

Data sharing in dementia research – the EU landscape



Contents

1. Preface	3
2. Executive summary	4
3. Introduction	6
3.1 Discussion paper aims and scope	9
4. European frameworks for data sharing	10
4.1 Policy frameworks: Open Science, data strategies and AI	10
4.2 Legal frameworks: the General Data Protection Regulation	16
4.3 Section summary: policy and legal perspectives	20
5. Data sharing in EU-funded dementia research projects	22
5.1 Dementia research projects in Horizon 2020	22
5.2 Data sharing in dementia research projects: case studies	31
5.3 Section summary: dementia research projects and data sharing	37
6. Data sharing – perspectives from researchers and participants	38
6.1 Researcher perspectives	38
6.2 Participant perspectives	39
6.3 Section summary: researcher and participant perspectives	42
7. Future perspectives	44
8. Conclusions: barriers and enablers for data sharing	47
9. References	48
10. Acknowledgements	52

1. Preface



In our 2019 report, “Estimating the prevalence of dementia in Europe”, we showed that the number of people with dementia in Europe is likely to double by 2050, increasing from 9,780,678 to 18,846,286 in the wider European region. However, despite its increasing incidence, and high health and social care cost, research on dementia receives a disproportionately low amount of funding compared to other disease areas. As a result, there is an urgent need to maximise the utility of data from dementia research.

Data sharing represents an important step towards meeting this need, and could help increase our understanding of the causes, treatment, prevention and care of dementia. However, there is still much to do to improve data sharing in dementia research - in particular, for clinical studies, where data sharing is not yet common practice.

The EU recognises that policy actions and governance frameworks play a major role in encouraging data sharing. With the launch of Horizon Europe just around the corner, we looked back at the policy developments that have shaped the data sharing landscape for dementia research in Horizon 2020, reviewing the perceptions and concerns of researchers and research participants.

In doing so, it has been possible to get a sense of the key obstacles for data sharing. It is evident that the dementia datascape has expanded substantially over the last decade, with digital biomarker and deep sequencing data now being collected alongside more traditional clinical measures. Many dementia research projects funded through Horizon 2020 involve complex clinical datasets which are both technically challenging and costly for researchers to share. In addition, researchers now have to navigate a risk-averse regulatory environment to ensure against the loss of privacy – a particularly challenging feat when transferring data across borders and between sectors.

On the other hand, there are signs that positive progress is being made. Surveys show that a majority of research participants agree with data sharing in principle. Open Access policies have been widely adopted by research stakeholders, and a number of projects are showcasing how data sharing can be a multiplier of impact for dementia research. Stakeholders from the whole research ecosystem – from participants to patient organisations, researchers to regulators and policymakers – are now working together to bring about the systemic changes needed to improve data sharing in dementia research.

We hope this report is helpful in summarising some of the main barriers and enablers to data sharing, whilst also providing a snapshot of the Horizon 2020 dementia research portfolio. We also hope it will stimulate discussions on how to embed data sharing principles in EU-funded research projects through cross-sectoral policy actions, to ensure people with dementia benefit from the progress made in recent years. Finally, we would like to thank Gates Ventures for supporting the development of this report by my colleague, Dr. Angela Bradshaw, in collaboration with our Policy Officer, Owen Miller.

Jean Georges

Executive Director
Alzheimer Europe

2. Executive summary

Dementia currently affects almost 8 million people in Europe and is a leading cause of disability and dependency in old age. With the share of people aged 65 and over projected to rise to 29% of the European population by 2070, dementia is an increasingly critical public health issue. As such, research into the causes, diagnosis, prevention and care of dementia is of enormous importance – particularly as there are no disease-modifying treatments currently available.

In December 2013, at the G8 Dementia Summit, ministers agreed to share information and data from dementia research studies, to get the best return on investment in research. While numerous initiatives have been launched to make this promise a reality, data from many dementia research studies still remains siloed, stored behind the firewalls of research institutions, drug companies and medical centers. Since 2013, the European policy landscape has evolved substantially; the Horizon 2020 Framework Programme for Research and Innovation, initiated in 2014, is drawing to a close, and digital technologies and personalised medicine are now firmly on the EU agenda. The legal landscape has also been reshaped; for clinical research data to travel between institutions and across borders, researchers must now show compliance with the General Data Protection Regulation (GDPR).

In this report, we outline the policy and legal landscapes that dementia researchers have had to navigate since the launch of Horizon 2020, identifying the key barriers and enablers for data sharing. We have also mapped the portfolio of projects on dementia funded by Horizon 2020, assessing the scale of EU investment in dementia research and the proportion of projects that involve the use of clinical research data. Finally, we reviewed recent surveys of researchers, research participants and patients, evaluating their perceptions, motivations and concerns regarding data sharing.

Our key findings:

- To date, EUR570 million has been invested by Horizon 2020 in 222 research projects on dementia and/or Alzheimer's Disease. Of this, over EUR403 million was allocated to 65 multi-partner, multi-country consortium projects, 75% of which involve the use of clinical data – underlining the importance of systems that support secure data sharing between partners and across borders
- EU Open Science policies have helped drive the development of data repositories, also embedding Open Access principles for publicly-funded research. However, concerns around intellectual property rights have prevented broad uptake of Open Science practices by the private sector; the implementation of FAIR practices also varies between member states
- The GDPR has not yet fully delivered on its aim of facilitating safe data flows for research datasets. There are many unresolved issues that impede data sharing between partners and across borders, caused in part by a perceived lack of clarity and in part by the regulatory divergence that has been built into the GDPR
- For patients and research participants, the benefit of data sharing comes with a privacy trade-off; loss of privacy is the most widely-reported concern for this group, which reduces their willingness to consent to data sharing
- For researchers, the high value placed on publications and grants by academic reward systems means that many perceive a “reputation cost” to data sharing. This, along with the financial and time cost, as well as data protection concerns, decreases their motivation to share data.

Our key recommendations:

- To fully embed Open Science principles in practice, policymakers and other research stakeholders should work on co-creating a “shared research knowledge” system, with clearer policies and intellectual property frameworks as well as greater transparency and reciprocity between actors
- To mitigate aspects of the GDPR that impede research data sharing, policymakers and legislators should develop pathways for faster, secure sharing of research data between sectors and across borders, including GDPR Codes of Conduct and tailored standard contract clauses for transfers beyond the EU
- To address the perceived “reputation cost” of data sharing, policymakers, funders and research institutions should promote academic reward systems that place a greater value on data sharing, transparency and openness, incentivising the sharing of data by adopting measures that ensure data generators are credited when their data are reused
- To ensure datasets and platforms can continue to be used and shared, funders and research institutions should provide support to maintain these valuable resources when their research funding period ends
- To increase awareness and trust in data use, reuse and sharing, policymakers at EU and member state level should take concerted actions to increase data and digital literacy, ensuring that older adults and vulnerable groups are not left behind
- To help ensure that decisions on research data sharing are relevant, transparent and ethically sound, researchers should involve people with dementia in the design and conduct of research, and in data governance; this is also a valuable step towards increasing trust.

Beyond these and other measures to facilitate data sharing, there is a need for increased investment in dementia research. People with dementia have also been disproportionately affected by the COVID-19 pandemic; as well as being at higher risk of mortality and morbidity, many have experienced a worsening of symptoms due to social isolation and lack of access to care. Despite this, due to the tightening of EU research budgets, there is a significant risk of dementia research being deprioritised, threatening its future viability and long-term sustainability. The EU and its member state governments must ensure that there is continued investment in dementia research, so that the gains from research – and data sharing – are not lost.



3. Introduction

The modern definition of data is “information, especially facts and numbers, collected to be examined, considered and used.” Data originates from the Latin word “datum”, a singular term which means “that which is given”. Aptly, data sharing entails the act of giving: sharing research data involves making it available for use by other investigators or stakeholders.

The principle of data sharing has rapidly gained traction across the research spectrum, with researchers, institutions, publishers, funding bodies and government agencies agreeing that data sharing can accelerate and improve science. By facilitating collaboration, increasing transparency and enabling reproducibility, data sharing paves the way for important new research findings: elucidating disease mechanisms, developing new diagnostics and testing innovative therapies. It is increasingly clear that data sharing also makes research more efficient, minimising the overall time and resource cost. For clinical research, this is a particularly important consideration; modern-day clinical studies have multiple endpoints, generating vast quantities of complex data, but at a correspondingly high cost. On average, a pivotal drug efficacy trial will cost USD19 million, although this figure increases substantially for longer or larger trials involving participants with complex diseases.

Data sharing holds great potential for accelerating and advancing dementia research. Dementia is a neurodegenerative disorder that is a major cause of disability and dependency in older people. With no disease-modifying treatment or cure, research on the causes, diagnosis, prevention and care of dementia is of enormous importance. As an ageing-associated syndrome with multiple underlying causes, dementia is also a growing health concern worldwide. Indeed, our 2019 report, “Estimating the prevalence of dementia in Europe”, shows that the number of people with dementia in Europe is likely to double by 2050, increasing from 9,780,678 to 18,846,286 in the wider European region. Currently, the global annual cost of dementia exceeds USD800 billion – a figure that is expected to increase substantially over the next few decades. However, despite its high health and social care cost, research on dementia receives a disproportionately low amount of research investment. In the US, research into Alzheimer’s disease (the most common cause of dementia) received [USD550 million in funding in 2010](#)(1). In comparison, USD5.7 billion was allocated to cancer research. In the UK, a [2015 study](#) estimated that for every GBP10 of health and social care costs attributable to the dementia, only GBP0.08 in research funding is provided – compared to GBP1.07 for cancer and GBP0.65 for coronary heart disease(2). These figures underline the urgent need to maximise the utility and societal benefit of data from dementia research. Data sharing represents an important first step towards meeting this need, allowing researchers to obtain better insights into the causes, development, care, treatment and prevention of dementia.

The dementia research datascape: from bench science to Big Data

The concept of data sharing is not a new one: long before the advent of computers, health economists performed secondary analyses on data from government reports, and meteorologists shared information on weather patterns. However, as our technological ability to handle data has increased over time, so has the size of research datasets – particularly those from clinical research studies. In his landmark 1906 lecture, Alois Alzheimer described the case of a 51-year old woman severely affected by the symptoms of a condition we now know as Alzheimer’s disease (AD). Alois Alzheimer’s first study on AD had a single participant: Auguste Deter. Nowadays, clinical studies commonly involve hundreds and sometimes thousands of participants, depending on study phase and design. For example, the recent [Phase II trial](#) of BAN2401, an

anti-amyloid immunotherapy, enrolled 856 participants with mild cognitive impairment(3). Cohort studies can be larger still: the European Prevention of Alzheimer’s Dementia (EPAD) Longitudinal Cohort Study recruited over 2000 participants, aiming to characterise the earliest stages of AD(4). Both are dwarfed by genome-wide association studies (GWAS): a [recent GWAS study](#) identified new risk factors for AD by analysing genetic data from 314,278 individuals in the UK Biobank cohort(5).

As well as growing in size, datasets from clinical research studies have also become more complex; a [2018 report](#) showed an 86% increase in the average number of clinical endpoints in trials registered between 2001–2005 and 2011–2015(6). This means that a single research participant may undergo MRI scans, neuropsychological assessments and biomarker tests; they may have their DNA sequenced, their blood pressure measured or their physical activity mapped – all this at multiple intervals throughout a study. Together, the data from these assessments, scans and tests can add up to gigabytes and terabytes of data per research participant. [Illustrating](#) the sheer scale of research data being generated, the US National Cancer Institute’s genomic data commons received 4.5 petabytes (the equivalent of 4.5 million gigabytes) in one year alone, and it is estimated that healthcare institutions managed 8.4 petabytes on average in 2018(7).

Data sharing in dementia research: the why and the how

With such breadth and scale, however, comes meaning. Linking data from these analyses helps clinical researchers to identify connections between biomarkers and cognitive symptoms, link alterations in brain scans to behavioural changes, and understand which genetic risk factors may contribute to dementia. Sharing this data adds a further, essential dimension: research findings can be compared, contrasted – and corroborated. Corroboration and validation of research is particularly important for interventional studies aiming to develop and test new treatments. Indeed, validation – and by extension, data sharing – is embedded into the clinical trial pipeline, with Phase 3 trials designed to confirm the preliminary evidence accumulated in earlier trial phases. For example, Biogen’s anti-amyloid drug, aducanumab, was tested in two synchronous, early AD Phase 3 trials (EMERGE and ENGAGE) to build on and validate the results of the Phase 1b PRIME [trial](#)(8). Aducanumab is currently under review at the US Food and Drug Administration (FDA). If approved, it would be the first disease-modifying treatment for AD to reach the market.

Validation is also an important consideration for dementia cohort studies which follow a defined group of individuals over time, elucidating links between behaviour, background and disease development. An inherent challenge in cohort studies is the representativeness of the participants with respect to the general population. For example, the landmark Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort has unparalleled data depth and longitudinal follow-up, however the recruited participants are primarily drawn from older, white and highly-educated groups. By comparing ADNI findings with data from the European AddNeuroMed cohort study (representative of a different population to ADNI), [researchers](#) were able to validate an AI-based model for dementia risk prediction(9). Similarly, [analyses](#) performed across data and samples from the ADNI and EMIF-MBD (multimodal biomarker discovery) cohorts allowed researchers to identify three biological AD subtypes, associated with different disease progression paths and outcomes(10). Together, these examples illustrate how data sharing today can lead to better diagnostics, treatments and care for the patients of tomorrow. Crucially, data sharing also honours the generosity of research participants, who contribute their time and undergo risky procedures for the sake of medical progress – a point that is addressed in greater detail in section 5.2 below.

Acknowledging the societal and ethical imperatives for data sharing, national and transnational research organisations have developed policies that either mandate or strongly recommend making clinical research

data available. In 2013, [EFPIA and PhRMA](#) jointly published their Principles for Responsible Clinical Trial Data Sharing, making a commitment to enhanced sharing of clinical trial data with researchers(11). In parallel, many intergovernmental organisations including the Organisation for Economic Co-operation and Development (OECD) and the World Health Organisation (WHO) have called for the public disclosure of results from all clinical trials. In 2014, the European Commission mandated that all trials registered in the EU must publish trial result summaries on the EUCTR database within 1 year of the end of the trial. In October 2016, the European Medicines Agency (EMA) went one step further, becoming the first major regulator to publish all clinical trial data submitted by pharmaceutical companies applying for drug approval. Although these clinical study reports do not yet contain individual patient data (IPD), the EMA foresees making anonymised IPD available at a later date.

Democratising science: the Dementias Platform UK (DPUK) data portal

The DPUK [data portal](#) was developed in collaboration between the DPUK public-private partnership and research teams leading dementia cohort studies in the UK. As a [data repository](#), the DPUK data portal facilitates access to [data](#) from over 3 million participants in studies such as Generation Scotland, LBC (Lothian Birth Cohort)1936 and GERAD (Genetic and Environmental Risk in Alzheimer's Disease). Data from these cohorts cover a broad spectrum of variables, including genetic test results, brain imaging data and neuropsychological assessments of memory and brain function. Once access to data is approved, researchers are able to work with the curated data in a secure, remote access environment, allowing them to develop new research questions, and test or validate novel hypotheses on dementia.

“The DPUK data portal was built to democratise science. By providing remote access to data from 42 cohorts (n>3.4m) anyone, from Botswana to Brussels, can access some of the world’s best data. All that’s required is a good idea, internet connectivity, and an academic or industry email. Our data discovery tools and streamlined access procedures enable rapid access decisions (median time to decision: 23 days), with most datasets and most computational facilities, including analytical software, being free at point of use. Bona Fide researchers work on approved projects within a secure and fully auditable multi-modal environment. Findings, but not data, may be exported for publication.”

Integrating imaging, genomic and health data is challenging, even for experts. So our Datathon and Summer School programme, targeting early career researchers, is designed to introduce up-and-coming analysts to good data management and rigorous longitudinal analysis. Further details on the data can be found [here](#), with details of our training programmes listed on our website.

In a spirit of collaboration, DPUK is working alongside data platforms around the world, including Dementias Platform Korea, Dementias Platform Australia, EMIF-AD, GAAIN, IALSA, SCAI, Cohen Veterans Bioscience, the Krembil Brain Institute, and ADDI, to create a global data alliance for dementia research. Interested?”

Professor John Gallacher, Director of Dementias Platform UK, University of Oxford

Despite these moves by funders, regulators and governmental bodies, research data is often siloed, stored behind the firewalls of research institutions, pharmaceutical companies and medical centres. To further encourage and facilitate data sharing, the last decade has seen the creation of several platforms for clinical data sharing (see inset box for a description of one such platform, the DPUK data portal(12, 13). The Global Alzheimer's Association Interactive Network platform (GAAIN) houses data descriptors (or “metadata”) from almost 500,000 participants in clinical dementia studies, provided by 51 GAAIN data partners. This federated model of data sharing means that the data partners control access to the raw data, whilst enabling

researchers to explore metadata and create new cohorts across multiple data sources. Conversely, the [UK Biobank](#) platform exemplifies a more open model of data sharing(14). Established in 2006 by Sir Rory Collins, UK Biobank collects longitudinal health data and biological samples from 500,000 participants aged between 40–69 years. The electronic health records of participants are directly linked to the UK Biobank database, providing a record of disease events, drug prescriptions and deaths. This database also contains the DNA sequencing data from all 500,000 participants, facilitating large-scale genetic studies. Of particular relevance to AD, UK Biobank collects imaging data and is aiming to obtain brain MRI scans from 100,000 participants. From the outset, UK Biobank has made all its anonymised data openly available to bona fide researchers, subject to verification that the research is health-related and in the public interest. Since 2012, 2,551 applications for data access have been approved: of these, 152 relate directly to dementia. To date, over 1,500 papers using UK Biobank data have been published in peer-reviewed journals, identifying new genetic risk factors for disease and paving the way for improved, targeted therapies.

3.1 Discussion paper aims and scope

The invention of the World Wide Web by Tim Berners-Lee in the late 1980's set the stage for more widespread sharing of research data. By creating tools and pathways for data assessment and exchange, the internet has opened new horizons for researchers, catalysing innovation and supporting collaboration on a global scale. The benefits of sharing data are clear: research can be validated, the returns on investment are increased and, importantly, new hypotheses can be generated by linking datasets from different studies, accelerating scientific innovation and leading to improvements that directly benefit citizens and patients.

Despite this consensus view, sharing of research data – and, in particular, sharing of data from clinical research – is far from being common practice. In a 2017 [Springer Nature survey](#), 39% of the 2,683 academic respondents working in the medical sciences did not share data via repositories or other data sharing platforms(15). Despite legal and ethical requirements to report clinical study data, a [2018 study](#) showed that only 49% of European clinical trials were in compliance with EU regulations on timely reporting(16). Meanwhile, only **33%** of large pharmaceutical companies have made the data from novel drug trials accessible to external investigators(17).

The main aim of this Discussion Paper is to evaluate the barriers and enablers to data sharing by European dementia research projects, with a particular focus on projects that involve human research participants. In [section 4](#), we describe recent EU policy developments that have set the stage for data sharing, analysing the policy and legal frameworks that regulate the use, re-use and sharing of sensitive personal data. In [section 5](#), we present the results of a mapping and surveying exercise of dementia research projects funded by the Horizon 2020 Framework Programme, evaluating the scope, geographic spread and data use of the dementia research portfolio. In closer focus, we present three case studies of data sharing in dementia research projects, including two landmark projects co-funded by the EU via the Innovative Medicines Initiative. In [section 6](#), we examine the perspectives of researchers and research participants on data sharing, summarising the results of recent studies in this area. [Section 7](#) looks to the future, reflecting on our learnings from the COVID-19 pandemic. Finally, in Section 8, and based on our findings, we identify some of the key obstacles to data sharing for dementia research, outlining recommendations to improve the data sharing ecosystem throughout the EU and beyond.



“We could enable faster progress on all fronts of the Alzheimer’s fight by facilitating more data-sharing. Scientists all around the world are working hard to generate new discoveries every day. The data they’re collecting in the process are a tremendously powerful tool that can be harnessed better to understand and reduce the impact of the disease.” **Bill Gates**

4. European frameworks for data sharing

The EU recognises that policy actions and governance frameworks can play a major role in encouraging data sharing. Since 2002, the European Commission has proposed [several initiatives](#) to enhance the sharing of data that is generated by EU Member States and in EU-funded research projects, aiming to cement the position of the EU as a leader in the global data economy(18). These include the creation of platforms such as the [EU Open Data Portal](#) (which provides access to the data of the EU Institutions, agencies and bodies(19)) alongside measures such as the [Recommendation](#) on access to and preservation of scientific information, which calls on Member States to implement policies that promote data sharing(20). Recognising harmonised data protection frameworks as an important enabler of data sharing, the EU has replaced the 1995 Data Protection Directive with the more expansive General Data Protection Regulation (GDPR), aiming to establish a single law across the EU to provide greater clarity and transparency for citizens, researchers and businesses alike(21).

4.1 Policy frameworks: Open Science, data strategies and AI

The Recommendation on access to and preservation of scientific information was published by the Commission in 2012. Two years later, in his November 2014 [mission letter](#) to Carlos Moedas (the incoming Commissioner for Research, Innovation and Science), President Jean-Claude Juncker emphasised the links between research and the digital economy. He also asked for the value added and Impact of Horizon 2020 to be maximised, by ensuring the effective use of project results. In 2015, Carlos Moedas set out his [vision and priorities](#) for his 5-year tenure(22). This vision was structured around three strategic priorities: Open Science, Open Innovation and Open to the World. Open Innovation and Open to the World addressed two interlinked challenges: firstly, the need to involve more actors in the innovation process to create an ecosystem that would drive technologies to market; and secondly, the need to engage in scientific diplomacy and collaboration on a global scale, to ensure European research remained relevant and competitive.

Open Science: data sharing takes centre stage

Open Innovation and Open to the World were both broadly focused on stakeholder engagement and collaboration. However, [Open Science](#) required systemic changes affecting all aspects of the research cycle – with data sharing taking centre stage(23). To increase the efficacy of science and amplify the impact of research, Commissioner Moedas called for research practices to become Open and collaborative, from the point of scientific project design all the way through to the publication, communication and sharing of results. He reasoned that this would improve the quality of research (by enabling researchers to build on previous results), avoid duplication of effort (by increasing awareness of research outputs and encouraging collaboration), increase the impact of research (by speeding up innovation and its translation to market) and improve transparency of the scientific process (by involving citizens and society in discussions).

European Open Science Policy

The OECD definition of Open Science is “Science that encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives”. In 2016, following consultation with two expert groups, the European Commission published its [Open Science policy](#), designed to meet Commissioner Moedas’ overarching goals(24). The policy was framed around 8 ambitions:

- **Open Data:** the results of EU-funded research projects should, by default, become Findable, Accessible, Interoperable and Reuseable (FAIR) and Open
- **European Open Science Cloud:** the creation of a federated ecosystem of research data infrastructures, to enable researchers to share data across borders and research domains
- **New Generation Metrics:** to develop Open Science metrics that go beyond conventional indicators of research quality and impact
- **Future of scholarly publication:** peer-reviewed publications of EU-funded research projects should be freely accessible, and the early sharing of research outputs should be encouraged
- **Rewards:** research career evaluation systems should acknowledge Open Science activities
- **Research Integrity:** EU-funded research should adhere to agreed standards of research integrity
- **Education and skills:** researchers in Europe should have access to the training required to apply Open Science practices
- **Citizen Science:** the general public should be able to contribute to European research and knowledge generation

By 2016, when the Open Science policy was published, the Horizon 2020 Framework Programme for Research and Innovation had been running for 2 years. However, Open Science policies were already embedded in Horizon 2020, in alignment with the 2012 Commission Recommendation on access to and preservation of scientific information. For example, the Horizon 2020 [Model Grant Agreement](#) states (29.2) “..researchers must ensure outputs are Open Access ...as soon as possible, and at the latest upon publication; deposit a machine-readable electronic copy of the published version...in a repository for scientific publications together with bibliographic metadata providing the name of the Action, project acronym and grant number.”

A similar provision was adopted in the [grant agreement](#) for the Innovative Medicines Initiative-2 (IMI2) Joint Undertaking, which was launched at the same time as Horizon 2020. To comply with these provisions, research outputs from projects funded through Horizon 2020 and the IMI must be published in fully Open Access journals, and/or be made available through repositories such as Zenodo, bioRxiv or PubMed-Central. To further support Open Access, the European Commission has recently launched Open Research Europe, a publishing platform for Horizon 2020 beneficiaries that provides submitted publications as pre-prints before they undergo Open peer review and revisions. The first submissions to this platform will be published in March 2021.



Horizon 2020 and FAIR data

To complement its Open Access policy, Horizon 2020 included provisions aimed at facilitating data sharing. The data guidelines for Horizon 2020 that were published in 2013 state: “Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable and wherever possible, interoperable to specific quality standards.” In a 2014 Lorentz Workshop, participants formulated a set of high-level principles along these lines, using the acronym FAIR: Findable, Accessible, Interoperable and Reuseable. These data principles have been widely endorsed by European research funders, and are described in detail in an [article](#) entitled “The FAIR guiding principles for scientific data management and stewardship”, which has been cited over 2,100 times since its publication in 2016(25).

In line with the FAIR principles, a restricted Open Research Data (ORD) Pilot was launched as part of the 2014 Horizon 2020 Work Programme. Based on the concept that data should be “as Open as possible, as closed as necessary”, the ORD pilot required the underpinning data for scientific publications to be deposited in a research data repository, together with the information necessary to analyse and interpret the data. Optionally, Horizon 2020-funded beneficiaries could provide further raw or curated data, such as unprocessed image files or databases. Importantly, all funded projects were asked to provide a Data Management Plan (DMP), an essential step towards embedding data sharing principles in projects at an operational level. Guidelines were also developed to help projects manage data in a FAIR way. To mitigate concerns around intellectual property loss, data privacy and national security issues, the Commission allowed opt-outs from the ORD Pilot – but only if projects gave a reasonable explanation for opt-out. In addition, projects were allowed to apply different degrees of data sharing, from fully Open Access data, to restricted/controlled access, or fully Closed data. From 2017, the ORD Pilot was extended and made the default option for all Horizon 2020-funded projects, paving the way towards widespread sharing of data from EU Research & Innovation projects.

In 2018, 4 years into the Horizon 2020 Framework Programme, the European Commission published a [cost-benefit analysis](#) for FAIR research data(26). In this report, the Commission cited a figure of EUR10.2 billion as the annual cost of **NOT** having FAIR research data, formulating 36 policy recommendations on how to practically implement the FAIR principles in a cost-effective way. Shortly after this report was published, the 2018–2020 European Commission Work Programme for European Research Infrastructures allocated EUR375 million to implement the European Open Science Cloud (EOSC), in turn aimed at supporting the implementation of FAIR. EOSC is a federated ecosystem of research data structures aimed at facilitating data sharing between researchers and across borders, from Horizon 2020 to Horizon Europe and beyond. The groundwork for EOSC was laid between 2016–2017, when the EOSC declaration was endorsed by over 70 institutions, identifying provisions for FAIR data, research data services, architecture and governance.

Open Science in the EU: where are we now?

To assess the global development of Open Science, the EU Open Science Monitor (OSM) performed a [study](#) of Open Science practices and adoption across the EU28 and G8 countries, assessing the extent of Open Access (OA) to publications, Open Research Data and Open collaboration(27). Between 2009 and 2018, the number of Open Access publications grew from 361,000 to over 684,000, with EU28 countries leading the way in terms of OA publication share, with 52.5% (UK), 50% (LU) and 49.9% (NL) of publications appearing in OA journals. By October 2019, over 3,400 data repositories were listed on the Re3data registry of research data repositories, over 94% of which provide Open Access to the data they house. Focusing on EU-funded research projects, a [survey](#) of Horizon 2020 projects funded between 2014–2016 revealed that

68% of core area projects participated in the ORD Pilot, with an average opt-out rate of 32%(28). The main reasons cited for opting out were issues relating to intellectual property (36% of opt-outs) and data privacy (18% of opt-outs). Beyond the core areas included in the ORD Pilot, a further 9% of Horizon 2020-funded projects voluntarily opted-in.

However, there is still a long way to go to ensure Open Science is embedded across all EU-funded research projects. In its [final report](#), published in May 2020, the Open Science Policy Platform provided an update on progress towards achieving the 8 ambitions laid out in the EU Open Science Policy(29). They acknowledged that substantial progress has been made towards implementing rewards and incentives for Open Science, and that Open practices for publication are much more widely adopted. They were also encouraged by recent developments of the EOSC and its work on FAIR data. However, they felt that more progress was required on research integrity, and on the adoption of Open Science practices by the private sector, with more defined frameworks to address concerns around intellectual property rights. They finished by calling for Member States and relevant stakeholders to help co-create a “shared research knowledge” system, with clearer policies and governance frameworks and greater transparency and reciprocity between actors.

Echoing some of these findings, the “FAIR in practice” Task Force of the EOSC identified several key challenges in implementing FAIR principles, publishing a [report](#) on this topic in October 2020(30). From a technical perspective, the Task Force highlighted interoperability as a particular challenge for certain research disciplines, also citing a lack of funding for skills development and data stewardship. Interestingly, the report also identified regional differences in FAIR adoption, which could impede EU-wide implementation of FAIR practices. Western European countries, in particular France, Germany, the Netherlands and the UK, have made more progress in embedding FAIR principles in research: for example, the Dutch National Plan for Open Science promotes FAIR access to research data, while the German National Research Data Infrastructures require that all hosted data is managed according to the FAIR principles. To improve and accelerate the adoption of FAIR principles, the Task Force called for greater efforts in recognising and rewarding improvements in FAIR practices, as well as greater alignment between top-down, policy-driven actions with bottom-up, community-driven initiatives.

From Horizon 2020 to Horizon Europe – and beyond

Although Carlos Moedas was replaced as Commissioner for Research and Innovation by Mariya Gabriel in 2019, Open Science remains a priority for the new European Commission. One of the overarching objectives of Horizon Europe, the successor to Horizon 2020, is “*fostering Open Science and ensuring visibility to the public and open access to scientific publications and research data, including appropriate exceptions.*” Indeed, early drafts of the proposal for Horizon Europe listed Pillar 1 as “Open Science”. Now named “Excellent Science”, this Pillar will support frontier research projects (European Research Council grants), Fellowships and researcher exchanges (Marie Skłodowska-Curie Actions) as well as infrastructures to support world-class research.

Horizon Europe has also retained and extended many of the Open Science provisions in its draft [Model Grant Agreement](#), including the obligation to publish in OA formats, Open Data by default, and mandatory DMPs that are in line with FAIR. [There may also be incentives](#) for adherence to Open Science principles, and obligations to use EOSC for storage and access to data, although these provisions have yet to be confirmed(31). Horizon 2020 funding for the implementation of EOSC will last until [2021 and 2022](#), optimising the interfaces for accessing EOSC and defining the rules and governance frameworks that will ensure EOSC data sharing services can be delivered in a secure and legally compliant manner. Beyond this

period, Horizon Europe will fund EOSC under Pillar 1 of the Programme, as a co-programmed European Partnership alongside several European Institutes of Innovation and Technology (EITs). The final budget for Horizon Europe is EUR95.5 billion, of which EUR24.8 billion has been allocated for Pillar 1 activities.

Moving beyond Open Science, the last decade has seen data sharing gain increasing prominence in EU health policy frameworks. While health policy is not the focus of this report, it is important to note that policy developments in this area often have direct and indirect impacts on dementia research. Electronic health records (EHR) from primary and secondary care provide “real-world evidence” on a person’s health status, treatments and health resource utilisation. As such, EHRs represent a hugely valuable resource for dementia research, with the potential to help researchers evaluate the prevalence and incidence of dementia, and assess the [long-term value of treatments](#)⁽³²⁾. In its 2018 [Communication on the digital transformation of health and care in the Digital Single Market](#), the Commission recognised that sharing and use of health data could deliver tangible benefits for citizens, “*making it possible to tackle major health challenges such as cancer or brain disease.*”⁽³³⁾ A number of actions were proposed, including the creation of infrastructures for the secure sharing of health data between stakeholders, adoption of measures to enable cross-border health data transfers, as well as enhanced cooperation to stimulate the supply and uptake of digital health across member states. If fully enacted, these health policy actions could greatly benefit dementia research across the clinical spectrum, from fundamental mechanisms to diagnosis, progression, prevention, treatment and care.

Data sharing also features prominently in one of the headline priorities of Ursula von der Leyen’s Agenda for Europe: “A Europe fit for the Digital Age”. In her [mission letter](#) to Margrethe Vestager, the Executive Vice-President for A Europe fit for the Digital Age, Ursula von der Leyen mandated her to coordinate a European strategy for data and a European approach for Artificial Intelligence (AI), “*looking at how we can use and share non-personalised big data to develop new technologies and business models that create wealth for our societies and businesses.*” On 19 February 2020, the European Commission published a Communication on the European Data Strategy and a White Paper on AI, setting the EU on a path towards greater sharing of public and private sector data.



A European Strategy for Data

The Data Strategy and White Paper are the first pillars of the new digital strategy of the European Commission, building on previous data initiatives and recommendations. The European Data Strategy communication highlights the importance of data – and data sharing – for the economy and for society, stating that “*the value of data lies in its use and re-use.*” It acknowledges that there is insufficient sharing with, and use of privately-held data by public bodies – and vice-versa – and that there are issues regarding data interoperability, quality and governance that have yet to be resolved.

To address these challenges, the [European Data Strategy](#) proposes actions based on four pillars:

- 1. Data Governance:** creating a cross-sectoral governance framework for data access and use;
- 2. Data enablers:** investing in data and strengthening Europe’s data infrastructures and capabilities;
- 3. Competences:** empowering individuals, investing in skills and SMEs
- 4. Data spaces:** creating common data spaces for using and exchanging data within and across strategic sectors and priority areas

As part of the actions laid out in the Strategy document, the European Commission proposes large-scale investments in FAIR data spaces and EU-based, federated cloud infrastructures to facilitate data exchange. The Data Strategy also proposes several legislative actions, including the creation of a Data Act and an implementing act on high-value datasets. To increase digital literacy and expand capacity for SMEs and start-ups, there are also proposed investments in education, training and recruitment. Supported by a new Regulation for European data governance (for which the [draft text](#) was published in November 2020⁽³⁴⁾), the EU hopes that this Data Strategy will create a single market for data, in which data can flow – securely – across sectors and between stakeholders, for the benefit of all European citizens. In the Data Strategy, the Commission envisions a funding requirement of between EUR4–6 billion for the creation of European data spaces and federated Cloud structures, to be obtained through co-investment from Member States, industry and the Commission.

“More data should be available for the common good”

Following its publication, the European Commission held an open public consultation on the Data Strategy, aiming to collect views on the strategy from public and private stakeholders⁽³⁵⁾. Of the 807 [responses](#), almost all (97.2%) confirmed the need for an overarching digital strategy to enable the digital transformation of society, agreeing that “*more data should be made available for the common good.*” 84.6% of respondents considered that it should be easier to give secure access to existing data held on them, with a similar proportion identifying data literacy as an issue that complicates data use and re-use. There was broad consensus in the need for standardisation and greater interoperability, and almost 70% of respondents indicated that they would be happy to make their data available for the public interest, and particularly for health-related research purposes.

Data, trust and Artificial Intelligence

The [White Paper on AI](#) was published alongside the Data Strategy, outlining how the frameworks described in the latter will support the development of trustworthy AI to maximise the sharing, impact and value of data. Indeed, the concept of Trust has a prominent place in the White Paper, in which the European Commission shows strong support for human-centric, ethical AI that is grounded in European values and human

rights such as the protection of human dignity and privacy. The White Paper identifies two main pillars: an ecosystem of excellence, and an ecosystem of trust. Building on and extending the 2018 Plan for the development and use of AI in Europe, the “ecosystem of excellence” will propose a revised Plan to Member States, and facilitate the creation of centres of excellence in AI, to support the testing and deployment of new applications. Similar to the Data Strategy, the White Paper also proposes actions to increase skills and competencies in AI, including improved collaboration with SMEs and startups.

Meanwhile, the “ecosystem of trust” will build on the [Guidelines](#) on Trustworthy AI produced by a High-Level Expert Group to develop a new regulatory framework for AI⁽³⁶⁾. These guidelines include four ethical principles that must be ensured, including respect for human autonomy, prevention of harm, fairness and explicability. The primary aim of the proposed regulatory framework is to build trust among consumers and businesses in AI whilst also laying the groundwork for a harmonised approach to approving and marketing AI applications across Member States. To further increase public trust in AI, the Commission also proposes to empower authorities to check AI applications, to ensure that high-risk AI systems are transparent, traceable and under human control.

4.2 Legal frameworks: the General Data Protection Regulation

A common thread linking many parts of this report is the concept of trust, and its role as an enabler of data sharing. The protection of personal data is an important concern for people – especially when it comes to sensitive data about health. As well as being a fundamental right, enshrined in the EU Treaties and Charter, data protection is also a core issue for research ethics. Many dementia research projects involve the use of personal data¹, ranging from projects that re-use existing clinical datasets to those that include research studies generating new clinical datasets. Data protection is therefore a central concern for dementia research: done right, data protection can guarantee the privacy of individual patients and research participants, increasing trust in data use, re-use and sharing.

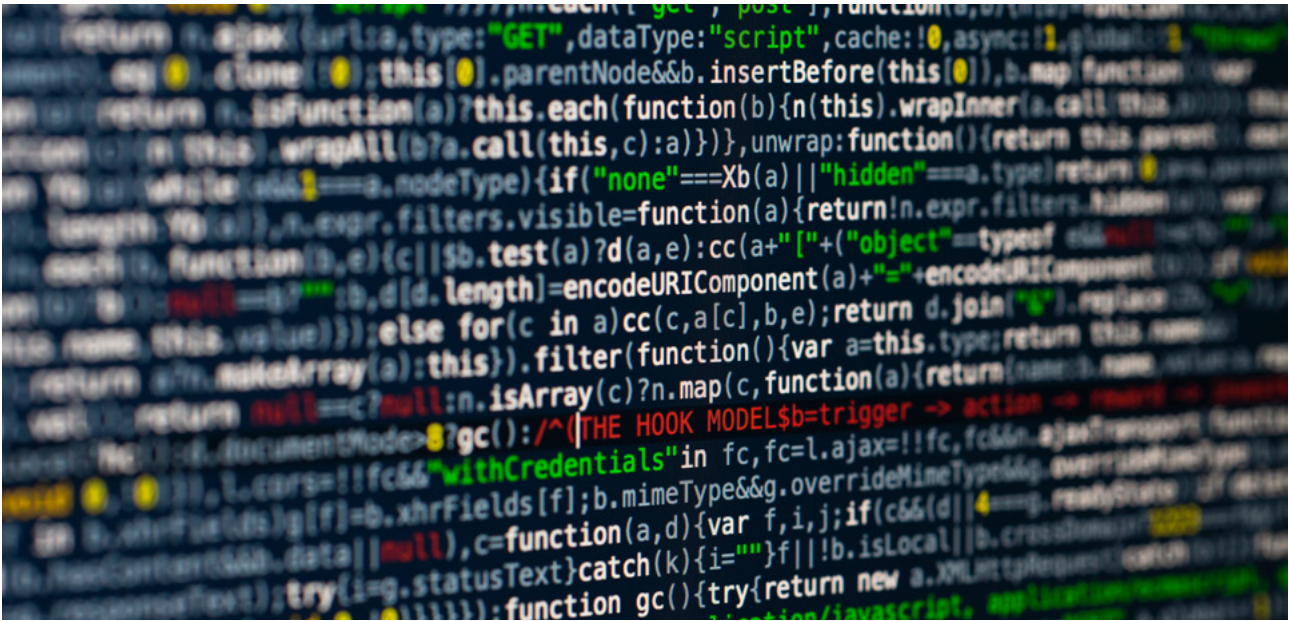
From 1995, the use, re-use and sharing of personal data within the EU was regulated by the Data Protection Directive ([95/46/EC](#))⁽³⁷⁾. In May 2016, this Directive was replaced by a new Regulation (EU) 2016/679 on the protection of personal data. Designed to boost innovation within the EU, and framed as a cornerstone of the Digital Single Market, the General Data Protection Regulation ([GDPR](#)) came into application on 25 May 2018⁽²¹⁾. One of the three core objectives of the GDPR is to ensure the free movement of data throughout the EU, whilst also guaranteeing the right to personal data protection within and beyond the EU. It does so by laying down rules on the use (termed ‘processing’), re-use and sharing of personal data.

The ins and outs of the GDPR

The concept of “data protection by design and by default” has been woven into the fabric of the GDPR. In practice, this means that data controllers and processors²(e.g. clinical trial sponsors and principal investigators of research projects) must integrate data protection measures into every aspect of their personal data processing activities, from the design stage through to data collection, analysis and, eventually, sharing. To reinforce the concept of data protection by design and by default, the GDPR lays out six founding principles for processing personal data, which require that data processing be lawful, confidential, purpose-limited,

¹ Defined as “any information relating to an identified or identifiable natural person”; sensitive health data is defined as “personal data relating to the physical or mental health of a natural person[...]which reveals information about his or her health status”

² A data controller is a person or organisation which has full authority to decide why and how data is processed, and has overall responsibility for the data; a data processor is a person or organisation which processes personal data on behalf of the data controller



accurate, minimal and secure. Underlying these principles are the six legal bases (or scenarios) for data processing; unless organisations can demonstrate that the proposed data processing activity fits within one or more of these scenarios, that activity can be deemed unlawful. Of particular relevance to dementia research, additional safeguarding conditions apply to the processing of sensitive types of data, such as data about health. To further increase trust, rules on consent are also tougher: consent for data processing must now be unambiguous and not assumed from inaction, hence the proliferation of data protection notices online and elsewhere. Finally – and of particular relevance to dementia research – the so-called “Research Exemption” (laid out in Article 89 of the GDPR) enables some of the rights of the data subject to be derogated if they are likely to seriously impair scientific research.

The GDPR was designed, in principle, to support research and innovation. Research communities lobbied EU policymakers at length to alter parts of the draft text that were viewed as excessively restrictive, such as clauses precluding broad consent and obliging researchers to seek renewed consent to reuse data collected for another purpose. Recontacting and reobtaining consent from research participants for new data processing operations can be problematic, particularly when a long time has elapsed since data collection or when the ability of participants to provide “*freely-given, specific, informed and unambiguous*” consent has become impaired – as can occur in dementia research. The modification of the draft text to include the “Research Exemption” (Art.89) was viewed as a positive outcome of negotiations with the EU. Art.89 provides researchers with a means to circumvent the need to obtain specific, explicit consent to particular data processing operations – but only if “appropriate safeguards” (such as pseudonymisation) are in place to ensure the rights and freedoms of the data subject. A related exemption, Art.85, gives Member States the option to derogate from the GDPR “*to reconcile the right to protection of personal data with the right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression*”.(38)

Derogations and divergence

Allowing member states the freedom to legislate at national level in certain areas is clearly important. However, the inclusion of exemptions and derogations in the GDPR has led to a degree of regulatory divergence between member states, which complicates data use, re-use and sharing in practice. Researchers

are [critical](#) of the fact that there is no clear lawful basis for secondary data processing (i.e data reuse) in the GDPR – instead, the choice of basis is left for individual researchers or institutions to determine⁽³⁹⁾. In addition, there is no clear definition of “scientific research”; Recital 159 lists a series of examples, but whether the rights of data subjects are likely to “[render impossible or seriously impair the achievement of these specific \[research\] purposes](#)” is left open to interpretation. As such, researchers wanting to avoid falling foul of the GDPR have to strike a delicate balance; between the rights of the data subject against the importance of using their data for healthcare or research.

A similar criticism of the GDPR is the room that is left for interpretation when it comes to pseudonymisation, anonymisation and consent. As a result, many Member States have introduced derogations, fragmenting the data protection landscape for research. For example, Spain has mandated a technical and functional separation between the investigation team and the person(s) who pseudonymise the data. Meanwhile, the [UK Health Research Authority](#) has suggested that the GDPR could be interpreted to consider pseudonymised data as effectively anonymised (and therefore beyond the scope of the GDPR³) when used by a processor who will never have the key for reidentification⁽⁴⁰⁾. On consent, Italy will accept an approval from the relevant research ethics authority when it is impossible or hard to recontact the data subject to obtain consent for data processing. In [Ireland](#), however, the Health Research Regulation has tightened rules on consent, requiring approval from the National Consent Declaration Committee when it is not possible to obtain explicit consent⁽⁴¹⁾.

A related complication is the fact that there are no approved solutions to pseudonymisation that works for all approaches in all possible scenarios – and across all Member States. Effective pseudonymisation requires a high level of competence in order to reduce the threat of discrimination or re-identification, whilst also maintaining the degree of utility necessary for the processing of pseudonymised data. Without a consistent framework to manage pseudonymisation, anonymisation and consent, and hefty fines for GDPR non-compliance, many research institutions and data protection officers are understandably hesitant to share data for secondary research.

In a [Preliminary Opinion](#) published in January 2020, the European Data Protection Supervisor, Wojciech Wieworowski, sought to highlight some of the challenges in the application of the GDPR to scientific research⁽⁴²⁾. Although this report stated that ‘there is no evidence that the GDPR itself hampers genuine scientific research’, it did identify a lack of convergence in data protection practices as an emerging issue. Potential remedies included the creation of EU Codes of Conduct and certification of research activities, to increase harmonisation and confidence in compliance with the GDPR. Efforts are currently underway to draw up Codes of Conduct for [Health Research](#) (led by the European Research Infrastructure for Biobanking/BBMRI-ERIC⁽⁴³⁾) and [clinical trials](#) (led by the EU Federation for Contract Research Organisations⁽⁴⁴⁾). [ENISA](#), the European Union Agency for Cybersecurity, along with some [national data protection agencies](#), have proposed best practices and techniques for pseudonymisation.

3 The GDPR does not apply to truly anonymous data, or to data on persons who are no longer alive

“I see positive signs that the GDPR is acting as an enabler of trust in personal data use and sharing. The GDPR is also helping citizens - and this includes researchers and research institutions - to become digital natives, although there is still a long way to go in this respect. However there are legal uncertainties that we are trying to address with the Code of Conduct for Health Research. A Code of Conduct is a powerful instrument of soft law, that can help researchers and research institutions identify, on a case-by-case basis, which practices are legally acceptable.

While Codes of Conduct, standard contract clauses and adequacy decisions are useful enablers, it is important to understand that there is no single solution or silver bullet when it comes to data protection. Greater harmonisation and shared vocabularies are needed, as well as systemic changes that embed practices and vocabularies across stakeholder communities. It is also important to understand that research silos extend to people and communities, not just data. There are barriers to data sharing, but none of them are insurmountable if these silos are broken down.”

**Dr. Michaela Th. Mayrhofer, Head of ELSI Services and Research
Biobanking and BioMolecular resources Research Infrastructure – European Research Infrastructure Consortium (BBMRI-ERIC)**

Transatlantic data sharing under the GDPR

It remains to be seen how these guidelines and codes of practice will be applied – and how effective they will be. Meanwhile, commentaries from US-based researchers in [Nature](#)(46) and [Science](#)(47) have drawn attention to a further unintended consequence of the GDPR: the restriction of certain data transfers between EU and non-EU countries. Under the GDPR, transfers of personal data to countries outside the EU are only permitted when a country-wide Adequacy Decision, or organisational ‘standard contractual clause’, is in place, to ensure an adequate level of data protection and safeguarding. In principle, these are straightforward arrangements. However, the laws of certain countries are not in compliance with the provisions required by the GDPR, and agreeing standard contractual clauses (SCC) can be a time-consuming and costly process. The July 2020 Schrems II decision (which concerned the legality of EU-US data transfers by Facebook) has created additional challenges: Schrems II invalidated the Privacy Shield Adequacy framework, and placed new restrictions on SCCs. [In practice](#), this means that researchers may have to carry out case-by-case assessments of whether the destination country can provide an equivalent level of data protection to the EU, and if not, which safeguards need to be applied(48).

Evidencing the material impact of the GDPR on transatlantic research, [Francis Collins](#) (Director of the US National Institutes of Health) describes how a decades-long collaboration with the Finnish National Institute for Health and Welfare has effectively been put on ice since May 2018, as the NIH can no longer guarantee that the stringent Finnish data protection requirements will be met(47). Citing similar issues, Sudha Seshadri – the founder of the International Genomics of Alzheimer’s project – now carries out separate analyses of data from either side of the Atlantic. This increases the cost of data analysis, whilst also making it harder to identify rare genetic variants that might increase AD risk. To Robert Eiss (Senior Advisor to Francis Collins) this is the tip of the iceberg: in his 2020 [Nature commentary](#), he states “...the GDPR has stalled at least 40 clinical and observational studies on risk factors and exposures for cancer.”(46)

Acknowledging the importance of addressing the barriers to data transfers caused by the GDPR (both pre- and post-Schrems II), the EU has been working on refining and improving their data protection guidelines. In November 2020, the European Data Protection Board published [recommendations](#) on safeguarding measures to ensure international data transfers comply with the EU data protection requirements(49). It is hoped that these and other guidelines will help researchers identify the supplementary measures required to complement SCCs. In the longer term, [potential solutions](#) could include the creation of sector-specific adequacy decisions or model SCC that could cover transfers of personal data for research purposes(50). Building on the increased general awareness of data protection rights and regulations, it is hoped that these strategies will result in the safe, unimpeded data flows originally envisaged by the architects of the GDPR.

GDPR and trust

Notwithstanding the challenges described above, there have been some positive outcomes of the GDPR. It is important to note that the GDPR was designed as a “general”, all-encompassing Regulation, and not solely as a data protection framework for research. One of the main goals of the GDPR was to give individuals greater control over their personal data, aiming to increase trust in data use, sharing and re-use. There are [encouraging signs](#) that the GDPR is having a beneficial effect in this area(51). For example, the [Fundamental Rights Survey](#) in 2019 showed that almost 70% of Europeans have heard about the GDPR, and are more aware of their rights to privacy and data protection(52). Research carried out by the DRG (Decision Resources Group) in 2018 found that [33% of surveyed respondents](#) were less concerned about sharing their personal data with pharmaceutical companies as a result of the increased privacy protections in the GDPR(53), while a recent [poll](#) of European businesses showed that almost 75% believed that the GDPR had increased consumer trust(54).

4.3 Section summary: policy and legal perspectives

The advent of the Big Data era at the start of the 21st century, accompanied by technological advances in computing and informatics, has led to considerable changes in research. The benefits are clear: high volumes of data can be turned into actionable knowledge for researchers, drug developers and clinicians, with the potential to transform healthcare systems and yield substantial improvements for patients and citizens. However, these benefits can only be fully realised if data is shared. The EU, recognising the value of data sharing, has embraced the principles of Open Science in its research policy frameworks and in the research projects it funds. Beyond research, Europe is developing a Data Strategy that includes governance frameworks and data spaces that further support the sharing of public and private sector data – all for the benefit of society.

However, these benefits do not come without risk. Personal datasets – and health datasets in particular – contain sensitive information that require a high level of protection, to ensure that data subjects are not exposed to ethical risks such as breach of confidentiality or social harm. Re-identification of individuals in the Big Data era is of particular concern; for example, each person’s DNA sequence is unique and a DNA sample can arguably never be truly anonymised. Consequently, the last decade has seen the development of a more stringent regulatory environment for health research, materialised in the GDPR. This dynamic regulatory environment can be a challenging one for research projects to navigate – particularly in projects such as those funded by the EU, where data often needs to cross public-private boundaries, travelling through Europe and beyond. As outlined in this section, the GDPR is still working towards its aim of facilitating safe data flows for personal datasets. Several unresolved issues for data sharing remain, caused in part by a perceived lack of clarity and in part by the regulatory divergence that has been built into the GDPR.

However, [recent reports](#) indicate that the Commission has recognised these challenges, and is actively developing strategies to address them in concert with a wide range of stakeholders(51).

Obstacles to data sharing

- **Lack of data literacy:** a lack of digital skills and data literacy can exclude citizens from participating in the digital economy, and impede trust in data sharing
- **Uneven adoption of Open Science and FAIR practices:** concerns around intellectual property rights has prevented broad uptake of Open Science practices by the private sector; the implementation of FAIR practices varies between member states
- **Regulatory divergence:** GDPR derogations in different member states has created a fragmented data protection landscape that can be hard to navigate for multi-site studies
- **Lack of methodological clarity:** absence of defined standards and methods for pseudonymisation and consent
- **Issues with cross-border and international data transfers:** when sharing data with collaborators outside the EU, researchers have to perform case-by-case assessments of whether data protection standards are adequate

Facilitators for data sharing

- **Education, training and public engagement:** increasing levels of data and digital literacy will increase awareness of, and trust in data use, reuse and sharing
- **Support centres & funding for data stewardship and sharing:** embedding support and funding from the start of EU-funded research projects
- **Recognition and reward:** recognising and rewarding adoption of Open Science and FAIR practices will increase awareness and spur further adoption
- **Best practice guidelines:** guidelines for pseudonymisation and consent will help researchers identify and apply appropriate techniques in their studies
- **GDPR Certification and Codes of Conduct:** EU Codes of Conduct and certification of research activities could increase harmonisation and increase confidence in compliance with the GDPR
- **Resources for data protection authorities:** Greater resources and research expertise within national data protection authorities could accelerate approvals and advice
- **Smoother paths for data transfers beyond the EU:** the creation of sector-specific adequacy decisions or model SCC could facilitate data transfers and sharing

5. Data sharing in EU-funded dementia research projects

The Horizon 2020 Framework Programme for Research and Innovation was launched by the European Commission in January 2014. With a budget of more than EUR77 billion available over 7 years (2014–2020), Horizon 2020 was the largest EU Research and Innovation (R&I) programme to date, exceeding the budget of the preceding FP7 Framework Programme by almost EUR25 billion. Designed around three pillars – Excellent Science, Industrial Leadership and Tackling Societal Challenges – Horizon 2020 aimed to drive economic growth and job creation by funding a portfolio of projects spread across a wide range of instruments and Actions. These include collaboration-based Actions such as Research & Innovation Actions (RIA) and Innovative Training Networks (ITNs), alongside single beneficiary grants such as the Marie-Sklodowska Curie Actions (MSCA), European Research Council (ERC) and SME Instrument actions. Horizon 2020 has also contributed funding to existing EU Joint Programming Initiatives such as the JPND (Joint Programming Initiative on Neurodegenerative Diseases Research), which was formed in 2010 and pools national efforts and funding from 28 countries to support research projects on neurodegenerative disease. Since 2010, EUR 100 million in funding has been committed by JPND members, with EUR 10 million in co-funding from the Horizon 2020 budget.

In addition, Horizon 2020 provides funding for non-grant actions such as prizes, as well as funding for public-private partnerships (PPP), which bring together academic institutions and private sector organisations in collaborative research projects. The PPP funded by Horizon 2020 include the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking, which became the world's largest life sciences PPP when it was launched in 2014 with a budget of EUR3.27 billion. EUR1.63 billion of the IMI2 budget was sourced from the “Health, Demographic Change and Wellbeing” Societal Challenge pillar of Horizon 2020, with the vast majority of the remaining budget being committed by pharmaceutical companies (represented by EFPIA/European Federation of Pharmaceutical Industries and Associations). Similar to the JPND, IMI2 organises its own research agenda, with a multi-annual strategic research agenda that identifies specific priority areas for investment.

For this section of the report, we surveyed the dementia research projects funded by Horizon 2020 and the IMI, to complement the 5-yearly internal mapping exercise performed by the JPND (the most recent of which was published in 2016). The charts and results in this section were generated based on an analysis of a dataset downloaded from the EU Community Research and Development Information Service (CORDIS) platform in July 2020, which includes the details of all EU R&I activities supported by Horizon 2020 and the IMI since the Framework Programme and IMI2 Joint Undertaking were launched in 2014.

5.1 Dementia research projects in Horizon 2020

As indicated above, the Horizon 2020 Framework Programme has a budget of over EUR 77 billion, deploying these funds to support Research and Innovation activities across a broad spectrum of social, physical, engineering and life sciences. At the time of writing, over 29,000 projects were listed in the CORDIS database of Horizon 2020-funded projects. CORDIS provides details on the project acronym, title, objective, duration, funding scheme, call, amount of funding provided, and coordinator country (among other categories). To identify relevant projects for our analysis, keyword searches were performed in the ‘Objective’ category, which describes the project aims and goals. Keyword searches for the terms “Alzheimer” and/or “dementia” identified 222 projects (projects identified as “closed” or “terminated” are not included in this

number) that have been funded since 2014. The total funding for all dementia projects amounted to EUR 573,633,790.28, of which EUR 445,623,027.76 was contributed by the Horizon 2020 Framework Programme (see table below).

5.1.1 Horizon 2020 dementia research: country participation

Project coordinators (defined as the legal entity or individual acting as the central contact point for the project) were based across 24 countries, including 20 of the EU28 Member States and four Associated Countries to Horizon 2020: Israel, Norway, Switzerland and Turkey. A further 6 countries were identified as participants (but not coordinators) in Horizon 2020-funded dementia research projects, including Canada, Cyprus, Estonia, Slovakia, Romania and the US. No Horizon 2020 dementia research project coordinators or participants were identified for Bulgaria, Lithuania and Malta.

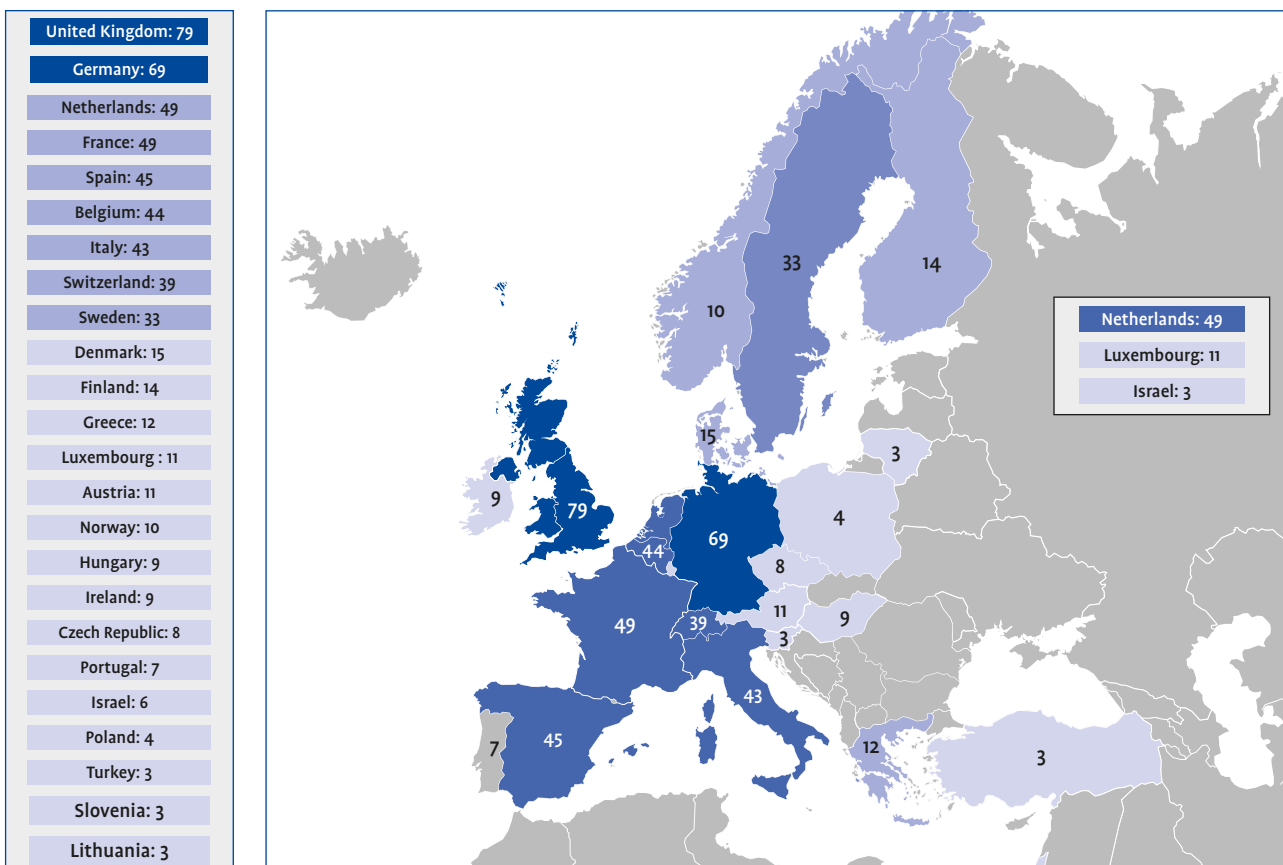
Number of AD & dementia projects	222
Total budget (EUR)	573,622,790.28
EC budget contribution (EUR)	445,623,027.76
Participating countries	30



Partners in 9 EU countries are involved in 78% of H2020 dementia research projects

Horizon 2020 is open to participation from legal entities based in the EU Member States and in non-EU countries associated to the Framework Programme. Across all projects, 78% involved partners based in just 9 countries in western Europe, suggesting a concentration of research expertise in dementia in these countries: Belgium, France, Germany, Italy, Netherlands, Sweden, Switzerland and the UK (Figure 1). Partners in the UK were involved in over 35% of projects as a coordinator and/or participant (79), closely followed by Germany, with German organisations participating in over 31% of projects (69). Partners based in France and the Netherlands participated in the same number of projects (49), equivalent to 22% of all Horizon 2020-funded dementia projects. At the other end of the scale, partners in Lithuania, Slovenia and Turkey each participated in 3 projects, with countries including Denmark (15 projects), Finland (14 projects) and Greece (12 projects) in an intermediate position.

Figure 1. Country participation in H2020 dementia projects (number of projects)



UK-based partners coordinate a large proportion of H2020 dementia research projects

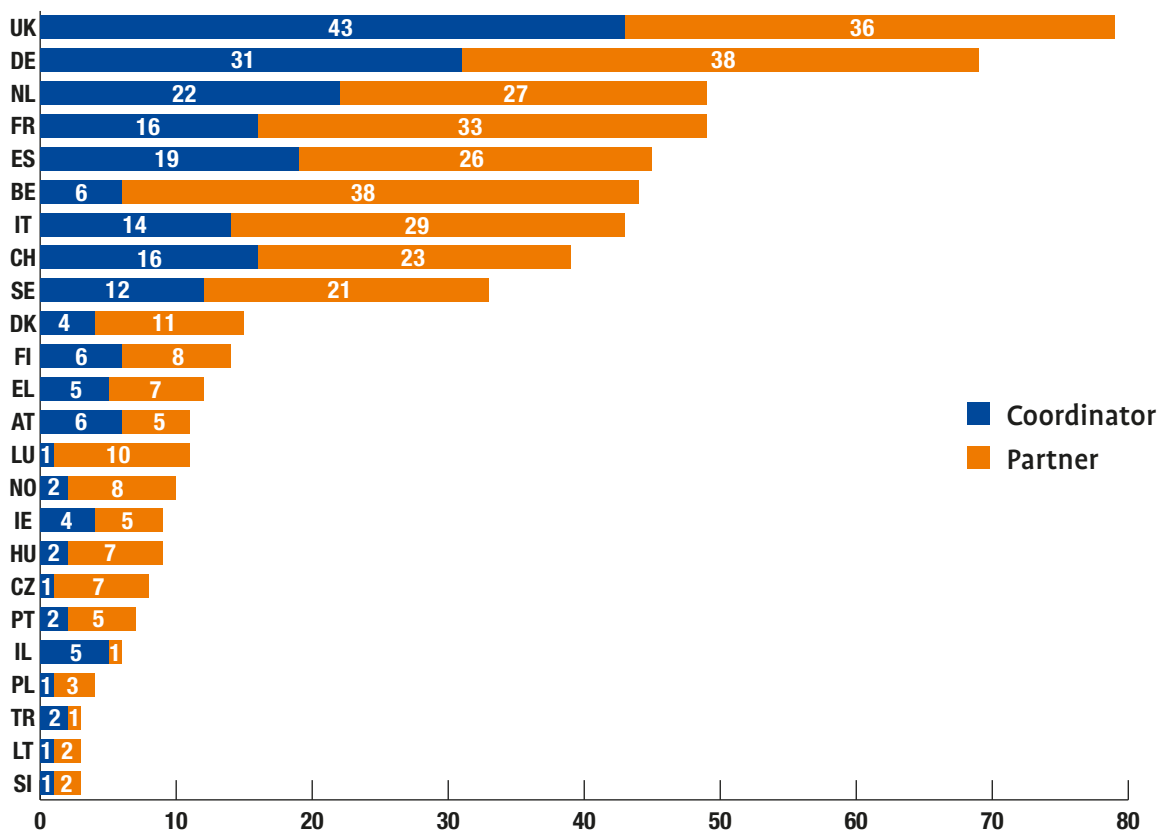
Closer analysis of the CORDIS data revealed subtle differences between countries. Focusing on the mode of project participation (i.e as project coordinator or partner), organisations based in the UK were most frequently named as coordinators of Horizon 2020 dementia projects (43 projects: Figure 2). The UK was also identified as a leading funding recipient (EUR139,588,450.0⁴). In comparison, organisations in Belgium

⁴ It should be noted that funding for consortium projects is usually divided up between partners, so the funding amounts identified for consortium projects will not necessarily reflect the funding amount allocated to UK-based partners within the consortia

and Germany were more frequently involved in Horizon 2020 dementia projects as participants (38 each). However, Belgium-based organisations coordinated substantially fewer projects than organisations based in Germany (6 projects vs 31 projects), and received less Horizon 2020 funding compared to organisations in Germany (EUR10,841,605 [BE] vs. EUR76,087,897[DE]).

Similarly organisations in, Luxembourg (10 projects), Norway (8 projects) and the Czech Republic (7 projects) were more frequently involved in Horizon 2020 dementia projects as partners, rather than coordinators. Of note, Luxembourg-based organisations partnered in a large number of dementia research projects relative to the size of this small country, participating in a total of 10 projects. At the other end of the scale, Austria (6 of 11 projects), Israel (5 of 6 projects) and Turkey (2 of 3 projects) were more frequently involved in Horizon 2020 dementia projects as coordinators (Figure 2).

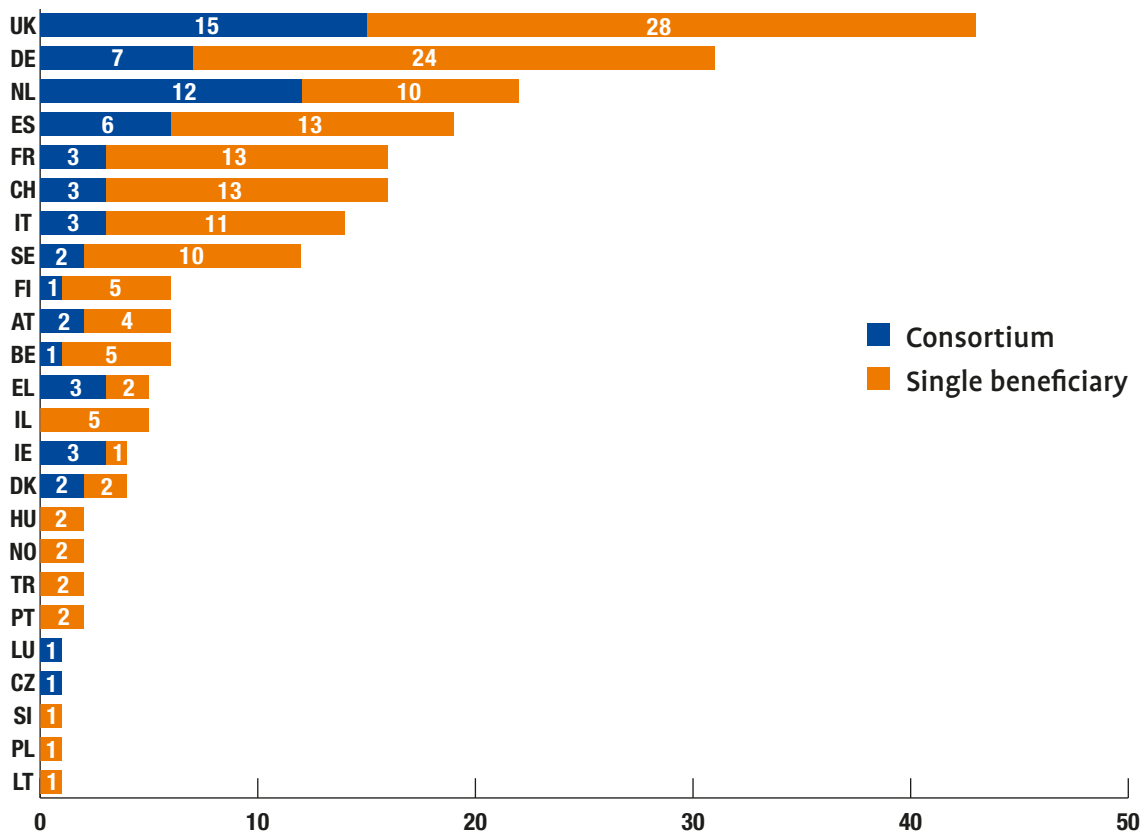
Figure 2. H2020 dementia project participation as coordinator (blue) or partner (orange)



Centres of dementia research excellence are prominent among coordinating institutions

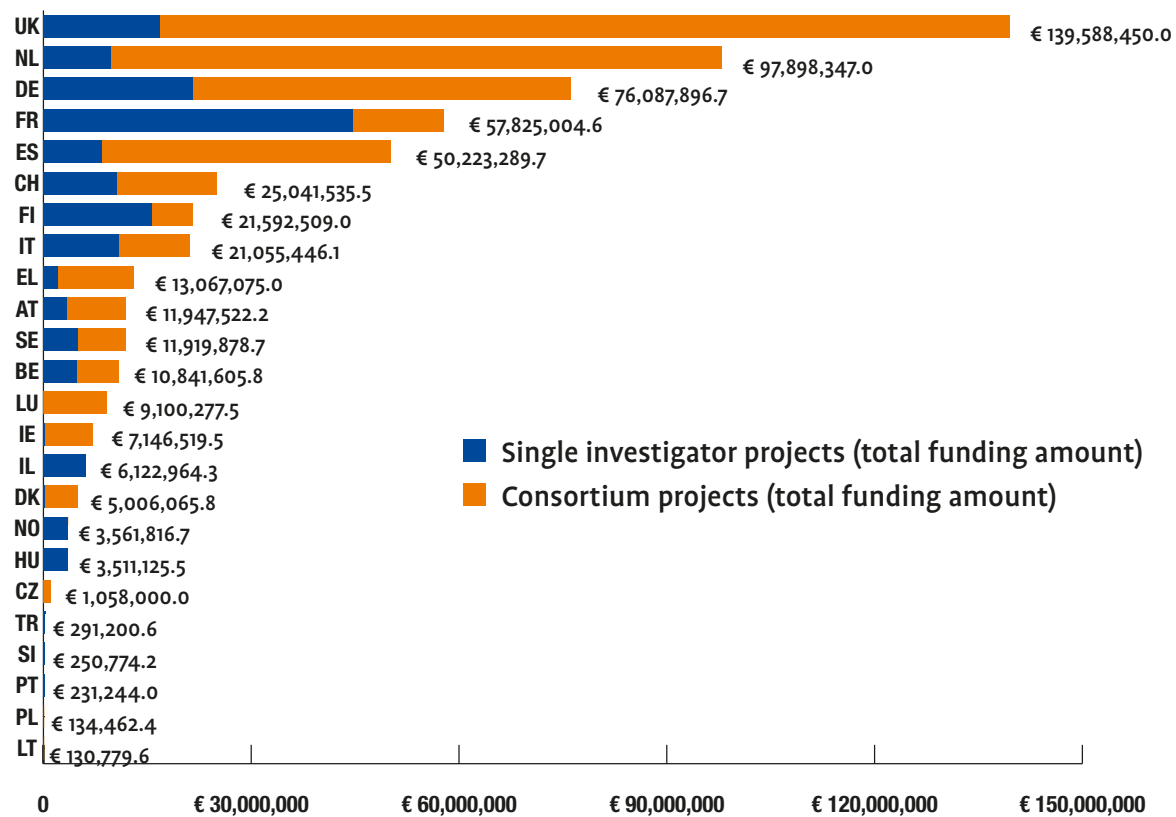
Horizon 2020 projects can broadly be classified into single beneficiary grants (awarded to individual investigators, institutions or organisations) or consortium grants, which are coordinated by a single institution but involve 2 or more partnering organisations. Focusing on project coordination, UK-based institutions coordinated the largest number of consortium projects (15 projects), followed by the Netherlands (12 projects) and Germany (7 projects) (Figure 3). 8 countries did not coordinate any consortium projects, including Lithuania, Poland, Slovenia and Hungary. Partners in the UK also coordinated a substantial number of single beneficiary grants (28 projects), similar to Germany (24 projects) (Figure 3).

Figure 3. Coordination of consortium (blue) or single-beneficiary (orange) projects, by country



Interestingly, 11 projects were coordinated by the same UK institution (University of Oxford), with a further 12 projects coordinated by institutions in Cambridge and London. In terms of Horizon 2020 funding budget, EUR 122,828,509.72 was attributed to the 15 consortium projects coordinated by UK-based partners, with the remaining 29 single-investigator projects receiving EUR 16,759,940.32 (Figure 4). Perhaps unsurprisingly, given the large number of consortium projects coordinated by Dutch partners (12 projects), the Netherlands was the next largest funding recipient, with a total budget of EUR 97,898,347.03. Of this, only EUR 9,761,446.48 came from single-beneficiary grants, with the remaining budget allocated to consortium projects coordinated by Dutch partners. Amsterdam-based institutions such as Amsterdam UMC were most highly represented among the coordinating partners (6 of 12 projects). At the other end of the scale, a substantial proportion of the funding budget allocated to French partners resulted from the coordination of single-beneficiary projects, amounting to EUR 44,703,394.18 of the total EUR 57,825,004.6 for all projects coordinated by organisations based in France (Figure 4).

Figure 4. Total funding allocation by country, for single-beneficiary grants (blue) or multi-beneficiary grants (orange)



5.1.2 Horizon 2020 dementia research: funding schemes

As indicated above, Horizon 2020 provides grants to single beneficiaries as well as multi-partner consortia. Examples of schemes that provide grants to single beneficiaries include the MSCA (via its Individual Fellowship schemes) and SME Instruments actions; in addition, all the funding schemes operated by the ERC apart from its Synergy grant scheme are aimed at individual investigators. On the other hand, Horizon 2020 funds multi-partner collaborations via its Research & Innovation Actions (RIA), via the IMI Joint Undertakings, as well as through the MSCA Innovative Training Network (ITN) and Research & Innovation Staff Exchange (RISE) actions.

Multi-partner collaborations receive six times more funding than single-beneficiary grants

Of the 222 Horizon 2020 projects on Alzheimer's and/or dementia, 65 were collaboration-based actions, involving consortia with partners based in 2 or more countries and funded through the RIA, IMI or other collaboration-based actions. In total, these 65 projects received EUR 403,365,957.68 in funding, with an average of EUR 6,205,630.11 in total funding per project. Of particular relevance to data sharing, and highlighting the importance of efficient and secure data flows across borders, two of the largest consortium projects involved partners based in 16 different countries, with an average of 6 countries represented in each collaborative dementia research project. The remaining 157 projects were single beneficiary grants, allocated to individual investigators, with a total funding allocation of EUR 170,256,833.60. Each single-beneficiary

project receiving on average EUR 1,084,438.43 in total funding – a sixth of the average budget allocated to consortium projects.

Supporting the next generation of dementia researchers and public-private collaborations

Looking more closely at the different Horizon 2020 actions, over 50% of the dementia research projects funded by Horizon 2020 received funding through the MSCA Individual Fellowship (75 projects) or ERC actions (61 projects) (Figure 5, upper panel). These single-beneficiary funding awards are primarily focused on training junior researchers (MSCA) and supporting “frontier” research driven by early-career investigators (ERC), showing an investment of the EU in developing the next generation of dementia researchers. Among all the funding actions represented in the Horizon 2020 dementia project dataset, public-private partnership projects funded through the IMI received the largest proportion of budget, amounting to EUR 181,811,982.2 in total (approximately half of this was contributed by EFPIA companies) for 14 projects, followed by the RIA actions, which received EUR 139,444,207.4 for 21 projects (Figure 5).

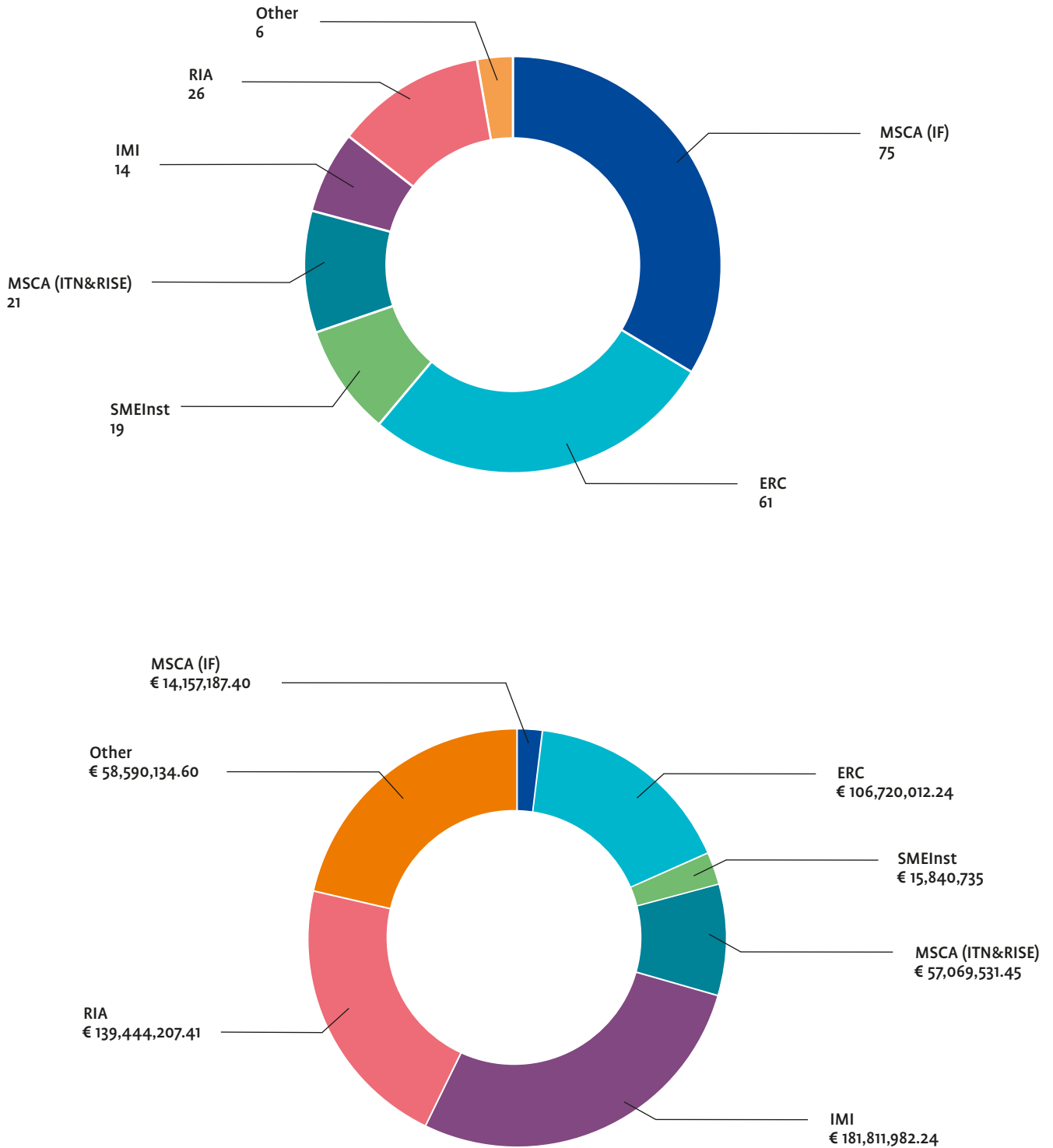
The Horizon 2020 Framework Programme included a mandate for SME participation in collaborative projects, stating that 13% of budgets for the “Leadership in Enabling Technologies” and “Societal Challenges” two-year Work Programmes should be allocated to SMEs. While the amount of Horizon 2020 dementia research funding allocated to the SME Instrument actions is relatively small (EUR15,840,735.0; Figure 5, lower panel), the substantial budget allocated to the IMI and RIA actions emphasizes the support of the EU for dementia capacity-building in the private sector, as well as the value placed on public-private collaborations in dementia research. From a data sharing perspective, the relatively large number of dementia research projects that include SMEs and Industry alongside academic participants underscores the importance of addressing the ethical and data governance challenges associated with the transfer of data between public and private sector institutions.

5.1.3 Data use in Horizon 2020 dementia research projects

The dementia research projects funded by the Horizon 2020 Framework Programme cover the full research spectrum from fundamental research studies in cells, model proteins and animals, to clinical or psychosocial research involving human participants, epidemiological, public health and health systems research projects that process and analyse Big Data. As such, the dementia research projects supported by Horizon 2020 generate a diverse range of datasets, from quantitative data on cellular responses to disease-associated proteins, to DNA sequences, cognitive test results, brain imaging datasets and beyond.

As outlined in the previous section of this report, the technical, legal, motivational and ethical barriers and facilitators for data sharing vary substantially depending on the type of data that is being shared. For example, there are significantly greater legal barriers to sharing clinical data (due to concerns around patient privacy) compared to data arising from *in vitro* studies of cell behaviour; while brain imaging data from EEG scans cannot be shared or reused without specially structured databases and defined metadata. To gain an understanding of the different types of data being used and generated in the dementia research projects funded by Horizon 2020, we performed a detailed analysis of each project abstract in our CORDIS-derived database, assigning an identifier to denote whether the project involved the use of clinical or preclinical data, or whether it was an informatics project.

Figure 5. Funding schemes for H2020 dementia research projects. Upper panel: number of projects by funding scheme. Lower panel: total budget by funding scheme.

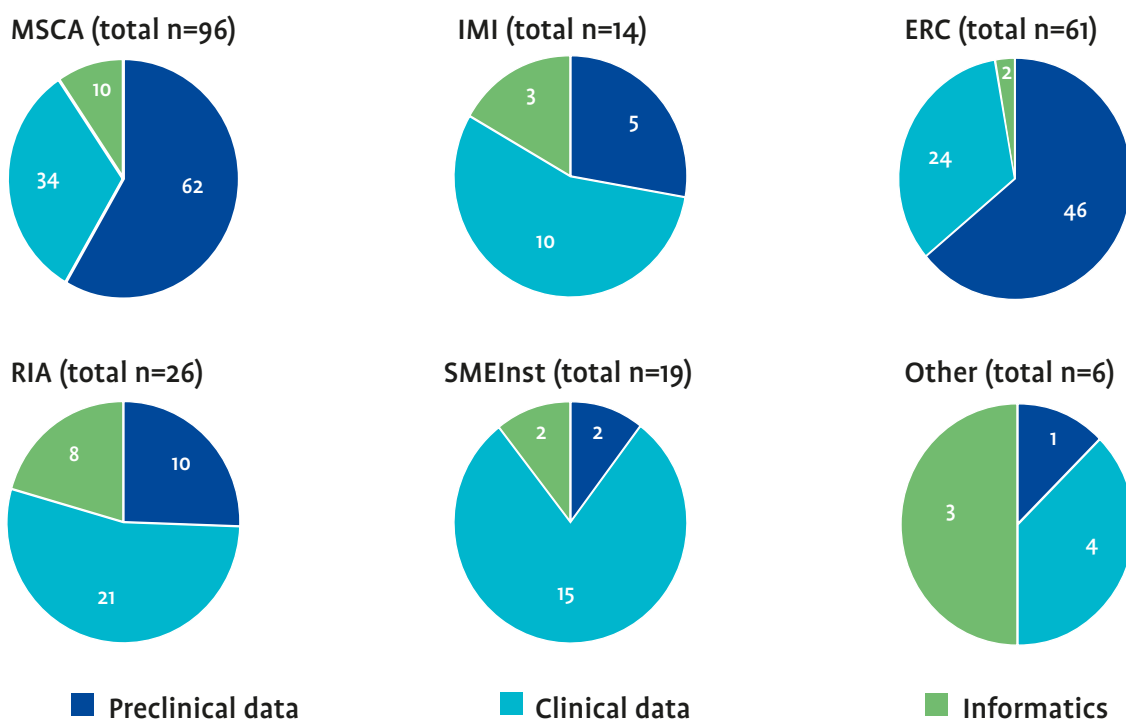


Clinical data is used in almost 50% of Horizon 2020 dementia research projects

According to this analysis, of the 222 dementia research projects funded by Horizon 2020, over half (n=126) used or generated preclinical data from *in vitro*, *in vivo* or *in silico*⁵ studies. A slightly lower number of projects (n=108) used or generated data from patients or research participants (termed “clinical data”), while 25 projects were categorised as informatics research that involved the development of data platforms or analytical processes. There were a number of translational, interdisciplinary research projects that involved the use of two or more different types of data, such as the IMI-funded PHAGO project, which is performing detailed analyses of brain-derived stem cells *in vitro* whilst also assessing the PET scans of people with Alzheimer’s disease.

Looking more closely at data use in dementia research projects funded through different Horizon 2020 actions, we observed that projects funded through the MSCA and ERC actions most frequently involved the use of preclinical data (62/106 projects and 46/72 projects, respectively), in line with the fact that these actions often focus on “frontier”⁶ and fundamental science (Figure 6). Conversely, dementia research projects funded by the IMI, SME Instrument and RIA actions more frequently involved clinical data (71%, 79% and 80% of projects, respectively) (Figure 7). Interestingly, a substantial proportion of the IMI dementia projects (7 of 14) and RIA dementia projects (10 of 26) involved the use of two or more classes of data, underlining the translational and interdisciplinary nature of the research funded through these two actions.

Figure 6. Numbers of H2020 dementia research projects using preclinical, clinical or informatics data, by funding scheme.

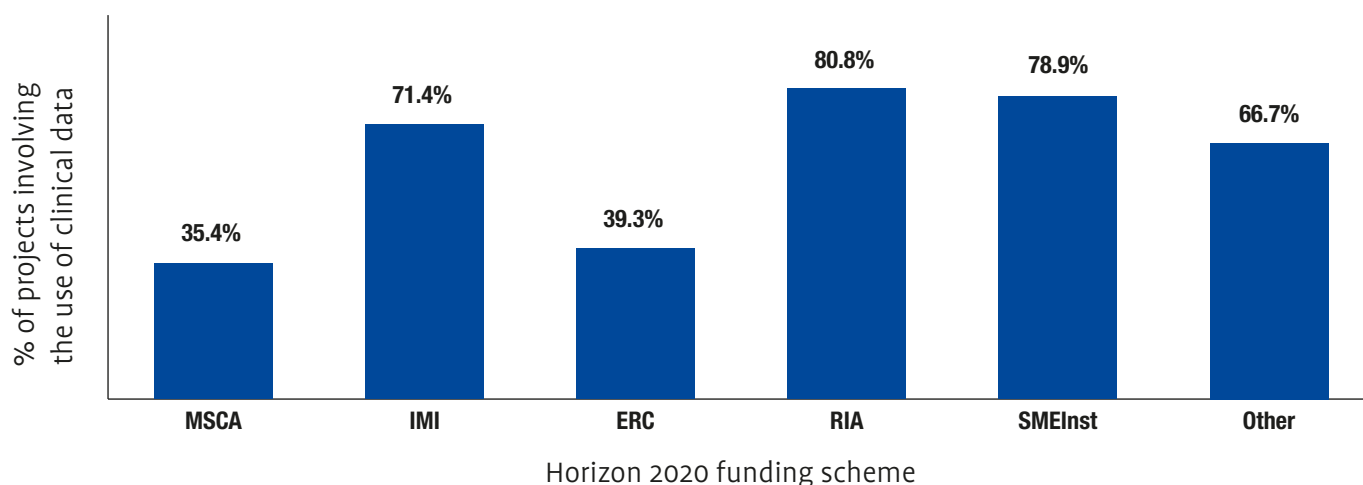


MSCA = Marie-Sklodowska Curie Actions; ERC = European Research Council; RIA = Research & Innovation Actions; IMI = Innovative Medicines Initiative; SMEInst = SME Instruments

⁵ *In vitro*, *in vivo* and *in silico* studies refer to studies that are carried out in laboratory-cultured cells (*in vitro*), using preclinical animal models (*in vivo*) and using virtual or simulated approaches (*in silico*)

⁶ Defined as research that takes place at the limits of existing knowledge

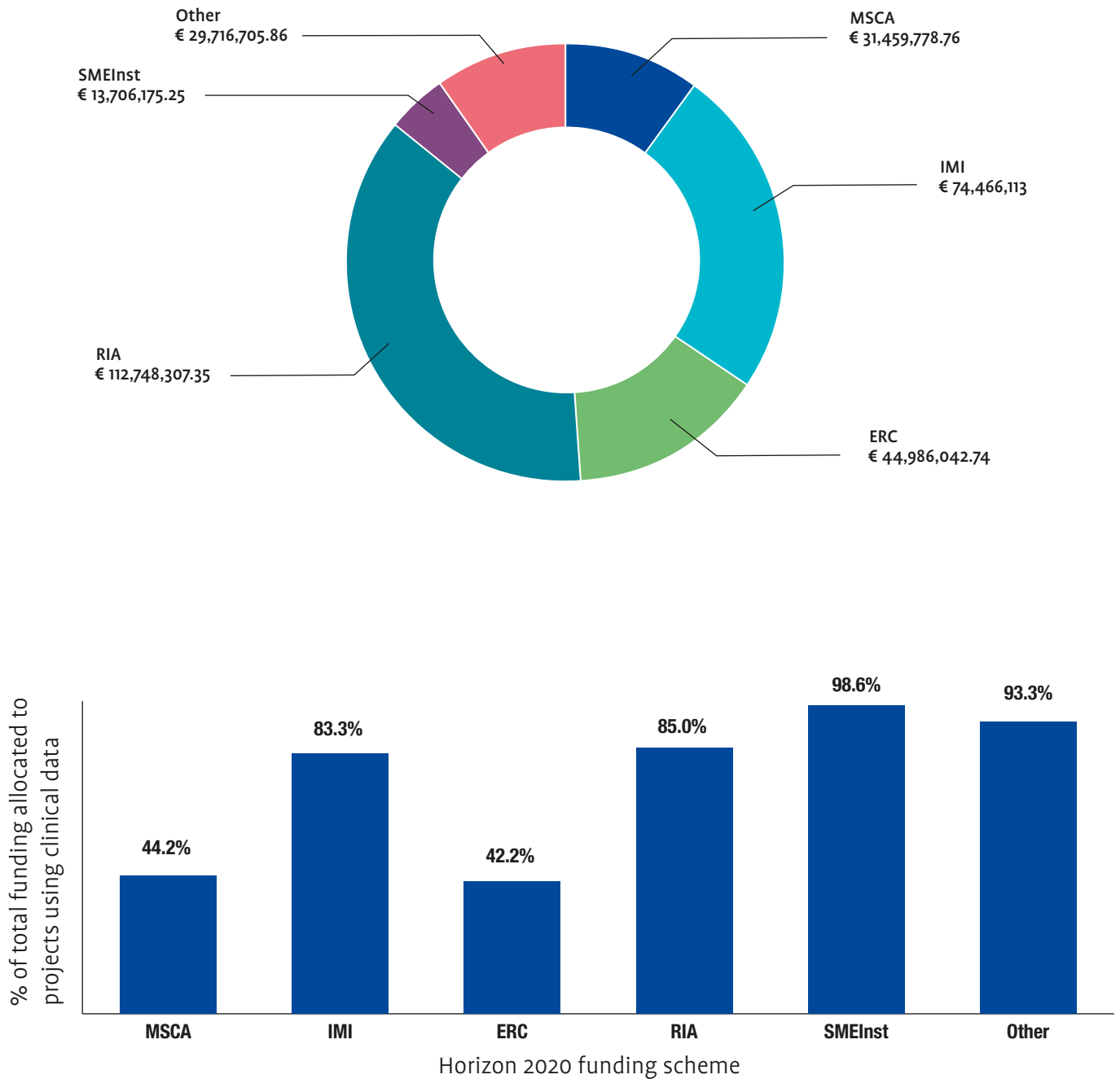
Figure 7. Percentage of dementia research projects in different H2020 funding schemes that involve the use of clinical data.



73% of funding allocated to dementia research projects involving the use of clinical data

The sharing of clinical data poses particular challenges due to a number of factors, including the sensitive nature of these datasets, the potential identifiability of the data subjects, and the rigorous ethical and legal requirements that need to be met. Looking at the 108 dementia research projects involving the use of clinical data, we observed that these projects received a large proportion of the total funding allocated through Horizon 2020 (EUR 417,116,441.30) compared to the 114 projects that did not involve the use of clinical data, which received EUR 156,517,348.9. Focusing on the budget allocated by the European Commission to different Horizon 2020 actions, a substantial majority of funding for dementia projects using clinical data was provided via the IMI and RIA actions (Figure 8, upper panel). Indeed, over 80% of the European Commission funding provided via the IMI, RIA actions was allocated to projects involving the use or generation of clinical data (Figure 8, lower panel). As such, a large number of Horizon 2020 dementia research projects funded through the IMI and RIA actions will be faced with data sharing challenges at several levels, due to the need to share clinical data, across borders, and between public and private sector participants – some of whom may be based outside the EU.

Figure 8. H2020 funding allocation to projects involving clinical data, by funding scheme (upper panel; funding amount) and percentage of total budget awarded to dementia research projects (lower panel).



5.2 Data sharing in dementia research projects: case studies

Data sharing is often viewed as the final stage of the research lifecycle, coming after research planning, project execution, data analysis and dissemination of findings. However, data sharing can also be the starting point for research: in the introduction, we described some of the ways researchers have used shared data to validate dementia risk prediction models and develop new biomarkers for dementia progression. Researchers are increasingly aware of the benefits of data sharing, and the growing adoption of data sharing policies has paved the way for new dementia research projects that embed data sharing by design. In this section, we describe three such projects: the US-based Alzheimer’s Disease Neuroimaging Initiative (ADNI), the European Medical Information Framework (EMIF) and the European Prevention of Alzheimer’s Dementia (EPAD) project.

5.2.1 ADNI: the Alzheimer’s Disease Neuroimaging Initiative

By the early 2000s, researchers were realising that the clinical and neuropsychological measures commonly used as outcomes in AD trials could not always distinguish between disease-modifying and symptomatic effects of trial drugs. In addition to these “traditional” measures, more sensitive biomarkers were necessary: fluid-based biomarkers, detected in blood or cerebrospinal fluid (CSF) samples, and neuroimaging biomarkers, obtained using brain imaging scans. In addition, new AD disease models were required, in which these biomarkers could be traced and mapped across the timecourse of AD development – from its earliest, pre-symptomatic phases to the final, advanced AD dementia stages.

Against this backdrop, the US National Institute on Ageing (NIA) of the National Institutes of Health (NIH) launched the Alzheimer’s Disease Neuroimaging Initiative (ADNI), one of the first and largest public-private partnerships worldwide. The overarching goal of ADNI was to characterise how biomarkers change during the development of AD, to enable earlier diagnosis and improve monitoring of disease progression. Since its inception in 2004, ADNI has launched four study phases, starting with ADNI-1 and followed by ADNI-GO, ADNI-2 and ADNI-3, with funding amounting to almost USD 220 million. During the first three ADNI phases, researchers gathered and analysed brain scans, genetic profiles, fluid biomarkers and clinical data from people with AD, early or late MCI, or significant memory concerns (SMC), as well as elderly people without cognitive impairment. ADNI-3 continues the work of the earlier phases, adding PET scans that detect the presence of tangled Tau proteins to the brain imaging workup that included MRI, FDG-PET and in some participants, amyloid PET scans. To date, the 59 clinical centres for ADNI in North America have enrolled almost 2000 individuals, representing the entire Alzheimer’s disease continuum(55).

The extensive, longitudinal characterisation of each ADNI participant, managed by the Clinical, PET, MRI, Biomarker, Genetics, Neuropathology, Biostatistics and Informatics Cores, makes ADNI an unparalleled resource for scientific discovery and clinical innovation. ADNI has also been a game-changer when it comes to data sharing; the ADNI-1 grant stipulated that all data collected in the ADNI database would be made available, without embargo, to all scientists who requested it. By embedding data sharing at the heart of ADNI and by reducing inconsistencies via the adoption of harmonised protocols and procedures, ADNI has generated data that has been downloaded over 144 million times and has been used in over 1800 scientific publications.

The impact of ADNI extends to the validation of imaging modalities for improved clinical trials; the development of new tools for measuring fluid biomarkers; and the identification of novel genetic risk factors for AD, to name but a few of the innovations that have resulted from ADNI data re-use(56). ADNI has also paved the way for similar neuroimaging initiatives in Europe, such as Swedish ADNI, Italian ADNI and AddNeuroMed, which was funded by InnoMed (the precursor to the IMI) and used the same MRI sequences as ADNI(57).

5.2.2 EMIF: the European Medical Information Framework

On a structural level, ADNI has showcased how the public-private partnership model can drive research and innovation in the AD field. Building on the success of ADNI, the [EMIF](#) project was launched in January 2013, aiming to develop a common information framework of patient-level data that would enable scientists to develop and test new hypotheses on the causes, prevention and treatment of disease(58). One of the largest projects funded by the first IMI Joint Undertaking, EMIF had a total budget of EUR55,784,311.0, divided across [three sub-projects](#): EMIF-Metabolic (EMIF-MET), EMIF-Alzheimer's Disease (EMIF-AD) and EMIF-Platform (EMIF-PLAT)(59). EMIF was funded for a 5-year period, bringing together 57 partners from academic research institutions, EFPIA companies, SMEs and patient organisations, including Alzheimer Europe.

The substantial efforts invested by EMIF in data sharing were primarily centred on EMIF-PLAT, which created structures and tools to enable researchers to find, evaluate, use and re-use health data from different sources in Europe. On one level, EMIF-PLAT was developed as a unified platform to support the wide range of clinical studies being conducted as part of EMIF-MET and EMIF-AD; to facilitate the use of these data for large-scale research on biomarkers and risk factors for disease. On another level, EMIF-PLAT aimed to support a federated network of electronic health record (EHR) data sources, by integrating healthcare databases and EHR cohorts in different countries. This vast resource of real-world data, including EHRs from over 60 million Europeans, holds enormous potential to increase our understanding of AD in the real-world setting.

To make these goals a reality, huge efforts were invested in ensuring EHR and EMIF cohort metadata were harmonised, to enable researchers to combine and aggregate datasets from different sources. EHR databases were mapped to the OMOP common data model, with similar harmonisation efforts applied to EMIF clinical data, such as the data collected from the 11 cohorts making up the EMIF-AD [multimodal biomarker study](#)(60). To facilitate access to these data resources, EMIF-PLAT created the [EMIF Catalogue](#), a web portal designed as an interface between data custodians and researchers wishing to access and evaluate [data](#)(61). In addition, EMIF-AD and EMIF-PLAT developed an AD database on tranSMART to store and share harmonised individual level data from over 12 cohorts, including omics data generated as part of the EMIF-AD multimodal biomarker study.

Using the EMIF Catalogue, bona fide researchers can browse through an extensive library of federated EHR and cohort databases, querying their metadata to identify datasets that could help answer their research questions. EMIF-PLAT also supported researchers in their efforts to obtain datasets from custodians, providing template legal agreements and giving researchers access to a private remote research environment for secure data analyses. To allow the level of data governance to be tailored to the specific requirements of different datasets, the EMIF Catalogue was structured around individual data communities. 10 such data communities are currently hosted in the EMIF Catalogue, including the EMIF-AD community, which provides detailed information about 65 AD cohort studies, and the EMIF-EHR community, which federates healthcare databases from several EU countries.

To date, results from EMIF-AD and EMIF-PLAT have been published in over [140](#) peer-reviewed [scientific articles](#), substantially impacting the field of dementia research(62). In particular, both projects have made major contributions to our understanding of early biomarkers for AD; thanks to the sharing and re-use of AD cohort data, researchers have been able to identify novel fluid-based and neuroimaging biomarkers, showing how these biomarkers could be deployed for AD risk assessment and diagnosis. Alongside, the analysis of -omics data from people across the AD spectrum has allowed researchers to gain deeper insights into the pathophysiology of AD, identifying how different genetic risk factors contribute to the development of disease. Finally,

analysis of the EMIF EHR resources have provided insights into the incidence of mild cognitive impairment and dementia in Europe, and on the prevalence of ageing-associated comorbidities across populations.

“During its 5-year funding period, EMIF developed a number of assets for the dementia research community. Using our data catalogue, researchers are able to identify cohorts for their own projects. The EMIF data platform allows us and others to export clinical research data to collaborators that wish to analyse that data. And the dataset from the EMIF-MBD study continues to yield new insights into the causes of, and contributors to dementia.

Although ongoing work with the assets from projects such as EMIF shows that data sharing is both achievable and beneficial, there are still some challenges to overcome. A fundamental problem is that the landscape of dementia research data remains fragmented: at the platform level, cohorts can be listed in several catalogues, providing data to different extents, through various access pathways. Platforms employ varying ontologies and data dictionaries, which complicates harmonisation. Arranging data access and transfers can also be challenging, involving time-consuming legal and administrative processes.

In an ideal world, there would be a single, centralised access point for cohort data, with harmonised datasets and simple access models that are not costly to set up, or time-consuming to navigate. To help us reach that point, we need more local support for researchers to prepare and share data; clearer regulations for legal departments to follow; and greater harmonisation between platforms.”

Prof. Pieter-Jelle Visser, University of Maastricht, Department of Psychiatry and Neuropsychology

5.2.3 EPAD: the European Prevention of Alzheimer’s Dementia project

In 2015, IMI launched the [EPAD](#) project as part of Call 11, which was focused on the creation of a European platform to facilitate proof of concept trials for prevention of Alzheimer’s dementia. With funding of EUR58,986,698, EPAD was designed to address the dual challenges of identifying people who are likely to develop Alzheimer’s dementia, and our relatively poor understanding of the early stages of AD. A further aim was to lay the foundations for adaptive trials in AD, a more flexible trial design in which parameters of the trial protocols and/or statistical procedures can be modified during the trial’s course depending on the interim results.

To achieve these goals, [EPAD](#) developed three core clinical research strategies: a registry, a cohort and a trial(63). First, to help reduce clinical study screening failures it created the EPAD Registry, a pan-European register of over 500,000 people across the risk spectrum for dementia(64). Described as a participant discovery platform, the EPAD Registry includes minimal data on preselected participants from ongoing cohort studies, which can be queried using the PrePAD software tool. As well as providing a tool to facilitate pre-screening, the EPAD Registry also laid the foundations for an EPAD Trial Delivery Network composed of 29 centres across [Europe](#)(65). Cohort participants from the EPAD Registry were invited to join the EPAD Longitudinal Cohort Study ([LCS](#)), which served two research goals: firstly, to act as a readiness cohort for the EPAD trial (PoC; a platform, adaptive Proof of Concept Phase 2 trial), and secondly to generate data for disease modelling of the preclinical stages of AD(66).

Designed as a prospective, multicentre, pan-European study, the EPAD LCS recruited participants without dementia, aiming to cover the entire “probability spectrum” for AD dementia development. Participants

in the EPAD LCS underwent regular health checks, standardised tests and brain scans, returning for yearly follow-up visits. Between May 2016, when the first participant was recruited, and February 2020, when recruitment was halted, the EPAD LCS screened 2,096 participants, generating extensive patient-level datasets that include many parameters essential for accurate disease modelling. These include cognitive outcomes, neuroimaging and fluid biomarker assessments (including cerebrospinal fluid/CSF), physical examinations and sociodemographic information.

One of the founding principles of EPAD was to make the LCS data FAIR and [openly available](#) to bona fide researchers. After a brief embargo period, LCS datasets were periodically released to the research community every 6 months from May 2019 onwards, culminating in the release of the final dataset (v.IMI) in October 2020. To facilitate data sharing, EPAD partners Aridhia used its Digital Research Environment to harmonise data from the 29 European centres participating in the EPAD LCS. By creating a secure, online workspace for researchers to use, Aridhia was able to ensure that the sharing and reuse of EPAD datasets meets the information governance protocols established at the start of the project. This data was the first to be incorporated on the ADDI platform, the pilot phase of which was undertaken jointly by Aridhia and the EPAD team at the University of Edinburgh.

By embedding a culture of data sharing, EPAD has not only created an unparalleled data resource for AD disease modelling; it has also developed a secure infrastructure that will enable researchers to access and analyse the LCS datasets as well as an array of software tools and processes for data exploration, tracking and harmonisation. In addition, while EPAD did not meet its objective of initiating a secondary AD prevention trial (the EPAD PoC), in creating the EPAD Registry and developing a network of trial delivery centres it has laid the groundwork for more streamlined, effective and robust studies to trial therapies for AD dementia.

“Our philosophy in EPAD was to view the Longitudinal Cohort Study (LCS) data and samples as gifts from the people who participated in the LCS. The LCS research participants got involved in EPAD because they wanted to advance the knowledge of Alzheimer’s disease. As custodians of their data, it is therefore our responsibility to ensure this data generates as much knowledge as possible. This means making the data openly accessible, at scale, and in the long-term. During the four years of the LCS, data on a wide range of cognitive, biomarker and neuroimaging outcomes was gathered. We also collected over 100,000 biosamples, which are securely stored in the EPAD biorepository. Together, these assets will continue to generate new knowledge on the causes, diagnosis and prevention of Alzheimer’s disease.”

The EPAD cohort recruited over 2,000 individuals across the risk spectrum of Alzheimer’s disease in Europe. However, there is often a disconnect between research cohorts and the general population. By their nature, cohorts tend to be more homogeneous, and do not always capture the diversity present in society. Conversely, data collected in real-world clinical practice can lack measures of value for dementia research – particularly in mid-life age groups, a key target group for dementia prevention. To look beyond the treatment of established disease, and towards the maintenance of brain health, we now need better mechanisms to collect, analyse and share real-world data across the lifecourse. This underpins the establishment of Brain Health Services as is the case through Brain Health Scotland, a Scottish Government backed initiative which commenced in 2020. This will pave the way for early detection, risk profiling and personalised dementia prevention, empowering people to reduce their risk of dementia through positive preventive measures.”

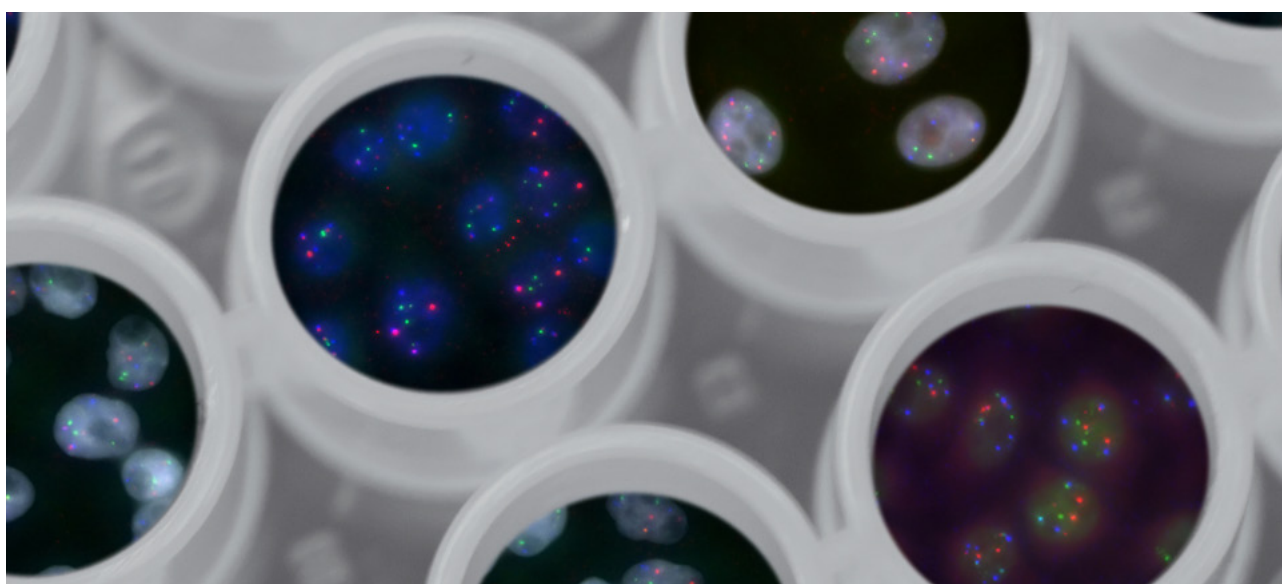
**Professor Craig Ritchie, Director of Edinburgh Dementia Prevention, University of Edinburgh
Director of Brain Health Scotland**

5.3 Section summary: dementia research projects and data sharing

Together, our analyses of the dementia projects funded by the Horizon 2020 Framework Programme reveal a diverse research portfolio, covering frontier research on the underlying causes of dementia to clinical studies of dementia risk and epidemiological analyses of electronic health record data. They also reveal a diverse range of funding beneficiaries, from early-career researchers receiving Marie Curie fellowships, to biotech spin-offs, SMEs, large, research-intensive universities and medical centres, many of which collaborate in multi-partner consortium projects.

Totalling just over EUR570 million, the EU investment in dementia research projects has particularly benefited research institutions in 9 western European countries, with 78% of project partners based in the UK, Germany, the Netherlands, France, Spain, Belgium, Italy, Switzerland and Sweden. UK-based partners coordinate a large proportion of Horizon 2020-funded dementia research projects, with centres of dementia research excellence such as the University of Oxford and Amsterdam UMC featuring prominently on the list of coordinating institutions. Although 50% of the dementia research projects funded through Horizon 2020 are single-beneficiary Fellowships or ERC grants, the public-private partnership projects of the IMI received the largest proportion of funding, at over EUR180 million. Of note, the majority of these collaborative, multi-partner and cross-sector projects involve the use of clinical data, and almost 75% of the total funding for dementia research is allocated to projects with a clinical research component. Many, if not all, of these projects may be faced with data sharing challenges at several levels, underlining the importance of developing robust, secure mechanisms to share clinical data across borders, and between public and private sector participants both inside and outside the EU.

The three projects presented in the previous section showcase data sharing as a multiplier of impact for dementia research. By embedding principles of data sharing by design, ADNI, EMIF and EPAD have contributed to important new discoveries on genetic risk factors, biomarkers and preventive treatments for AD and dementia. As large-scale, public-private partnerships, projects like these are ideally placed to facilitate translation of this new knowledge into improved diagnostics and therapies for people living dementia. As projects that exemplify data sharing in practice, ADNI, EMIF and EPAD have ensured that generations of dementia researchers to come will be able to benefit from their valuable datasets and resources.



6. Data sharing – perspectives from researchers and participants

6.1 Researcher perspectives

There is broad acceptance of data sharing in principle across researcher communities, with numerous studies showing strong recognition of the benefits of data sharing – as well as an increasing willingness to share and reuse data. In part, this is thanks to the adoption of data sharing policies by funders and publishers of research, which often require the data underlying published research to be shared or, at least, made available upon request. The US National Institutes of Health were among the first funding agencies to introduce policies on sharing data in 2001, and since then policies adopted by the International Committee of Medical Journal Editors ([ICMJE](#)), [Wellcome](#) Trust, [Nature Research Journals](#) and others have strongly incentivised researchers to share their data. For example, since the Nature research journals introduced a requirement for data availability statements, [88% of authors](#) have stated that they will make data available upon request(67).

Moving beyond research policy, a majority of researchers understand that there are substantial scientific, financial and ethical rewards to be gained from sharing data. For example, over 68% of respondents to a [2015 survey](#) of biomedical researchers stated that they shared data to further a collaboration, while 64% cited a desire to advance science in a particular area(68). Most participants in a [2020 study](#) involving genomics researchers highlighted the importance of data sharing for societal benefit, to “*reduce disease or reduce suffering due to disease*”(69), and researchers are increasingly aware of the [ethical imperative for data sharing](#), to honour the contribution of those participating in clinical research.

Motivational barriers to data sharing

However, researchers also report several obstacles to data sharing in practice. A recurring theme in surveys are social or motivational barriers to data sharing. In particular, researchers draw attention to the fact that current academic reward systems place a particularly high value on grant funding and high-impact publications, preferably as a first or senior author. A 2015 [article](#) on academic data sharing labelled these systems as “reputation economies”, highlighting a perceived “reputation cost” for sharing data(70). While most researchers agree that data should be shared, 80% of the 2661 researchers surveyed cited concerns that “*other researchers could publish before me*” and 78% wanted to “*publish before sharing*”. Similarly, clinical trialists at a data sharing Summit hosted by the New England Journal of Medicine (NEJM) in 2017 [argued](#) that an expectation that data should be swiftly shared could disincentivise clinicians from conducting trials, as their substantial efforts would not be sufficiently recognised in the “reputation economy”(71). Linked to this, trust – or lack thereof – was identified as a barrier to data sharing. Traditionally, research data has been shared through tight-knit, collaborative networks, ensuring that data generators could exert a degree of control over when and how their data was being reused. Many trialists at the NEJM summit worried that open sharing of complex data could lead to misinterpretations or misuse by researchers with no direct links to the original study.

Researchers – and in particular, researchers working on clinical studies – also cite the financial and time cost of data sharing as a key challenge to overcome. As described in the introduction, clinical research datasets are becoming increasingly complex, with multiple variables that need to be suitably annotated,

cleaned, and anonymised where necessary. Ethical and data protection requirements add a further level of complication: for individual participant data from clinical studies, data access committees need to manage requests and ensure that participants are not re-identified from their data. Although many data sharing platforms and repositories now exist, a degree of technical skill is often required to navigate them. To share data well takes time and effort – and not a small amount of cost, depending on the scale and type of data to be shared. An example of particular relevance to dementia research is neuroimaging data from MRI scans. To effectively reuse MRI data, researchers need to have detailed information on experimental design, data acquisition, image pre-processing and analysis parameters. Due to these and other challenges, data handling and sharing is often a fragmented activity, with individual researchers, research groups and institutions adopting independent strategies. Unsurprisingly, a recent [meta-synthesis](#) of studies on academic data sharing identified the provision of data storage infrastructure and data stewardship support as an important enabler for data sharing(72).

*“In biomedical research we face a number of challenges that are related to data sharing. First, a **reproducibility crisis**: Scientific studies are often difficult to replicate and findings often do not generalize in light of additional data. Second, **complexity** of the biological processes prevents us from understanding the mechanisms underlying disease. We need better theoretical frameworks to infer the underlying causes from disparate observations. Thirdly, a lack of **technical infrastructure** for sensitive data processing in compliance with GDPR and means of its certification – to protect the privacy of personal data.*

The Horizon 2020-funded Project “Virtual Brain Cloud” targets these challenges - thereby contributing to the implementation of the European Open Science Cloud (EOSC). To understand the complexity of biological processes, we are developing multi-scale simulations to provide a theoretical framework for multimodal data integration, using mathematical models. To enable reproducibility, we are sharing the data and computational steps that underly our research findings as well as the explicit workflows describing how to generate the results.

World-wide exchange of data is crucial. Data privacy is also crucial. There is no simple solution to simultaneously and efficiently meet both these needs. The solution proposed by Virtual Brain Cloud is to use encryption, sandboxing and access control as technical means to protect personal data. Inside containers, the data is always encrypted, and access rights to the containers are controlled via access tokens. A registry is set up that associates data sharing agreements with data containers and data controllers or processors. In this way, we can manage data access - and protect the rights of the people to whom the data belong.”

**Professor Petra Ritter, Director of the Brain Simulation Section
Charite University Hospital, Berlin
Coordinator of the Virtual Brain Cloud Project**

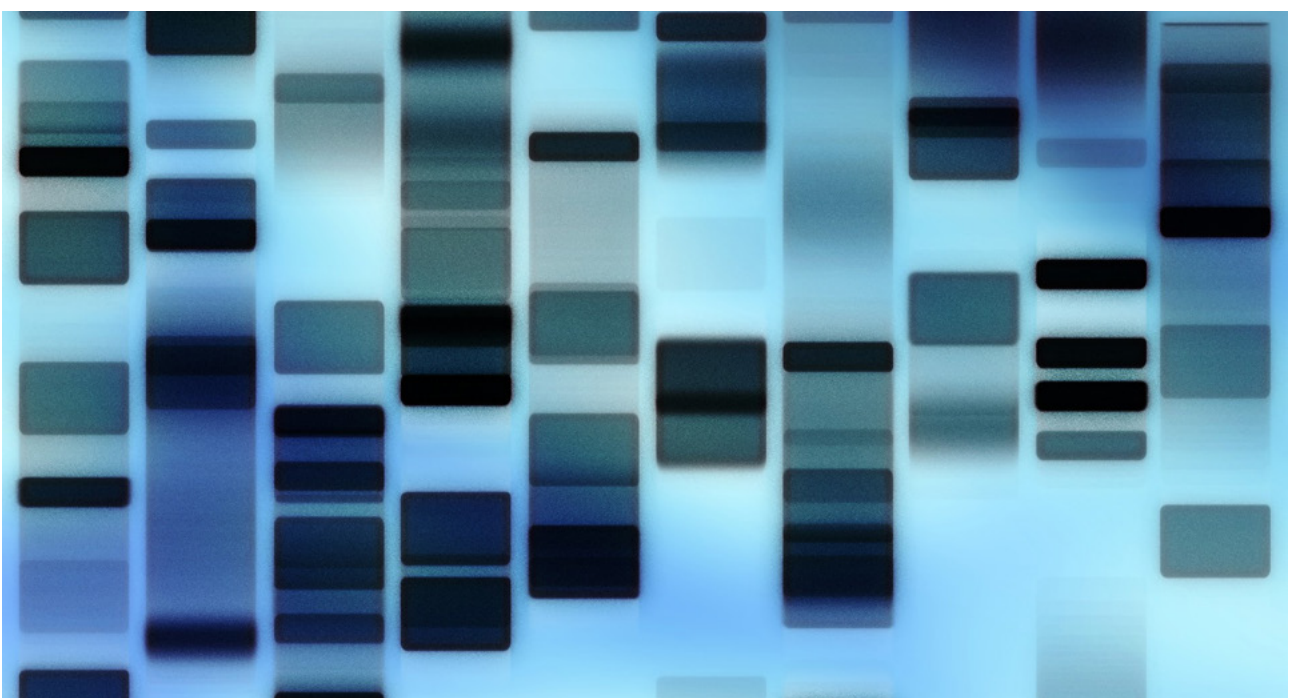
6.2 Participant perspectives

Similar to surveys of research communities, studies on participant and patient preferences have identified broadly supportive attitudes towards clinical data sharing. Speaking alongside clinical trialists the NEJM [data sharing summit](#), trial participants voiced their belief in sharing data and experiences in order to help themselves and fellow patients, encouraging researchers to look beyond concerns around loss of data

authorship and patient confidentiality(73). A [follow-up survey](#) confirmed the willingness of many trial participants to share data: fewer than 8% of the 771 respondents felt the potential negative consequences of data sharing outweighed the benefits(74). The desire to help others as much as possible was a dominant theme in the survey, with several respondents urging greater cooperation and less competition among researchers. An earlier [focus group study](#) of participants in the ACT aging and dementia cohort identified scientific advancement, research efficacy and health improvements as important outcomes from data sharing: one participant stated “...I think there does have to be an open exchange of information in order for some of these really significant things to happen for peoples’ benefit”(75). Similarly, a [systematic literature review](#) of healthcare consumers found that respondents across studies recognised the importance of research and its benefit to society, and the role that data sharing can play in advancing research(76). Together, these studies indicate relatively broad acceptance of the principles of data sharing for individual and societal benefit.

Trust and transparency

But what about attitudes to data sharing in practice? Despite broad agreement on the value of data sharing, patients and participants have voiced concerns about the potential loss of privacy and a perceived lack of transparency in how and when data is shared. These concerns are rarely black and white, instead existing on a continuum that varies depending on the type of data being shared, and the individuals or organisations it is being shared with. For example, systematic reviews indicate high levels of trust in using data from disease registries, which hold clinical data about people diagnosed with a specific disease or condition. The Rare Disease Barometer Data Protection and Sharing [survey](#) of 2,013 people living with a rare disease was published by EURORDIS in 2020, revealing that 97% of respondents would be happy to share their disease registry data for research purposes(77). Similarly, a majority of respondents to a [survey](#) conducted with members of the European Leukodystrophies Association agreed with the principle of sharing disease registry data(78). Conversely, while a 2019 [survey](#) of 1,246 hospital patients also revealed broad agreement with data sharing in principle, 76.6% of survey respondents identified one or more electronic health record data items they were not prepared to share(79).



Patients and research participants also hold strong views on who they would like their data to be shared with, particularly when these data are not anonymised. [Studies](#) frequently noted a reduced willingness to share data with pharmaceutical companies and insurers, with lower levels of trust in the ethical use of data by these organisations(76). For example, a 2016 Wellcome Trust survey on public attitudes to commercial access to health data showed that 17% of respondents objected to private companies having access to their health data under any circumstances. Worryingly, a [recent survey](#) indicates that data scandals involving organisations such as Cambridge Analytica and Google have damaged public trust in data sharing, with over 23% of US-based respondents stating that they are unwilling to share their health data for any reason(80). Data scandals have also damaged trust in Europe; a [2018 survey](#) conducted with over 8,000 respondents from Germany, Finland, France and the Netherlands revealed that almost 70% had changed their privacy settings and/or use of digital services as a result of data leaks, with 42% of respondents distrusting digital service providers with their personal data(81). Indeed, trust – or rather, distrust – is a common theme across studies, particularly regarding the potential loss of privacy; data privacy is addressed in greater detail in section 4.2 above.

Similar opinions about data sharing have been elicited during focus group consultations carried out with the European Working Group of People with Dementia (EWGPWD), which was launched by Alzheimer Europe in 2012. The [EWGPWD](#) is composed of people with dementia who are nominated by their national Alzheimer associations, and currently includes 14 members from Austria, Germany, Ireland and the UK, among other countries. The members of the EWGPWD are actively involved in many of Alzheimer Europe's activities, projects and meetings, helping to generate ideas for research, advising researchers and contributing their views through focus groups and consultations. EWGPWD members participating in a [2018 focus group](#) for the [ROADMAP](#) project expressed conditional acceptance of many data sharing activities, stating that only necessary data should be shared, with the necessary safeguards(82). Similar views were expressed in a 2019 [consultation](#) for the RADAR-AD project, which explored issues linked to the use of remote monitoring technology (RMT)(83). EWGPWD members participating in this consultation raised concerns about the breadth of data collected using RMT, and the possibility of RMT data ending up in the wrong hands.

Loss of privacy and misuse of data were identified as key risks in both ROADMAP and RADAR-AD focus groups, with many citing concerns about their data being accessed by individuals without a legitimate reason to do so. In the ROADMAP consultation, Linda⁷ stated “*You wonder about in a care home or in a hospital [...] how much protection you have, because everybody can tap into [...] your records. So I'd have a question about that, you know, that it's free for all – that it's free for all staff. So should that be right? I don't think it should be.*” Similar to previous surveys, the concept of trust figured prominently in focus group discussions, and the need to have a degree of trust in healthcare and research systems to underwrite participation in research. “*Complete transparency*” at each stage of research was identified as an enabler of trust, along with public engagement and involvement activities.

Patients as partners in the research process: public involvement

While similar themes have emerged across studies, surveys and focus group consultations on data sharing, individuals have varying opinions on the associated benefits and risks. As such, it is important not to presume there is a social licence for data sharing; the diverse values of patients, participants and the public should also be incorporated in data governance and sharing frameworks. A common theme across studies is the importance of considering the patient voice in debates around data sharing, and the value

7 Names were changed to maintain confidentiality

of involving research participants in decisions on sharing their clinical data. At the 2017 NEJM Data Sharing Summit, [Sharon Terry](#) (a patient advocate and the president of Genetic Alliance) stated “*Trial participants are not patients in the traditional sense of the word. It really should be looked at as a partnership.*”(84)

A cross-sectional [survey](#) of participants in the European DIRECT (Diabetes Research on Patient Stratification) raised a similar point, with over 50% of respondents stating that they would like to be involved in decisions on how and with whom their data should be shared(85). Going beyond the views of research participants, the 2019 Alzheimer Europe [publication](#), “Overcoming ethical challenges affecting the involvement of people with dementia in research”, discusses how public involvement of people with dementia can contribute to the quality, relevance and ethical conduct of research, including questions around data sharing and governance(86). Indeed, questions on data sharing are frequently raised during public involvement (PI) consultations with the EWGPWD, reinforcing the value of PI for data-driven dementia research.

Similarly, a [recent report](#) on maximising the impact of National Health Service (NHS) data on the health and wealth of the UK highlighted the value of engaging with the public (and particularly people from under-represented groups) to gain a cohesive view of the acceptable uses of health data, and the trade-offs that people are willing to make between sharing data for societal benefit and potential loss of privacy(87). To facilitate these discussions, [DataSavesLives](#) and [use MY data](#) are involving people living with disease in consultations on data sharing, and have developed tools such as patient data citations to acknowledge the important contribution of patients and participants to research.

6.3 Section summary: researcher and participant perspectives

Overall, the body of literature on researcher, patient and research participant views of data sharing reveals a broadly positive picture. However, there are areas of shade amidst the light. Although there is widespread acceptance of the principle of data sharing for societal benefit, more nuanced views exist when it comes to what data is shared, who it is shared with – and when. For patients and research participants, the benefit of data sharing comes with a privacy trade-off, and the willingness of individuals to accept this privacy risk will vary depending on their perception of the benefit(s), and their level of trust in the systems used to share and re-use data. For researchers, there are important motivational considerations for data sharing, linked to the “publish or perish” culture that characterises many academic reward systems. In addition, data sharing comes with a financial and time cost, a burden that many academic researchers in particular are unable to bear.

There are several potential enablers that could help overcome the aforementioned challenges. Tipping the privacy risk balance towards societal benefit could be achieved through the adoption of more harmonised data protection frameworks that provide standardised tools to ensure patient and participant anonymity (a point that was addressed in greater depth in section 4.2). Obtaining and incorporating the views of patients and research participants on data sharing is also a valuable step towards ensuring transparency – and increasing trust. To overcome motivational barriers to data sharing, [research stakeholders](#) are now working on developing systems that link data generators with the reuse of datasets that they generate, using approaches such as data citations and persistent data DOIs (digital object identifiers) to ensure data generators are credited when their data are reused(88). Interestingly, [evidence](#) from the International Neuroimaging Data sharing Initiative (INDI) indicates that there may be a net career benefit to data sharing; publications using shared data are well-represented in high-impact journals, and include more junior researchers alongside senior investigators(89).

From a financial and technical perspective, open-access platforms such as the Alzheimer’s Disease (AD) Workbench have an important role to play in supporting researchers to share their data. Developed by the [Alzheimer’s Disease Data Initiative](#), a medical research organisation dedicated to advancing dementia research, the AD Workbench provides an accessible and secure way for researchers to find, use and share their data. Increasing awareness of the positive impact of openly-shared data on the scientific literature – and of the platforms, tools and support available to facilitate data sharing – will hopefully encourage more positive researcher attitudes towards data sharing.

Obstacles to data sharing

- **“Reputation economies”**: academic reward systems that primarily value senior authorship, high-impact publications and grant income
- **Lack of researcher trust**: concern that data will be misinterpreted or misused
- **Financial and time cost of data sharing**: researchers may not have the resources or time to prepare and share their data
- **Technical challenges**: data curation, harmonisation and annotation are complex processes requiring technical expertise
- **Privacy concerns**: patients and research participants worry about a loss of confidentiality and privacy when their data is shared
- **Lack of transparency**: many patients and research participants would like to know how and when their data is used

Facilitators for data sharing

- **Data sharing policies**: funders and publisher policies which mandate or endorse data sharing
- **Established methods to attribute data authorship**: data DOIs and citations
- **Systems for monitoring data sharing and reuse**, to further incentivise data sharing
- **Technical and administrative support** for data stewardship and storage
- **Increased transparency for patients and participants** on how their data are used and shared
- **Data privacy tools and methods** that ensure patient and participant anonymity
- **Public involvement**: involving patients and the public in research, to ensure that data is managed, used and shared in ways that are worthy of trust

7. Future perspectives

In 2019, when the work for this discussion paper was started, infectious threats and pandemics were far from our minds. At the time of writing, however, over 2 million people have lost their lives to COVID-19, making it the worst public health crisis in over a century.

From a public health perspective, experience from previous pandemics has shown that effective actions to restrict virus transmission are dependent on a continual supply of disaggregated data, to ensure that decisions are based on the best available evidence. The unprecedented, global scale of the COVID-19 pandemic has therefore magnified the importance and necessity of data sharing for public health purposes. The huge investments in research on COVID-19 and SARS-CoV-2, the coronavirus which causes the disease, have also reinforced the importance of sharing data generated by these research projects. For example, the [UKCDR/GloPID-R](#) database, which maps funded research projects across the world related to COVID-19, currently includes 5,046 studies across 102 countries. In the six months between January and June 2020 alone, over 23,000 scientific articles on COVID-19 were published in peer-reviewed journals – an astonishing rate of almost 1,000 papers per week.

Recognising the need for tools to manage, share and leverage this abundance of data, a large number of organisations have created COVID-19 data and knowledge sharing platforms. Many of these platforms aim to facilitate the secure exchange of data between stakeholders and across borders, to improve public health responses and accelerate the development of treatments and vaccines. A recent report by the International COVID-19 Alliance identified almost 100 COVID-19 data repositories, platforms, databases and libraries, such as the EU COVID-19 data portal and platforms for clinical trial datasets such as [Vivli](#), [IDDO](#) and [ISARIC](#).

Thanks to the rapid and open sharing of thousands of SARS-CoV-2 genome sequences through platforms like GISAID, epidemiologists have been able to track the geographic spread and transmission dynamics of COVID-19, improving our understanding of how different measures can reduce viral spread. From a clinical perspective, trials for COVID-19 treatments have progressed at an unprecedented pace, thanks to the creation of large, international multi-site trials that generate and share enormous volumes of research data. For example, the [SOLIDARITY](#) trial launched by the World Health Organisation and partners has enrolled over 12,000 participants in 500 hospital sites across 30 countries, aiming to evaluate the efficacy of repurposed drugs. By working in unison with standardised clinical protocols, trial procedures and data collection modalities, feeding into a single trial database, SOLIDARITY has conclusively identified which treatments have little or no benefit in terms of mortality – all within a period of 6 months. Finally, on the societal scale, surveillance dashboards and data trackers created by organisations such as the [WHO](#) and the [ECDC](#) (European Centre for Disease Control) have helped inform policymakers at local and national level, informing decisions on when and how to institute physical distancing interventions.

As such, the response to the COVID-19 pandemic has illustrated how data can become a pillar in the fight against a global threat. It has shone a light on best practices for data-driven policymaking, and has shown that data **can** be shared in a spirit of open collaboration across sectors and between partners.

As a result of the pandemic, healthcare systems have taken large steps towards the adoption of telehealth, which fundamentally relies on data sharing to enable remote consultations with patients. The development of digital biomarkers (as well as the sensors and tools required to measure them) has also been accelerated. Building on smartphone-based symptom trackers for COVID, companies such as Apple

It should also not be forgotten that dementia, like the COVID-19 pandemic, is a public health crisis. The numbers of people with dementia are continually increasing, as are the associated health and social care costs. People with dementia have been disproportionately affected by COVID-19. As well as being at higher risk of mortality and morbidity, many have experienced a worsening of symptoms due to social isolation and lack of access to care. Despite this, due to the tightening of EU research budgets, there is a risk of dementia research being deprioritised, threatening its future viability and long-term sustainability. The [position statement](#) of Alzheimer Europe, published in July 2020(90), therefore calls on the EU and national governments to ensure that there is continued investment in dementia research, so that the gains from the projects showcased in this report are not lost.

“The COVID pandemic has changed a lot of things for researchers. We have witnessed much higher preparedness to share data, with researchers and policymakers working in unison to advance science. We are also going beyond sharing just data: in our research centre, we are sharing data models and AI workflows, which can be tested and trained behind institutional firewalls without compromising patient privacy. This is helping us to develop more representative synthetic datasets, based on real-life, anonymised patient data.

Looking to the future of dementia research, I would like to see more widespread use of remote, digital measures of health and disease, to complement traditional clinical measures such as neuropsychological tests. Smartphone apps that analyse speech and monitor gait are much less costly and invasive than lumbar punctures and PET scans, and could potentially be used as surrogate measures of clinical disease. This could also address inequalities in research participation, reducing the ethnic bias that currently exists in many clinical studies. Rather than multiple platforms using different data models, we need more unified semantics and data ontologies, that are broadly adopted across research communities. But it’s not just about collecting more and more data – it’s about what you do with it. So that we can capitalise on these advances, we need to ensure that informatics experts have a place at the top table of decision-makers, to help clinicians and scientists deliver innovations and medical progress.”

**Prof. Martin Hofmann-Apitius, Bioinformatics Group Leader,
Fraunhofer Institute for Algorithms and Scientific Computing**

8. Conclusions: barriers and enablers for data sharing

Researchers, research participants, funders and policymakers broadly agree that responsible data sharing can accelerate scientific progress, leading to medical improvements that directly benefit patients and citizens. Despite this consensus view, and as outlined in this discussion paper, data sharing is far from being common practice, particularly in the context of clinical research. So, what can be done to encourage data sharing?

The first step is to understand what data is being shared, and what isn't. As outlined in the introduction, a 2018 study revealed that only 50% of clinical trials registered in the EU had reported results summaries in the EU clinical trials database (EUCTR) within a year of completion, as required by law. In response, several EU and US groups have set up [Trials Trackers](#) to independently monitor compliance with clinical trials reporting rules. To assess data sharing from Industry-sponsored drug trials, researchers in the US have recently developed a [scorecard](#) to measure clinical trial data sharing policies and practices in pharmaceutical companies(17). Encouragingly, at the time of writing 61.5% of due trials had reported results to the EUCTR within the 1-year timeframe, up from 50% in 2018. The European Commission is also monitoring participation in the Open Research Data component of Horizon Europe. It is hoped that both data monitors will show a similar upward trend when it comes to data sharing.

The next step is to understand why data isn't being shared. As outlined in the section summaries of this report, frequently-cited reasons include privacy concerns and the technical difficulties of sharing data, as well as the high financial cost of data sharing. When studies end, so do the associated funding streams, which means that data sharing can be financially unsustainable in the long term. Worryingly, many researchers feel that academic systems do not adequately incentivise data sharing (section 5.1: researcher perspectives). In addition, researchers have to navigate a complex, risk-averse regulatory environment to ensure against the loss of privacy (section 4.2: GDPR).

Using this knowledge, we need to develop improved environments and methods for data sharing. To overcome motivational barriers, methods to better incentivise data sharing have been proposed, such as ensuring researchers act as [data stewards](#) and are [fully credited](#) when their data are re-used(91). When it comes to patient privacy and data protection concerns, GDPR Codes of Conduct and tailored standard contract clauses may create paths for faster yet secure sharing of research data between sectors and across national borders. Artificial Intelligence may provide novel solutions: computer scientists on the [AETIONOMY](#) project are aiming to create a [virtual dementia cohort](#) that could simulate real patients, while the [VirtualBrainCloud](#) project is developing a GDPR-compliant cloud platform for brain simulations and dementia diagnostics. Curation of older, incomplete or non-interoperable datasets, as recently published for AddNeuroMed (92), could enhance the utility and accessibility of valuable data resources that are currently under-used. Linked to this, a number of [initiatives](#) to support data sharing from a technical perspective are well underway: the [EHDEN](#) project is creating a federated data network to enable access to harmonised health data, while platforms such as the Alzheimer's Disease Workbench will enable researchers to work with multiple datasets in a secure environment. Finally, stakeholders from the whole research spectrum should work together to provide continued support and funding for data governance, management and sharing, to ensure that data resources remain findable, accessible and shareable long beyond their project funding periods.

Scientific progress thrives when the evidence base is complete and openly accessible, allowing researchers to build on, challenge and refine the findings of their peers. As our data processing capabilities increase, so will the rewards for responsibly sharing that data, paving the way for new dementia diagnostics, treatments and care.

9. References

1. Luengo-Fernandez, R., Leal, J. & Gray, A. UK research spend in 2008 and 2012: comparing stroke, cancer, coronary heart disease and dementia. *BMJ Open* **5**, e006648 (2015).
2. James, B. D. *et al.* Contribution of Alzheimer disease to mortality in the United States. *Neurology* **82**, 1045–1050 (2014).
3. BioArctic announces positive topline results of BAN2401 Phase 2b at 18 months in early Alzheimer's Disease. *BioArctic* <https://www.bioarctic.se/en/bioarctic-announces-positive-topline-results-of-ban2401-phase-2b-at-18-months-in-early-alzheimers-disease-3600>.
4. EPAD | European Prevention of Alzheimer's Dementia Consortium. <http://ep-ad.org/>.
5. Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 1–7 (2018).
6. Press Release: Tufts CSDD Impact Report July/August 2018, Vol. 20 No.4, Available Now. *Tufts CSDD* <https://csdd.tufts.edu/csddnews/2018/7/12/press-release-julyaugust-2018-tufts-csdd-impact-report-just-released>.
7. Banks, M. A. Sizing up big data. *Nat. Med.* **26**, 5–6 (2020).
8. S, B. H. *et al.* Clinical Development of Aducanumab, an Anti-A β Human Monoclonal Antibody Being Investigated for the Treatment of Early Alzheimer's Disease. *J. Prev. Alzheimers Dis.* **4**, 255–263 (2017).
9. Birkenbihl, C. *et al.* Differences in cohort study data affect external validation of artificial intelligence models for predictive diagnostics of dementia - lessons for translation into clinical practice. *EPMA J.* **11**, 367–376 (2020).
10. Tijms, B. M. *et al.* Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain* **143**, 3776–3792 (2020).
11. PhRMA Principles for Clinical Trial Data Sharing. <https://www.phrma.org/Codes-and-guidelines/PhRMA-Principles-for-Clinical-Trial-Data-Sharing>.
12. Bauermeister, S. *et al.* The Dementias Platform UK (DPUK) Data Portal. *Eur. J. Epidemiol.* **35**, 601–611 (2020).
13. DPUK Data Portal. <https://portal.dementiasplatform.uk/>.
14. UK Biobank - UK Biobank. <https://www.ukbiobank.ac.uk/>.
15. Astell, M. *et al.* Practical challenges for researchers in data sharing - Springer Nature survey data (anonymised). 655626 Bytes (2018) doi:10.6084/M9.FIGSHARE.5971387.
16. Goldacre, B. *et al.* Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ* **362**, k3218 (2018).
17. Miller, J., Ross, J. S., Wilenzick, M. & Mello, M. M. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *BMJ* **366**, l4217 (2019).
18. trinoma. Data policies and legislation - Timeline. *Shaping Europe's digital future - European Commission* <https://ec.europa.eu/digital-single-market/en/data-policies-and-legislation-timeline> (2020).
19. Welcome | European Union Open Data Portal. <https://data.europa.eu/euodp/en/data/>.
20. Recommendation on access to and preservation of scientific information (2012/417/EU). <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32012H0417&from=EN>.
21. General Data Protection Regulation (2016/679). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
22. Open Innovation, Open Science, Open to the World. *European Commission - European Commission* https://ec.europa.eu/commission/presscorner/detail/fr/SPEECH_15_5243.
23. *Open innovation, open science, open to the world: a vision for Europe*. (Publications Office of the European Union, 2016).
24. EU - Open Science Policies. *European Commission - European Commission* https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en.
25. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

26. Cost-benefit analysis for FAIR research data : policy recommendations. <http://op.europa.eu/en/publication-detail/-/publication/d3766478-1a09-11e9-8d04-01aa75ed71a1/language-en> (2019).
27. About the Open Science Monitor. *European Commission - European Commission* https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/about-open-science-monitor_en.
28. Open Research Data (ORD) - the uptake in Horizon 2020 - Datasets. <https://data.europa.eu/euodp/en/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020>.
29. Open Science Policy Platform: final report. *openscience.eu* <https://www.openscience.eu/open-science-policy-platform-final-report/> (2020).
30. Six Recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group. <http://op.europa.eu/en/publication-detail/-/publication/4630fa57-1348-11eb-9a54-01aa75ed71a1/language-en> (2020).
31. Burgelman, J.-C. *et al.* Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Front. Big Data* **2**, (2019).
32. Gallacher, J. E. J. *et al.* Challenges for Optimizing Real-World Evidence in Alzheimer’s Disease: The ROADMAP Project - IOS Press. *J. Alzheimers Dis.* **67**, 495–501.
33. Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. *Shaping Europe’s digital future - European Commission* <https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering> (2018).
34. Proposal for a Regulation on European data governance (Data Governance Act). *Shaping Europe’s digital future - European Commission* <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act> (2020).
35. Summary report of the public consultation on the European strategy for data. *Shaping Europe’s digital future - European Commission* <https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data> (2020).
36. Ethics guidelines for trustworthy AI. *Shaping Europe’s digital future - European Commission* <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (2019).
37. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. 281 vol. OJ L (1995).
38. Mourby, M., Gowans, H., Aidinlis, S., Smith, H. & Kaye, J. Governance of academic research data under the GDPR—lessons from the UK. *Int. Data Priv. Law* **9**, 192–206 (2019).
39. Peloquin, D., DiMaio, M., Bierer, B. & Barnes, M. Disruptive and avoidable: GDPR challenges to secondary research uses of data. *Eur. J. Hum. Genet.* **28**, 697–705 (2020).
40. Controllers and personal data in health and care research. *Health Research Authority /planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/data-controllers-and-personal-data-health-and-care-research-context/*
41. Clarke, N. *et al.* GDPR: an impediment to research? *Ir. J. Med. Sci.* 1971 - **188**, 1129–1135 (2019).
42. PIRONET, B. Preliminary Opinion on data protection and scientific research. *European Data Protection Supervisor - European Data Protection Supervisor* https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en (2020).
43. Learn More – A Code of Conduct for Health Research. <https://code-of-conduct-for-health-research.eu/faq>.
44. 27/11 – GDPR Code of Conduct for Clinical Trials in progress. <https://www.eucrof.eu/news-eucrof/latest-news/27-11-gdpr-code-of-conduct-for-clinical-trials-in-progress>.
45. Pseudonymisation techniques and best practices. <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>.

46. Eiss, R. Confusion over Europe's data-protection law is stalling scientific progress. *Nature* **584**, 498–498 (2020).
47. Rabesandratana, T. Researchers sound alarm on European data law | Science. *Science* **366**, 936.
48. Hallinan, D. *et al.* International Transfers of Health Research Data Following Schrems II: A Problem in Need of a Solution. *SSRN* (2020) doi:10.2139/ssrn.3688392.
49. Olbrechts, A. Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data. *European Data Protection Board - European Data Protection Board* https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/recommendations-012020-measures-supplement-transfer_en (2020).
50. Bradford, L., Aboy, M. & Liddell, K. International transfers of health data between the EU and USA: a sector-specific approach for the USA to ensure an 'adequate' level of protection. *J. Law Biosci.* (2020) doi:10.1093/jlb/l5aa055.
51. Data as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the GDPR. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0264>.
52. Your rights matter: Data protection and privacy - Fundamental Rights Survey. *European Union Agency for Fundamental Rights* <https://fra.europa.eu/en/publication/2020/fundamental-rights-survey-data-protection> (2020).
53. Health Data in the GDPR era. *Data Saves Lives* <https://datasaveslives.eu/public-opinion/gdpr>.
54. GDPR is Transforming Consumer Trust and Data Security in Europe According to a New Study. *Check Point Software* /press/2019/gdpr-is-transforming-consumer-trust-and-data-security-in-europe-according-to-a-new-study/.
55. ADNI | About. <http://adni.loni.usc.edu/about/>.
56. Weiner, M. W. *et al.* Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimers Dement.* **11**, 865–884 (2015).
57. Frisoni, G. B. Alzheimer's Disease Neuroimaging Initiative in Europe. *Alzheimers Dement.* **6**, 280–285 (2010).
58. EMIF. <http://www.emif.eu/>.
59. Lovestone, S. The European medical information framework: A novel ecosystem for sharing healthcare data across Europe. *Learn. Health Syst.* **4**, e10214.
60. Bos, I. *et al.* The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. *Alzheimers Res. Ther.* **10**, 64 (2018).
61. Oliveira, J. L., Trifan, A. & Bastião Silva, L. A. EMIF Catalogue: A collaborative platform for sharing and reusing biomedical data. *Int. J. Med. Inf.* **126**, 35–45 (2019).
62. PUBLICATIONS | EMIF. <http://www.emif.eu/publications/>.
63. Ritchie, C. *et al.* Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project - The Lancet Psychiatry. **3**, 179–186.
64. Vermunt, L. *et al.* Prescreening for European Prevention of Alzheimer Dementia (EPAD) trial-ready cohort: impact of AD risk factors and recruitment settings. *Alzheimers Res. Ther.* **12**, 8 (2020).
65. Ritchie, C. W. *et al.* The European Prevention of Alzheimer's Dementia (EPAD) Longitudinal Cohort Study: baseline data release v500.0. *J. Prev. Alzheimers Dis.* **7**, 8–13 (2020).
66. Solomon, A., Kivipelto, M., Molinuevo, J. L., Tom, B. & Ritchie, C. W. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. *BMJ Open* **8**, e021017 (2018).
67. The importance and challenges of data sharing. *Nat. Nanotechnol.* **15**, 83–83 (2020).
68. Federer, L. M., Lu, Y.-L., Joubert, D. J., Welsh, J. & Brandys, B. Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff. *PLoS ONE* **10**, (2015).
69. Heather Nick. Researcher Knowledge, Attitudes, and Communication Practices for Genomic Data Sharing.
70. Fecher, B., Friesike, S., Hebing, M. & Linek, S. A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Commun.* **3**, 1–10 (2017).

71. Rosenbaum, L. Bridging the Data-Sharing Divide – Seeing the Devil in the Details, Not the Other Camp. *NEJM* **376**, 2201–2203.
72. Perrier, L., Blondal, E. & MacDonald, H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLOS ONE* **15**, e0229182 (2020).
73. NEJM: Bringing Together Ideas for Sharing Clinical Trial Data. *NEJM Library Hub* <https://libraryhub.nejm.org/article-hub/2017/07/nejm-bringing-together-ideas-sharing-clinical-trial-data/> (2017).
74. Mello, M. M., Lieou, V. & Goodman, S. N. Clinical Trial Participants’ Views of the Risks and Benefits of Data Sharing. *N. Engl. J. Med.* **378**, 2202–2211 (2018).
75. Trinidad, S. B. *et al.* Genomic research and wide data sharing: Views of prospective participants. *Genet. Med.* **12**, 486–495 (2010).
76. Hutchings, E., Loomes, M., Butow, P. & Boyle, F. M. A systematic literature review of researchers’ and healthcare professionals’ attitudes towards the secondary use and sharing of health administrative and clinical trial data. *Syst. Rev.* **9**, 240 (2020).
77. Courbier, S., Dimond, R. & Bros-Facer, V. Share and protect our health data: an evidence based approach to rare disease patients’ perspectives on data sharing and data protection - quantitative survey and recommendations. *Orphanet J. Rare Dis.* **14**, 175 (2019).
78. Darquy, S. *et al.* Patient/family views on data sharing in rare diseases: study in the European LeukoTreat project. *Eur. J. Hum. Genet.* **24**, 338–343 (2016).
79. Kim, J. *et al.* Patient Perspectives About Decisions to Share Medical Data and Biospecimens for Research. *JAMA Netw. Open* **2**, e199550 (2019).
80. Public perceptions on data sharing: key insights from the UK and the USA - The Lancet Digital Health. [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30161-8/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30161-8/fulltext).
81. The use of digital services. *Sitra* <https://www.sitra.fi/en/publications/use-digital-services/>.
82. D8.3 Brief on findings of ELSI Focus Groups for a RWE approach in AD. <https://roadmap-alzheimer.org/downloads/deliverables/>.
83. Welcome to RADAR-AD | Radar-AD. <https://www.radar-ad.org/welcome-radar-ad>.
84. Haug, C. J. Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate? *N. Engl. J. Med.* **376**, 2203–2205 (2017).
85. Shah, N. *et al.* Motivations for data sharing—views of research participants from four European countries: A DIRECT study. *Eur. J. Hum. Genet.* **27**, 721–729 (2019).
86. Alzheimer Europe - Ethics - Ethical issues in practice - 2019: Overcoming ethical challenges affecting the involvement of people with dementia in research. <https://www.alzheimer-europe.org/Ethics/Ethical-issues-in-practice/2019-Overcoming-ethical-challenges-affecting-the-involvement-of-people-with-dementia-in-research>.
87. Ghafur, S., Fontana, G., Halligan, J., O’Shaughnessy, J. & Darzi, A. *NHS data: Maximising its impact on the health and wealth of the United Kingdom*. <http://spiral.imperial.ac.uk/handle/10044/1/76409> (2020) doi:10.25561/76409.
88. Pierce, H. H., Dev, A., Statham, E. & Bierer, B. E. Credit data generators for data reuse. *Nature* **570**, 30–32 (2019).
89. Milham, M. P. *et al.* Assessment of the impact of shared brain imaging data on the scientific literature. *Nat. Commun.* **9**, 2818 (2018).
90. Alzheimer Europe - Policy - Our opinion on ... - Dementia Research and COVID-19. <https://www.alzheimer-europe.org/Policy/Our-opinion-on/Dementia-Research-and-COVID-19>.
91. Ohmann, C. *et al.* Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* **7**, e018647 (2017).
92. Birkenbihl, C. *et al.* ANMerge: A Comprehensive and Accessible Alzheimer’s Disease Patient-Level Dataset. *J. Alzheimers Dis.* **79**, 423–431 (2021).

10. Acknowledgements

Angela Bradshaw, Alzheimer Europe

Owen Miller, Alzheimer Europe

Jean Georges, Alzheimer Europe

This report was developed by Alzheimer Europe thanks to support from Gates Ventures. Alzheimer Europe would like to gratefully acknowledge contributions from Professor John Gallacher (University of Oxford, UK), Professor Martin Hofmann-Apitius (Fraunhofer SCAI, Germany), Dr. Michaela Th. Mayrhofer (BBMRI-ERIC, Austria), Professor Craig Ritchie (University of Edinburgh, UK), Professor Petra Ritter (Charite University Hospital, Germany) and Professor Pieter-Jelle Visser (Maastricht University, the Netherlands).



Alzheimer Europe • 14, rue Dicks • L-1417 Luxembourg
Tel.: +352 29 79 70 • Fax: +352 29 79 72 • info@alzheimer-europe.org
www.alzheimer-europe.org



www.facebook.com/alzheimer.europe



[@AlzheimerEurope](https://twitter.com/AlzheimerEurope)

ISBN 978-2-919811-00-7