NCBIO
ERA CHAIR

Data Management Plan

This project received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement nº. 951923

Data Management Plan

NCBio Deliverable D1.2

Project Documentation Sheet	
Project	NCBio: Unlocking Excellence in Research and Innovation in Neurobiology and Neurological Disorders at IBMC/i3S
Acronym	NCbio
Grant Agreement nº	951923
Call identifier	H2020-EU.4. C. - ESTABLISHING ‚ERA CHAIRS' WIDESPREAD-06-2020 - ERA CHAIRS
Start date of the project	1.1.2021
Duration	72 month
Project Officer	Katherine Quezada
Coordinator	Mónica Sousa
Partners	Instituto de Biologia Molecular e Celular- IBMC

Deliverable Documentation Sheet	
Number of deliverable	D7.1
Title	Data Management Plan
Related WP	WP1. Project Management
Lead Beneficiary	IBMC
Author(s)	Mónica Sousa (IBMC)
Contact email	msousa@ibmc.up.pt
Nature of the deliverable	DEC
Dissemination level	Public
Due Date	30.06.2021 (M6)
Date of submission	30.06.2021
Status of the document	1st draft by Susana Magalhães & Gabriela Almeida (IBMC) on 21 June 2021 2nd draft review by Mónica Sousa (IBMC) on 25 June 2021 Final Version approved by Mónica Sousa (IBMC) on 30 June 2021
Version	Version 1.2

Abbreviations and Acronyms

Abbreviation Acronym	Definition
DMP	Data Management Plan
DPO	Data Protection Officer
IBMC	Institute for Molecular and Cell Biology
WP	Work package

Index

Executive summary

This Data Management Plan (DMP) outlines the research data the project will generate, collect and process. It also aims to describe whether and how data will be shared and how it will be curated and preserved, in line with ´FAIR' (findable, accessible, interoperable and re-usable) data management guidelines. This is a working document and will be updated upon recruitment of the ERA Chair holder, according to the particularities of the research data to be generated/used by their research team. Further updates will be made alongside periodic project evaluation and whenever significant changes arise, such as inclusion of new data types, changes to national/institutional/funder policies, identification of new innovation potential/patent filing decisions and/or change of personnel. IBMC/i3S Research Data Management policy is under development.

1. Data Summary

1.1. Purpose of the data collection/generation and its relation to the objectives of the project

To obtain information, To share information, To keep on record.

The main objective of NCbio is to recruit and maintain an experienced researcher and research project manager in the field of Neural Cell Biology (ERA Chair Holder), as well as her/his scientific team that will establish a new research group at i3S/IBMC. The Neural Cell Biology research group (also referred as "NCbio group") led by the ERA Chair Holder will provide excellence and trigger innovation in this particular area and will modify the research and innovation (R&I) landscape in PN&ND at the institute with a positive impact on the region and Portugal, and contributing to the achievement of several UN SDG goals, such as Goal 3: Good health & well-being and Goal 4: Quality education.

1.2. Types of data that the project will generate/collect

Data and Information collected, processed and/or generated within the scope of NCBio may include the following types:

i) **Non-human Biological Data:** Biological data resulting from experiments, in its raw and/or processed forms. Formats and size: Data formats and volume will be highly dependent on experimental techniques and tools employed to collect, generate and analyse the data. For example, if the research employs bioimaging techniques, a high data volume can be expected, particularly at high spatiotemporal resolution (potentially TBs of data per experiment);

ii) **Human Patient Data:** Human data collecting or generated in a clinical setting. Formats and size: To be determined, and may include medical imaging and/or biological data. Data of the Center for Predictive and Preventive Genetics Center-CGPP (integrated in IBMC/i3S) has gathered a large collection of human samples (DNA and cell lines) that are complemented by clinical and pedigree data to which NCbio can gain access;

iii) **Secondary Research Output Information-Institute Wide:** This covers all secondary research output information, including Publications and quantitative and qualitative information regarding Patents, Grants/Fellowships, Collaborations, Personnel, Public

Engagement. Formats and size: Information collected, used and generated will be stored in widely-adopted and accessible text, image, audio, video and database formats, including ISO 19005-1-compliant PDF/A where possible (aiming for maximum long-term preservation).The total volume of information collected, used and generated will not be significant (Gigabytes). Secondary output information will be collected automatically from internal databases, fed by internal administrative units and international databases *e.g.* PubMed.

1.3. Origin of the data

The majority of biological data will likely result directly from experiments conducted by the ERA chair holder and their team on-site as part of (inter)national and/or intersectoral collaborations.

1.4. Data utility

The data collected will be of utility to Researchers, Research communities, Decision makers, Education, The public, and Industry.

It is likely that biological data will be of long-term value to academic and non-academic researchers. Clinical datasets will be of long-term value to clinicians, academic and non-academic researchers working in similar fields, policymakers and will be preserved, in line with ethics review recommendations. Collated secondary project output information will be of long-term value to academic and non-academic researchers and research managers, funding bodies, policymakers and the general public.

1.5. Reusable Data

Data from previous research projects conducted by the researcher, collaborators or other parties may be re-used only following granting of appropriate ethical approval and any required data sharing agreements.

2. FAIR Data

2.1. Making data findable, including provisions for metadata

Metadata will be used to describe the data, according to standardised vocabularies for metadata. We will make use of globally unique and persistent identifiers such as Digital Object Identifiers (DOIs), which are automatically assigned upon deposition of data records into certified repositories.

2.2. Provide searchable metadata for data

Keywords will be in English and/or respecting naming conventions. Data in centralised storage (as part of Biodata.pt) will be analysed by curators and keywords will be extracted automatically. In the case where other external repositories are used alongside publication, the data owner will choose keywords manually, guided by repository and/or journal keyword guides, to confirm appropriateness and to optimise findability to interested parties with intent on re-use.

2.3. Ethical or legal issues that can impact sharing the data

Datasets underpinning research papers will be made openly available at the time of publication of results in a scientific journal, guided by journal policy, unless there is an explicit reason why this cannot be the case (for example, due to Information Security or Intellectual Property considerations). Whenever data requires protection due to its sensitivity, *e.g.* when a study involves potentially identifiable personal data about research participants, it will be only be shared if the identity of those participants is fully protected through anonymization, in line with the General Data Protection Regulation.

2.4. Data openly accessible

The IBMC/i3S Data Management Policy (under development) will be informed by national policy on management and sharing of data arising from research funded by The Foundation for Science and Technology (FCT), the primary national evaluator and funder of scientific research activities in Portugal. FCT policy encourages researchers to share primary data and other data with the scientific community, within the shortest time possible. Datasets underpinning research papers will be made openly available at the time of publication of results in a scientific journal, guided by journal policy, unless there is an explicit reason why this cannot be the case (for example, due to Information Security or Intellectual Property considerations). Where a managed access process is required the access mechanisms established will be proportionate to the risks, and will not restrict or delay access unduly. Specific biological and/or human datasets to be made available will be confirmed upon recruitment of the ERA Chair holder. Data of Research participants will only be shared if their identity is fully protected. All publications resulting from NCBio will additionally be deposited in OpenAIRE portal, in line with EU funding requirements and potentially also bioRxiv.org, the preprint server for Biology, at the author's discretion. Further research data-specific policies are under development, including specific sharing guides, which will depend on the type of data in question. We will aim for information regarding software needed to access the data or replicate analyses to be included alongside data. If appropriate, original software itself will be included (*e.g.* in open source code).

2.5. Storage security for the data and provision of storage backup and recovery procedures

Storage will be secure with backup and recovery.
An institution-wide policy on retention time and long-term offsite backup is under development.

2.6. Methods or tools required to access the data

We envisage no additional restrictions to access once data is deposited in a repository and access is set to open. In the case data is restricted, contact details for the data owner will be provided. Data users will be registered.

2.7. Making data interoperable

Data will be made interoperable through the use of controlled vocabulary.

2.8. Procedures for quality assurance of data

The acquisition of quality data will be assured through local immediate validation by trained researchers, supervised by internal support teams. Quality of analysis will be assured by the principal investigator and research peers. In addition, specific technical support units (equipment, software, animal platforms) will act as a quality assurance. For laboratory generated data, instruments are subject to frequent calibration and verification tests and so any potential data quality breaches can be prevented and/or traced. The ultimate accountability for the data and metadata quality lies with the data owner (Principal Investigator).
Specific processes will be determined by i3S policy.

3. Allocation of resources

3.1. Cost of making data findable, accessible, interoperable and reusable be covered

Use of institution infrastructure, Infrastructure Grant Costs may include repository storage, computational cloud services and/or personnel costs for necessary systems development and may be funded through the project or internally.

The Data Protection Officer (DPO) of i3S is not responsible for any data management but rather is appointed to ensure the compliance of personal data processing with the General Data Protection Regulation (GDPR), ensuring efficient communication with data subjects and cooperation with the regulatory authorities. Day-to-day management of research data will remain the responsibility of the Principal Investigator (in this case the ERA Chair holder). Day-to-day management of non-research data remains the responsibilities of the internal support unit managers (for institute-wide data) and the project manager (for project specific data).

3.2. Identify a data manager to manage data

A Data Manager, responsible for managing the projects' data will be identified.

3.3. Identify the people or roles that will be responsible for the management of the project data

Project manager to be hired.

4. Data Security

The in-house Storage Data provides a centralized storage solution for all researchers, allowing users to copy data from their workstation to a user folder hosted in the data centre accessible in any computer connected to the IBMC/i3S network.

An institution-wide policy on retention time and long-term offsite backup is under development.

Transfer of sensitive data will be subject to rules and guidelines specified in the project Ethics Review. For data deemed appropriate to share, external field-specific data repositories will be chosen based on their relevance to the field as described, but also on their storage and security capacities -- for example, the general purpose open-access repository Zenodo, which securely stores data plus metadata through cloud infrastructure operated by CERN's data centre.

5. Ethical aspects

i3S is committed to upholding responsible conduct of research and of service provision. With these standards in mind, IBMC/i3S has already implemented a series of codes or abides by regulations and decrees, such as:

1. Compliance with Decreto-Lei 80/2018 (revision of Decreto-Lei nr. 97/95); Clinical Research Ethics Committees
2. General Data Protection Regulation (GDPR – EU)
3. Compliance with Código Ético de Conduta Académica da UPorto
4. Strict observance of the IBMC/i3S Authorship Guideline, in line with the Authorship Guidelines by ICMJE

Other guiding documents are:

1. Compliance with The European Code of Conduct for Research Integrity (ALLEA)
2. Declaration on research integrity in responsible research and innovation (UNESCO)

As of now, IBMC/i3S runs a supervising and counselling committee for ethical and responsible conduct of research.

i3S has a Unit for Responsible Conduct in Research (an Integrity Office) aims to:

- Work close to the researchers, answering their queries and promoting their training in Ethics and Responsible Conduct of Research;
- Be the contact point for those wishing to report, in confidence, cases of research misconduct, before any formal allegation is made;
- Support the i3S community in the implementation of international codes of conduct and best international practices in research ethics and integrity;
- Work in collaboration with other national and international institutions of excellence in the field of Ethics and Responsible Conduct in Research;
- Work in collaboration with the Committee for Ethical and Responsible Conduct of Research (CECRI) concerning the following shared points:
- Promote training actions on vital areas pertaining to bioethics and responsible conduct of research, as well as other scientific activities and consequent results dissemination;
- Promote other discussion initiatives centered on ethical and responsible conduct of research issues relevant to the research carried out at i3S;
- Look into allegations of scientific misconduct and advise the i3S Board of Directors on the best course of action for such situations.

CECRI (the Committee for Ethical and Responsible Conduct of Research) at i3S aims to establish and ingrain in IBMC/i3S a culture of responsible conduct; this is mostly an advisory and educational body, which combines ethical aspects and issues of scientific integrity and should soon expand by including a management office appointed to handle matters related to ethics and integrity. CECRI and corresponding management office are bound to:

- Offer an expert perspective on the observance of ethical regulations and responsible conduct of research whenever research projects or protocols are up for approval that entails the collection and/or use of human material or data;
- Coordinate with the existing Animal Welfare and Ethics Committee;
- Identify other organs with authority to follow through with approval procedures, such as data protection, animal use, research in a clinical setting and so forth;
- Review approval processes carried out by external structures of projects including significant participation of i3S groups or researchers;
- Oversee the development of projects and procedures authorized by the Committee or external entities;
- Oversee the implementation of internal guidelines in responsible conduct of research;
- Promote training actions on vital areas pertaining to bioethics and responsible conduct of research, as well as other scientific activities and consequent results dissemination;

- Promote other discussion initiatives centred on ethical and responsible conduct of research issues relevant to the research carried out at IBMCi3S;
- Look into allegations of scientific misconduct and advise the Board of Directors on the best course of action for such situations.

The ERA Chair Holder and her/his Team will have access to all the IBMC/i3S facilities such as the Animal Facility, that provides training on laboratory animal experimentation (FELASA A, B and C levels) and complies with all ethical and regulatory framework (National and European).

Further to this, the ERA Chair Holder may also leverage CGPP where the DNA/cell-line bank stores thousands of human samples with clinical and pedigree data. Research with human subjects or samples at IBMC/i3S is carried out in collaboration with hospitals and reviewed by the ethics committee of the hospitals in question.

The Data Protection Officer (DPO) at i3S will monitor human personal data collection. Personal human data is processed in accordance with national and European legislation on the protection of individuals, as well as other legal, regulatory and good practices. Methods used for processing sensitive/personal data will entail: Pseudonymization or Anonymization of data where necessary, Privacy constraints and applicable ethical norms, Data accompanied by informed consent statements, Privacy policies, National laws.

6. Other

We identify no additional procedures for data management at this time. This section will be updated whenever any new policies or procedures come into effect.