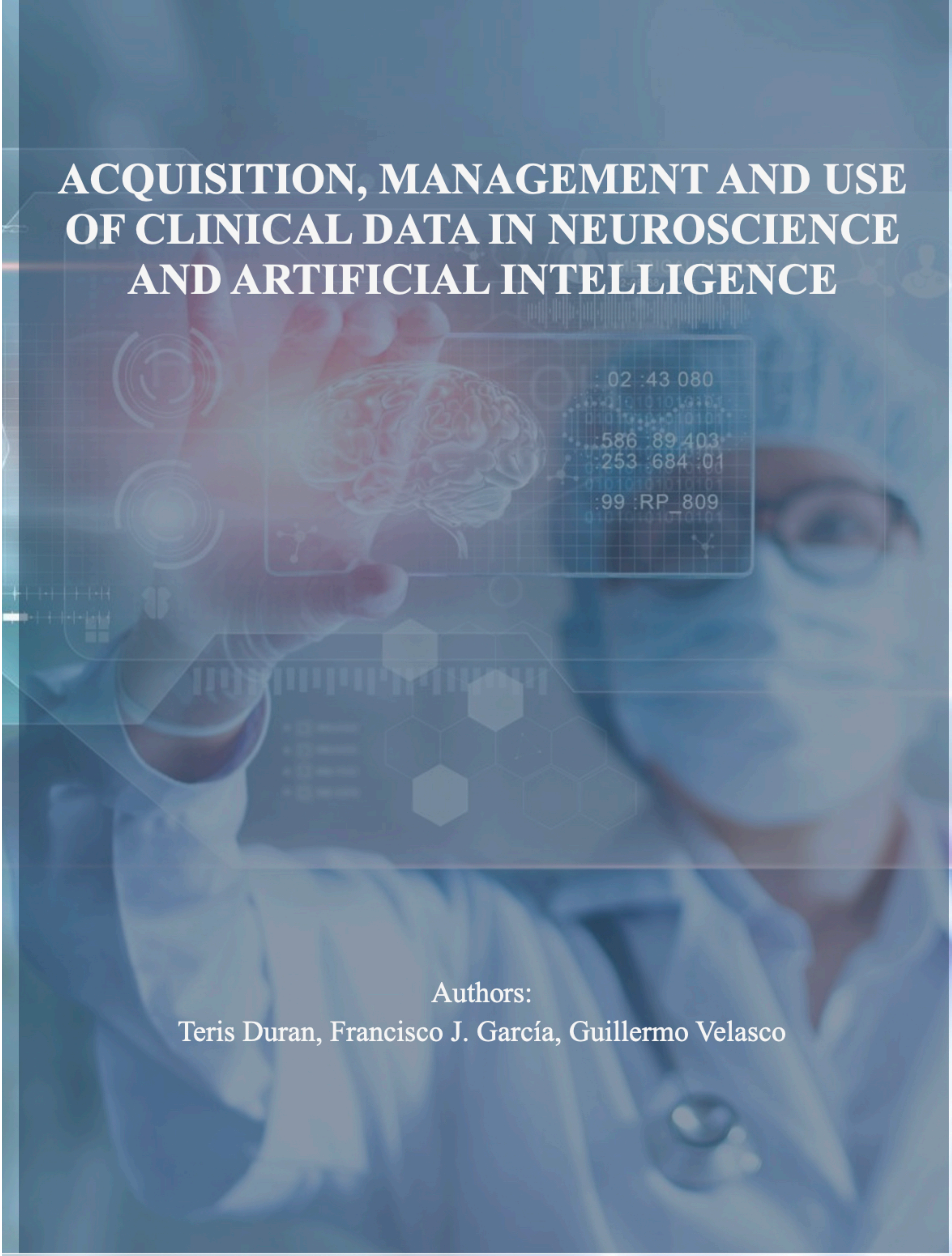


ACQUISITION, MANAGEMENT AND USE OF CLINICAL DATA IN NEUROSCIENCE AND ARTIFICIAL INTELLIGENCE

Authors:
Teris Duran, Francisco J. García, Guillermo Velasco

Human Brain Project
EBRAINS
POLITÉCNICA
UNIVERSIDAD POLITÉCNICA DE MADRID
UE 23
MINISTERIO DE CIENCIA E INNOVACIÓN
Instituto de Salud Carlos III
itemas isciii
Plataforma de dinamización e innovación de las capacidades industriales del Sistema Nacional de Salud

Acquisition, management and use of clinical data in neuroscience and artificial intelligence

September, 2023

Human Brain Project Innovation and Technology Transfer Node
Polytechnic University of Madrid
&
Knowledge Transfer Office
Carlos III Health Institute

Authors:
Teris Duran, Francisco Javier García and Guillermo Velasco

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	Acquisition, management and use of clinical data in neuroscience and artificial intelligence.
Document File Name:	Data management_neurotechnologies_AI.pdf
WP(s)/Task(s):	WP8, T8.5
WP Objective(s):	WPO8.1
Output(s):	OP8.19
Dissemination Level:	PU = Public
Delivery Date:	M37 / September, 2023
Author(s):	Teris DURAN, Francisco Javier GARCIA, Guillermo VELASCO (P68)
Abstract:	The Polytechnic University of Madrid (UPM) and the Carlos III Health Institute (ISCIII) have combined efforts to study, in depth, the needs that health research institutes in Spain have regarding the acquisition, processing and analysis of brain data for the development and use of neurotechnologies, including the service provided in this area by the tools of the European research infrastructure EBRAINS. To this end, a methodology was defined that incorporates questionnaires and in-depth interviews with working groups at national level, adhering to the networks of the Platform for Dynamisation and Innovation of the Industrial Capacities of the National Health System, part of the ISCIII. The results show, among other aspects, the need for human and physical resources for data analysis and storage, the obstacles to finding specialised databases, and the need to bring EBRAINS services closer to researchers. Furthermore, trends are identified in relation to intellectual and industrial property in the field of artificial intelligence in neuroscience are shared.
Keywords:	HBP, Human Brain Project, EBRAINS, Brain Data Management, Healthcare Data, Neurotechnologies, Artificial Intelligence
Target Users/Readers:	Hospitals, biomedical research institutes, neuroscience centres, health technology companies, research and technology organisations, universities

INDEX

ACKNOWLEDGEMENTS

We thank the persons shown here below for their efforts, time, knowledge and experience contributing to the preparation of this report:

Judit Domingo
ITEMAS-ISCIII Platform

Pilar Gayoso
Deputy Director General for Research in Cell Therapy and Regenerative Medicine at ISCIII

Petra Ritter
Berlin Institute of Health - Charité Universitätsmedizin Berlin

Bryan Strange
Polytechnic University of Madrid

Lluís Blanch
ITEMAS-ISCIII Platform

María Inmaculada Aymerich
Galicia South Institute for Health Research (IISGS)

Nerea Alonso
Galicia South Institute for Health Research (IISGS)

Carlos Spuch Calvar
Galicia South Institute for Health Research (IISGS)

Exuperio Diez-Tejedor
Institute for Health Research Hospital Universitario La Paz (IdiPAZ)

María Alonso de Leciñana
Institute for Health Research Hospital Universitario La Paz (IdiPAZ)

Estrella Morenas
12 de Octubre Hospital Research Institute (i+12)

David Ezpeleta
Spanish Society of Neurology (SEN)

Maria Josep Martí
August Pi i Sunyer Biomedical Research Institute (IDIBAPS)

Jaime de la Rocha
August Pi i Sunyer Biomedical Research Institute (IDIBAPS)

Jesus Plays
Research Institute - Vall d'Hebron University Hospital
(IR-HUVH)

Carlos Molina
Research Institute - Vall d'Hebron University Hospital
(IR-HUVH)

ABOUT THE AUTHORS

Teris Duran is a researcher at the Polytechnic University of Madrid in the Human Brain Project and EBRAINS. She holds a Master in Neuromarketing and Consumer Behaviour from the The Complutense University of Madrid, a Master in Economics and Innovation Management from the Autonomous University of Madrid, with a degree in International Trade and a specialisation in International Business from the Universidad Alejandro de Humboldt and the Metropolitan University, Venezuela. Currently studying a degree in Psychology at the National Distance Education University, given her interest in human behaviour, she has diversified work experiences in international trade, finance and international business.

Francisco Javier García is an A1 civil servant of the Institute of Health Carlos III (ISCIII) in the Scale of Specialised Senior Technicians of the Public Research Bodies. He holds a degree in Biochemistry from the University of Seville and Malaga, specialising in Molecular and Applied Biochemistry. With a master's degree in Regenerative Biomedicine and various training courses in intellectual and industrial property, he is currently in charge of the Knowledge Transfer Office and is involved in several European projects in which the ISCIII participates, such as the Human Brain project and the European Personalised Medicine partnership (EP Permed).

Guillermo Velasco is leader of the Innovation and Technology Transfer Unit of the Human Brain Project, at the Polytechnic University of Madrid and coordinator of the Spanish EBRAINS Node. He is an honorary research fellow at the Manchester Innovation Research Institute, part of the Alliance Manchester Business School. With a PhD in Innovation and Technology Foresight from the University of Manchester, and an MPhil in Economics and Innovation Management from the UPM, Guillermo is currently an Engineer in Industrial Organisation and Economics and an EC expert for the evaluation of research and technological innovation projects and proposals. For this, he has participated in numerous foresight, future scenarios and innovation projects for the EC, UNIDO and other international organisations.

Introduction

Health research is a fundamental pillar for the development of society and the improvement of citizens' quality of life. The integration of research with clinical practice guarantees a higher quality of health services and allows the implementation of scientific advances for the prevention, diagnosis, and treatment of diseases to be faster and more effective (Carrasco, 2007).

Health research in Spain is mainly carried out through health research institutes (HRIs) formed by the association of the National Health System (NHS) hospitals with universities, public research organisations and other public or private research centres in the area.

The Institute of Health Carlos III (ISCIII the acronym in Spanish) has developed a programme for the evaluation, accreditation and monitoring of HRIs aimed at the consolidation of research centres, the core of which are NHS health centres, as centres for the generation and transfer of knowledge to promote innovation, responding to health priorities at both the national and European level. Currently 35 HRIs are accredited by the ISCIII, in 13 Autonomous Communities, involving more than 24,000 researchers.

Within the research carried out in the HRIs, and more specifically in the area of Neurosciences, the use of data obtained in hospitals and health centres is a key factor. The procedure by which these data are acquired, managed and used is full of obstacles and difficulties that all doctors and researchers must try to overcome on a daily basis.

This study has been funded by the FET Flagship initiative 'Human Brain Project' (HBP) of the European Commission, and prepared by the Polytechnic University of Madrid (UPM) and the ISCIII with the collaboration of the platform named Dynamisation and innovation of the industrial capacities of the NHS and their effective transfer to the productive sector (commonly known in the healthcare environment as ITEMAS. It addresses the most relevant aspects of data acquisition, curation and use in the NHS, exploring the main tools used by researchers in different Spanish hospitals for this purpose and the obstacles they face on a daily basis, defining current trends in the sector, and highlighting the importance of standardising the way in which these processes are carried out.

Part I. Context

1.1 Objectives of the study

The main objective of this study is to understand the current situation of the acquisition, management and use of data for neurotechnology in the NHS, with the aim of being able to find possible solutions through the European platform EBRAINS.

The specific objectives of this study include:

- Understand the obstacles faced by NHS researchers regarding data acquisition, management, and use.
- Know the tools commonly used in the NHS for the acquisition, management and use of data.
- Define trends in data acquisition, management and use.
- Disseminate the results of the HBP project and the services that EBRAINS can offer to the NHS.

1.2 Human Brain Project (HBP)

The Human Brain Project (HBP) was a long-term, large-scale initiative pioneering digital brain research. It was selected by the European Commission as one of its flagship projects in future emerging technologies "FET Flagship".

HBP aimed to understand in depth the extremely complex structure of the human brain, and its functions, with a unique interdisciplinary approach at the interface of neuroscience and digital technology (Human Brain Project, 2020).

The HBP project ran for ten years (2013-2023), engaging more than 80 European and international research institutions.

The Spanish representation was comprised of 25 research laboratories belonging to 13 scientific organisations, which represented 9.8% of the "Flagship". Within this, noteworthy are the inputs of the Polytechnic University of Madrid (UPM), with a total of six research laboratories, which represented 24% of the national representation (UPM, 2013).

1. 3 EBRAINS Infrastructure

The completion of HBP in 2023 has given way to EBRAINS, a new digital research infrastructure, created by the HBP and funded by the EU, which brings together a wide range of data and tools for brain-related research. EBRAINS will capitalise on the work done by HBP teams in digital neuroscience, brain medicine and brain-inspired technology (EBRAINS, 2023).

EBRAINS' mission is to revolutionise the way neuroscience is conducted, providing a digital ecosystem to enable breakthroughs in brain research that translate into innovations in neuroscience, healthcare and technology.

EBRAINS’ three priority areas include:

- Advancing scientific understanding of the brain
- Improving diagnosis and treatment of brain diseases
- Translating brain knowledge into technological breakthroughs.

1.4 ISCIII-UPM Agreement

On 17 February 2021, the UPM and the ISCIII signed an agreement to support the innovative development of the HBP, the aim of which was to channel the technological results obtained by public HBP researchers towards the NHS and the productive sector.

This agreement establishes that the ISCIII would participate in the SGA3 phase of the project for the performance of certain activities related to the project once it was admitted as a third party, "third party", of the UPM in HBP, and in accordance with the Horizon 2020 (H2020) expenditure eligibility concepts for "third parties". BOE-A-2021-3132.

Once the ISCIII was admitted as a "third party", an addendum to the agreement was drawn up and signed on 20 February 2023, detailing the activities to be carried out by the ISCIII and the schedule of deliverables (see BOE-A-2023-5356).

This report describes the development of one of the research activities proposed in the addendum and includes all the deliverables set out in the addendum.

1.5 ITEMAS-ISCIII, nodes and affiliated centres ITEMAS

The ISCIII is an autonomous body created by Law 14/1986 of 25 April 1986 on General Health, which currently has a dual functional and organic dependence on the Ministry of Science and Innovation and a functional dependence on the Ministry of Health.

Article 47 of Law 14/2011, of 1 June, on Science, Technology and Innovation, grants the ISCIII the status of Public Research Organisations (PROs), a status also held by the Spanish National Research Council (CSIC), the National Institute for Aerospace Technology (INTA), the Centre for Energy, Environmental and Technological Research (CIEMAT), and the Institute of Astrophysics of the Canary Islands (IAC).

The PROs are defined in the Law 14/2011 as organisations created to directly carry out scientific and technical research activities, technological service provision activities, and those other complementary activities necessary for the appropriate scientific and technological progress of society, which are attributed to them by this law or by their rules of creation and operation. In addition, this law assigns the ISCIII the task of financing scientific and technical research, something that sets it apart from the other PROs.

Among the initiatives carried out by the ISCIII are the ISCIII Platforms of support for R&D&I in Biomedicine and Health Sciences, which are a set of infrastructures and professionals belonging to different institutions with the capacity to provide transversal support services to the R&D&I system for the generation and transfer of high-quality knowledge in the NHS.

Currently, there are 3 ISCIII platforms to support R&D&I in Biomedicine and Health Sciences: the Biobanks and Biomodels platform, the Clinical Research Support platform and the ITEMAS platform.

ITEMAS, which has played a key role in the development of this study, is currently made up of 18 health entities and HRIs funded by the ISCIII (nodes). As part of the work and organisation of ITEMAS, each node must provide innovation services to other entities that have not been funded by the ISCIII (member centres). These adhering centres are health entities and Research Institutes adhered to one of the nodes in order to receive innovation services and to be able to access the submission of projects to the Platform (ITEMAS 1, 2022).

ITEMAS’s main objective is to boost, effectively, the industrial capacity generated by the NHS and the real transfer of the results of research carried out within the NHS to the productive fabric (ITEMAS 2, 2022).

1.6 HBP Innovation and Technology Transfer Team and EBRAINS

HBP and EBRAINS innovation and technology transfer activities are designed and developed by a multidisciplinary team at the Polytechnic University of Madrid. UPMs’ team is specialised in providing support to scientific and research groups for the exploitation of their research results.

Since practically from the start of HBP, the team's activities have focused on developing market analysis and technology roadmaps, evaluating exploitation plans for a wide variety of technologies, mentoring on intellectual property and industrial property issues, strengthening relationships with industry, training on entrepreneurship issues, and connecting HBP start-ups and spin-offs with potential investors.

HBP and EBRAINS' commitment to technological innovation is firmly aligned with the goals of scientific and industrial collaboration promoted by the CE, and it is in the spirit of the UPM’s team to foster close cooperation with the neuroscience community, industrial technology leaders, the medical and clinical sector, digital developers, and neural computational researchers, among others. The overarching goal is to continue to advance scientific knowledge transfer and innovation. To learn more about our activities, we invite you to visit the following link:

https://www.humanbrainproject.eu/en/collaborate-hbp/innovation-industry/

1.7 Importance of data for neuroscience research

One of the main challenges facing NHS researchers is to obtain homogeneous data that can be used by different research groups in different hospitals regardless of their location. This would speed up research projects considerably and reduce the time it takes for new drugs and therapies to reach the population.

However, to achieve this end, proper data curation and processing is essential. Data curation is a complex process where NHS researchers face different obstacles on a routine basis. Some of these obstacles are the demanding regulations, the lack of homogeneity of the data collected, the lack of both resources and qualified personnel for data processing, plus bureaucratic procedures.

Data protection regulations (Organic Law 3/2018, of 5 December, on Personal Data Protection and guarantee of digital rights and Regulation (EU) 2016/679 of the European Parliament and of the Council, of 27 April 2016) or the growing and necessary awareness of neuro-rights also increase the requirements of the different research ethics committees, regulated by Law 14/2007, of 3 July, on Biomedical Research.

The current research ethics committees in the centres that carry out biomedical research are responsible, among other functions, for assessing the feasibility of the projects and the qualifications of the research teams, and for weighing up the methodological, ethical and legal aspects. Obtaining a favourable report from this committee is necessary for the implementation of the different research projects involving human subjects.

On the other hand, the particular situation in Spain, with the transfer of health care competencies to the Autonomous Communities and the particularities of each hospital, this leads to a lack of homogeneity that hinders the use of research data from other researchers.

Although current initiatives aim at overcoming the above-mentioned obstacles, such as federated medical data platforms, there is still a long way to go.

1.8 Methodology

1.8.1 Creation of the Working Group

The development of this research has been made possible thanks to the access to hospitals and associated centres that are in the platforms and networks promoted by the ISCIII itself. The creation of a working group for achieving the objectives marked the beginning of all activities.

The working group was made up of the UPM as a member of the HBP project, the ISCIII as a "third party" of the UPM and in charge of developing the tasks described above, and the ITEMAS platform, an ISCIII initiative, which through its network would allow access to neuroscience research groups from different hospitals and health centres distributed throughout Spain.

The first steps of the UPM-ISCIII-ITEMAS working group consisted of developing a working methodology, defining the tasks to be carried out by the group over the coming months and drawing up a timetable that included the deadlines for completing the tasks.

1.8.2 Work methodology diagram

The methodological sequence - figure Nº1 - defined by the working group to meet the objectives set is as follows:

- Hold a seminar for ITEMAS nodes and member centres to present the HBP project and the EBRAINS’ infrastructure. Through this task it was possible to set up initial contacts with doctors and researchers from the

NHS who are developing research projects in the field of neurosciences; present the tasks to be carried out by the working group; and, to disseminate HBP and EBRAINS.

- Design, dissemination and analysis of the questionnaire on the acquisition, management and use of data in the field of neurotechnologies in Spain. This task provided a general map of the situation in the National Health System and made it possible to select potential profiles for planning interviews in order to go more deeply into the different issues raised.
- Make plans for the interviews with the selected profiles. The interviews allowed us to add a richer content to the work and to identify more accurately the obstacles, challenges and trends encountered in this field.
- Celebrating a final HBP event. This event, to be held in September 2023, will disseminate the results obtained in this study and will allow further dissemination of the HBP and EBRAINS brand in both the public and private sectors.

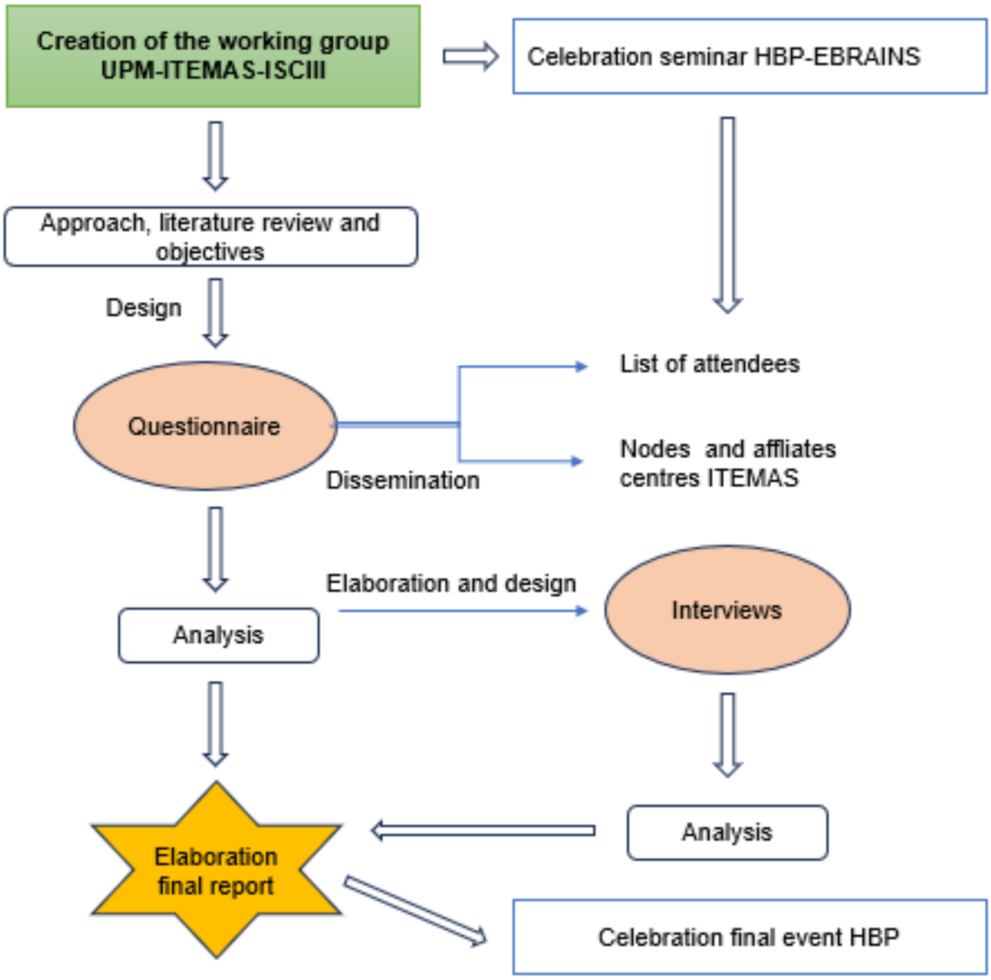


Figure 1. Diagram of the working methodology carried out by the UPM-ITEMAS-ISCIII working group.

1.8.3 HBP-EBRAINS Seminar

The seminar entitled *"HBP-EBRAINS: new opportunities for research in Neurosciences" took place* on 25 January 2023 and was disseminated through the nodes and centres linked up to the ITEMAS platform. The modality of the event was on-line, counting on ISCIII's resources for this type of virtual engagement. Both the registration and the seminar were carried out through the Zoom platform and, once ended, the event was published on the official ISCIII YouTube page.

The total number of people registered for the seminar was 150, including doctors, researchers, and R&D&I personnel in the neurosciences from all over Spain. The seminar lasted 62 minutes and in the end 92 people connected to the session. For all those who, for various reasons, were unable to connect, are here provided with the You Tube link:

(https://www.youtube.com/watch?v=yz2PHuRU4SE&t=28s)

Significantly, the seminar made it possible to disseminate the importance of the tools developed over the last few years by HBP - already included as services for the neuroscientific community in the EBRAINS infrastructure - to the different nodes and member centres, as well as to inform attendees about the tasks to be carried out by the UPM-ISCIII-ITEMAS working group.

1.8.4 Questionnaire

Following the seminar, the working group sent out a questionnaire (See Annex I) containing a series of questions related to the acquisition, management and use of neuroscience data in Spain. The questionnaire consisted of 19 questions that sought to obtain information on the profile of the respondents, the types of data they handled, the tools they used to acquire and curate data, the needs and obstacles they encountered, and the type of collaboration they carried out in their projects.

For the dissemination of the questionnaire, the ITEMAS network of nodes and member centres and the list of seminar attendees were again used. Several reminders were made during the months of January to June, resulting in a total of 54 surveys received and satisfactorily completed. The main regions contributing surveys were Madrid, Catalonia, and Andalusia, which together accounted for approximately 70% of the responses. A possible territorial bias is therefore assumed in the analysis.

Regarding the response rate to the questionnaire, it is difficult to stipulate an exact percentage. In the case of the ITEMAS nodes and centres, the questionnaire was sent to the innovation units of 18 nodes and 100 member centres. However, we do not have details on the number of neuro research groups in these centres and whether all of these units transferred the questionnaires.

As for the total list of seminar attendees, only those profiles that could have or had had experience in research projects in the area of neuroscience were selected. Of the 72 people selected, a total of 22 (30.56% of the total) replied.

Given that the response rate was not particularly high, the data extracted should be taken with sufficient caution to avoid extrapolating the results of our analysis to the entire National Health System.

The analysis of these results was carried out by means of Excel graphs and tables from which the different conclusions were obtained.

1.8.5 Interviews

Once the analysis of the results had been carried out, a selection for interviews was made based on the profiles that had completed the questionnaires and who showed greater interest due to their experience in the handling and use of data. The UPM-ISCIII-ITEMAS working group designed the content of the interviews with a total of 15 questions (see Annex II), which aimed at diving more deeply into the questions and to addressing more specific issues The format of the interviews was mostly face-to-face, although three of them were conducted in online format due to the impossibility of keeping a face-to-face appointment. The face-to-face interviews took place in Madrid, Barcelona and Vigo during the months of June and July 2023.

A total of 7 interviews with an average duration of 90 minutes were conducted with 5 women and 6 men, including neurologists, researchers, and innovation managers.

Part II. Questionnaire results and analysis

2.1.1 Respondent data

The questionnaire on the acquisition, management, and use of neuroscience data in Spain was completed by NHS staff as shown in Figure 2. The majority of respondents appear as Principal Investigators (PIs) in neuroscience-related research projects (74.07% of respondents) and are therefore familiar with the processes and tools used in this area, as well as the main needs and obstacles faced by researchers in their day-to-day work.

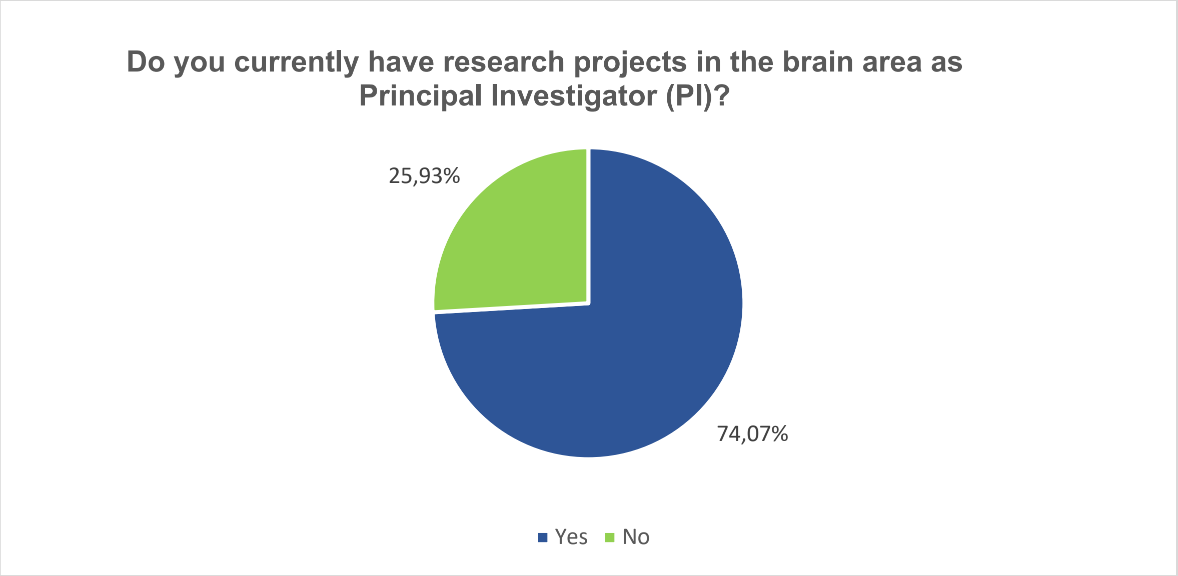


Figure 2. Pie chart indicating whether respondents are involved as Principal Investigator in a neuroscience-related research project.

As can be seen in Figure 3, the profile of these respondents is predominantly senior, which ascribes more experience to the questionnaire. The average age of those consulted was 49.7 years, with the 36-45 years window having the highest number of respondents. In addition, the survey provides for a broad parity in relation to the gender of the respondents.

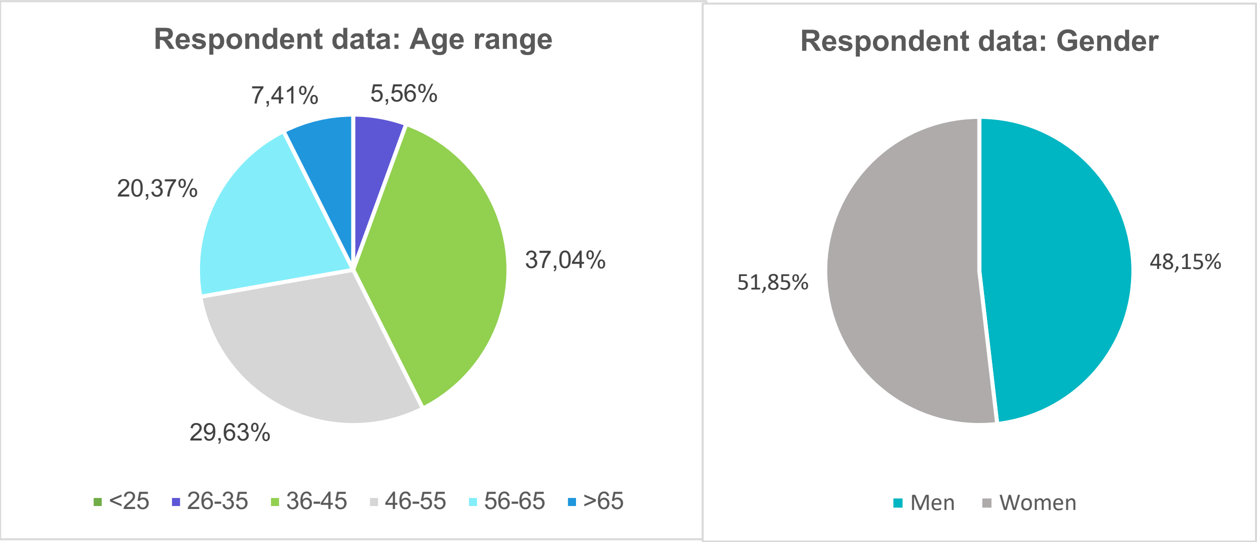


Figure 3. Age and gender of respondents

As for the job positions of the respondents, Figure 4 shows that the majority of respondents are doctors (50%) who combine their healthcare tasks with research work and researchers attached to the HRIs who are exclusively dedicated to research projects (38.89%). Similarly, university professors and lecturers with experience in the field of neurotechnologies also participated, although to a lesser extent (11.11%).

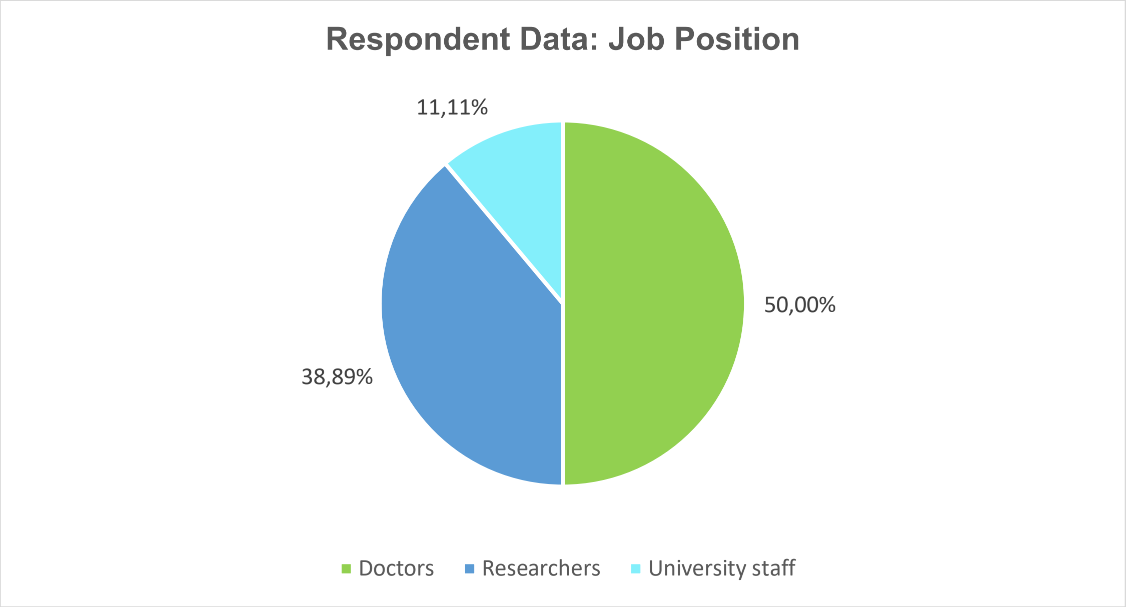


Figure 4. Respondents' employment position.

2.1.2 Opinion on data types

To know the type of data that respondents handle and their opinion about it, three different types of data were defined, as shown in Figure 5.

- **Open data:** according to Law 37/2007, of 16 November, on the reuse of public sector information, this is data that anyone is free to use, reuse and redistribute, with the only limit, where applicable, being the requirement to attribute its source or acknowledge its authorship.

- **Sensitive data:** according to the Commission Implementing Decision (EU) 2021/914 of 4 June 2021 on standard contractual clauses for the transfer of personal data to third countries in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council, these involve personal data revealing racial or ethnic origin, religious beliefs, genetic data or biometric data intended to uniquely identify a natural person, data concerning health or data concerning the life or sexual orientation of a natural person.

- **Synthetic data**: This is a new area of data processing in which data are produced in such a way that they realistically resemble real data (both personal and non-personal), but do not actually refer to any specific identified or identifiable individual, nor to the actual measurement of an observable parameter in the case of non-personal data (ENISA, 2022).

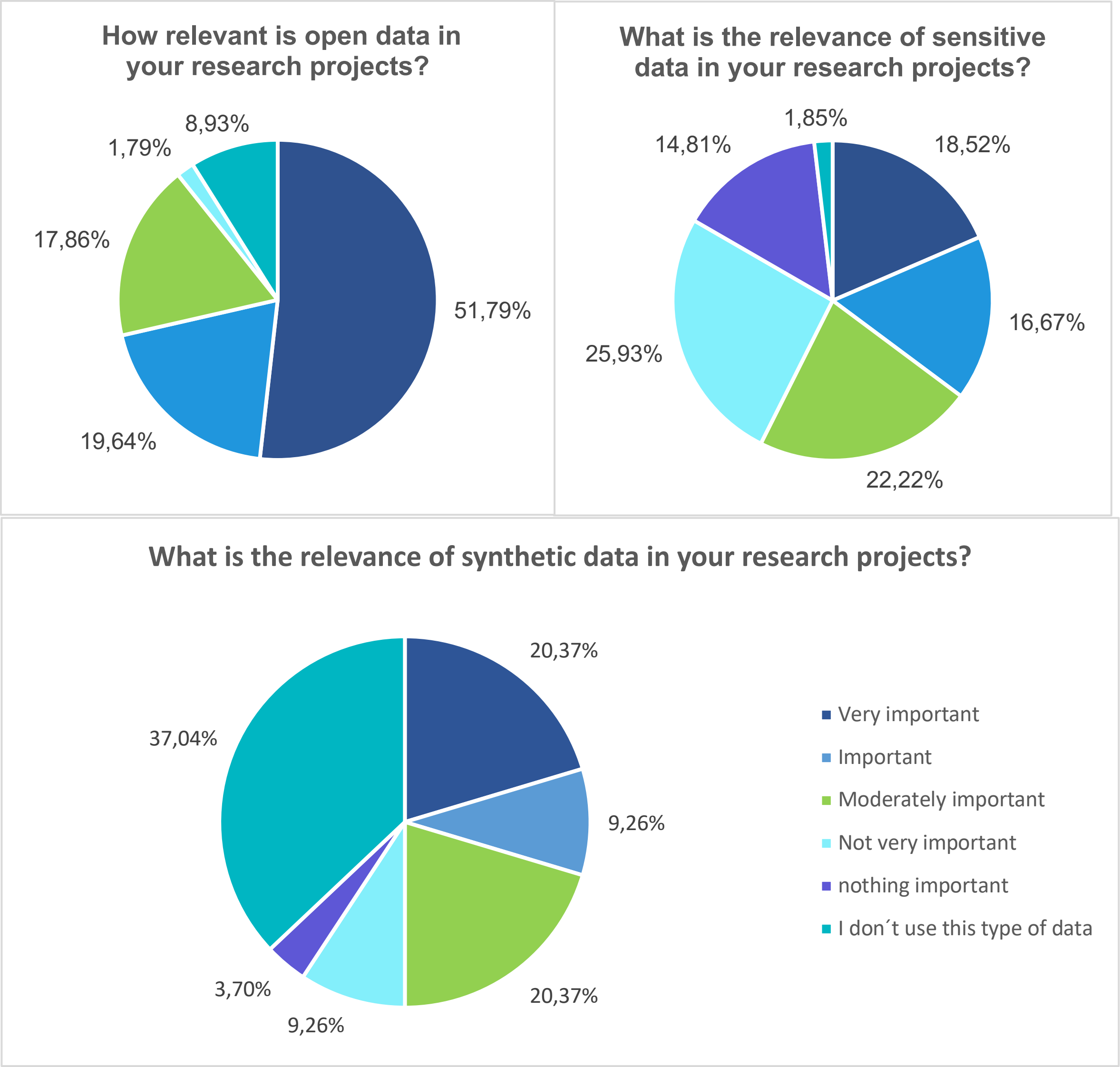


Figure 5. Pie charts representing the experience and opinion of respondents in relation to the types of data defined in the questionnaire: open data, sensitive data and synthetic data. The legend is the same for all charts.

In a general analysis of the different types of data, open data is most important for respondents. As regards sensitive data, the diversity of opinions may be a consequence of the way in which each respondent has interpreted the question. This is because many respondents do not need to use sensitive data to carry out their research tasks since they generally receive the data already anonymised or without needing data that can identify the donor. Finally, with regard to synthetic data, the most remarkable thing is that still a large part of the respondents have not worked with this type of data, and it remains the most unknown type to them.

2.1.3 Data acquisition and curation tools

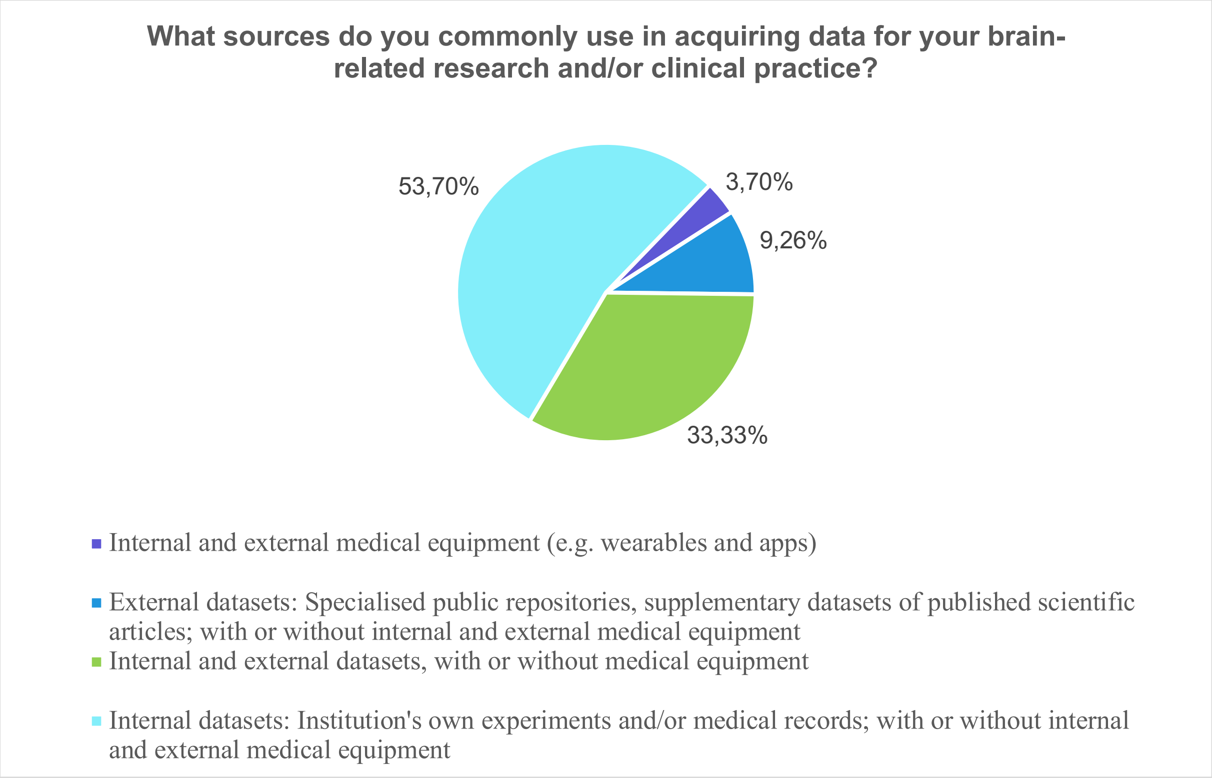


Figure 6. What sources do you typically use in acquiring data for your brain-related research projects and/or clinical practice?

More than half of the sample (53.70%) use *internal databases* as a regular source for data acquisition, including those participants who in parallel obtain data through external medical equipment (e.g. wearables and apps) and/or internal medical equipment. This, in contrast to those who regularly use external databases (specialised public repositories), supplementary datasets of published scientific articles, with or without medical equipment, which represent only 9.26%.

On the other hand, **33.33% of the participants regularly use a combination of resources**: internal and external databases, with or without medical equipment. This suggests a preference of the groups to work with their own resources and to a lesser extent they would be motivated to look for alternatives in external data acquisition. Only 3.70% of the sample suggested only having internal or external medical equipment to acquire data, excluding the regular use of databases. The responses from this last segment are not large enough to know whether they have the resources to create their own databases with the devices they use.

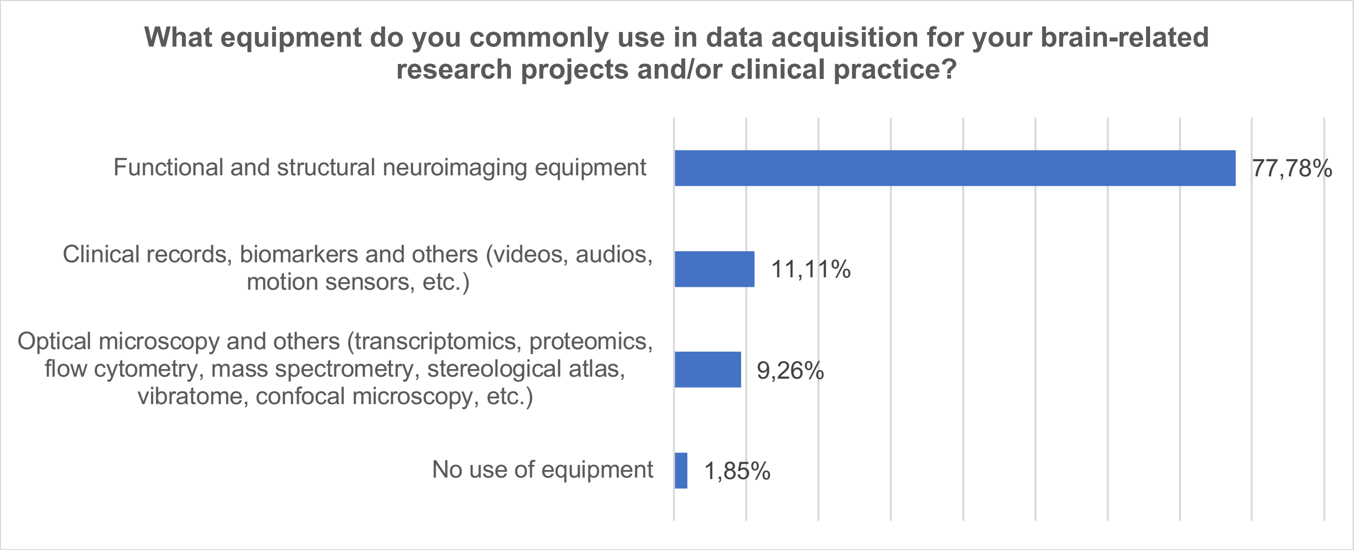


Figure 7. What equipment do you typically use in data acquisition for your brain-related research projects and/or clinical practice?

The vast majority use functional and structural neuroimaging equipment (77.78%), such as CT, PET, MRI, functional MRI, SPECT, EEG, MEG, ECoG, EMG (Figure 7). Some of them also use video, audio, motion sensors or other systems, such as patient-reported outcomes (PROM/ERCP), clinical scales (NIHSS, mRS, TOAST, BARTHEL), cellular analysis, behavioural analysis software, etc.

This is followed by those who, instead of using functional and structural neuroimaging equipment, obtain data from clinical records, biomarkers and others (e.g. videos, audios, movement sensors), representing 11.11% of the sample. In a smaller proportion are those using optical microscopy and others such as transcriptomics, proteomics, flow cytometry, mass spectrometry, stereological atlases, vibratome, confocal microscopy (9.26%) and less than 2% do not use equipment.

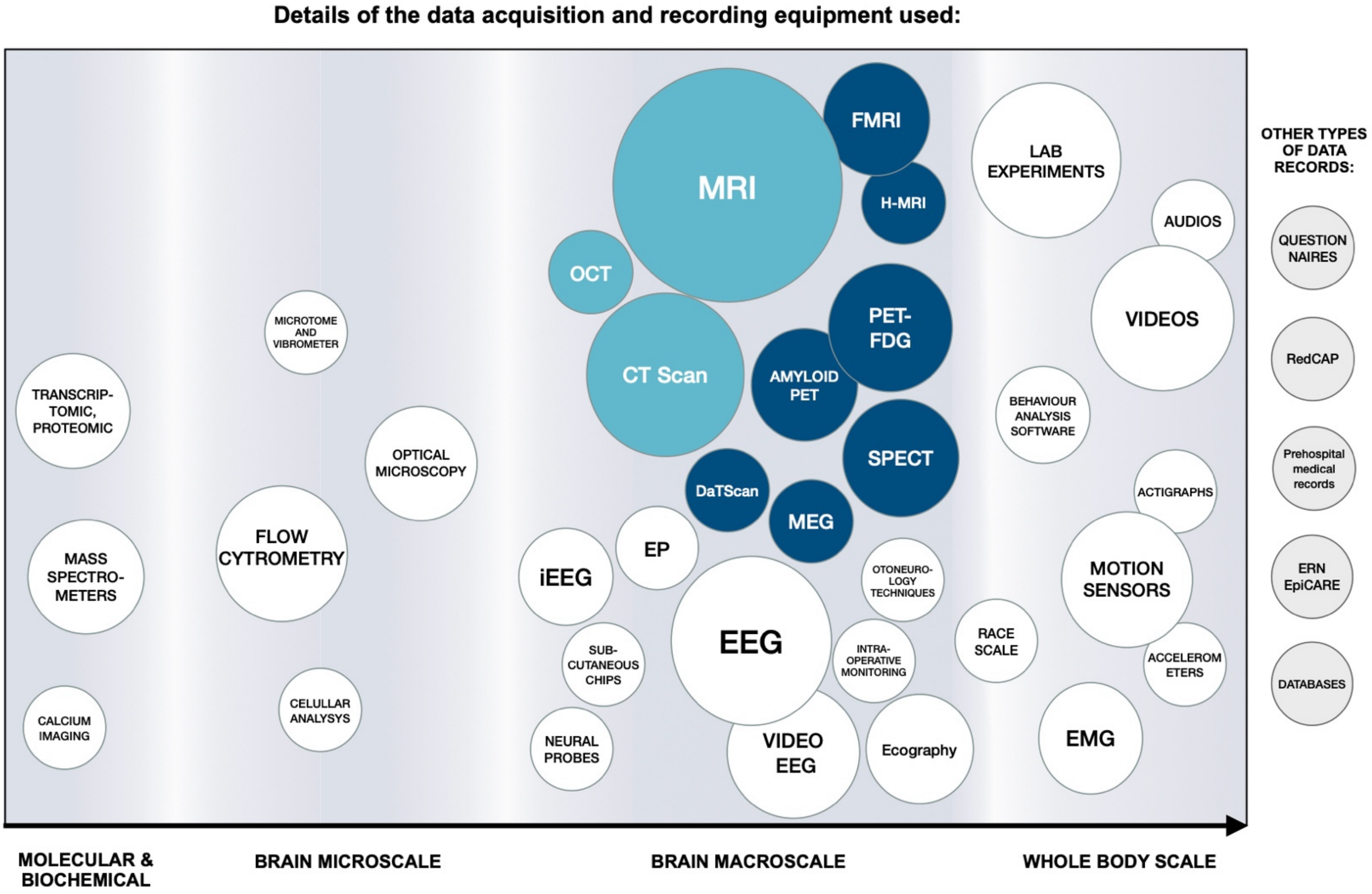


Figure 8. Diagram of open responses, detail of data acquisition and recording equipment. Authors' elaboration.

Figure 8 shows a snapshot of the techniques used by the participants. In this diagram, the tools are distributed according to the depth of analysis they offer, from the molecular and biochemical level, brain microscale, brain macroscale, to a broader scale. Note that on the right-hand side, applications or software for data acquisition and clinical decision support are shown.

Although more participants in neuroscience and its related specialties will be needed to get a more complete picture, some insights can be drawn: A good part concentrates on the use of brain macro-scale analysis tools, particularly functional and structural neuroimaging techniques, as seen in the diagram. To a lesser extent, microscopy techniques, cellular analysis, transcriptomics, proteomics are used, also because a large part of the sample is dedicated to healthcare practice.

In turn, each circle involves the inclusion of other methods of data collection, e.g. otoneurology for hearing disorders, or the use of computer software behind each of the techniques provided. At the level of body movement recordings, a common factor is the use of videos, audios, movement sensors, accelerometers, actigraphs, which are devices that generate a large amount of information.

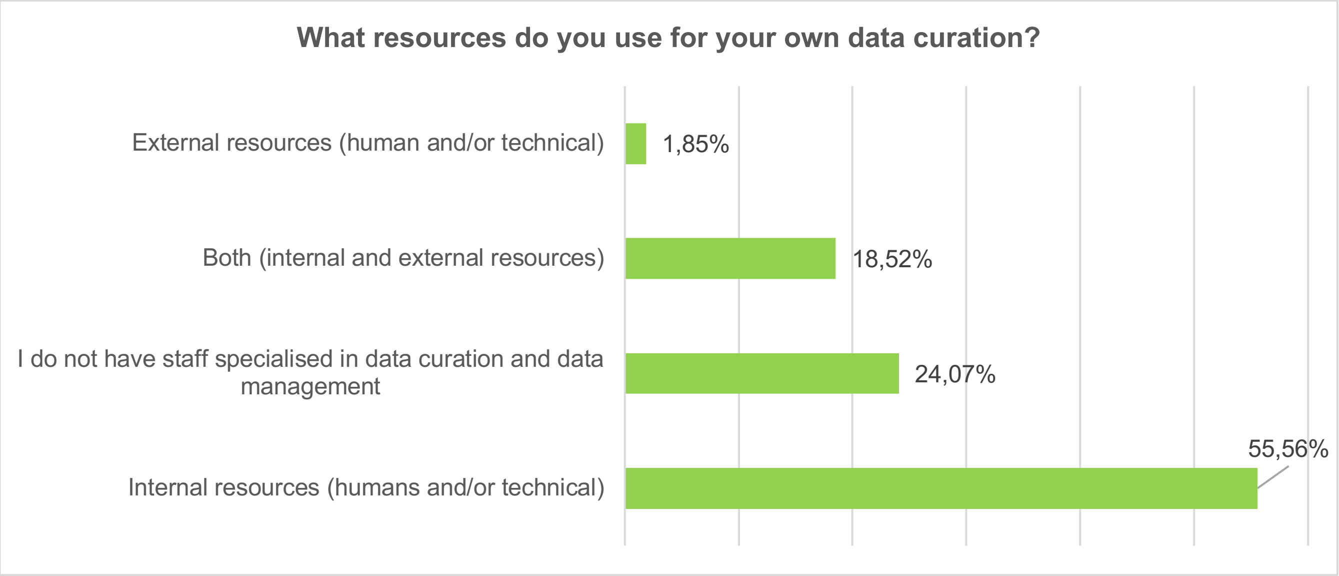


Figure 9. What resources do you use for your own data curation?

As can be seen in Figure 9, **more than half of the participants (55.56%) use their own resources for data curation and management**, information that was also corroborated in the interviews. This is followed by those who do not have specialised staff (24.07%), and a smaller proportion (18.52%) who use both modalities - internal and external resources - and occasionally use external channels. Using only internal resources and not having specialised staff are not mutually exclusive conditions. However, it could be inferred that the participants - especially those dedicated to healthcare practice - have much more limited time for data curation and management, in addition to relying on other members of the research group, who may (or may not) have specialised training in this area.

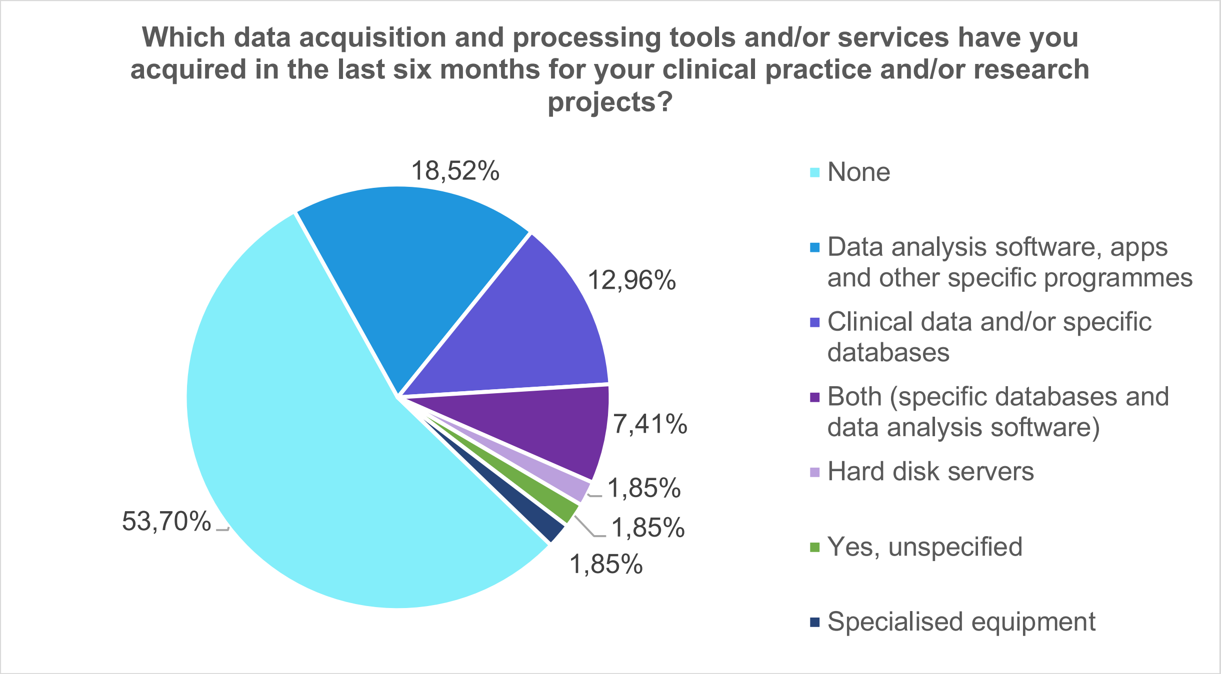


Figure 10. Which data acquisition and processing tools and/or services have you purchased in the last six months for your clinical practice and/or research projects?

In the last six months more than half of the participants have not obtained new data acquisition and processing tools and/or services (53.70%), as we can see in figure Nº10. In other words, **part of the sample has not had the need to renew or acquire new data analysis software, and keep to current programmes**. Some reasons might be associated with this, such as the bureaucracy that slows down the acquisition of new tools, the nature of the projects that possibly do not require them, or the scarcity of alternatives that would really make a difference to what currently exists, among other reasons. The needs would lean towards training in the use of these programmes and the analysis of the data obtained, as shown below.

On the other hand, 18.52% suggest having acquired different data analysis and management software: Excel, SPSS, Matlab, Freesurfer, Redcap (although it can perform some analysis, this is more management-oriented), R, Python, and other software that apply machine learning algorithms or cluster analysis. These participants also describe the tools with a higher level of specialisation, such as Flowjo, for the analysis of acceleration signals, video and microplate readers. Knowing that most participants have their own registers and databases, **the handling of multiple computer programmes - from the most comprehensive to the most specialised - becomes a basic requirement in their research projects**. The next closest group (12.96%) refers to those participants who have their own clinical data or who have obtained specific databases, and a smaller group that has acquired both analysis software and databases in the last six months. In smaller proportions, as well, are those participants who have acquired specific technologies such as functional MRI or hard disk servers.

2.1.4 Needs and challenges for data storage, acquisition and protection

Regarding support services in medical routine and/or research projects, **more than half of the participants consider the *analysis and processing of clinical data*, *data sharing platforms with other researchers*** (62.96%) **and *AI applied to medical data*** (50%) **to be very useful, as shown** in figure Nº11. Regarding *neurorobotic platforms* such as **virtual simulators with robotic models** for experimentation in neuroscience*,* the responses are heterogeneous, and there may be doubts about how these could be transferred to their studies.

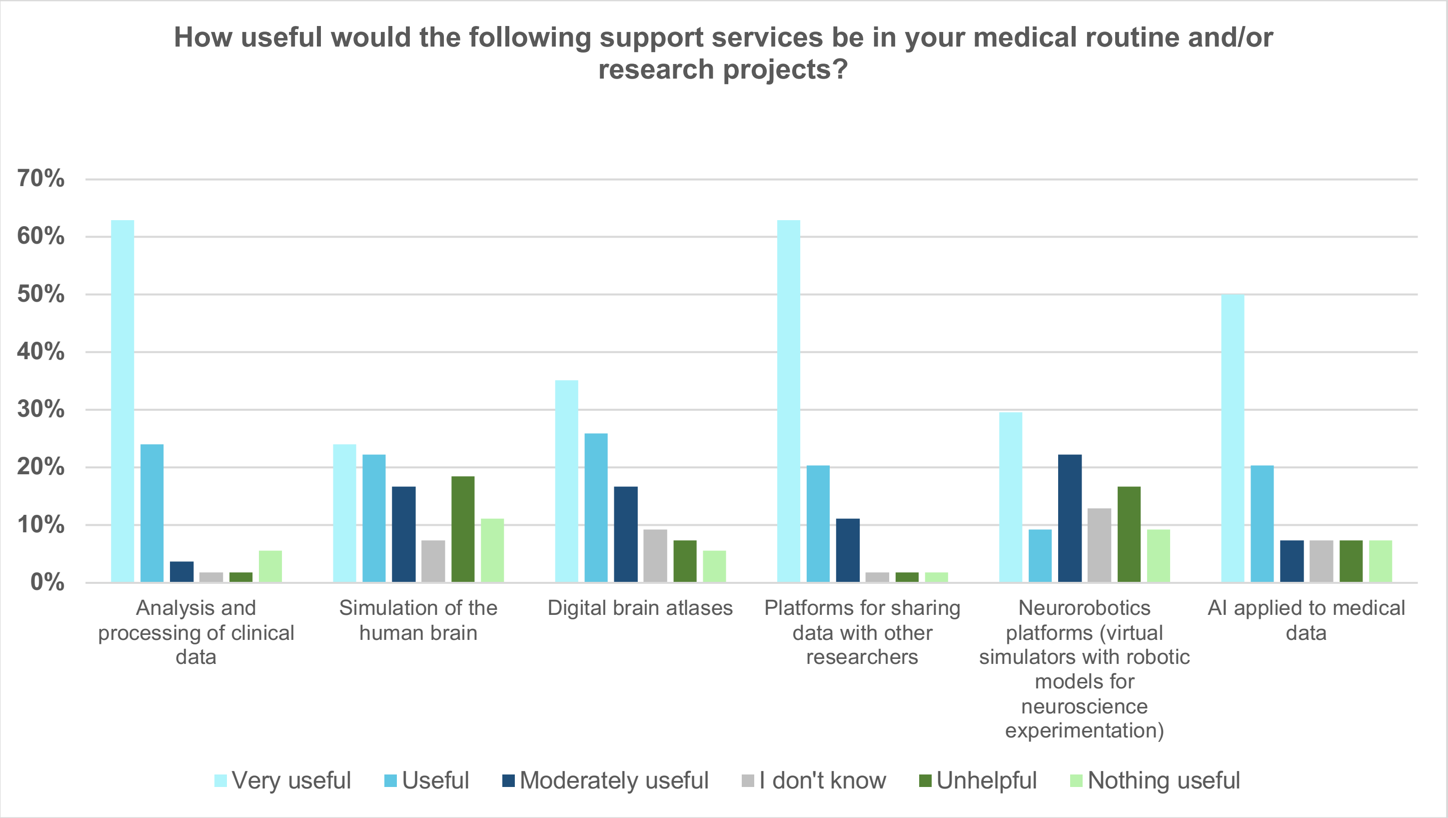


Figure 11. How useful would the following support services be in your medical routine and/or research projects?

Another aspect of analysis is the **evaluation of *digital brain atlases versus human brain simulation* tools***.* The former are considered *very useful* or *useful* (61.12%) compared to the latter, which have a joint rating of 46.29% in the sample. A part of the sample considers human brain simulation tools as not very useful (18.52%) or not at all useful (11.11%) in their studies, in contrast to digital brain atlases which show a rating of not very useful (7.41%) or not at all useful (5.56%). Meanwhile, 16.67% consider both services *moderately useful.* Factors related to this assessment could be, for example, the lack of technical knowledge of the possibilities offered by some tools developed in collaborative projects in neuroscience, whose accuracy and reliability make it readily possible to transfer innovations to medical diagnosis in real contexts, or the difficulty that could be associated with their use.

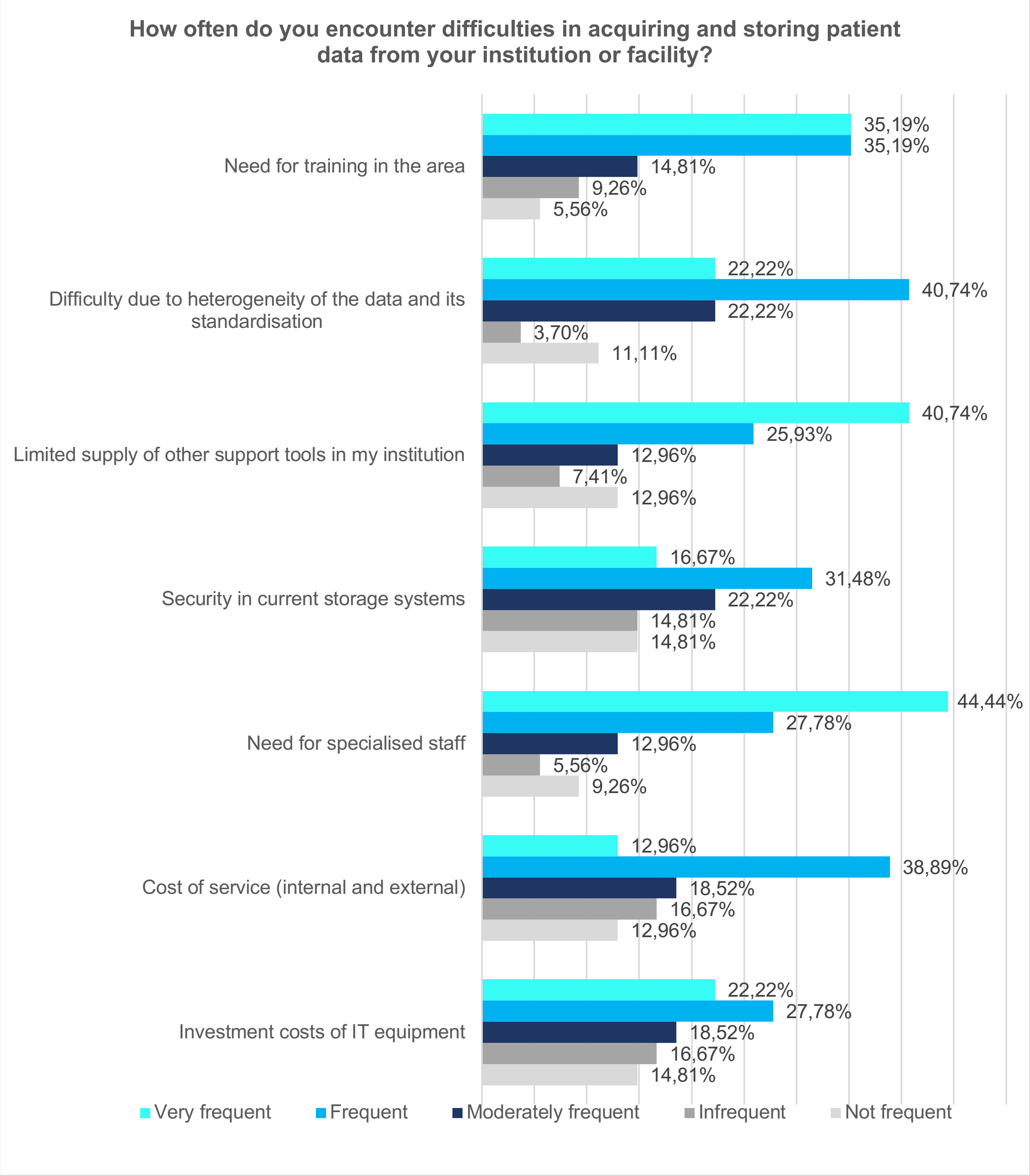


Figure 12. How often do you encounter difficulties in acquiring and storing patient data from your institution or facility?

With regards the frequency of difficulties encountered by the participants when acquiring and storing patient data from their institution, the most frequent difficulties were *the* ***need for specialised staff*** (44.44%), ***the limited supply of other support tools in their institution*** (40.74%) with the ***need for training in the area*** (35.19%), selected as *very frequent*. Among the most *frequent* were *the difficulty due to the heterogeneity of the data and their standardisation* (40.74%), *the cost of the service both internally and externally* (38.89%) and again the need for training (35.19%), as shown in Figure 12. The need for training in the area favours the search for services to accompany staff in the acquisition and storage phases, especially in those groups where specialised staff dedicated exclusively to these processes is scarce.

On the other hand, in terms of *infrequent* and *not frequent* difficulties, the *security of current storage systems* and *the investment cost of IT equipment* were mentioned. With regards the former, although in specific cases security threats were described, the most likely factor would be related to the restrictions on using external programmes or those not supported by the institution, making data acquisition and storage difficult, especially for those users who work mainly with sensitive data. In general, it is observed that security in current storage systems is not a very frequent obstacle.

Regarding the investment cost of IT equipment, assessments are mixed. This could be related to differences in equipment procurement processes between institutions, internal planning, provisioning and frequency with which equipment is purchased, i.e. whether they would have a lower level of periodicity in being renewed. In contrast, the ***costs of services for data acquisition and storage (internal and external)*** are rated as frequent (38%), as these are specialised services that involve qualified technical human resources to support these procedures.

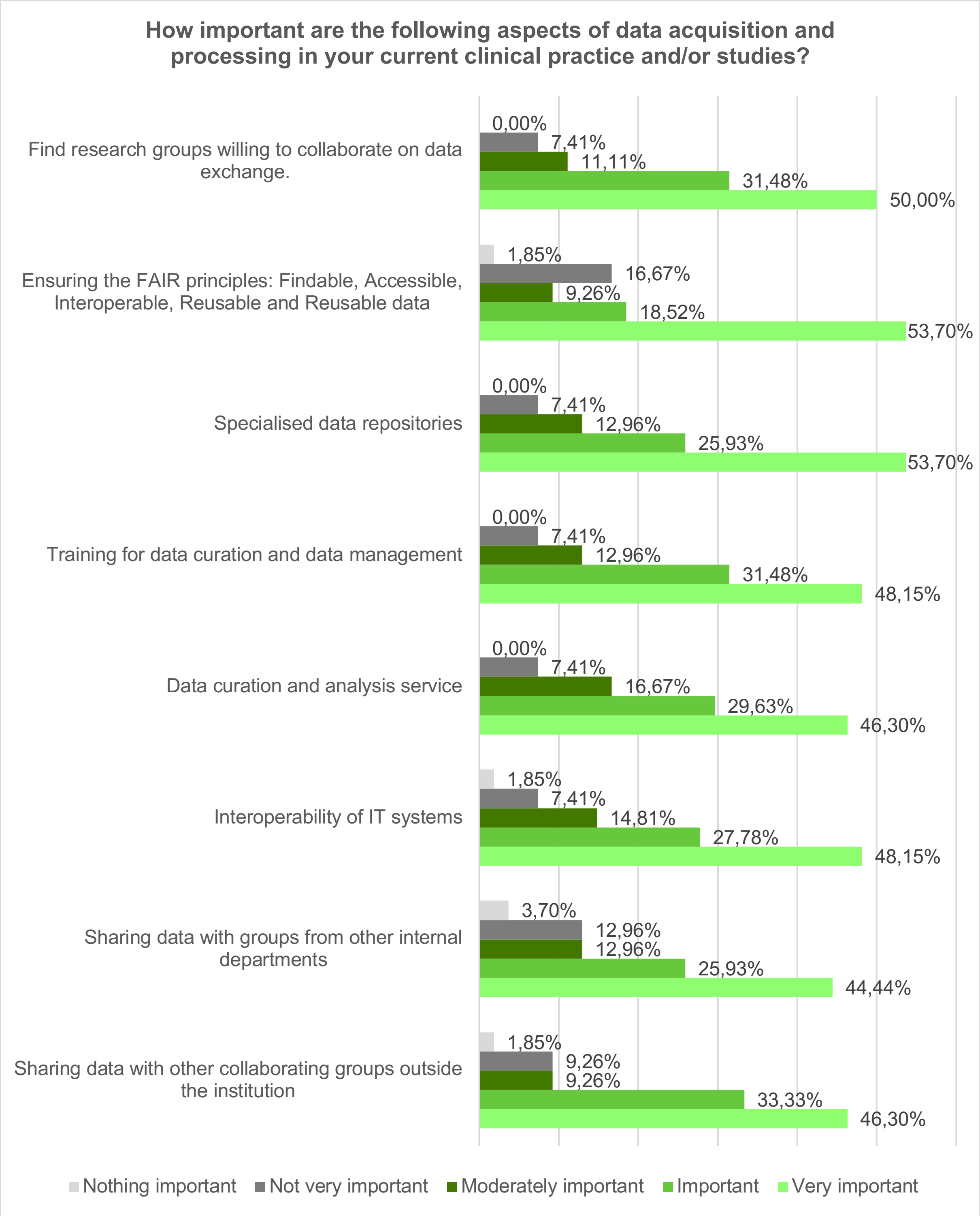


Figure 13. How important are the following aspects of data acquisition and processing in your current clinical practice and/or studies?

As shown in Figure 13, among the main factors that were rated as *very important* and *important* by more than half of the participants were: *specialised data repositories* (53.70%), *ensuring FAIR principles* (53.70%), *research groups willing to collaborate in data exchange* (50%). Nearly half of the respondents indicated as *very important* the need for training *in data curation and management* (48.15%) and the interoperability of IT systems (48.15%). Only in some cases is it *not very important* the assurances of FAIR principles (16.67%) or data sharing with other collaborating groups outside the institution (9.26%).

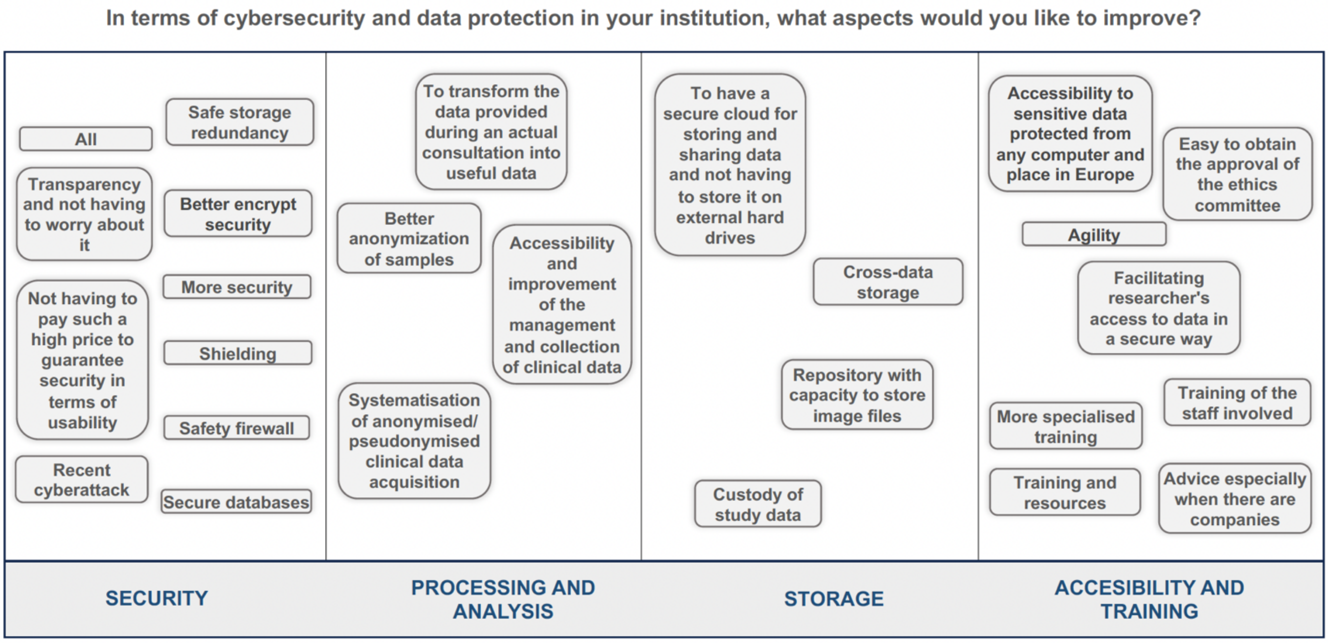


Figure 13. Cybersecurity and data protection of your institution. Authors' elaboration

To identify the most important and common needs of the participants from the perspective of cybersecurity, data protection and points for improvement, the diagram in Figure 14 has been created based on responses to the questionnaire. Aspects such as strengthening the **security** of IT systems, improvements in terms of data **processing and analysis**, **storage** alternatives, **accessibility** of data in a secure way, request for specialised **training** and speed in terms of approval by ethics committees, etc. are key needs.

Regards ***security**,* it is worth noting that some centres have suffered attacks on their IT systems, an increasingly common problem both in Spain and internationally. In fact, the report published by the European Union Agency for Cybersecurity (ENISA, 2023) on the *cyber threat landscape of the healthcare sector in the EU* states that 54% of *ransomware* attacks were detected between January 2021 and March 2023, which led to the closure of emergency departments and the suspension of surgical interventions in two of the affected centres.

In the ***processing and analysis*** segment, innovative ideas were put forward such as the systematic analysis and preparation of information obtained in medical consultations, which would not only reduce the computer workload of health care staff when assisting patients, but increase the efficiency of transcription and processing of data. Likewise of importance are using software and implement procedures whereby the anonymisation of data is a safe process, both for the patients and the researcher, with special attention to systematisation and accessibility.

In terms of ***storage*** needs*,* improvements include cloud services for data sharing, cross-data storage of individuals involved in multiple research projects, storage of image archives and data custody. The independence of external hard drives is perceived as an obstacle by some participants, however, there are centres that see them as a means of safeguarding information or maintaining backups, so the use of the cloud would be an additional benefit to these needs.

In terms of ***accessibility and training***, there is a need for agile data-sharing platforms, speedy interaction with ethics committees and easy access to information from other sites. We note that the handling of sensitive data and current regulations - such as the Organic Law on Personal Data Protection (LOPD) - are clear limitations to implementing some of these demands. ***Training*** is a recurring element in the survey, so it would be an advantage for staff to be able to manage, analyse, standardise and make appropriate use of clinical data. Training courses that, in addition to including health documentation technicians, could reach all health research staff.

2.1.5 Ownership of unique databases and collaboration with other organisations

One set of issues we were most interested in finding out was whether respondents shared data, with what type of entity they carried out these collaborations and whether they had their own databases (DB) that they were also willing to share. To find this out, we defined different types of collaboration that could occur in the healthcare environment first:

- Collaboration at individual level: This refers to one-off researcher-researcher collaboration in the health sector.
- External collaboration with the public sector: This refers to the collaboration that takes place at the institutional level through the signing of a data transfer agreement with other public sector entities, including health research institutes, PROs, public sector foundations, universities, etc.
- External collaboration with the private sector: This refers to the collaboration that takes place at the institutional level through the signing of some kind of data transfer agreement with other public sector entities including companies, start-ups, private sector foundations, etc.
- Internal collaboration: This refers to the collaboration that takes place with other departments and/or groups within the organisation itself.

Based on the results obtained, only 5.56% of the respondents have not carried out any type of collaboration on data sharing issues. On the other hand, the majority of respondents have indeed carried out some type of collaboration on data sharing issues, the most frequent being internal collaborations (75.93%), collaborations on an individual level (68.5%) and external collaborations with the public sector (62.96%). However, the percentage of respondents who have carried out some kind of collaboration with the private sector on data sharing issues is very low, only 18.52%. Public-private collaboration is essential for innovation and the exploitation of research results. These data suggest that more incentives should be provided to encourage this type of collaboration.

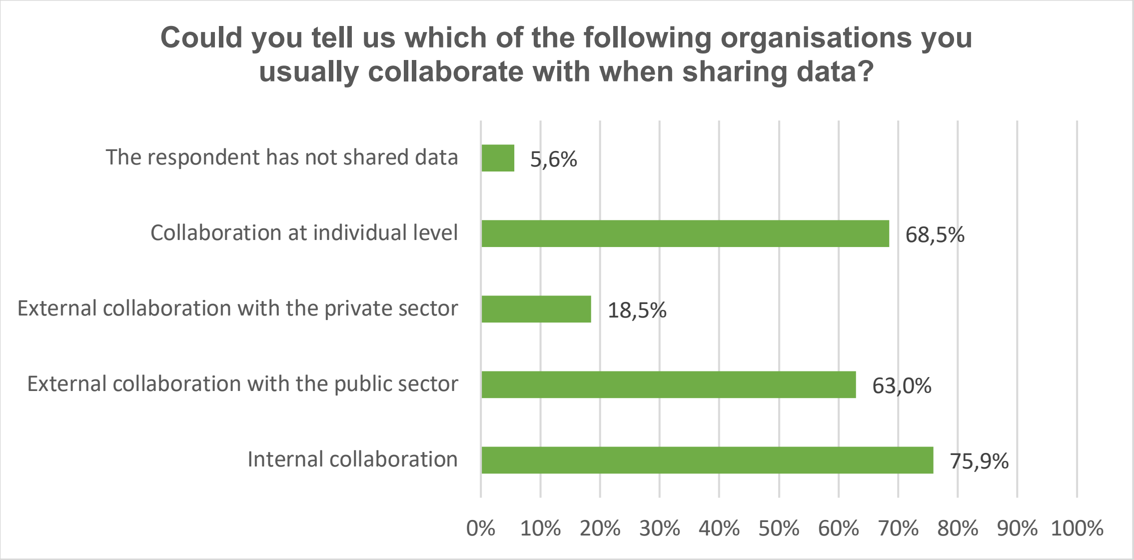


Figure 14. Type of collaborations made by respondents on data sharing issues

In continuation, we explored whether the researchers surveyed had unique DBs that were made up of data obtained in the different research projects carried out by their groups, and what their position was on the possibility of sharing these DBs with the scientific community.

In Figure 16 below, the results show that 50% of the researchers surveyed do not have unique DBs. The other 50% do have unique DBs, but the results vary depending on the willingness of each researcher to share data. The preferred option among these researchers (33.33%) is that they while they are interested in sharing data, the difficulty is in the way it is shared. On the other hand, 12.96% of the respondents would be willing to share their data without discussing how to share it. Only 3.70% of respondents are not interested in sharing data, mainly because of the obstacles and limitations imposed by research ethics committees and their own centres given that sharing such data might contain some kind of sensitive information.

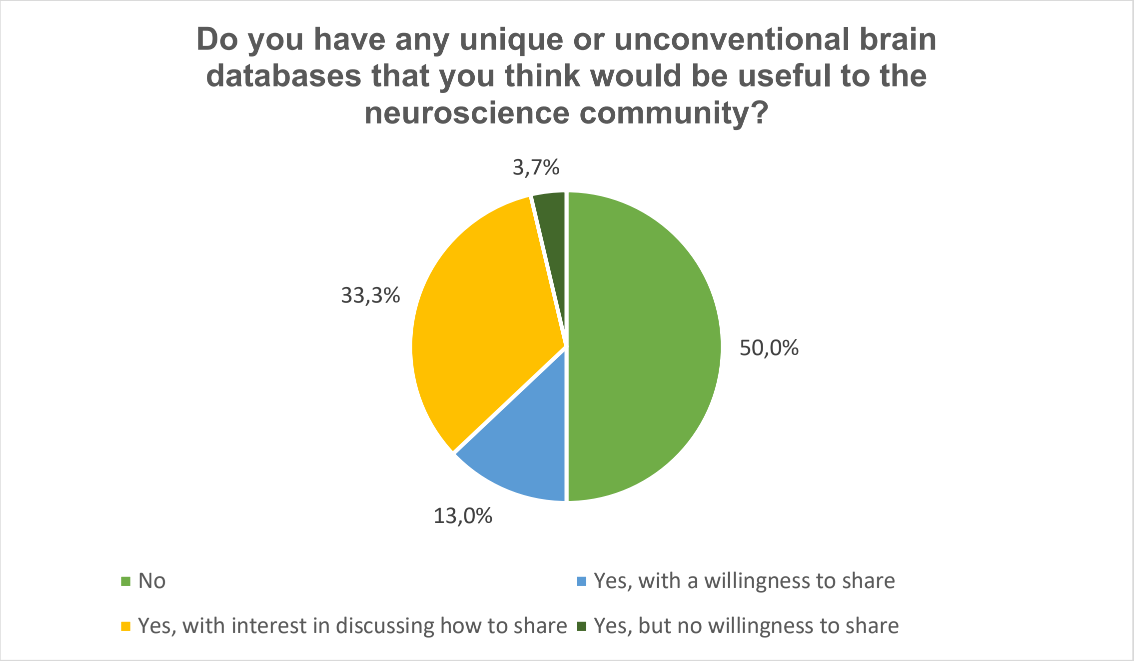


Figure 15. Ownership of databases and respondents' willingness to share this data

To conclude this section, we asked respondents which companies or start-ups in the field of neuroscience they have worked with or collaborated with. Given the low percentage that in the previous section acknowledged collaborating with the private sector in the area of data sharing, it was to be expected that the vast majority 68.52% could not answer the question. Among those who did answer, however, 24.07% provided the names of the companies they worked with, while only 7.41% did not provide these names, presumably for reasons of confidentiality.

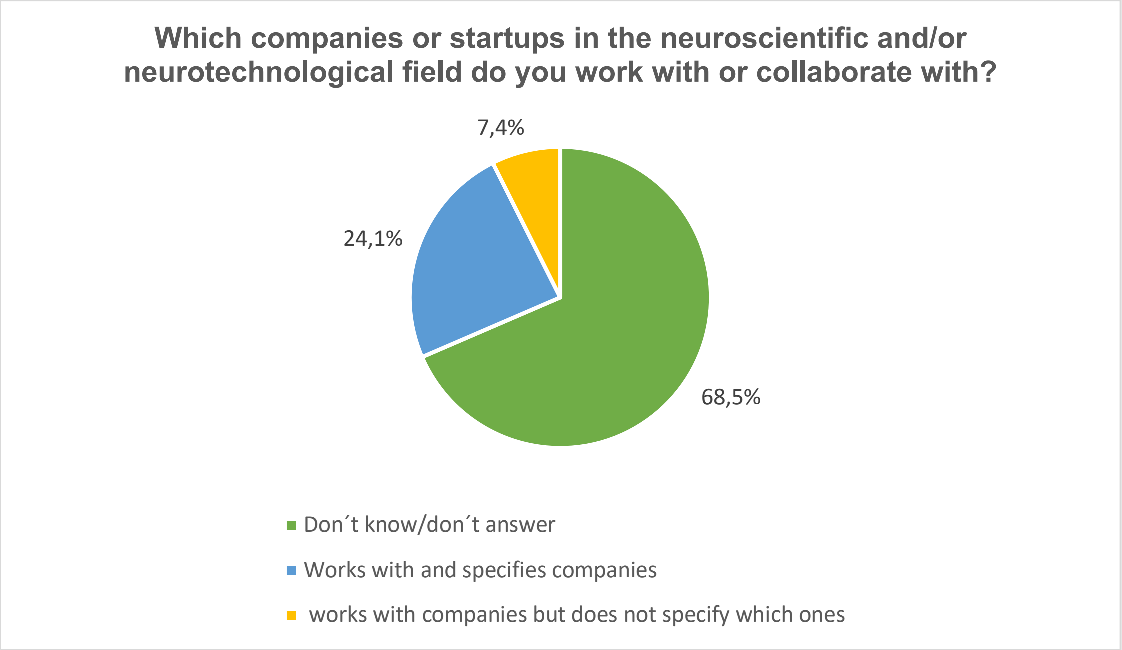


Figure 16. Percentage of respondents working or collaborating with companies or start-ups in the field of neuroscience

Part III. Interview results and analysis

3.1 Introduction and general information about the respondents

The **Health Research Institute of the Hospital 12 de Octubre** in Madrid **(i+12),** created in December 2009, "*is a multidisciplinary and translational biomedical research structure oriented towards basic, clinical, epidemiological and health services research"* (imas12, 2023). The Hospital Universitario 12 de Octubre is the core of the Institute, with teaching and research staff from the Complutense University of Madrid , the Autonomous University of Madrid, the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas and the Universidad Europea, as well as research groups from Primary Care in the Community of Madrid, being the Fundación para la Investigación Biomédica of this hospital the managing entity of the Institute. The Institute encompasses care, research and teaching under one umbrella (imas12, 2023). Among the different research areas interviewed was the Neurosciences and Mental Health group, focusing on neurodegenerative diseases, addictions and comorbidity, cognition and psychosis, neurovascular diseases, behavioural and applied neuropharmacology, neurotraumatology and subarachnoid haemorrhage, and translational psychiatry.

Figure 17. Health Research Institute of the Hospital 12 de Octubre (i+12 Institute). Source: imas12, 2023

The **Spanish Society of Neurology (SEN)** is a non-profit scientific association focused on research, promotion, advancement of Neurology and interaction between professionals in the field (see SEN, 2023). Founded in 1949, it is made up of two foundations: the Private Foundation Spanish Society of Neurology and the Spanish Foundation for Neurological Diseases (Fundación Española de Enfermedades Neurológicas (Fundación del Cerebro). Its scientific committee is composed of 12 members of the SEN, with its associates now numbering 3,563, including neurologists - one of whom was interviewed - and other health professionals. The SEN is an active association, regularly organising conferences, online and face-to-face courses, study group meetings, e-poster sessions, among others.

Figure 18. Neurodesafío, LXXIV Annual Meeting of the SEN, Seville. Source: SEN, 2023.

The **Vall d'Hebron Research Institute (VHIR)** promotes biomedical research, innovation and teaching, with a team of more than 1,800 staff (VHIR, 2023). VHIR is a public sector institution that has been promoting and developing biomedical research, innovation and teaching at Vall d'Hebron University Hospital since 1994. VHIR is made up of 5 reference institutions: Vall d'Hebron Hospital, Vall d'Hebron Recerca, Vall d'Hebron Institute of Oncology, Cemcat Centre d'Esclerosi Múltiple de Catalunya and the Universitat Autònoma de Barcelona (UAB). In terms of care, they receive 7,000 patients a day, 9,000 professionals, 95 research groups, 17 buildings and 575 residents. In the field of neurosciences, they have about 100 researchers distributed in 13 working groups. Our interview took place with the research group specialised in stroke. In terms of clinical trials, 399 of these are oriented towards oncology services (Vall d'Hebron Barcelona Hospital, 2023).

Figure 19. Vall d'Hebron University Hospital (2023)

The **HRI Galicia Sur** is a multidisciplinary research space in Biomedicine, based at the Álvaro Cunqueiro Hospital in Vigo, which provides services to more than 600,000 people in the city (Xunta de Galicia, 2023). It brings together clinical research groups from the Health Areas of Southern Galicia and biomedical groups from the University of Vigo, with the aim of promoting translational research and innovation that can ensure the effective application of research results in terms of health benefits for citizens. Interviewed at its facilities were professionals from the area of innovation and transfer, and researchers focused on the area of Translational Neuroscience and Neuroimaging. The research lines of the former are related to psychiatry and to neurodegenerative diseases such as Alzheimer's, cognitive impairment, neurorehabilitation, addictions, etc. The second group specialises in diagnostic imaging, with research lines on tumour treatment, neurointerventionism and the use of artificial intelligence.

Figure 20.Álvaro Cunqueiro Public Hospital. Source: SERGAS, 2023

The **Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)** is a health research institute dedicated to biomedical research of excellence that addresses the most common diseases in our environment. Employing around 2,000 professionals organised into a hundred multidisciplinary groups, it publishes more than 1,500 original articles per year and is one of the leading biomedical research centres in Spain, with a clear international vocation. The Hospital Clinic de Barcelona is the core of the institute, and is also associated with the University of Barcelona. IDIBAPS specialises in areas such as cerebrovascular diseases, Parkinson's and other neurodegenerative movement disorders, clinical neurophysiology, pathogenesis of autoimmune neuronal diseases, Alzheimer's disease, and other cognitive disorders (Clínic Barcelona, 2023). In the Clinical and Experimental Neuroscience research area of IDIBAPS, the **Brain Circuits and Behaviour Lab,** the result of the union between the Theoretical Neurobiology group and the Cortical Circuits group (Brain Circuits and Behaviour Lab, 2023), was interviewed.

Figure 21.IDIBAPS (2023)

In Barcelona, the Neurology service associated with the **Clinical Neurosciences Institute (ICN)** attends 1,300 patients a year, 20,000 outpatient visits are made, 4,000 hospital sessions a day, 9,000 cognitive or physical intervention sessions for patients with neurodegenerative diseases and 7,500 functional tests (EMG, EEG, etc.) (Clínic Barcelona, 2023). The care service is deployed by the Units of

Cerebral Vascular Pathology, Epilepsy, Neuroimmunology and Multiple Sclerosis, Parkinson's and Movement Disorders, Alzheimer's and other cognitive disorders. Interviewed were the groups focusing on Parkinson's disease and other neurodegenerative movement disorders.

Figure 22. Hospital Clínic de Barcelona

The Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), in Madrid, is dedicated to preclinical, clinical and translational research, as well as deepidemiological and health services. It is made up of centres of recognised prestige such as the Hospital Universitario La Paz and the Autonomous University of Madrid and is also associated with other centres such as the Hospital Universitario de Getafe and the Universidad Europea de Madrid. It is legally supported by the Fundación para la Investigación Biomédica del Hospital Universitario La Paz (FIBHULP). The Research Groups in the Neurosciences area include: 1) Neurology and Cerebrovascular Disease (CVD, neurodegenerative diseases, movement disorders and ataxias, epilepsy, cephaleas, neuromuscular-ELA, Neurosonology and Computational Neurology, Translational Research Laboratory); 2) Psychiatry and mental health; 3) Neuroprotective Strategies in Neurodegenerative Diseases; 4) Implication of the Glycergic and GLutamatergic Systems in Central Nervous System Pathologies; 5) Molecular mechanisms and biomarkers in neurodegenerative diseases; and, 6) Research in Multiple Sclerosis. Both the Neurology and Cerebrovascular Diseases Group and the Innovation Area were interviewed.

Figure 23. La Paz University Hospital. Source: UAM, 2023

The information and observations of the interviewees provided in section 3.2 are deliberately presented anonymously for privacy reasons. The order of presentation of the institutions provided above does not, therefore, correspond to the sequence of information in this section.

3.2 Knowledge of and interest in the Human Brain Project (HBP) and EBRAINS

Institution interviewed No.1: The groups interviewed are aware of both projects and, although they have collaborated with other researchers associated with HBP/EBRAINS, they have not used these services.

Institution interviewee No.2: The neurologist associated with the neurodegenerative diseases research institute learned about the HBP/EBRAINS services from the online presentation organised by the UPM and ISCIII. She had previously explored the project website, interested in the data services.

Institution interviewed No.3: Participants interviewed share that they have knowledge about BPH and have participated in several online activities, including Calls that they have received from the innovation unit of the centre. They perceive the resources offered as excellent. However, they have not been able to access these services, either due to a lack of knowledge about the opportunities they offer, or due to a partial vision that has been transmitted to clinicians. More outreach from HBP/EBRAINS to them is needed. On this, as participants who are members of a research institute and potential users, they have felt more like spectators than participants.

Institution interviewee No.4: The neurologist is familiar with both HBP and EBRAINS. He finds the ethical commitment of these projects outlined on the website *key.* He has followed several of the presentations over the last few months, and has been surprised by how far HBP has come. He shares that he could not choose just one tool among the variety of solutions they offer and believes that it is a wonderful set of resources compared to what they are normally used to seeing in clinical neurology.

He is interested in brain virtualisation, and explains that the development of biological models of the brain has had a strong influence on predicting how the brain works, from the creation of the first perceptron, a mathematical formula that basically replicates Santiago Ramón y Cajal's polarised neuron, to advanced convolutional brain network systems and generative AI. Brain virtualisation models can return the favour to us in turn, he goes on, with their ability to simulate how the brain works in many ways. In addition to his clinical work, the interviewee shares his keen interest in this area, understanding how the study of the brain inspires the creation of neural networks and neuromorphic computing, and the influence of brain knowledge for the advancement of AI systems. He congratulates the European HBP and EBRAINS initiative, and the North American BRAIN Initiative, for their important contributions to the advancement and development of virtualisation.

Institution interviewee No.5: The neurologist interviewed learned about HBP recently and as a result of the communications sent by the UPM-ISCIII. She finds the project very interesting.

Institution interviewee No.6: The physics PhD shares that he has heard news about the HBP project since its inception. Among other aspects, he is of the opinion that the project deviated from the initial promise of simulating the entire human brain. He thinks it is appropriate for neuroscientists to be self-critical internally and to acknowledge to the scientific community not only the successful work done, but also what has not worked, in line with the funding obtained. At the same time, he points out the importance of evaluating the image that HBP currently has within the neuroscientific community, beyond the groups that make it up, and the future impact that EBRAINS services will have on the community.

Institution interviewed No.7: Neurologists were aware of HBP and EBRAINS through neurology forums, AI-related professionals, and through multidisciplinary forums. In terms of decentralised data sharing and exploitation, they recognise that it is useful to find platforms that help optimise these processes. They were not aware that EBRAINS offers these services, finding it interesting to unify these efforts for this, as well as in terms of funding, consulting and pathways for harmonisation.

FOOD FOR THOUGHT

- Most of the interviewed specialists, PIs in the area of neuroscience and neurology, were aware of the Human Brain Project (HBP) in general. However, many of the details of the EBRAINS project services became known through the UPM-ISCIII online presentation and the interviews.
- Dissemination and communication efforts should be increased to bring EBRAINS basic and clinical neuroscience services closer to clinicians and researchers, with an emphasis on participation.
- Increase the project's communication and dissemination activities to reach out to users who are unaware of basic and clinical neuroscience data services and tools.

3.3. Type of data used in clinical practice and/or research

Institution interviewed No.1: The Translational Neuroscience group focuses on the study of neurological diseases such as Alzheimer's, cognitive impairment, and mental diseases such as schizophrenia, addictions, depression, bipolar disorder. The data he uses as a leader range from proteomic and molecular level to clinical data managed by the rest of his team. Proteomic and molecular data are not particularly sensitive.

The neuroimaging group specialises in neuroradiology and works mainly with imaging data to identify lesions or extract variables from them (MRI, CT, etc). The centre's expert in transfer and innovation means explains that sometimes there is a certain reluctance to use public databases by other researchers because they do not know for sure how the data has been processed nor is it clear where the data comes from or what equipment has been used in the investigation.

Institution interviewed No.2: The neurologist works with (sensitive) clinical data obtained through examination and anamnesis, in addition to biomarkers- a field in which she is a specialist. She obtains laboratory data, encephalorachidian fluids, and blood tests. She uses other neuroimaging data from patients in nuclear medicine.

Institution interviewee No.3: From the hospital, they work with their own data, largely from their own healthcare practice and translational research projects. They also want to know how to access project databases and how they could contribute to them. In other words, they perceive a gap in communication about the resources offered by the platform and what they can contribute. They work with different projects and lines of research, including those focused on brain damage repair therapies, biological protection mechanisms, injury markers, plus non-invasive technologies (in animal models and humans).

Institution interviewed No.4: The neurologist does not work in a health research institute nor does he lead a research team. However, he does collaborate with many people in the sector, conducts studies, and contributes to decision-making on aspects that can influence society. Thus, from the committee specialised training is offered to the medical community.

Institution interviewed No.5: The neurologist uses both hospital and third-party data. The data she uses are open, while sensitive data are pseudonymised.

Institution interviewed No. 6: In his laboratory, the doctor analyses data that is neither very sensitive nor pseudonymised. He does not work with data or clinical records because the experiments are on animals or volunteers - not patients. His need is for medium to long-term *storage of raw data*, i.e. the storage of large amounts of gigabytes of data. He explains that although his institute is discussing how to maintain this service to researchers, in the meantime they have to find ways to solve this storage issue, either through their own server, third party servers, etc. In short, they require cloud storage services and in large volumes.

For this, we suggested the supercomputing services of HBP and EBRAINS. The project has an agreement with the European supercomputer network FENIX, including the *Barcelona Supercomputing Center* (BSC-CNS), which provides computing servers, with utilisation and service levels agreed by the consortium. The interviewee explains that his institute has acquired its own high-powered servers, which require system-level maintenance and specialist technicians, a machine hosted at BSC-CNS, which they in turn offer to other groups, as networked machines. He finds it interesting that EBRAINS can offer storage or computing services through the consortium, so that neuroscience researchers can carry out simulations or brain analysis.

Institution interviewed No. 7: The European project this institution leads within the framework of the European Health Initiative, follows a very similar structure to EBRAINS through digital information platforms. Being focused on strokes, it takes into account the entire process of the disease - not only in the hospital, but also in the patient's own journey, in the ambulance, at home, etc. He points out that they have developed a modular function that integrates data - clinical, biomarkers, genetic data, epigenetics, imaging, as well as the care and follow-up process - in a harmonious and standardised way, all coordinated by his research group.

Similarly, their records feed into the Catalan Agency for Health Quality and Assessment (AQUAS). For example, in the HARMONICS project, they not only standardise the stroke process in Catalonia, but also systematically collect health outcomes efficiently. He points out that more than 80% of the outcomes data of stroke patients are available, feeding in turn the Servei Català de Salut (Catsalut), and generating a *pay-for-value* model, as a way to incentivise researchers associated with annually-defined targets. With this model, i.e. based on patient outcomes, as patients recover, the variable fee improves, too. This scenario is still under discussion.

The institution specifies that the most easily reproducible data are *outcomes* or treatment impact, based on important patient information. This reproducibility is not possible in all centres, though. Complex data, such as those from biomarkers, depend on the techniques used to obtain them, and standardising them across hospitals is still a technical challenge. In this centre, several lines of research use, in function of purpose, different biomarkers, sample analysis, etc.; so, the aim here is to ensure that the researcher develops a line of work that also be reproducible in other scenarios.

The aim is not for the researcher to keep the data and use it solely in her/his own centre, but for it to be reproducible. The associate neurologist says that he himself records samples from patients undergoing thrombectomy and tries to reproduce them in other settings. He recognises that these are difficult processes, especially finding teams or ecosystems of people who want to get involved in these kinds of processes. Thus far, fruits of his efforts are limited to the contacts he has obtained or professionals with whom he has collaborated in projects, congresses, publications, etc. It would therefore be interesting for HBP/EBRAINS to identify common areas of interest in stroke, centres that want to get involved and generate nodes that can build on, or replicate, each others' data.

FOOD FOR THOUGHT

- In general, research groups generate their own data and have the capacity to work with sensitive and open data. Each group manages its own acquisition, storage and analysis process, so there is some reluctance to use third-party data or databases.
- The EBRAINS data storage, management and sharing services are, in general, unknown to the interviewees, who want to know more about the type of data they could obtain and share on the platform.
- Long-term data storage needs are growing rapidly, and typically involve significant investments in monetary, physical, energy and human resources.
- *Pay-for-value* is a payment modality that could incentivise researchers to use federated data platforms in healthcare, reducing healthcare costs and targeting patient recovery.

3.4. On clinical data analysis and processing services

Institution interviewed No1: The Translational Neuroscience group carries out its own analysis and processing of data, games for neurorehabilitation, biomarker development, memory studies, etc. However, in terms of development and imaging studies, they are supported by researchers from outside the institution who use other equipment such as magnetoencephalography. The programmes and evaluation equipment managed by the leaders interviewed are also used by the rest of the team, made up of more than 30 professionals.

In the neuroimaging group, around 50 staff specialised in medical imaging maintain a transversal service, with lines of research in radiomics in abdominal pathologies, and work with a large volume of data. They also work with their own data and seek to validate them within their own centre or carry out external validations.

Institution interviewee No.2: This participant explains that there are many members of the neurodegenerative diseases group, about 3 PIs in the lab, while the clinical team is larger. Each group has its own way of processing data. A database is being built in the memory unit, developed using a platform centralised by the research institute. Their cohort is just starting, although previously they have used databases from international platforms, such as ADNI or GAAIN, since they were focused on neurodegenerative diseases such as dementia, and other disease biomarker databases. In her previous lab, they used centralised samples from other locations, which were then analysed by her own group who provided more data; subsequently, they received information from patients, which they then worked with.

Institution interviewed No.3: They handle a variety of programmes such as SPSS, Stata, R. However, they have no specific training, being self-taught. Nor do they have additional dedicated staff or the necessary knowledge to access services such as those offered by HBP/EBRAINS. They are interested in receiving support and training in data processing and analysis. They also value the figures of a data expert and of biomedical engineering and statistics, so that they can know how to extract data and obtain useful information.

Institution interviewed No.4: The participants explain that the institution can offer entrepreneurial services, given the fact it has years of experience in handling technologies and making them *intelligible* to neurologists, organising courses, seminars and carrying out activities with other study groups. From their organisation, they have transferred to clinical neurologist aspects that were unknown to them before. For instance, such basic proposals as the use of shared, secure, robust databases through the REDCAP platform. Not only does the institution provide access, but more and more neurologists are using it, asking it for access to programmes, or to become partners in the field of data, etc. In these courses, the institution teaches health professionals how to *handle, tabulate and visualise* data, and has created a manual with this information. These are demanding courses, each having only about 20 students (among the 4,000 members), and 80 new residents join each year.

In his particular experience in a private hospital, the dementia research area used advanced analytical tools, although mainly supported by clinical trials. In the department one or two people advised or performed statistical analysis, however, they did not have a platform with tools for on-demand use. He suggests that neuroimaging services should have a different approach, because they use datasets of normal brains, at rest or doing certain activities, which they then compare with other patients.

Institution interviewed No.5: The group is made up of about 7 people including neurologists, nurses, administrative staff, biologists and geneticists. As examples of data collection, they have built a database of a neurological disease (multiple system atrophy), with the support of the statistics team and following the hospital's regulations. The database has not been requested by any researcher, and even if it was, he explains that they would not know how to share it, they do not have that possibility.

Institution interviewed No.6: Their laboratory needs for calculation and storage have been covered, either with its own resources in the laboratory, or with the institute and with agreements between the institute and third parties, as in the case of BSC-CNS. The laboratory employs 2 PIs. They work with 5 PhD students, 4 postdocs, master students, technicians. In total, about 10 people. Half the time they do experiments and the other half they work analysing the data they generate.

They analyse data on behaviour, electrophysiology, and use programming and statistical techniques to evaluate all the information in more depth. If the data analysis is sophisticated, they require specific programming, not just using Excel, but specific programmes that are shared within their lab. In terms of data acquisition, they obtain data from patients through the hospital, the product of their own experiments, not from the doctors. Collaboration with the hospital involves approval from ethics committees, and once approval is received, schedules are coordinated to receive participants and conduct the studies. The resulting data are sometimes stored on the institute's servers, but usually on their own servers.

Institution interviewed No.7: The studies they carry out take into account the entire stroke process, recording the variables of the disease, all through the acute, sub-acute and ambulatory phases, identifying the problems in these pathologies. They underscore the priority that all data are correctly traced, which is why - to be able to know the problem in depth - they handle data from biomarkers, neuroimaging, clinical, biomechanical variables, devices, etc. They obtain all possible variables and quantify them in order to know what can be done at each stage of the process. To do this, they draw on the expertise of a multidisciplinary team, from neurologists specialised in stroke, forming a small ecosystem of innovation and research. They are also developing a multimodal approach to the pathology: in addition to neurologists, they are therefore also involving engineers, occupational therapists, nurses, rehabilitation, physiotherapy and speech therapy.

It is often a challenge to get filtered, noise-free data. Depending on each need, the type of data to be obtained from other projects must also be verified. HARMONICS, the project this institution coordinates, seeks to standardise these data management processes so that these can be replicated in different scenarios. The institution thus finds it interesting to connect with other data nodes in Europe to identify trends in this area of clinical analysis.

FOOD FOR THOUGHT

- Each research team carries out its own method of data processing and analysis, with few standardised pathways between centres or between groups.
- A low proportion of researchers rely on third-party services for some support in data analysis.
- In medicine and neuroscience, expertise in statistical, mathematical and programming software is increasingly required, in particular for the use of multi-platforms for data analysis.
- The figure of Data Manager or data expert is not very common, although it is considered necessary. There is no staff dedicated exclusively to these activities, which are carried out by the researchers themselves.

3.5. On human brain simulation tools, digital atlas and federated medical data platforms

Institution interviewed Nº1: From their innovation and transfer team, they have heard about other federated medical data platforms. So far, they have not used them for neuroscience studies, although they have used them for colon samples, which do not carry a risk of identifying a patient. They understand the usefulness of these systems in protecting sensitive data without the need to destroy it. They have been able to upload videos and images of colonoscopies to biobanks, so that these resources can be used by other researchers. Both the Translational Neuroscience research group and the Neuroimaging group have not used human brain simulation software.

Institution interviewed No.2: She has not used human brain simulation tools or federated medical data platforms. She would, however, like to have a better understanding of how these tools work, where the simulation results come from and what you can do with them.

Institution interviewed No.3: They do not use them in clinical practice. They are interested in having access as a research group and being part of the project. However, the activities initiated with the medical IT platform were not followed up, with the result being that these collaborations came to naught.

Institution interviewee No.4: He has not used virtual brain simulation tools or federated data platforms. While acknowledging that they can be useful, he would have to analyse the purpose for having them. Although the technologies may be advanced, the question is - for a neurologist - how it can be brought into clinical practice. He suggests that the examples of HBP/EBRAINS epilepsy or Parkinson's may be key. He would find it interesting to simulate certain lesions in a virtual brain, in multiple sclerosis or cognitive impairment. He explains, though, that first the *imbalance* has to be compensated for given that it is a virtual brain, and secondly, that the clinician has to learn how to use it for research. According to the interviewee, this is an *anti-natural gap* resulting from a technological and computer age that is advancing faster than the needs and research areas of the clinicians themselves.

Institution interviewee No.5: The respondent has not used human brain simulation software, federated medical platforms or those created in HBP/EBRAINS as explained in the interview. It is possible that at some point they will be taken into consideration for use in the studies they carry out.

Institution interviewee No.6: They have experience in neural network simulation and software tools, some of them like The Brian Simulator. They find these tools useful, especially when the learning curve is simple, meaning that researchers who come from other fields do not have to master programming languages nor spend too much time learning to use the codes. Though these kinds of tools are not used everyday use, only on occasion, perhaps in the future they will become a fundamental tool.

They have not used brain atlases, but they are interested in starting to use them. While they have no experience with federated medical platforms such as HBP/EBRAINS, they do find it very positive that different hospitals have joined this project. In their lab, they work with specific aspects of brain function (e.g. memory) to assess how this is altered in patients. Doctors usually test these patients, using the institution's own lab designs and tools that can be accessed from a mobile phone or tablet. For the, it would be useful to explore possibilities to disseminate their assessments to other patients using these medical platforms.

Institution interviewed No.7: They have extensive experience in federated medical data platforms. The projects they are developing are focused on stroke. The Stroke project is managed as an internal platform. HARMONICS focuses on the standardisation of the stroke process, although they also have records on epilepsy. They believe that regards these diseases they achieve cost-effectiveness of the data as they are frequent pathologies. They can segment the data according to disease progression so as to be able to make adjusted predictions. In this aspect, they obtain global and holistic health outcomes, redefining disability in the patients they assist. The Stroke data platform is a federated network at the European level, coordinated by their institute.

FOOD FOR THOUGHT

- Almost none of the research groups interviewed have used human brain simulation tools and brain atlases, neither in clinical practice nor in experimentation. They have had no need to use them.
- Only one research team has made occasional use of neural network simulators.
- Few groups are aware of federated data platforms or are part of one.
- In general, all groups show interest in learning about such tools and exploring possible collaborations.

3.6. Relationship with data-sharing platforms

Institution interviewed No.1: The expert from the innovation and transfer unit at the institute believes that the health system in her autonomous community has a very efficient intranet that helps to organise data. However, some drawbacks do arise, such as the fact that some of the data cannot be opened on a computer, or the limitations on the use of software from external providers for communication. Using open tools may be possible in the laboratory context; however, in the clinical setting the platform is strictly under the control of the doctor. Researchers may share some data from their research projects with other computers outside the health system. In general, the public health system is restrictive in terms of installation and usage of other software, and these processes are usually run by the IT departments.

Another aspect discussed is the formation of health *data lakes*, as a data repository initiative. Once researchers have completed their studies, they can share the data in this *system* so that it can be accessed by other groups in the future. However, in practice this processing is done by the researchers themselves, with the drawback being that standardisation in one area - for example, in biobanks - has a different level of demand. In the long run this means more work for the team and discourages making such data available to the whole community. Bureaucracy makes it difficult to share data, so they end up being destroyed or safeguarded in the public health system.

Institution interviewed No.3: The specialists have stressed that they agree with an open data policy in science and have also shared their research data in other repositories, although not with those created by HBP.

Institution interviewed No.4: The respondent noted that the main challenge for data sharing has a lot to do with taxonomy and conforming to a universal syntax. It is essential to denote diseases by name, and to have data with common *tags.* Spanish hospitals, in general, have not reached an agreement on this, so that a database of a disease such as Alzheimer's, for example, can be homogeneously shared using a common syntax. Without a common language, the *matching* processes will never be exact. This prior work of homogenisation and common language between the data is necessary.

Another important aspect for the neurologist is that the datasets of each hospital are sufficiently similar so that other hospitals or centres can evaluate and validate their machine learning systems, whether searching for new prognostic factors of a disease or using virtual diagnostic equipment. Data - not just the platforms - must be similar so that they can be shared. He believes that the definitions of the main data must be clear, following the appropriate taxonomies and syntaxes so that everyone can work on them in a common way. So that they can be shared with other hospitals, with other countries, and not be wasted.

Institution interviewee No.5: This respondent explains that in her Autonomous Community each hospital has its own data management systems or programmes, and many are adopting SAP. With the Michael J. Fox Foundation, they are carrying out a study known as the Parkinson's Progression Markers Initiative (PPMI) and are contributing to the construction of a database. They are also leading an initiative called Healthy Brain Aging (HeBA), a multi-centre project between four cities: Kassel, Innsbruck, Barcelona and Luxembourg. In this case, the acquired data are stored in the hospital according to internal rules and approval of its ethics committee before being shared to the technological infrastructure of another city, in this case - Luxembourg. Thus, they can have internal databases and at the same time work with other international platforms, while sensitive information is managed according to a set of codes known only to the researcher.

Interviewee Institution No.6: Its laboratory sometimes shares (anonymised) data, usually with other collaborating researchers and in some cases when they publish. The data from its experiments first have to be stored according to a protocol whereby each patient is identified with a code, and this information is stored separately. The laboratory does not have access to this identification data.

Institution interviewed No.7: They point out that they do not share data, it does not leave the hospital. To work with other groups on European projects, they use federated platforms, such as the Horizon Europe-funded Stroke initiative. The neurologist (and principal investigator of the project) explains that the initiative is made up of European public and private partners, in a federated learning structure based on algorithms for prediction throughout the stroke process, which they use to evaluate a series of events such as response to treatment, readmissions, etc. The data does not leave the hospital as a way of treating patient information with security.

Regarding access to clinical data, given that each Autonomous Community has its own electronic computerised records, one can only access the public ecosystem or the network of public hospitals in the community strictly for clinical reasons. On the other hand, for research and projects, the application goes through the ethics committee.

FOOD FOR THOUGHT

- The health system differs significantly between Spanish Autonomous Communities in terms of data storage, organisation and sharing.
- A universal ontology or syntax needs to be followed within hospitals and research centres to facilitate data sharing and accelerate scientific advances. In addition, there is a need for similar databases across centres to validate AI techniques.
- Federated platforms are seen as a useful and attractive alternative for data sharing in neuroscience.

3.7. Experience using virtual simulators for neuroscience experimentation or clinical practice

Institution interviewed No.1: They would like to use this type of tool. They do not currently use them in their studies because they need to generate enough data to do so. It is not an activity they have done at the centre thus far.

Institution interviewee No.2: It is not the institution’s speciality. As a physician, she finds the simulation services offered by EBRAINS a bit complicated, but also quite interesting.

Institution interviewed No.3: They do not have experience in the use of HBP simulators, but they do have experience with others. They are interested in incorporating these platforms as educational and training tools for other researchers. For clinical trials, virtual simulators would be interesting for generating hypotheses, although it would be difficult to transfer the results directly to clinical practice.

Institution interviewee No.4: It has not used these tools. The respondent explains that in classical medicine, they look for a way to solve a clinical problem (for example, to better rehabilitate a stroke patient) by searching among different contrasted alternatives. The use of virtual reality and simulation, as emerging alternatives, requires the neurologist to make an extra effort to understand what the tool is specifically for. Here, training is necessary so that its potential can be channelled to resolve or provide answers to the real questions posed by neurology.

On the one hand, he finds it very interesting to be able to simulate neurodegenerative diseases such as Alzheimer's, Parkinson's, Multiple Sclerosis, and other diseases that atrophy the brain. On the other hand, using these technologies, in principle, seems very complex to reproduce what we can detect directly in patients or to simulate the disease. It is also important to know how to carry out such a simulation and who to count on for support and back-up.

Institution interviewed No.5: They have no experience in the use of virtual simulators for experimentation, as they have not had the need to apply them in their studies.

Institution interviewee No.6: As explained above, they have experience in neural network simulation and software tools. They are not tools they use daily, but only occasionally. Perhaps in the future they will be fundamental tools, but not now.

Institution interviewed No.7: They do not use these techniques in their care practice.

FOOD FOR THOUGHT

- The groups interviewed have no experience using virtual simulators for neuroscience experimentation or clinical practice. However, they want to know which ones would be available to them and to learn how to use them.

- Efforts should be increased to raise awareness in the medical community of the usefulness of these technologies in care practice and research, as well as the capacity to solve real problems in neuroscience.

- If, despite still needing technological improvements and greater precision in today's virtual simulators, medical and research staff were to show greater flexibility and willingness to explore and test them, the technological offerings of their own research centres would most likely be strengthened.

3.8. Data curation resources: challenges, actions and recommendations

Institution interviewed No.1: The translational neuroscience group points out that since proteomics generates a lot of data, the biggest difficulty is being able to able to share and adapt the data for other researchers because there are no solutions developed for this. The neuroimaging group has similar difficulties, mainly due to the lack of standardisation when using different diagnostic and evaluation equipment. Unification and AI algorithms help to homogenise data.

Other difficulties have to do with their staff, who have a high care load, limiting the time they can devote to research. In addition, it is difficult to find repositories of data on rare pathologies, such as, for example, brain lesions as a result of COVID-19, at the imaging level. They also point out that there are very specific pathologies for which repositories of data do not exist.

Institution interviewed No.2: The main difficulty they have is related to the size of their research group, which is quite small, making their capacity for analysis rather limited. They use statistical analysis tools in their studies and have no restrictions for third-party programmes. She points out that the centres do not dedicate human resources especially for data curation, based on their experience in Spain and Germany.

In her first laboratory, a *data manager* was hired to manage the data. In Germany, they collaborated with a group of engineers, who supported them in this management, and when this support was sometimes a bit lacking, they contacted colleagues in the United States. Regarding public database initiatives on dementia, the interviewee reported that some of these had an associated statistical service for the data. However, other databases did not, which meant having to download the different datasets and needing the help of a technician to manage the information. The professionals with whom they collaborated were also not solely dedicated to this management, so again they had to rely on American colleagues.

Finally, she points out that having a data manager in each institution not only calls for sufficient financial resources, but also involves assessing possible work overload, given that she or he should in general support different medical specialties, while also making sure that assistance is provided to each area, with its own different specifications and particularities.

Institution interviewed No.3: They value the support of engineers and technicians, as well as the formation of multidisciplinary teams. Data processing requires people who know about, and can manage, information technology, who have a very clear idea of the clinical needs in relation to the data. He believes that the data analytics expert does not necessarily have to be a medical expert.

Institution interviewee No.4: The respondent says he has participated in meetings together with hospitals, AI companies, and corporate foundations where it has been really difficult to reach a consensus on what is health data, a *data lake*, or how to design data governance. Data that is handled in bioinformatics or genetics in practice requires large and secure databases. In clinical neurology, however, most clinical or research staff in small centres use their own databases for small studies. These data may or may not be shared, and they themselves perform the corresponding statistics.

Regarding the courses they give, they have difficulty in getting young neurologists to sign up to learn about programming and statistics. They offer everything from an introduction to AI to the practical use of the different algorithms. They introduce programming fundamentals, and organise annual activities. He says that there is still very little information on how neurologists and patients can benefit from generative AI tools, metaverse or virtual reality.

Institution interviewed No.5: They comment that, in general, there have not been many problems in using data management software. Other aspects have been more complex, e.g. refining the HeBA survey to determine the risk of developing Parkinson's disease. They are satisfied with the SAP resource for data management. Internally, each member of their team is assigned a project and manages the corresponding data. They also recognise that a *Data Manager* would be very useful in their hospital.

Interviewee Institution No.6: The main problem they face when sharing data is the lack of unification. In general, each researcher stores data differently, and although some projects are trying to implement a harmonised form of management, currently no standards to facilitate such common operations exist. For researchers, this management requires more labour-intensive processes; so while data sharing is certainly beneficial, there must be systems in place to facilitate these processes. The same is true for codes, which need to be documented, tested, structured. Even though there are codes, created by researchers, that are *clean*, commented and easy to share, others still lack these qualities. Finally, it should be noted that standardising and sharing data is not yet an obligation for institutions, but rather a recommendation.

The laboratory interviewed also took the initiative to motivate its members with dissemination of good practices for *good coding.* Researchers have been gradually trained in this area. Weekly workshops have been organised to discuss specialised manuals, and activities have been developed that have helped to reduce the current limitations.

Data curation is a time-consuming process, though, in realty, the priority for laboratories is to perform and complete experiments. Perhaps data curation is not properly managed because of lack of time or lack of knowledge how to do it.

Institution interviewed No.7: In relation to the federated *Stroke* data platform, cited beforehand, it is noted that there should always be, in the retrospective phase, a harmonisation and standardisation of data, as was done in HARMONICS. This means making the efforts to harmonise and structure data from different regions and from different electronic health records. In other projects, algorithms are not only developed with the retrospective data, but also validated, to avoid working with data that are discriminatory. Once the retrospective phase is over, they continue with proofs of concept stored in an internal repository, with data that are treated locally, and then sent to the federated network server. The data travels from hospital to hospital in a structure that *feeds into* a central node, with progressive enhancements to form such a federated data analysis structure.

The neurologist explains that these technologies are increasingly important because they make it possible to work with data from common neurological diseases, such as stroke, as well as with rare diseases. Each disease has its own casuistry, each being the focus of highly qualified centres or centres of reference for specific pathologies. Therefore, the type of federated data analysis desired, the commitment of the centres, confidentiality, and above all the precise definition of the specific aspects of the pathology to be investigated, must all be evaluated beforehand. The concept of *data lakes*, such as hospital encryption systems, is no longer commonly heard of, since it is rarely used. Rather, the trend in the European Union - including EBRAINS - is towards the use of federated data, i.e. data that do not leave the hospital but is analysed together with other data in a decentralised way. These systems reduce or make it impossible for hackers to identify or capture specific patient data.

The interviewee shares the view that the problem is not really the technological management of the data, because there are harmonised databases, practical and contracted ways of integrating them into hospital platforms (national or international), and agreements with private companies. What makes these processes more complex and time-consuming in practice is the need to go through an ethics committee in order to use them in an orderly and legal way. This is one of the main problems observed in this area, and one that discourages many researchers from developing digital projects.

FOOD FOR THOUGHT

- The teams interviewed generate a large amount of data, the most important difficulty being to organise and adapt it for use by other researchers. These tasks are time-consuming and increase the workload, thus giving priority to experimentation.

- Some research groups are small and do not have specialised staff to curate data. On the other hand, the figure of Data Manager and the integration of multidisciplinary teams is valued.

- In general, there is a need for staff training on data curation and data analysis in the centres.

- The use of databases constructed by European consortia, including federated ones, is becoming increasingly complex due to the necessary involvement of ethics committees.

3.9. Key ethical issues and protocols for data processing and data sharing

Institution interviewee No.1: A key role of the ethics committee is to ensure that its protocols are rigorously adhered to. It is common for the committee to catalogue almost all data as pseudonymised, which makes the decision to share data even more difficult. The level of ethics committee stringency between different Autonomous Communities or between different institutions can vary substantially.

Institution interviewed No.2: All data have to be anonymised in order to be shared. Although it is stored in pseudonymised form, it is shared anonymised with other researchers. There are no major restrictions from the centre to share data. She has never had any difficulties with ethics committees regarding the sharing of anonymised data. There is no specific protocol used in a general way in the hospital to manage the data, with each group having its own method for this. They have received some training in this respect.

Institution interviewee No.4: At the moment, they do not work with sensitive data. In terms of scientific dissemination and ethical issues, the interviewee highlights the debate on neuro-rights, as well as the issues to come on identity and the metaverse. One of his recent publications, on *technologies and the T-generation*, has caught the attention of the media. He points out that any study related to data has to be useful and properly focused. Data must be fully anonymised and the associated technological tools must ensure this security, including the ethical principles of justice and beneficence, as applied to society and to the individuals who give up such data. From his perspective, data sharing could follow the path of decentralisation, such as using *blockchain* or similar technologies, with the information thus being secure and distributed.

Institution interviewee No.5: In general, they follow all the protocols required in terms of data protection law. All projects are sent to the ethics committee or to the hospital's legal department. The bureaucratic burden associated with these processes stands out.

Institution interviewed No.6: They follow all protocols strictly, complying with data protection rules to anonymise data, and ensuring accessibility control. Ethical protocols are increasingly important and they have a specialist to help with any queries. He notes that their data management is not complex, as once data is anonymised, they are generally allowed to share it.

Institution interviewee No.7: Procedures are strict and lengthy. The ethics committee can often become a bottleneck in the process. The interviewee shares that many of the European projects that have been reviewed and approved get bogged down in such committees, a problem which, in his opinion, is a common factor in other European countries. The problem is internal, and there is an urgent need to find ways to speed up these processes. The databases that are currently being created suffer from the same situation, which makes it difficult for disruptive projects, innovative and advanced, delaying their deployment while taking up to a whole year to be approved.

FOOD FOR THOUGHT

- The ethics committee remains one of the most important authorities within hospitals. Among the Autonomous Communities, the flexibility of the committees seems to vary significantly.
- Pseudonymisation and anonymisation of data is a priority for all research groups. Each centre has a specific protocol for these processes.
- Neuro-rights are becoming increasingly important in the neuroscientific context. Thus, the development of neuro-technologies must include ethical principles of justice and beneficence, following the principles of bioethics.

3.10. Training for data curation and data management

Institution interviewed No 1: They have not received specific training in this respect at their centre, although they are considering including data management, *open science* and other specific challenges in this year's training plan.

Institution interviewed No. 2: They require specific training for these processes. In any case, these influence practice no matter if a laboratory in the biological sector normally works with less data for experimentation or one in the clinical sector that normally works with larger cohorts. There is also a lack of training on how a database should be created technically, as it is complicated to transfer this information to other technicians and specialists in the centre who help to keep the records pseudonymised. It is also interesting to have training so that the samples are correctly organised with codes.

Institution interviewed No.3: They value positively the training opportunities and resources that enable them to understand the use of the EBRAINS project technologies, as well as the opportunity to receive practical and prior information on how to use its tools.

Institution interviewed No.5: No specific training in this area. The team has adequate training in statistics.

Institution interviewee No.6: There is no training available for data curation, but it would be useful to have it. It would also help if platforms such as EBRAINS that promote data sharing could propose a formula for sharing data, with a format and standards for its use. It is important to teach users how to use these formats, with didactic videos so that it can be shared among researchers.

Institution interviewee No.7: They usually try to train in applied biostatistics, although there are no predetermined training paths. At the doctoral level, statistical training or specialised skills are usually required. Normally, each professional finds a way to get trained in data management individually; a common and more homogeneous orientation among professionals would therefore be useful.

FOOD FOR THOUGHT

- There is a great need and interest in receiving training in data curation and data management, adapted to the lines of research, as well as in database creation.
- In particular, the interviewees would like to find trainings, didactic videos on EBRAINS, adapted to the requirements of the groups.

3.11. On medical technologies and innovations to be incorporated in clinical research

Institution interviewed No.1: The translational neuroscience group is interested in the use of tools that combine molecular data with imaging, especially the combination of both. The electronic medical record available in its Autonomous Community is very extensive, but its extraction is done by hand and can only be performed by a doctor.

The transfer and innovation team points out that this platform is a pioneer in mass data analysis: it applies AI in a system that responds to requests from researchers according to the variables to be investigated. It is centralised, complies with the current legal frameworks, and provides security in data storage and management. However, limitations exist related to accessing this tool. Some studies have used other authorised systems to massively extract data, but these are very limited cases. Finally, the neuroimaging group interviewed would like to be able to find repositories of databases of rare pathologies.

Institution interviewee No.2: In relation to brain simulation tools, it would be useful to study EEG-based neural networks, as well as to access initiatives with databases on dementia with Lewy bodies - which are limited - or reliable simulations of this disease. Among other technologies not yet created, the interviewee would love, during his medical routine, for AI to elaborate clinical courses, to extract information and place it in the database, to detect biases in studies, plus identify possible errors and mistakes in the process.

Institution interviewed No.3: As an innovative group that promotes the inclusion and creation of technologies aimed at research and clinical care in neurological diseases, it has incorporated a specific structure in the service to promote these activities in a transversal manner. For example, they have developed AI models for clinical management, segmentation and classification of ultrasound images, optical capture and automatic analysis of movement for the objective quantification of functional deficit, development of brain connectivity studies, etc.

Institution interviewee No.4: They are looking into the possibilities of generative AI, although they are not yet sure about its applications in medicine. During the current year they are planning two activities in their committee on the topic, inviting two specialists to share ideas on the potential applications of generative AI in health. He shares how difficult it has been to find professionals who can make inferences about how these applications could contribute to the medical field. They want to know how they can use these technologies in clinical practice and in neurology.

On the other hand, he makes reference to a published paper on generative AI in which they applied fMRI scanning to read people's thoughts. As described in the paper, scientists analysed the BOLD signal in people who were reading a text and were able to achieve an 80% semantic approximation about what people were reading. He sees these kinds of technologies as key to general artificial intelligence and the exploration of uncharted territory in the field of neurotechnologies.

Institution interviewee No.5: There is interest in projects where, for example, several radiological markers need to be identified in patients and AI applied to determine those at risk of developing a disease.

Institution interviewee No.6: They find magnetoencephalography (MEG) a very powerful and interesting technique for the project they are developing. They would like to find intracranial records of patients with epilepsy and electrocortigraphy (ECoG) data but they would need a large volume of data because records are usually limited.

Institution interviewed No.7: There is interest in AI techniques with Machine Learning (ML), predictive algorithms based on ML techniques integrating different data, both for clinical and for biological, mental, neuroimaging, biomarkers, in order to develop predictive models and clinical intervention.

SUMMARY OF INNOVATIONS RELEVANT TO CLINICAL RESEARCH

- Tools combining molecular data analysis with imaging.
- Automatic extraction of data from the healthcare system for use in clinical cases.
- Technologies that detect biases and errors in research studies.
- Applications of generative AI in neurology.
- Identification of radiological markers in patients and application of AI to determine the risk of developing diseases.

- Increased accessibility to techniques such as Magnetoencephalography (MEG), intracranial electroencephalogram (iEEG) and electrocorticography (EcOG) recordings.
- ML-based AI techniques to develop predictive and intervention models.

3.12. Future areas of interest in health and neurotechnology

Institution interviewed No.1: The translational neuroscience group opines that it is necessary to achieve true personalised medicine in practice, something that all groups currently claim to do, but no group has really achieved. As personalisation is fundamentally data-driven, powerful tools must be used to access this data. For its part, the neuroimaging group highlights the need for access to image data through *data lakes or* repositories large and meaningful enough to contribute to current research. Another important aspect is the role of clinical decision support systems in general, and it is essential that they are integrated into clinicians' workflows to access data easily and quickly.

Institution interviewed No.2: With interest in biomarkers, they consider it important that both those under study, and currently advancing, and those currently being developed, are soon implemented in clinical practice. With this, predictive models of neurodegenerative diseases can be developed, which can be shown to patients to help in their prevention and/or recovery.

Institution interviewed No.3: The experts state that, in addition to the above-mentioned initiatives to move towards precision medicine, interdisciplinary collaboration is crucial to address medical needs. For their part, they collaborate in the creation of the National Neurotechnology Centre along the lines mentioned above.

Institution interviewee No.4: In addition to the aforementioned generative AI, there are young neurologists who, instead of using regression models, are applying different ML models with their own datasets to assess which would achieve better levels of accuracy. He shares that the training provided by his organisation has begun to bear fruit, and in hospitals in Madrid, for example, research teams are being created that are using AI, gradually displacing classical statistics.

The team the neurologist works in studies the evolution of research groups applying or developing new technologies. The number of groups is increasing year by year, and awards are given to those groups that manage to implement the technology. Based on this information, together with the congresses held by the organisation, it is possible to detect trends ranging from AI applied to neuroprosthetics, including brain-computer interfaces, sensors, digital biomarkers, genomic analysis using advanced bioinformatics tools, the application of sensors in movement disorders for Parkinson's disease, biomarkers of Alzheimer's disease for the prediction of dementia, and the automation of diagnoses, among others. He explains that these advances will contribute to true precision medicine for the early diagnosis of certain diseases.

Institution interviewee No.5: Their area of research focuses on neurodegenerative diseases and, therefore, they focus on the earliest and preclinical stages. She explains that this has led them to contact specialists and multidisciplinary groups, with experience in different fields (proteomics, omics, etc.) and to work with people in basic or biological research, far from the clinical context.

Interviewee Institution No.6: The clinical sector of this institution shows interest in working with basic research groups that are developing new technologies. They have created a tool with which users can access laboratory data from a tablet or mobile phone, making it easier for patients to use at home to enable more personalised research.

Institution interviewed No.7: The interviewees point out that the integration of multidisciplinary teams and the classic clinic is not enough to deal with certain clinical problems. Clinicians should be trained in other areas, in order to obtain a more general vision of patients' priorities: it is likewise important to integrate professionals with other profiles to promote training in new techniques and integrate new technologies. Despite the large amount of data that can be extracted in clinical practice, it is more important to achieve a certain therapeutic outcome and that, in parallel, the intensive use of data does not end up feeding sterile exercises. There is a need to promote national and European funding streams, create cooperative networks and invest more in training to coordinate groups and obtain better clinical results.

FUTURE NEEDS AND AREAS OF INTEREST

- Facilitate access to neuroimaging data with comprehensive and useful repositories.
- Integrate clinical decision support systems into clinicians' workflows that work in a user-friendly way.
- Implement the use of biomarkers in clinical practice to predict neurodegenerative diseases.
- Further explore the scope of AI and its variants (such as generative AI) in neuroscience.
- Increase collaborative work with multidisciplinary teams, including basic and clinical research.
- Foster innovation and the use of AI in the field of neuroprosthetics, brain-computer interfaces, digital biomarkers, genomic analysis, sensors for movement disorders, and the automation of diagnostics.
- Support technological developments for patients that can be taken home, facilitating the collection of personalised data.

Institution interviewed No.1: The translational neuroscience group has proteomic and *raw data* databases that they could share. They generate molecular data with neuropsychological data and clinical data and are currently integrating data from different populations (in Spain, Portugal, Colombia, Mexico). Their goal is to manage these data to stratify them by country, region, age, etc. The neuroimaging group serves a wide population area as the only centre within the hospital; therefore, they have a large amount of data to share from multiple pathologies. Interested, too, in finding radiomic variables derived from images, they are keen and willing to share study data if they are provided with the infrastructures and services to do so.

Institution interviewed No.2: They are interested in biomarker databases, accessing the sample to be able to validate the results of the cohort, which is useful to find equivalent data to test hypotheses and analyse the data. They cite, as an example, amyloid analysis in cerebrospinal fluid. Other resources of interest are brain genetic data, which are not available. In terms of willingness to share, she explains that the database they are building is small and needs to be tested to verify accuracy before using other data. He can share general data on dementia, although in the area they are currently studying - dementia with Lewy bodies - the availability of data is still very limited and they need to do a first analysis before moving ahead.

Institution interviewee No.3: Proteomics databases could be useful for the scientific community, along with kinematic data, among other data. They highlight the need to create a solid network in which groups interact and truly get to know each other, as happens in other international projects. They are open to support researchers in computational neuroscience and any other area of neuroscience.

Institution interviewee No.4: Quite useful it finds to have complete datasets and series of disease studies or hospital-based trials. For example, if it were Parkinson's disease and you want to model it, you would need clinical data, neuroimaging data, digital biomarker data, etc., not just cross-sectional, but over time. In the case of epilepsy, it would be important to have complete datasets, e.g. electroencephalograms (EEGs), beyond the graphs.

Other types of data of interest to be obtained are those related to neurological diseases, where the studies follow international guidelines and clinical trials are determined by target variables. The relevance of virtual clinical trials, the development of in silico medicine, and the creation of digital twins is underscored in the interview. In addition, valuable information to start creating virtual patient models can be found not only in hospitals but also in the pharmaceutical industry.

For example, if information is available on the evolution of patients who have participated in recent migraine clinical trials and a placebo group, a virtual placebo group could be analysed, mathematised and modelled against which other drugs could then be compared. The pharmaceutical industry would then have an important role to play in creating consortia for the creation of virtual groups for different diseases for which data are historically available.

Institution interviewee No.5: Any researcher can, in principle, request data from the projects they are part of. If they need data related to Parkinson's disease, for example, they can contact the project coordinator, with this being approved (or not) beforehand by the hospital's ethics committee. Similarly, another avenue for data sharing comes from a request for samples from a pharmaceutical company. Apart, interviewees comment that they would be interested in finding databases of registries of patients with rare diseases or movement disorders, such as dystonia and atypical Parkinsonisms, given the fact these data are difficult to find.

Institution interviewee No.6: They would be interested in getting access and usage of large databases of diseases such as schizophrenia, as well as standardised behavioural tasks of these patients. Data that does not only include neuropsychological tests, but also the tasks designed by basic researchers, which are aimed at studying brain mechanisms and that can be quantified with their measures. They are also willing to share their data (every time they publish they share it) even before they publish (following Open Science practices like peer review, pre-registration in repositories, etc..). They recognise that they do not always invest enough effort to make these data available to other researchers, acknowledging that a mere link to the publication is not sufficient. They recognise that some data may not always be curated in the best way, but they are willing to improve to achieve the highest quality curation and dissemination as possible given their resources.

Finally, they mention as they carry out tests on patients, in the laboratory they are trying to adapt these tests in animal models, particularly in the mouse. By this, they seek to obtain a similar behaviour or response, in order to develop a model of the disease. They need to design tasks that can therefore be performed by both humans and animals, and basic science researchers need to be involved in these kinds of experiments.

Institution interviewed No.7: They show interest in biological sample data in patients with mechanical thrombectomy.

FOOD FOR THOUGHT

- The groups interviewed not only have a large amount and variety of data, but show a willingness to share it: from proteomic and molecular data to neuroimaging and neurodegenerative disease data.

- The data needed by most groups is very specific, given the type of disease or clinical assessments required, so in many cases databases are not available or are very small.

- There is significant interest in accessing samples and data from other cohorts to validate one's own results.

- The search for and identification of biomarkers is becoming increasingly common for the diagnosis and therapy of neurological diseases.

3.14. Opinion on the activity of Spanish companies in their line of research: collaborations and challenges

Institution interviewed No.1: The neuroradiology area is often very active at the corporate level. They have collaborated with companies on projects related to neurological diseases and the COVID-19 virus.

Institution interviewed No.2: They collaborate with the Polytechnic University of Madrid with the bioengineering and telemedicine group, in neural network studies. They also collaborate with international groups, such as Christian Haas' laboratory in Germany, and with initiatives such as GAAIN. He has not tried to contact companies in his field.

Institution interviewed No.3: They call attention to the possibility that there are technology companies that do not have a grounded understanding of the potential of clinical and research teams and all that they can offer. Many research groups, too, need support to overcome the transition from research to innovation. Even having a product at an acceptable technological level, and tested in real environments, there comes a time when development cannot continue - either due to lack of funding or lack of vision in the form of market research. These are aspects that are not generally mastered by researchers.

As medical doctors, the time available for research is limited. They need funding to hire staff to carry out the projects they develop. What's more, the administrative part is a huge management effort on top of everything else.

Institution interviewed No.4: The interviewee explains that increasingly startups are focusing on research and development of technology applicable to certain biological problems, while the need for even more companies to enter the arena continues. In the 5th edition of the institution's new technologies course, 2 or 3 Spanish companies with commercialised developments in AI, brain-computer interface sensors, etc. will take part. Much talent exists in Spain: more and more companies are up and coming in this field. His organisation wants to position itself in this ecosystem, including at the earliest business stages of initiatives.

Institution interviewee No.5: They have had contacts with some pharmaceutical companies and laboratories, as they need samples from specific patients for the development of treatments.

Institution interviewee No.6: They say that they feel that the number of Spanish companies working in these areas is still limited and that there should be more of them. In the past they have approached some companies, although they have not been able to collaborate closely. In their current project, they developed a mouse training system, and are working on ways to bring these innovations to the market.

Institution interviewed No.7: They agree that it is necessary to foster a culture of innovation with companies, to promote cooperation, act in forums more visible to industry, and to participate in calls for proposals that combine these visions.

They suggest that there is a disconnection between companies and the clinical sector in Spain, and that there is a need for tools to promote the creation of start-ups and the generation of patents. Keen to establish collaborations with new companies, they are open to innovative proposals with the private sector, as this facilitates technology transfer and the development of new research projects.

FOOD FOR THOUGHT

- Collaboration with companies has been most effective with pharmaceutical companies, and to a lesser extent with start-ups, spin-offs and the like.
- In general, the number of Spanish companies operating in the field of neuroscience and neurotechnology is perceived as small.
- There is great interest from research groups in receiving support in terms of innovation, market research and technology exploitation, complementing the work of their transfer offices.

3.14. Limitations of the study

There are several limitations to this study. On the one hand, the participants surveyed constitute a small part of the total number of research institutes and hospitals belonging to the National Health System. On the other hand, as noted above, the opinions expressed in the interviews are exclusive to the participants and do not necessarily represent the position of their affiliated centres. Finally, we recognise that there may be a regional bias since 33.33% of the responses came from researchers in Catalonia, 25.93% in Madrid, 14.81% in Andalusia and 25.93% in the rest of Spain.

While such methodological limitations may have influenced our analysis, we believe that the information provided, far from being generalisable, should be useful in generating constructive debates on the management and use of clinical data in neuroscience.

3.15 Trends

The literature review, the exploration of specialised databases of scientific articles, intellectual property registrations, as well as the reflections obtained in the interviews and questionnaires, allow us to outline some trends in neuroscience, artificial intelligence, new technologies and digital health. However, before exploring them, it is important to share some preliminary aspects such as, for example, the countries where the majority of scientific publications and industrial and intellectual property registrations in these branches of knowledge are concentrated.

We searched Scopus, one of the largest bibliographic databases of scientific publications, for keywords such as artificial intelligence, neuroscience and data analysis (we use *data analytics* instead of *data analysis* because of the semantic implications and differences in scope, the first term being appropriate, as it is a broad approach to data mining and analysis for decision making). We limited the results to publications in English only, excluding areas outside the scope of this report, and worked with results between 2017 and 2023.

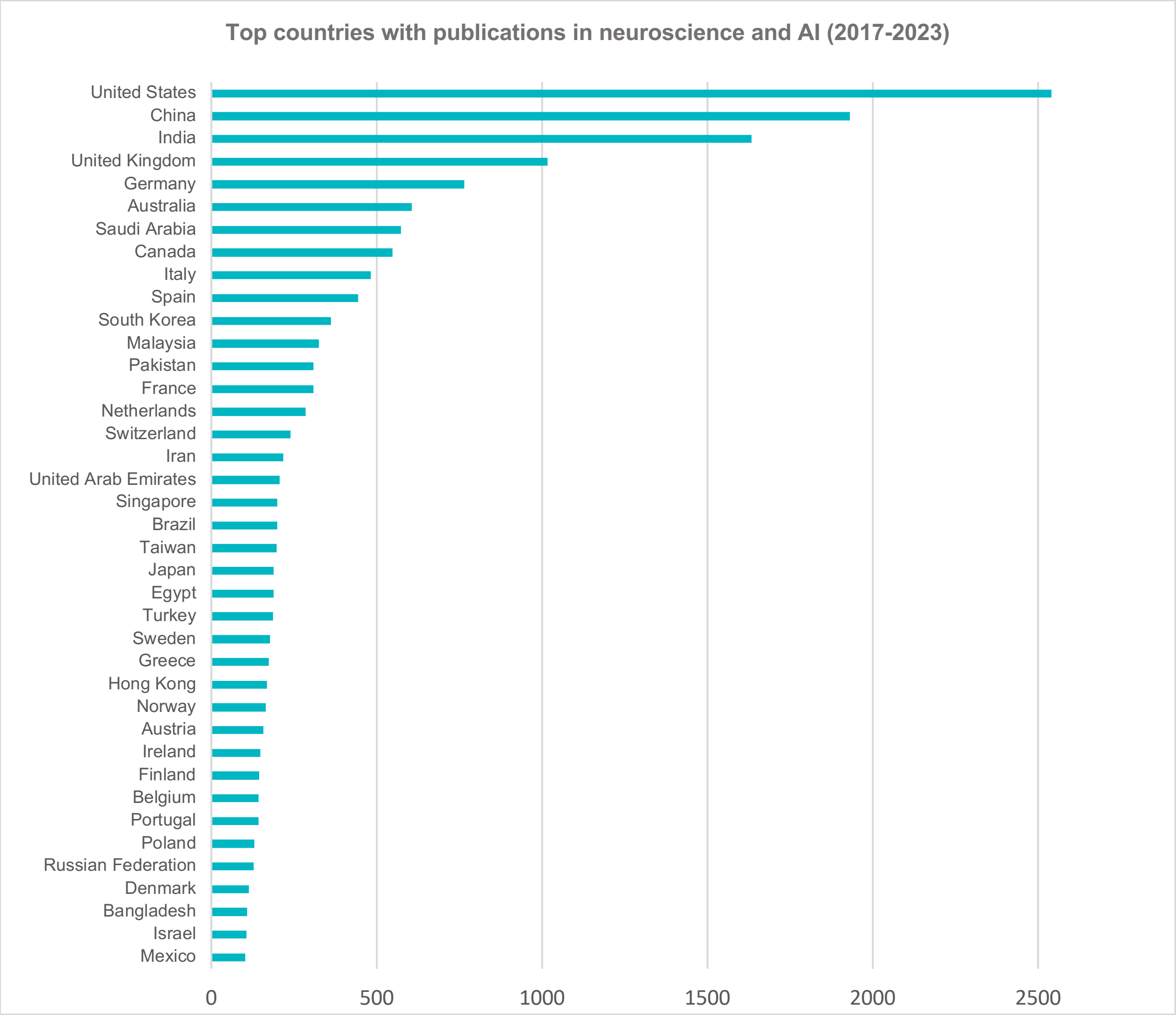


Figure 24. Authors' own elaboration. Data: Scopus

Figure 25: Among the top 10 countries with the most publications, the United States remains the leader, followed by the Asian giants China and India. In the United States, the major areas encompassing these publications are linked to *Computer Science (*26.50%), *Medicine* (16.40%), *Engineering* (11.5%), among others. In China, they are concentrated in *Computer Science (*32.80%), followed by *Engineering* (15.80%), *Mathematics* (11.9%), and to a lesser extent *Neuroscience* (8.8%). In India, *Computer Science* accounts for 37.70%, *Engineering* (19.60%), *Mathematics* (9.3%) and Medicine (7%). **This reflects the multidisciplinary nature behind these studies.**

The United Kingdom is in fourth position, followed by Germany, Australia, Saudi Arabia and Canada. Italy and Spain are in the ninth and tenth positions. The publications of these countries are concentrated in the same areas described above, computer science, engineering and medicine. In Italy, most of them are in the area of *Computer Science (*28.30%), followed by *Medicine* (12.70%) and *Engineering* (11.40%). In Spain the distribution is similar, *Computer Science* (30.10%), *Engineering* (14.30%), *Medicine* (8.90%) and *Neuroscience* with only 4.7%. As seen throughout the report, this highlights the efforts that leading countries, and Spain, in particular, must make to bring health specialties and research even closer to Computer Science.

Publications show an upward trend, with an average annual growth rate of **54.71%** between 2016-2022.

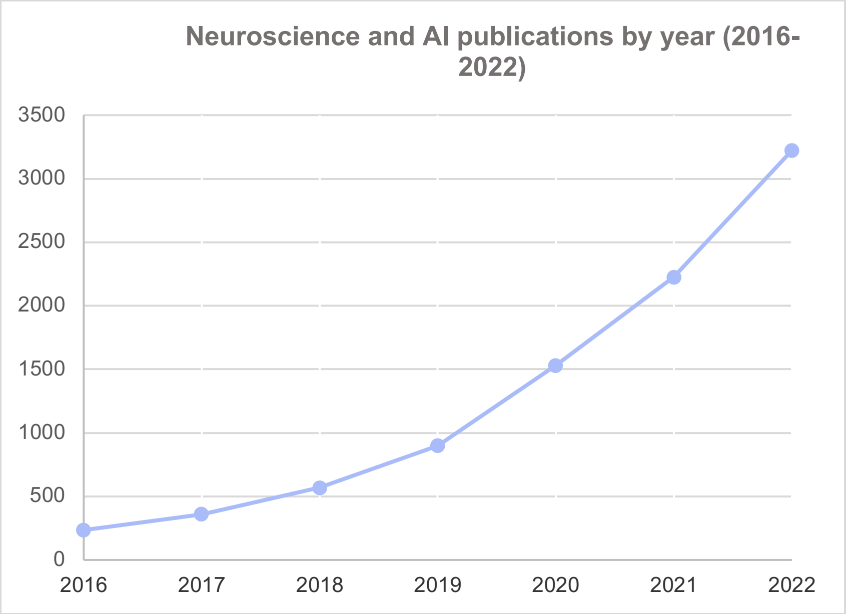
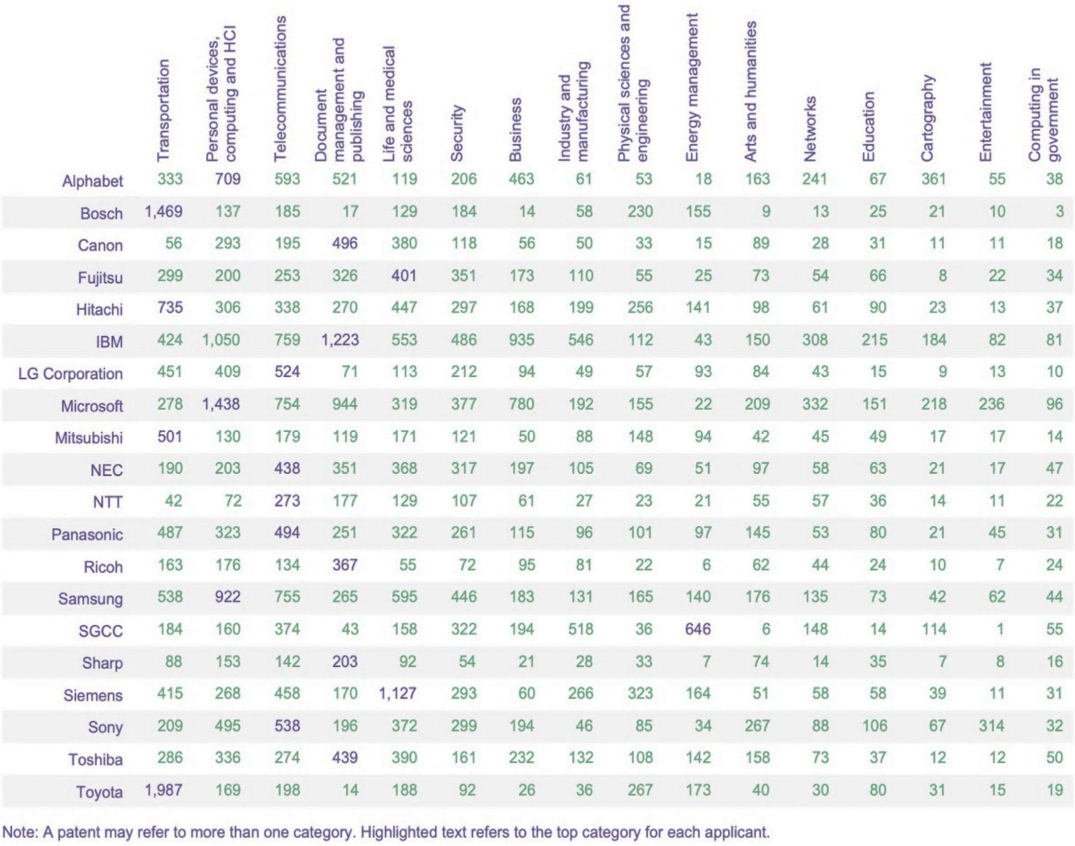


Figure 25. Authors' own elaboration. Data: Scopus

On the other hand, **in terms of intellectual and industrial property,** we consulted specialised international patent databases such as Patentscope of the World Intellectual Property Organization (WIPO). We refined the searches to focus on AI and neuroscience registrations considering the period 2016-2022. We also consulted the international classification system or IPC (SPTO, 2023), a classification mechanism that homogenises patent documents, streamlines search processes and facilitates greater control of related innovations.

A published report on trends in AI (WIPO, 2019) reports that the number of patents applied for in these techniques has reached 340,000 since 1950 and, although the report does not focus exclusively on neuroscience, it shows an interesting range across all fields, and delves into specific algorithms and applicants by sector (see Figure 27 below). This suggests that any search for AI patents in a particular area, in this case neuroscience, neurology or another medical sector, could be limited to a much lower number.



	Transportation	Personal devices, computing and HCI	Telecommunications	Document management and publishing	Life and medical sciences	Security	Business	Industry and manufacturing	Physical sciences and engineering	Energy management	Arts and humanities	Networks	Education	Cartography	Entertainment	Computing in government
Alphabet	333	709	593	521	119	206	463	61	53	18	163	241	67	361	55	38
Bosch	1,469	137	185	17	129	184	14	58	230	155	9	13	25	21	10	3
Canon	56	293	195	496	380	118	56	50	33	15	89	28	31	11	11	18
Fujitsu	299	200	253	326	401	351	173	110	55	25	73	54	66	8	22	34
Hitachi	735	306	338	270	447	297	168	199	256	141	98	61	90	23	13	37
IBM	424	1,050	759	1,223	553	486	935	546	112	43	150	308	215	184	82	81
LG Corporation	451	409	524	71	113	212	94	49	57	93	84	43	15	9	13	10
Microsoft	278	1,438	754	944	319	377	780	192	155	22	209	332	151	218	236	96
Mitsubishi	501	130	179	119	171	121	50	88	148	94	42	45	49	17	17	14
NEC	190	203	438	351	368	317	197	105	69	51	97	58	63	21	17	47
NTT	42	72	273	177	129	107	61	27	23	21	55	57	36	14	11	22
Panasonic	487	323	494	251	322	261	115	96	101	97	145	53	80	21	45	31
Ricoh	163	176	134	367	55	72	95	81	22	6	62	44	24	10	7	24
Samsung	538	922	755	265	595	446	183	131	165	140	176	135	73	42	62	44
SGCC	184	160	374	43	158	322	194	518	36	646	6	148	14	114	1	55
Sharp	88	153	142	203	92	54	21	28	33	7	74	14	35	7	8	16
Siemens	415	268	458	170	1,127	293	60	266	323	164	51	58	58	39	11	31
Sony	209	495	538	196	372	299	194	46	85	34	267	88	106	67	314	32
Toshiba	286	336	274	439	390	161	232	132	108	142	158	73	37	12	12	50
Toyota	1,987	169	198	14	188	92	26	36	267	173	40	30	80	31	15	19

Note: A patent may refer to more than one category. Highlighted text refers to the top category for each applicant.

Figure 26. Top patent applicants by IA application field. Source: Trends in AI, WIPO (2019).

AI has a wide range of applications and therefore is quite complex to size it into specific classifications. However, the IPC system has managed to adapt its codes for AI algorithms created for the life and biomedical sciences. The topics consulted together with the required classification approach 14,108 records, the main ones are shown in Figure 28.

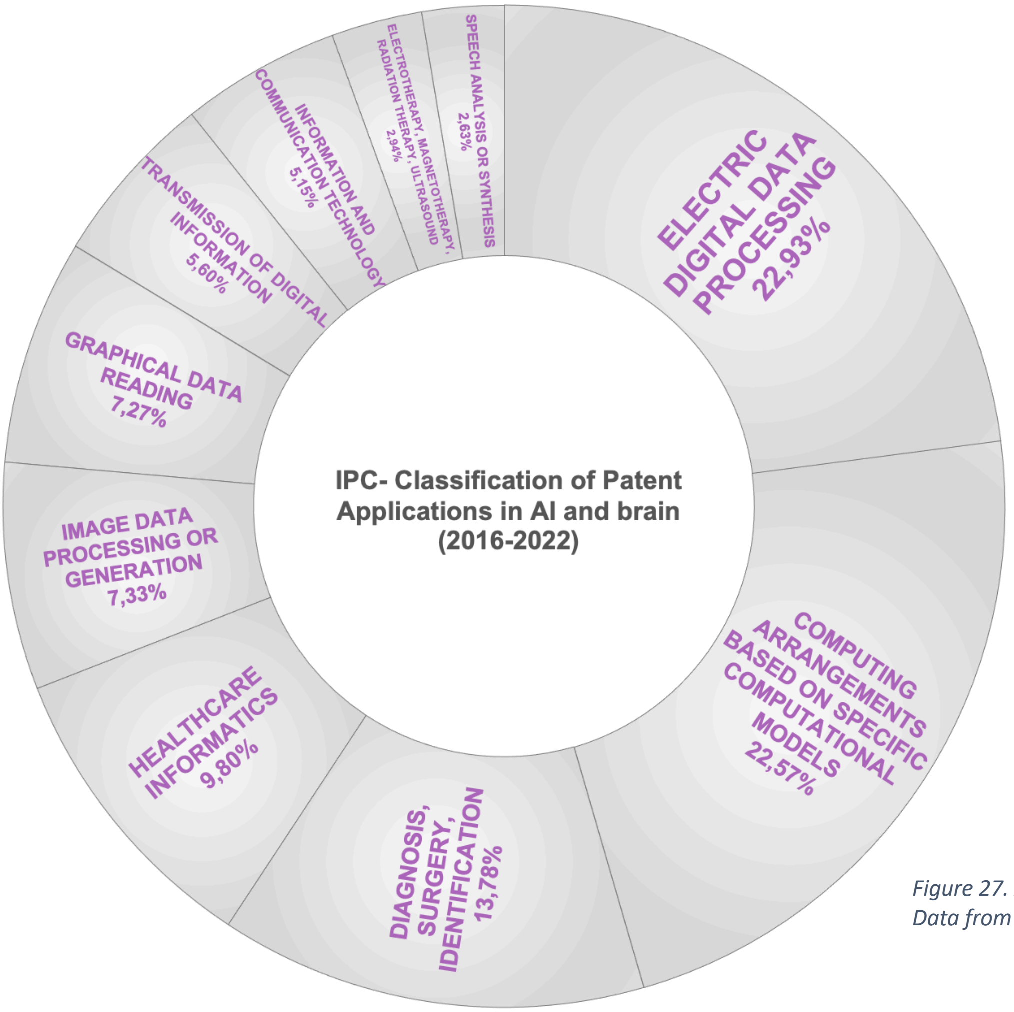


Figure 27. Authors' own elaboration. Data from Patentscope

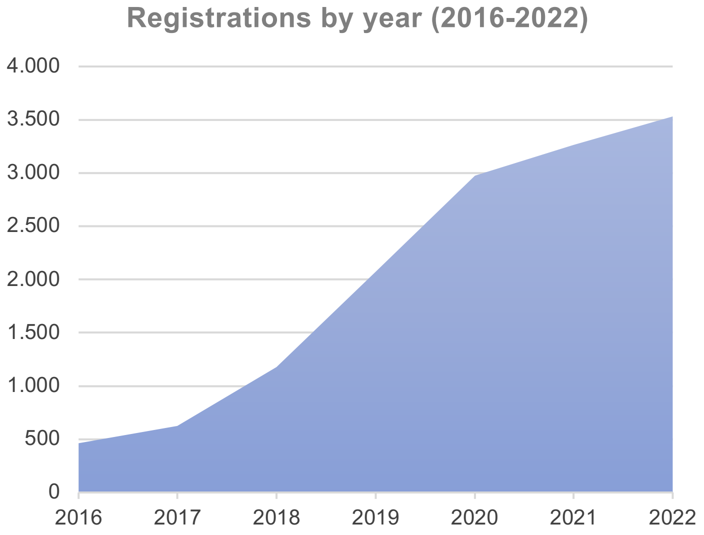


Figure 28. Authors' own elaboration. Data from Patentscope (2023)

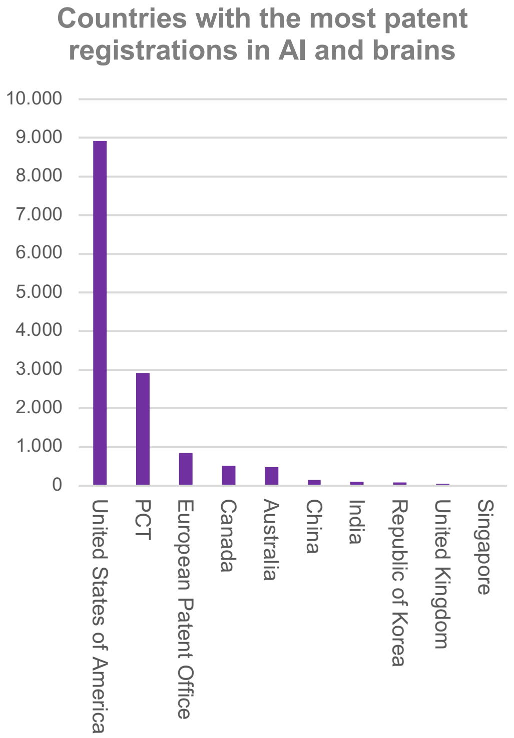


Figure 29. Authors' own elaboration. Data from Patenstcope (2023)

The growing trend of AI and brain patent filings is shown in Figure 29. In addition, applications through the Patent Cooperation Treaty (PCT) are becoming increasingly popular because it allows, once the patent is granted, protection in different countries. It also simplifies the entire process into a single application, instead of having to file for a patent in each country, and provides other benefits, too, such as more time for applicants to choose in which countries they wish to protect their invention (SPTO, 2023). Figure 31 shows the representative companies applying for patents in AI and the brain:

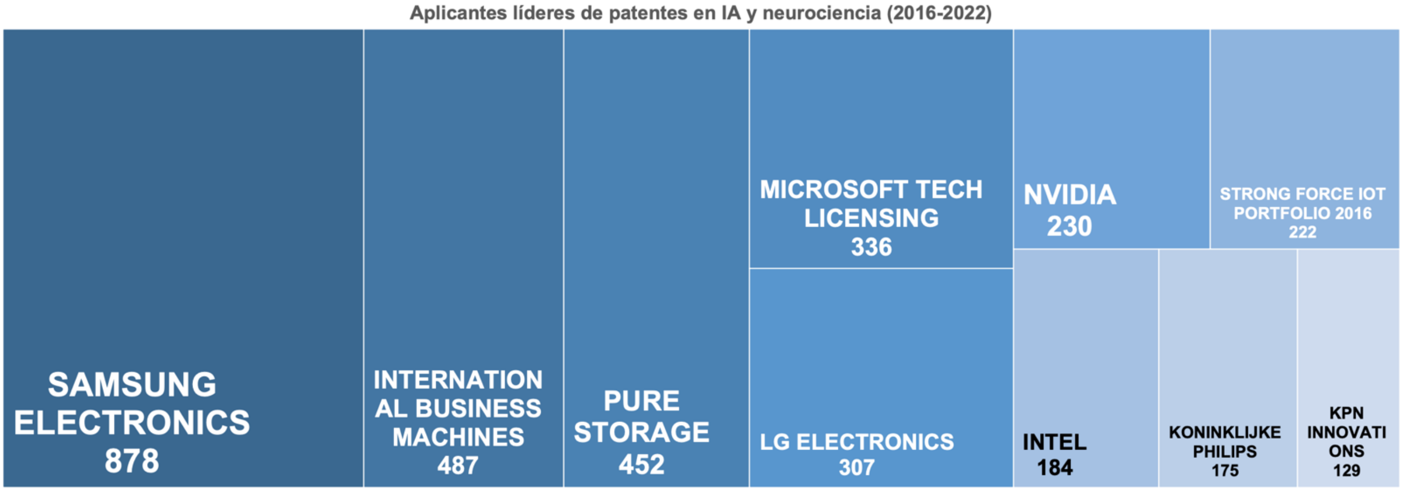


Figure 30. Own elaboration. Data from Patentscope (2023)

As the searches became more specified, interesting results came back from long-established companies in very diverse fields, such as the South Korean car manufacturer Hyundai and its subsidiary Kia Motors, who developed a programme that allows the reconstruction of images from the brain records of drivers by applying AI techniques, specifically with generative adversarial neural networks (GANs), in the year 2021. This technique is a variation of Deep Learning (DL) that gained popularity in text generation applications, such as Open-AI, shortly thereafter.

On *diagnostics, surgery and identification* - including analysis of biological material - the companies that applied for the most patents were: Phillips, ResMed, Samsung, Siemens Healthcare, International Business Machines, Medtronic, Biosense Webster, Brain Sentinel, Ablacon.

In the *health informatics sector,* in addition to some mentioned above, KPN Innovations, Tempus Labs, Rom Tech, Healthpointe Solutions, among others, stand out.

In the area of *electrotherapy, magnetotherapy, radiation therapy and ultrasound therapy*, examples of innovation leaders include companies such as Boston Scientific, Cochlear, Medtronic, Novocure, and Advanced Neuromodulation Systems.

3.15.1. Neurocomputing Platforms

Despite the existence of platforms for sharing and analysing neuroscience data in different modalities, other innovative projects are offering support in the standardisation, processing and analysis of datasets. As seen throughout this report, much of the data obtained by research institutes is generally processed according to the institute's own standards, both within and outside Spain. The value of neurocomputing platforms' proposals will continue to increase in order to obtain better performance from the data, while preserving privacy.

Examples of advanced neuroinformatics platforms include the HBP/EBRAINS *Medical Informatics Platform (MIP)*, which is based on federated technology for data exploration and AI, and involves more than 30 hospitals worldwide. The data provided by specialists must be pre-managed through specific procedures and applications provided by the project. Another interesting EBRAINS platform, called *HealthDataCloud* (HDC), provides a federated data research ecosystem with health data support and training tools.

Other powerful international initiatives, such as the Ontario Brain Institute's (OBI) *Brain-CODE*, aim to encourage researchers to share data from different neurological modalities and diseases, helping to efficiently manage these data in a way that follows FAIR principles, and involving 20 institutes. In Canada, another well-known initiative is the Canadian Open Neuroscience Platform (CONP) for data sharing (Poline et al., 2021). In Japan, the International Neuroinformatics Coordinating Facility node has contributed to the creation of 16 platforms in neuroscience subspecialties (Japan Node of the INCF, 2018).

3.15.2 Generative AI and AI-driven initiatives

Generative AI "uses a machine learning model to learn patterns and relationships from a dataset of content created by people" (Google Cloud, 2023). These are tools that allow new content to be created from previous inputs or data, which can be of various kinds. Recent advances are increasingly showing the great potential of generative AI in neuroscience, beyond the current offerings of pioneering companies such as OpenAI, Meta, Google, among others. For example, in the US, researchers at the University of Texas demonstrated the reconstruction of perceived and imagined stimuli with continuous natural language in three participants, using non-invasive fMRI recordings, and analysing the data with a generative neural network linguistic model and algorithms used in NLP (Tang et al., 2023). These applications will help expand our knowledge in neurocomputing and the study of functional brain relationships through BCIs.

Studies such as those by (Gong et al., 2023) assess the neuroimaging and computational benefits of neural networks, in particular how large amounts of data can be analysed from brain images. The molecular field is also engaged in these kinds of innovations; for instance, evaluating predictive models on different datasets for drug design (Langevin et al., 2023). Given the quantity, variety and complexity of the data generated in medical evaluations, different types of generative AI will provide useful alternatives to advance the study of neurological diseases. Another open-source application created in 2020 by NVIDIA offers a workspace for the application of DL including AI models like MONAI (MONAI, 2023).

3.15.3 Interfaces for data standardisation and analysis in neuroscience

Although there is still a long way to go in this field, the advances in applications and interfaces for data standardisation already mean less workload for researchers. Interfaces such as the *Neuroscience Data Interface (NDI)* make these advances in neuroscience experimentation a reality, being open-source and offering, among other benefits, the combination of data from different experiments (García Murillo et al., 2022). Software aimed at standardising data in neurophysiology, such as Neurodata Without Borders (NWB), promoted by the Brain Initiative Alliance, allows for sharing, storing and analysing data (Brain Initiative Alliance, 2023). In EBRAINS, tools such as *Knowledge Graph,* in conjunction with *OpenMINDS,* are also evolving rapidly, making it easier for users to find and share standardised datasets, thanks to metadata processing in the acquisition process. In parallel, KnowledgeSpace is envisioned as a neuroscience data-driven encyclopaedia connecting ontology and datasets, as well as an access point for data models from projects such as HBP, BRAIN, Japan Brain/MINDS, the aforementioned platform, CONP, among others (KnowledgeSpace, 2023).

3.15.4 Increasing regulations and ethical frameworks

In 2021, the European Commission proposed a Regulation on Artificial Intelligence to facilitate its adoption and, at the same time, address the risks of using these techniques, with all the ethical implications they entail. The proposed Regulation will set minimum requirements while keeping to the aim of not restricting technological development (European Commission, 2021). Final approval is expected by the end of 2023, which will obviously mean a change for any AI solution that may have an impact on human rights and safety. Such regulations will raise future concerns and will, in turn, also influence the design of strategies and policies to ensure that research groups do not fall behind in innovation.

Several organisations have been working in recent years on the ethical framework and regulatory aspects of neurotechnologies. In particular, the contributions of neurobiologist Rafael Yuste on neuro-rights in 2017, based on the BRAIN project in the United States, are widely known (Columbia Magazine, 2022). Such efforts open up challenging paths for addressing the ethical principles of technologies that influence our brain capacities. International bodies such as UNESCO have issued recommendations on ethical issues, calling on countries to regulate human-robot interactions as a way to avoid "*the possibility of using AI to manipulate and abuse human cognitive biases*" (UNESCO, 2023), among other issues.

Conclusions

Technological progress is having a decisive influence on the way national health systems manage clinical and research data. The increasing volume and ubiquity of data means that systems must assume and be prepared for greater and more significant IT security and confidentiality risks. However, the power of new data technologies also opens the door to new opportunities.

One conclusion that comes out from this report is that technological change in the area of acquisition, management and use of clinical brain data can be addressed more effectively and have a greater impact on neuroscience research if it is accompanied by personalised and regular training of scientific and medical staff, especially in relation to the use of sensitive data. Basic data management training should also reach a large part of the healthcare staff. The availability of human resources specialised in data management would in fact reduce the workload of researchers and favour the exchange of expertise. The use of the European EBRAINS platform for data management, in this sense, is an appropriate example of how collaborative data use tools can be widely supported by powerful and interactive training and learning modules.

Another important finding is the need to promote innovation in neuroscience by creating collaborative workspaces and further strengthening the existing contact between hospitals, universities and technology transfer offices. In this way, multidisciplinary projects can be generated to help research groups bring their data-driven innovations to the market, assessing the real market possibilities with the help of their respective transfer and exploitation offices.

Our analysis has also provided insight into the real dimension of the work of ethics committees in health research institutions. Their contribution is vital in the management of neuroscientific information. The contribution of the committees to debates related to the use of AI in neuroscience and clinical practice is particularly relevant. Articulated with flexible mechanisms and procedures that do not slow down innovation processes, the committees must ensure the privacy of patient data, responsible research and the proper use of clinical information.

Finally, the study notes the importance of identifying and strengthening links between health research institutes and the groups developing the technologies and services offered by the EBRAINS infrastructure, so as to accelerate progress both in data management and in the use of simulation and virtualisation tools for neuroscience research.

References

BOE-A-2021-3132 (27 February 2021). Agreement between the Universidad Politécnica de Madrid and the Instituto de Salud Carlos III, O.A., M.P., to support the innovative development of the Human Brain Project (HBP) https://www.boe.es/boe/dias/2021/02/27/pdfs/BOE-A-2021-3132.pdf

BOE-A-2023-5356 (28 February 2023). Addendum amending the agreement between the Universidad Politécnica de Madrid and the Instituto de Salud Carlos III O.A., M.P. to support the innovative development of the Human Brain Project (HBP). https://www.boe.es/boe/dias/2023/02/28/pdfs/BOE-A-2023-5356.pdf

Brain Circuits Behavior Lab (2023). https://braincircuitsbehavior.org

Brain-CODE (2023). https://braininstitute.ca/about-obi

Brain Initiative Alliance (2023). Neurodata Without Borders (NWB). https://www.braininitiative.org/toolmakers/resources/neurodata-without-borders-nwb/

Canadian Open Neuroscience Platform, CONP (2023). https://conp.ca

Carrasco Mallén, M. (2007) The future of Research in the National Health Care System. ISSN: 11372966

Columbia Magazine (2022). The Need to Protect the Data in Our Brains. Hond, P. https://magazine.columbia.edu/article/need-protect-data-our-brains

Commission Implementing Decision (EU) 2021/914 of 4 June 2021 on standard contractual clauses for the transfer of personal data to third countries pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council.

Elsevier (2023). Scopus. https://www.elsevier.com/solutions/scopus

EBRAINS (2023). Enabling breakthroughs in neuroscience and brain health. https://www.ebrains.eu/focus-areas

EBRAINS (2023). EBRAINS HealthDataCloud - Tackling the sensitive data challenge in brain research. https://www.ebrains.eu/news-and-events/ebrains-healthdatacloud-tackling-the-sensitive-data-challenge-in-brain-research

European Commission (2021). REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206

European Union Agency for Cybersecurity (ENISA) (2023). Health Threat Landscape. https://www.enisa.europa.eu/publications/health-threat-landscape

García Murillo D, Zhao Y, Rogovin OS, Zhang K, Hu AW, Kim MR, Chen S, Wang Z, Keeley ZC, Shin DI, Suárez Casanova VM, Zhu Y, Martin L, Papaemmanouil O, Van Hooser SD. NDI: A Platform-Independent Data Interface and Database for.

Gong C, Jing C, Chen X, Pun CM, Huang G, Saha A, Nieuwoudt M, Li HX, Hu Y, Wang S. Generative AI for brain image computing and brain network computing: a review.

Front Neurosci. 2023 Jun 13;17:1203104. doi: 10.3389/fnins.2023.1203104. PMID: 37383107; PMCID: PMC10293625.

Google Cloud (2023). https://cloud.google.com/?hl=es

Hermoso-Mendizabal, A., Hyafil, A., Rueda-Orozco, P.E. *et al.* Response outcomes gate the impact of expectations on perceptual decisions. *Nat Commun* **11**, 1057 (2020). https://doi.org/10.1038/s41467-020-14824-w

Vall d'Hebron University Hospital (2023). https://hospital.vallhebron.com/sites/hospital/files/2022-01/presentacion-corporativa-campus-ES.pdf

Human Brain Project (2020). Short Overview of the Human Brain Project. https://www.humanbrainproject.eu/en/about/overview/

Human Brain Project (2020). New openMINDS metadata models enhance EBRAINS services. https://www.humanbrainproject.eu/en/follow-hbp/news/new-openminds-metadata-models-enhance-ebrains-services/

INCF Japan Node (2018). Neuroinformatics Platform. https://www.neuroinf.jp/program/platforms.html

DATA PROTECTION ENGINEERING: From Theory to Practice. ENISA. DOI: 10.2824/09079

Instituto de Investigación Sanitaria Galicia Sur. http://www.iisgaliciasur.es/quienes-somos/

Carlos III Health Institute (ISCIII). 2023. Programme for the Evaluation, Accreditation and Monitoring of HRIs. https://www.isciii.es/QueHacemos/Financiacion/IIS/Paginas/Acreditacion_y_Contacto.aspx

Clinical Institute of Neurosciences (ICN) (2023). Neurology Service. https://www.clinicbarcelona.org/servicio/neurologia

ITEM 1. 2022. Organisational structure. https://itemas.org/sobre-itemas/mapa-de-cobertura/

ITEM 2. 2022. Mission, vision and values. https://itemas.org/sobre-itemas/mision-vision-y-valores/

Japan Node of the INCF (2018). https://www.neuroinf.jp

KnowledgeSpace (2023). https://knowledge-space.org

Langevin, M., Grebner, C., Güssregen, S., Sauer, S., Li, Y., Matter, H., & Bianciotto, M. (2023). Impact of applicability domains to generative artificial intelligence. ACS omega, 8(25), 23148-23167.

Law 14/2007, of 3 July, on Biomedical Research.

Law 14/1986 of 25 April 1986 on General Health.

Law 14/2011, of 1 June, on Science, Technology and Innovation.

Organic Law 3/2018, of 5 December, on the Protection of Personal Data and the guarantee of digital rights.

MONAI (2023). https://monai.io

Neuroscience Physiology and Imaging Experiments. eNeuro. 2022 Feb 9;9(1):ENEURO.0073-21.2022. doi: 10.1523/ENEURO.0073-21.2022. PMID: 35074827; PMCID: PMC8874953

Spanish Patent and Trademark Office, SPTO (2023). PCT International Patent Application. http://www.oepm.gob.es/es/propiedad_industrial/propiedad_industrial/que_se_puede_proteger_y_como/solicitud_de_patente_internacional_pct/

World Intellectual Property Organization (WIPO). https://www.wipo.int/patentscope/en/

World Intellectual Property Organization (WIPO). WIPO Technology Trends 2019 - Artificial Intelligence. Volume 1055 of WIPO Publication.

Poline, JB., Das, S., Glatard, T. et al. Data and Tools Integration in the Canadian Open Neuroscience Platform. Sci Data 10, 189 (2023). https://doi.org/10.1038/s41597-023-01946-1

Royal Decree 279/2016 of 24 June on the accreditation of biomedical or health research institutes.

Tang, J., LeBel, A., Jain, S. et al. Semantic reconstruction of continuous language from non-invasive brain recordings. Nat Neurosci 26, 858-866 (2023). https://doi.org/10.1038/s41593-023-01304-9

Polytechnic University of Madrid (UPM) (12 February 2013). *CEI Montegancedo, news in TVE for its participation in the HBP.* https://www.upm.es/UPM?fmt=detail&prefmt=articulo&id=d768581bd4ecc310VgnVCM10000009c7648a____

UNESCO (2023). The risks and challenges of neurotechnologies for human rights. https://www.unesco.org/en/articles/risks-and-challenges-neurotechnologies-human-rights?hub=83294

Vall d'Hebron Research Institute (VHIR). https://vhir.vallhebron.com/es/instituto/quienes-somos

ANNEX 1

Questionnaire

Link to the questionnaire: https://forms.gle/2qRjWZh7BCaqd6yZ7

ANNEX 2

Interview questions

1- Have you heard about the Human Brain Project and EBRAINS? Which services do you find most interesting? https://www.ebrains.eu/ ; https://www.humanbrainproject.eu/

2- *For participants who are PIs:* How many people does your team consist of? Can you state that the technologies you usually use for data acquisition (or the ones you have placed in the survey) are also used by other members of the group?

3 - What type of data do you use in your clinical or research practice? Do you use open data? Do you use sensitive data? Do you use synthetic data? Which platforms do you usually use for data sharing? Are they only internal? What is your opinion on federated platforms for medical data? Have you used these technologies?

4- What kind of clinical data would you be most interested in finding? Do you think that the unique databases you have would be useful to the scientific community? Would you be willing to share this data? Why?

5- You comment that clinical data analysis and processing services are very useful for you, so: do you use hospital software or can you use third-party services or tools? What are the main problems you face with these tools?

6- Do you have experience using virtual simulators for *neuroscience experimentation* or *clinical practice*? Would you like these tools to be part of your research project and why wouldn't you use them?

7- Regarding the usefulness of human brain simulation *software*, do you use any of them in particular or is it a current need in your research group? What aspects do you value most in these simulation tools?

8- Data curation is an important process when it comes to organising and integrating data, however, the internal human resources available for this purpose are usually scarce. Could you comment on what actions are taken to solve this problem and what recommendations would you give to speed up these processes?

9 Do you usually receive training in data curation and data management, and would your group be willing to collaborate with other researchers to support you in these services?

10- What ethical aspects do you consider most important or would you highlight in the acquisition, processing and sharing of data? What protocols do you follow in this respect?

11- What medical technology or innovation that you know or have read about recently would you like to incorporate into your medical routine?

12- In the field of data acquisition and management, what solutions would you like to use that you don't have right now?

13- What trends do you observe in your area of research that could bring Spanish healthcare closer to the medicine of the future?

14- Do you consider that there are enough companies in Spain working in your line of research? What barriers or difficulties do you find to carry out collaborations with the private sector?

15- As an ISR, how easy or difficult is it to access medical data? Do you have close contact with doctors in the hospital who provide you with research data? If you also consult, what percentage of your time is devoted to research and what part to care?