national laws as well as institutional policies. Examples of such countries include Denmark, France, Germany, Greece, Italy, and the Netherlands (Slokenberga, Tzortzatou, and Reichel 2021).

One of the main advantages of having a dedicated biobanking law is that it provides an opportunity to streamline various aspects of retrospective research on human biological samples under a single comprehensive specialised law. However, in practice, this goal is often undermined by the limited scope of national biobanking laws. In particular, such laws may not cover all types of biological samples used in research or may not apply to all organisations carrying out various biobanking activities. For example, in Estonia, some of the key provisions in the national biobank law do not apply to research involving human blood samples, which has been regulated under a separate Blood Act. Moreover, the Estonian biobank law specifically regulates the activities of the Estonian Biobank, a government-funded research entity established in 1999. Other legal entities carrying out substantively similar activities within the country, despite acting as research biobanks for all practical purposes, are not subject to this law (Pormeister 2021). [25] Similarly, in Sweden, the national biobank law only applies to a subset of biobanks, namely those whose sample collections were generated in a healthcare setting (also known as 'health biobanks'). By contrast, research biobanks are not explicitly covered by the law. The activities of Swedish research biobanks currently fall under a variety of other national laws, including the Swedish Research Ethics Committees Act and the national implementation of the GDPR (Stenbeck, Eaker Fält, and Jane 2021). Due to these scoping limitations, mere existence of a dedicated biobanking law in a country should not be interpreted to mean that the country has a streamlined, comprehensive framework for regulating research using human biological samples. While biobanking-specific laws will offer greater clarity around some issues unique to biobanking research, in other areas, they may need to be complemented by other laws and regulations.

The logic of complementarity also holds true in reverse: in countries where the practice of biobanking-based research is regulated in a fragmentary and incomplete manner (e.g., under data protection and research ethics laws, alongside other legal acts), a dedicated biobanking law can be designed to complement the existing framework specifically in the areas where additional legal clarity is needed. In this context, the development of a national biobanking law can be motivated by the legislator's recognition that the existing applicable legal framework does not adequately account for the unique characteristics of retrospective biomedical research, resulting either in uncertainty over researchers' compliance obligations, or in an overly burdensome regulation that is disproportionate to the minimal risks associated with this type of research. Designing a new biobanking law to complement an incomplete or inadequate national legal landscape has the potential to fill the missing pieces in the legislation, making it more fit for regulating retrospective biomedical research. However, special care must be taken to ensure that the new national biobanking laws don't conflict with the existing applicable laws in other areas, such as those regulating research ethics and privacy/data protection. As discussed in Chapter 3 (Governance View), potential conflicts among these pillars of the relevant legal framework can give rise to significant governance challenges. Apart from creating uncertainties over the precise compliance obligations of researchers, inconsistencies among the relevant national laws may complicate the separation of the roles and competencies of the key parties designated under different laws, including DPOs, data protection authorities, RECs/IRBs, and D(S)ACs.

[25] NB: Estonia is currently (2022) in the process of updating its national biobanking legislation, including abolishing special additional requirements for blood samples.

Finally, it is worth highlighting that national biobanking laws are particularly well-placed to provide legal guidance concerning operational and governance aspects specific to biobanks. This includes standard operating procedures (SOPs) of biobanks, biological sample quality assurance requirements, sample access and use conditions, as well as legal matters such as the ownership of human biological samples. Similar to other areas, there is some heterogeneity across the national biobanking laws along these issues; for example, with respect to the question of sample ownership, some countries (e.g., Finland) designate biobanks as legal owners of samples, whereas in other countries (e.g., Portugal), the sample donor is considered the legal owner (Southerington 2021; Barbosa and Andrade da Costa 2021). Notwithstanding this heterogeneity, the existence of national biobanking laws is generally helpful in clarifying the applicable legal conditions and requirements concerning the governance and day-to-day operations of biobanks. On the other hand, it should also be noted that irrespective of the national legislation applicable in their country, many biobanks choose to adhere to sector-specific international guidelines, best-practice recommendations and, increasingly, relevant ISO standards. This ensures that in practice, there is substantial alignment among biobanks on technical and operational aspects such as SOPs, sample Quality Assurance, and organisation-wide Quality Management Systems. In this respect, the critical role of international research infrastructures as key facilitators should be highlighted. In Europe, the Biobanking and BioMolecular resources Research Infrastructure - European Research Infrastructure Consortium (BBMRI-ERIC) has been instrumental in promoting greater interoperability across European biobanks (Haslacher et al. 2019). Thanks to these efforts driven largely by the research community, even in countries without a dedicated biobanking law, many of the day-to-day operations of biobanks[26] are effectively harmonised with those of their counterparts based in countries with biobanking-specific laws. However, existing differences at the governance level, including diverging sample access and use policies, remain an important challenge in the context of international research collaborations (Simell et al. 2019).

Privacy and data protection

Insofar as retrospective biomedical research makes use of personal data, national research ethics laws and biobanking/human tissue laws are inextricably linked with data protection laws, primarily the GDPR and, where available, its national implementations. Collectively, these laws form the complex legal framework governing retrospective biomedical research in European countries. The key legal issues surrounding processing of personal data for research purposes have been extensively discussed in Chapter 2 (Data Protection View) and its associated Annexes (A-D). The present section focuses on the interaction of the European data protection legislation with the other pillars of the national legal frameworks pertaining to retrospective biomedical research.

Owing to the broad definition of personal data under the GDPR, coupled with a high threshold for demonstrating anonymity, in Europe, processing of personal data should be generally seen as concomitant to retrospective biomedical research. Typically, this will also be true for the research that

[26] Strictly speaking, such entities are likely to be legally categorised as public or private research organisations, rather than biobanks. This is due to the fact that the national legislation in countries without a dedicated biobanking law typically does not recognise biobanks as distinct legal entities. However, for our purposes, the term “biobank” is used in a broader organisational sense that is not equated with its narrow legal meaning.

predominantly focuses on human biological samples, making only limited use of the data (e.g., using robustly pseudonymised identifiers to track the samples; or analysing samples in a targeted manner to generate limited biomarker data of interest). Consequently, the GDPR will apply to a wide range of research activities centred on the use of human biological samples. However, national research ethics laws and, where available, biobanking laws, may provide additional specifications governing certain aspects of data processing in retrospective biomedical research, particularly in the areas where discretion by Member States is allowed under the GDPR. For example, these laws may specify the maximum period during which samples and their associated (personal) data can be retained (e.g., by biobanks), as well as define various conditions and restrictions applicable to data processing for research purposes. Assuming these provisions do not conflict with the principles of the GDPR, they shall be seen as complementary to the Regulation, thus becoming an integral component of the national legal framework concerning data processing for research purposes. Importantly for controllers, where an envisaged data processing operation is explicitly addressed by national research ethics and/or biobanking laws, the controller will be able to rely on these laws when selecting a suitable legal basis for the processing operation under Art. 6(1) GDPR, coupled with a further legitimation under Art. 9(2)(j) GDPR, if processing special categories of data. The choice of the legal basis will be determined by whether the relevant national law gives the controller a mandate, or mission, to perform the processing in the public interest. In the event that such a mandate is defined by the national law, the controller can rely on Art. 6(1)(e) GDPR (performance of a task in the public interest), whereas in its absence, Art. 6(1)(f) GDPR (the controller's legitimate interest) will generally be a more appropriate legal basis. By contrast, in European countries where national research ethics and/or biobanking laws do not contain provisions meaningfully complementing the GDPR, the processing of personal data for research purposes will be predominantly, if not exclusively, regulated under the GDPR and its national implementations. Complex interrelatedness between the European data protection laws and other (especially national) laws regulating biomedical research is further highlighted in Annex F using concrete examples.

Annex F: Informed consent and ethics approval

Informed consent in retrospective biomedical research

Informed consent (IC) is a general requirement for ethically sound research, aimed at respecting core bioethical principles such as the dignity and autonomy of individuals, as well as the fundamental human rights of research participants, including their right to self-determination. The IC can be defined as a process whereby prospective research participants are asked to decide whether they are willing to take part in a scientific study after being provided with all the relevant information that may reasonably influence their decision (Dankar, Gergely, and Dankar 2019). In the context of biomedical research, this is generally understood to mean that an ethically valid IC must be voluntary and must clearly convey information that balances the potential benefits (societal and/or personal, as applicable) associated with study participation with the risks and disadvantages (McGuire and Beskow 2010).

Crucially, IC in biomedical research must be distinguished from consent as a legal basis for processing personal data under the GDPR. While IC typically covers all aspects of study participation (including any associated medical procedure, collection and research use of samples, collection and research use of data), consent within the meaning of GDPR only concerns processing of personal data. Unlike IC, which is a general ethical/legal prerequisite for compliant biomedical research, the GDPR consent is one of the several legal bases (namely, Art. 6(1)(a) GDPR, in conjunction with Art. 9(2)(a) GDPR) the controller may choose. Although the party acting as a controller for the purposes of the research project must have a valid legal basis to process personal data, the controller's legal basis need not be consent. On the contrary, as discussed in Chapter 2 (Data Protection View), the two alternative legal bases, those under Art. 6(1)(e) and Art. 6(1)(f) GDPR, are widely considered as more suitable for biomedical research than consent, albeit their availability varies across the European Member States. Consequently, it is not uncommon for European countries to generally mandate IC for biomedical research (under, for example, national research ethics laws), while at the same time providing biomedical researchers with a possibility, under the national implementation of the GDPR, to rely on a legal basis other than consent for the processing of research participants' personal data. This is, for example, the case in Sweden, where even though IC is generally required under both the national research ethics law and the national biobanking law, processing of personal data for research purposes is typically grounded in Art. 6(1)(e) GDPR, i.e., the performance of a task in public interest (Stenbeck, Eaker Fält, and Jane 2021). Under these circumstances, the researcher is not obtaining consent for data processing within the meaning of the GDPR and is under no obligation to comply with the GDPR requirements for a valid consent. However, where a researcher processes personal data based on Art. 6(1)(a) GDPR - a decision often dictated by the unavailability of another suitable legal basis in the researcher's country - the researcher must ensure that the IC also conforms to the requirements of consent under the GDPR.[27] Another crucial difference between the two types of consent lies in whether,

[27] Alternatively, when processing data based on consent (i.e., in accordance with Art. 6(1)(a) GDPR in conjunction with Art. 9(2)(a) GDPR), it is also possible for researchers to employ two different consents: i) a traditional, or "research ethics" IC for the participation in the research project; and ii) a separate "GDPR consent" for the processing of personal data. Whereas the former may provide a general description of the envisaged study - including its potential benefits and risks associated with the participation -, the latter will exclusively focus on the aspects relating to the processing of personal

and under what conditions, consent can be exempted or waived. With regards to the IC, in a particular jurisdiction, certain types of non-interventional and low-risk biomedical research projects may be explicitly exempted from the IC requirement. Moreover, irrespective of the nature of the research, there may be a possibility to waive the IC under certain circumstances, typically following a formal decision by a competent REC to grant such a waiver. By contrast, a valid consent under the GDPR must be explicit, expressed in the form of "a statement or by a clear affirmative action" by the data subject, and must constitute an "unambiguous indication of the data subject's wishes" in relation to the intended processing of personal data (Art. 4(11) GDPR). This means that, where biomedical researchers choose to process the personal data of their research participants based on Art. 6(1)(a) in conjunction with Art. 9(2)(a) GDPR, any exemptions or waivers from the IC requirements will likely not apply to the processing of personal data. The aforementioned differences between IC for biomedical research and GDPR consent, as well as several other characteristics along which the two types of consent diverge, are summarised in the table below.

Table 8 - Differences between IC for biomedical research and consent as the legal basis for data processing

	Informed Consent (IC) for biomedical research	**Consent as a GDPR legal basis for processing personal data**
Relevance	General ethical and, in many jurisdictions, legal requirement for biomedical research.	One of the six legal bases for processing personal data, under Art. 6(1) GDPR. That is, choice of consent (Art. 6(1)(a)) as the legal basis is optional in principle, provided that another suitable legal basis, such as 6(1)(e) or 6(1)(f), is available to the researchers processing personal data.
Material scope	Typically covers all aspects of study participation, including medical procedures, sample collection/use, and data collection/use.	Limited to the processing of personal data.
Exemptions/waivers	In jurisdictions where IC is generally required, certain categories of biomedical research (e.g., low-risk, non-interventional studies) may be exempt from the IC requirement by default. Additionally, IC may be waived based on an ethics approval, whereby a REC deems that the benefits of the proposed study clearly outweigh the risks it poses to the research participant.	Consent entails an "unambiguous indication of the data subject's wishes" based on "a statement or by a clear affirmative action" (Art. 4(11) GDPR). This effectively means that when processing of personal data is based on Art. 6(1)(a) GDPR, consent cannot be waived.

data, incorporating specific informational elements mandated by the GDPR. However, in practice, the two types of consent are typically combined into a single consent form.

Explicit consent	May not be required under certain circumstances. For example, in some countries (e.g., Belgium, Demark, Finland and Norway) samples and data collected in a healthcare context can be included in a biobank and/or a research project without an IC, typically subject to an ethics approval.	As per above, the data subject's consent must be explicit. An "implied"/"presumed" consent, whereby data subjects can opt out of processing but are not explicitly asked to opt in, is non-compliant under the GDPR.[28]
Specificity of consent	No consensus within the bioethics community. Various forms of IC have been proposed, ranging from "blanket" IC (i.e., for any future research uses of samples/data) to study-specific IC (whereby each future re-use of samples/data requires a new consent).	Must be obtained for specific purpose(s) of data processing.[29] Where data are collected to be processed for multiple distinct purposes, the data subjects must be asked to provide their consent for each purpose separately.
Mandatory information disclosure requirements	Varies across jurisdictions and institutions. May include: - a statement that the patient's refusal to study participation will not negatively affect the patient's medical care - (if applicable) a clarification that while the proposed study may deliver valuable scientific insights, the results are less likely to directly benefit the study participants personally	Various GDPR-mandated disclosure requirements, including: - A statement that the consent can be revoked by the data subject at any time, in accordance with Art. 17 GDPR (Right to Erasure). Consent should also outline the procedure in relation to the exercise of this right, and describe any restrictions or limitations thereof, if applicable. - A statement that the data may be transferred to parties located in third countries with inadequate data protection laws (if applicable). To ensure specificity of the consent, it is also recommended to explicitly ask data subjects whether they consent to such transfers. - Other disclosure requirements mandated under Arts. 13 and 14 GDPR.

[28] Interestingly, under the national implementation of the GDPR in some Member States (e.g., Italy), a supervisory authority designated by the applicable data protection legislation may have the authority to exempt scientific research projects from the consenting obligations under certain conditions, even where Art. 6(1)(a) GDPR is the controller's legal basis for data processing. This, however, constitutes an incongruity between the definition of consent as a legal basis for data processing under the GDPR on the one hand, and its operationalisation by some Member States on the other hand.

[29] As also previously discussed in Chapter 2 and Annex C, there is some uncertainty introduced by Recital 33 GDPR, according to which "data subjects should be allowed to give their consent to certain areas of scientific research". While sometimes interpreted as enabling a broad consent for scientific research purposes, this recital is not supported by any article within the GDPR, making its practical implications dubious. In any case, "certain areas of scientific research" should be understood to indicate that a "blanket" consent for any future research uses of the data would not be compliant under the GDPR.

These differences between the two types of consent, albeit subtle, may give rise to significant legal interoperability issues bearing on cross-border sharing of samples and data for research purposes. Generally, the most problematic cases are expected in the context of cross-border transfers where only one of the controllers involved in the transfer relies on consent as the legal basis to process the data. Their counterpart controller (e.g., the would-be transferring controller), using a GDPR legal basis other than consent, may not be able to obtain a GDPR-compliant consent from the data subjects. This effectively makes the research use of the personal data unlawful by the would-be recipient controller (that relies on consent as the GDPR legal basis), precluding the transfer of the data between the parties.

Consider a hypothetical example where human biological samples and data were collected without research participants' explicit IC. This is possible where, for example, the collections are generated in a healthcare context (i.e., as part of a medical procedure), and the research ethics or biobanking laws within the jurisdiction permit subsequent research use of the collections without the patient's explicit consent. In this scenario, the IC to study participation can be thought of as 'implied' or 'presumed': the patients typically have a legal right to opt out of subsequent research by submitting a request to the appropriate national or regional health registry, as is the case in the Nordic countries, particularly Denmark, Finland and, to a lesser extent, Norway. However, unless the patient exercises this right to opt out, the sample and its associated data are treated as if the patient has consented to their research use (Salokannel, Tarkkala, and Snell 2019; Hartlev 2021; Befring 2021). From an ethical point of view, this approach can be justified as a pragmatic solution that accomplishes two goals. On the one hand, it respects the essence of individuals' fundamental rights by allowing them to opt out of research if they so choose. On the other hand, by introducing practical barriers to opting out, this approach ensures that only the individuals who are most vigorously opposed to the utilisation of their samples and data for research purposes make use of this right, thus maximising the number of healthcare-derived biological samples (and the richness of the associated datasets) available for research. Although processing of personal data under this approach would have been unlawful if the controller relied on Art. 6(1)(a) GDPR legal basis in conjunction with Art. 9(2)(a) GDPR, the national implementations of the GDPR in such countries allow the controllers to process personal data under a different legal basis, typically Art. 6(1)(e) GDPR, coupled with Art. 9(2)(j) GDPR. Therefore, provided that the subsequent retrospective research takes place within the same country, the absence of a GDPR-compliant consent is not a legally relevant issue. However, the same is not true in the context of international research collaborations. Due to the uneven implementation of the GDPR by the Member States (and other countries where the GDPR applies), some biomedical researchers in other countries may not have an option to rely on a GDPR legal basis other than consent when processing personal data for research purposes. Strictly speaking, such researchers, using Art. 6(1)(a) GDPR legal basis coupled with Art. 9(2)(a) GDPR by default, cannot lawfully access and utilise human biological samples and data that were collected in another Member State under an implicit/presumed consent. In practice, sample and data sharing may still be possible at least to a limited extent, albeit doing so will require substantial practical assistance by the transferring party, i.e., the entity responsible for the sample and data collections. For example, the transferring party may agree to support the external researchers in recontacting the data subjects in order to obtain an explicit consent that would be valid under the GDPR. Alternatively, the transferring party could offer to provide samples and data in an irreversibly de-identified form that can be considered anonymous within the meaning of Recital 26 GDPR. Assuming that such irreversible de-identification is deemed technically

feasible by the transferring party, the resultant anonymised data would no longer be subject to the GDPR, obviating the need for using a valid GDPR legal basis by the recipient. Additionally, in some cases, it may be possible for the transferring party to share samples and data from deceased donors, as the GDPR does not apply to deceased individuals. However, this is only feasible where the national data protection laws in the countries of the transferring party and the intended recipient permit research use of data pertaining to deceased individuals, without imposing conditions that are substantively similar to those under the GDPR in relation to the processing of personal data.

Ethics approvals for data/sample use

Another key interoperability issue potentially hindering international data and sample transfers concerns the ethics (REC) approval for the intended research project as a prerequisite for transfer. As discussed previously, retrospective biomedical research may be exempt from the traditional ethical review under a particular country's research ethics law. This exemption, however, is potentially incongruous with biobanking laws in other countries, under which a pre-existing ethics approval is a relevant precondition for accessing and using biobank collections. As a caveat, it should be noted that virtually all national biobanking laws reviewed by the authors of this document allow for some discretion on the part of biobanks. For example, the biobank responsible for a particular sample and data collection may exempt researchers interested in utilising these resources from a REC approval, provided that the proposed research project satisfies certain criteria. These criteria will vary across biobanks and may include conditions such as immediate relevance of the proposed research project to the biobank's primary objectives, and/or researchers' demonstrable adherence to the international biobanking-specific standards, SOPs, and relevant ethical guidelines. Additionally, any transfer of samples or data is likely to take place under a legally binding contractual agreement, which serves as an additional accountability tool the biobank can rely upon vis-a-vis the external research organisation.

At present, there is no comprehensive survey of the practices of biobanks in relation to the conditions under which they provide samples and data to external researchers. Anecdotally, we are aware that the extent to which biobanks require additional assurances from external sample- and data-requesting researchers, including submission of a pre-existing REC approval for the intended retrospective research project, often depends on strategic and trust-related factors such as whether the biobank has an ongoing research collaboration with the requesting party. This is particularly true in the countries with dedicated biobanking laws since such laws tend to allow the biobanks some discretion in deciding under what conditions to grant access. On the other hand, the authors have also encountered biobanks with more rigid institutional rules with respect to the conditions for access, including requiring a pre-existing REC approval for the intended research project. A notable example in this respect is the German Centre for Neurodegenerative Diseases (DZNE), one of the partners of the EPND consortium. As part of the initial EPND case studies, several consortium partners needed to access human biological sample and data collections held by the DZNE. In order to perform the research activities under the case studies, the relevant DZNE collections had to be transferred to EPND consortium partners in other countries. However, under the applicable laws in the countries of the recipient partners, no ethics review and a formal REC approval was required for the proposed retrospective

research projects. Consequently, the initial request to the DZNE Biobank for the use of sample and data collections was made in the absence of a pre-existing REC approval. This, however, was found to conflict with the policy of the DZNE Biobank to only make sample and data collections available for projects that have received a formal approval from a competent REC. This incongruity between the legal frameworks in the recipient institutions' countries and the formal policy of the DZNE Biobank necessitated protracted negotiations among the EPND consortium members, delaying sample and data transfers by several months. Importantly, the EPND is not unique in this regard. It is common for large European research consortia to encounter similar challenges when pursuing cross-border sample and data transfers, even where the research collaboration is governed under a formal contractual framework among the partners (Simell et al. 2019).

Annex G: Regulatory guidance – Support tools offered by the European Medicines Agency for Innovations

Discovery, Qualification and Validation of biomarkers and methods

Scope:

The focus in this document is on support tools offered by the European Medicines Agency (EMA), as it is foreseen for EPND that contact is sought with the regulators at an European level. The advantage is that expertise from the whole of the European Medicines Regulatory Network is then incorporated.

The EMA has several tools to support the development of a biomarker, of which the following support tools are particularly relevant for the purposes of the EPND:

1. Innovation Task Force (ITF) meeting
2. Qualification Advice on novel Methodologies

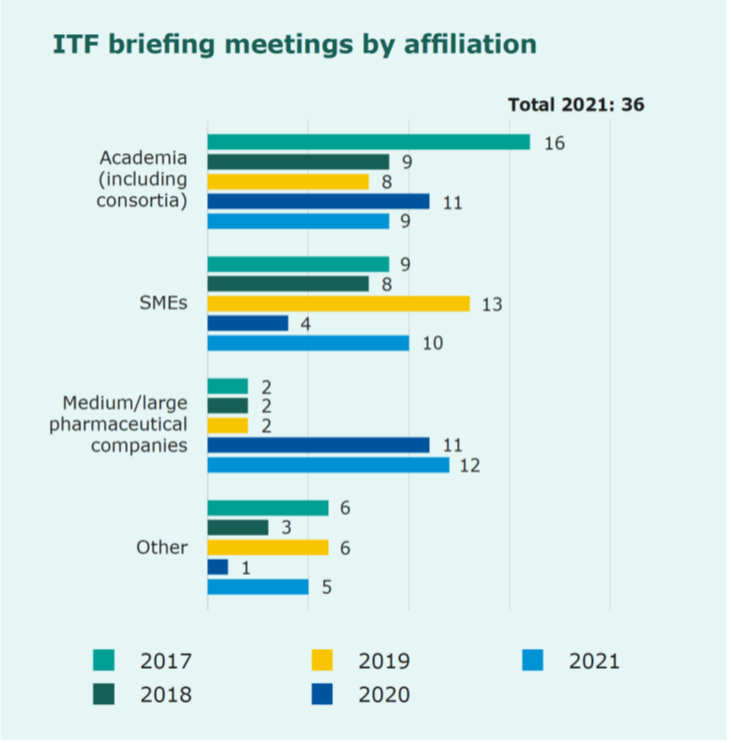
The most important characteristics of each of these two options are highlighted below, including references to the EMA website. EPND is planning to request an ITF meeting to be held in month 36 (October 2024; IMI project Deliverable 6.7). Further, the EPND project aims to go for a Qualification advice, as this is a formal point of view of the regulators bringing value to the project when, at the end of it, a path forward needs to be identified.

Discovery phase: Innovation Task Force meeting

The ITF offers a discussion platform for an early dialogue with regulators. It is a multidisciplinary platform for a preparatory dialogue and orientation on innovative methods, technologies (including biomarkers) and medicines. The ITF, due to its exploratory nature, is best suited for the discovery phase in the lifecycle of biomarker and method development.

In 2021, 9 out of the total 36 meetings (25%) where held with academic developers, including consortia (Figure 3). A third of the submitted requests concerned innovative methods (e.g., statistics, manufacturing, software, biomarkers), while around 15% were centred on technologies (e.g., 3D printing, e-health).

Figure 3 - ITF briefing meetings by affiliation, Source: EMA, annual report 2021



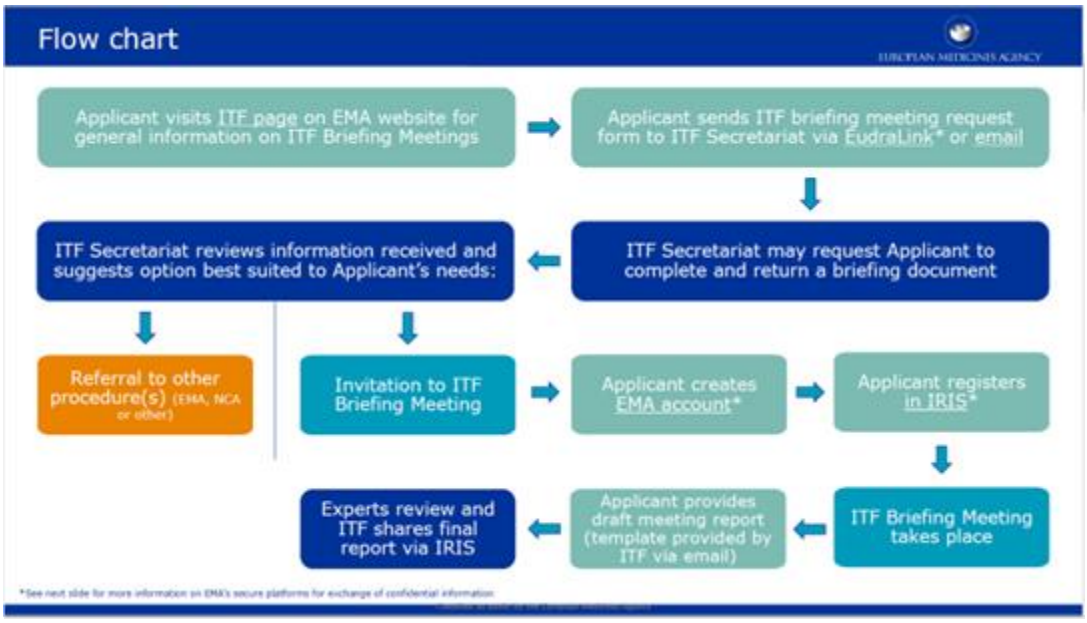
An ITF meeting is free of charge and can be considered as a complement or a preparation of the formal procedures (e.g., scientific or qualification advice). Participating in an ITF meeting can help with taking strategic regulatory decisions in the early development phase.[30] Available documents on the website are 1) Step by step guide including relevant links, 2) Briefing Document template and 3) ITF Briefing Meeting Report template.

The ITF meeting is set up as an informal dialogue with a flexible agenda. It is an educational exercise for both sides of the table, sharing new innovative approaches and identifying the gaps in knowledge. As it is an open exchange of ideas, it can be that different regulators express different views. This is different compared to the formal procedures (e.g., scientific and qualification advice), where even though there can be different views, the outcome will be a single EMA opinion or advice. On the EMA website the ITF Application process is described (see also Figure 4 below).

The timing of the ITF meeting is important for the EPND project. The consortium will need to be able to show data or concepts where the ITF can provide input on. The more details the consortium can show, the more nuanced the information exchange can be between the regulators and the EPND consortium. For the EPND project it is important to know at an early stage, which questions from the different work packages the consortium would like to address in an ITF meeting. Also, it is important to know whether various parties affiliated with the EPND have biomarkers that are likely candidates for a qualification procedure. As the tentative target date for the ITF meeting is month 36 (October 2024; project Deliverable 6.7), the EPND consortium will need to start actively preparing for the meeting from month 24.

[30] For more information on ITF, the mandate and how to apply for a briefing meeting, see: https://www.ema.europa.eu/en/human-regulatory/research-development/innovation-medicines#ema's-innovation-task-force-(itf)-section

Figure 4 - Flow chart for application for an ITF briefing meeting



Source: https://www.ema.europa.eu/en/documents/other/applying-innovation-task-force-briefing-meeting-itf-bm-step-step-guide-faq_en.pdf, accessed 15 September 2022

Qualification Advice on novel Methodologies

Endorsement of validated biomarkers by regulators in the form of a qualification advice provides drug developers with greater regulatory clarity, helps expedite the drug development process, and may be an important step in the adoption of precision medicine in clinical practice (Amur et al., 2015, CPT).

In Europe, a Qualification Advice on novel Methodologies (e.g., biomarker development) is a special form of Scientific Advice that can be requested from the EMA.[31] The advice is given by EMA's Committee for Medicinal Products for Human Use (CHMP) on the basis of recommendations by the Scientific Advice Working Party (SAWP).

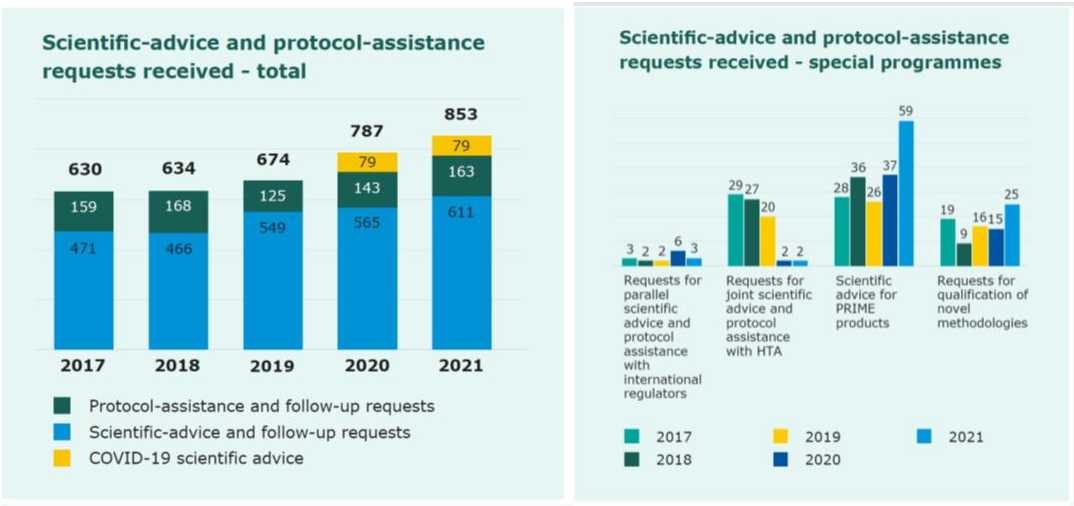
The qualification process leads to a CHMP qualification advice or CHMP qualification opinion. The qualification opinions are made publicly available and can be found on the EMA's website.[32] As an example of a qualification opinion relevant to the EPND's therapeutic area, see the qualification opinion on "Molecular neuroimaging of the dopamine transporter as biomarker to identify patients with early manifest Parkinsonism in Parkinson's disease". Unlike qualification opinions, qualification advices are confidential in nature and are not published.

[31] Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance

[32] Qualification opinions are available at: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/novel-methodologies-biomarkers/opinions-letters-support-qualification-novel-methodologies-medicine-development

In 2021 the number of Scientific Advices and Protocol Assistance (i.e. a type of Scientific Advice for developers of orphan medicines), was 853. Of these 853 advices, 25 were on the qualification of novel methodologies.

Figure 5 - Scientific advice and protocol-assistance requests received

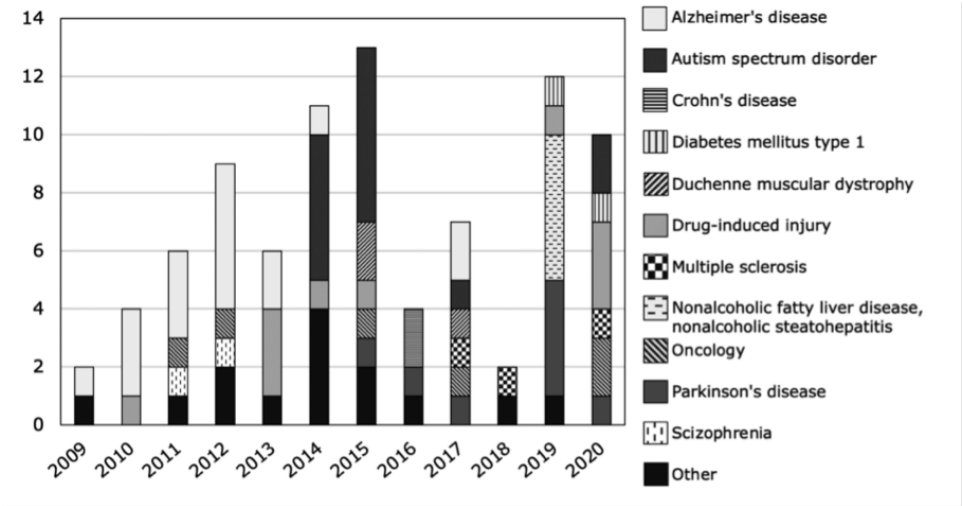


Explanatory note: A. Total. B. Special Programmes. Source: EMA, annual report 2021 (https://www.ema.europa.eu/en/documents/annual-report/2021-annual-report-european-medicines-agency_en.pdf)

A recent publication of Bakker et al. (CPT, 2022), has reviewed the Qualification procedure at the EMA from the start, spanning the period 2008 to 2020. In total, 86 biomarker qualification procedures were identified, of which 13 resulted in qualified biomarkers. Most biomarkers were proposed (45/86) and qualified (9/13) for use in patient selection, stratification, and/or enrichment, followed by efficacy biomarkers (i.e., pharmacodynamic/response markers; 37 proposed, 4 qualified). Overall, many issues were raised during qualification procedures, mostly related to biomarker properties and assay validation (in 79% and 77% of all procedures, respectively). Issues concerning biomarker properties were those related to sensitivity, specificity, baseline/reference measurements, thresholds of detection, cutoff values, and clinically relevant changes, among others. These findings are relevant to the EPND, providing the consortium with an overview of the common issues the consortium will need to address when of formulating its regulatory strategy.

In Figure 6 below, the main disease or disease areas over time from 2008 to 2020 are summarised, as reported by Bakker et al., 2022. Particularly relevant for the EPND are the qualification advices and opinions on Alzheimer's disease and Parkinson's disease.

Figure 6 - Biomarker qualification: main disease or disease areas over time from 2008 to 2020



Number of biomarker procedures per disease or disease areas for which they were prepared over time. Diseases or disease areas for which only one procedure was started are grouped in the category "other". In 2017 one procedure covered MS (multiple sclerosis) and two other diseases and was therefore assigned to both categories MS and "other". In 2020 one procedure covered MS, PD (Parkinson's disease), and several other diseases. This procedure was counted for MS, PD and "other". Source: Bakker et al., 2022

Validation

Unlike the ITF-procedure, a Qualification Advice/Opinion procedure is more strictly regulated regarding the briefing document, timelines, meeting format and procedures, as well as the documentation of the outcomes. In practical terms, this means that the party applying for a Qualification Advice/Opinion procedure requires more data and more preparatory work. On the other hand, a formal (positive) Qualification Advice/Opinion is extremely valuable in the context of Validating novel biomarkers/methods.

Building an in-house regulatory expertise around the EMA's procedures is also crucial for the sustainability of the EPND. Moreover, it is important that throughout the IMI project, the Regulatory team of the EPND consortium remains in close contact with the different work packages, in order to stay informed on the promising biomarkers that could be suitable candidates for the EMA's qualification procedures.