



D10.2 MID-TERM UPDATE OF DATA MANAGEMENT PLAN

Project Name	EmpoWering EDUC for Inclusive Development of the ERA
Project Acronym	EDUC-WIDE
Grant Agreement No.	101136533
Programme	HORIZON.4.1 - Widening participation and spreading excellence
Topic	HORIZON-WIDERA-2023-ACCESS-03-01 - European Excellence Initiative
Project Starting Date	1. March 2024
Project Duration	36 months
Deliverable No.	10.2
File Name	10.2. Mid-term Update of Data Management Plan
Work Package	WP10
Dissemination Level	Public
Contractual Submission Date	31 st August 2025
Actual Submission Date	29 th August 2025
Institution	MUNI
Key Words	Data Summary, Findable, Accessible, Interoperable, re-usable, Ethics, Open Access
Abstract	The report presents a summary of data management practices for 12 individual measurements conducted under the EDUC-WIDE 1st Call for Access to Research Infrastructures.



**Co-funded by
the European Union**

Funded by the European Union under Grant Agreement No. 101136533.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

History of changes

Version	Publication Date	Change
0.1	21.08.2025	Initial version
0.2	26.08.2025	Commented by partners
1.0	29.08.2025	Final version

List of Contributors

Beneficiary	Name	Author/Contributor /Reviewer
MUNI	Iva Sedlakova	Author
MUNI	Antonín Zita	Reviewer

Abbreviations

UNICA – University of Cagliari

UP – University of Potsdam

UJI – University Jaume I

USN – University of South-Eastern Norway

PNU – Vasyl Stefanyk Precarpathian National University

UNIVREN – University of Rennes

UPECS – University of Pécs

PD - Parkinson's Disease

Table of Contents

1.	Introduction	5
2.	Data Summary	6
3.	FAIR Data	9
3.1	Making data findable, including provisions for metadata	9
3.2	Making data accessible	11
3.3	Making Data Interoperable	13
3.4	Increase data re-use	15
4.	Allocation of resources	17
5.	Data security	18
6.	Ethics.....	19
7.	Other issues	20
8.	Conclusion	20

1. Introduction

The initial Data Management Plan (submitted in M6) included only administrative information, details about collaborative spaces, and administrative documents produced by the expert groups, whose work was primarily strategic and developmental.

The mid-term update of the Data Management Plan now includes details on how scientific data is managed, specifically data generated through the 1st Call for Access to Research Infrastructures. These small projects/measurements cover a wide range of interdisciplinary research topics across biomedical science, neuroscience, structural biology, genetics, materials science, and mental health.

Finally, 12 projects were implemented:

No	Author	Project name	Affiliation	Measurements
1	Antonio Luigi Manai	Predictive biomarker of Parkinson's Disease in intranasal rotenone-treated mice	UNICA	MUNI
2	Patrick Arnold	The genomic origin of Late Holocene elephants in western Asia	UP	UPECS
3	Judith Medina	Surface characterization of supported metal nanoparticles on different heterogeneous supports	UJI	MUNI
4	Jakob Ruickoldt	Structural elucidation of human COP9 signalosome species	UP	MUNI
5	Jörn Klein	Mapping the Microbial Composition of Healthy Hand Skin Flora in Two Different Age Group	USN	UPECS
6	Matteo Aroffu	Repurposing Drugs with Nano and microtechnology-Based Therapies for the Treatment of COVID-19: Challenges, Opportunities, and Innovations	UNICA	MUNI
7	Olha Strilbytska	Analysis of post-traumatic stress disorder gene expression profiles in mice models of neuroinflammation	PNU	MUNI
8	Paola Caria	Genetic profiling to assess malignancy risk in patients with indeterminate thyroid nodules	UNICA	UPECS
9	Romain Gilbeaux	In situ high-resolution characterization of spermatid manchette microtubules	UNIVREN	MUNI
10	Simone Pani	Ultrastructural investigation of nanosized therapeutic systems	UNICA	MUNI
11	Maria Angeles Marques Torrejon	Unraveling Glioblastoma: Molecular Study of Quiescent Cells	UJI	MUNI
12	Julia Compart	Elucidation of the structure of the glucan, water dikinase	UP	MUNI

2. Data Summary

- *Will you re-use any existing data?*
- *What types and formats of data are expected?*
- *What sizes of data are expected?*
- *What is the purpose and usage of the produced data?*

1	• Generative data will be entirely original. • .xls or .tsv in supplementary material of manuscript/s • The expected size of the data will be less than a GB • To clarify what is the involvement of the intestine in the development of Parkinson's Disease (PD); to reveal VGF (non-acronym) gene alterations in the intestine during PD development; To correlate VGF gene alterations with the other PD-related gene changes. The general aim of the project is to understand which genes, including VGF, (related to intestine proteins, neurotransmitters, or others) could play a crucial role in PD.
2	• Genome-wide sequencing data of modern elephants available from NCBI Sequence Read Archive is reused for comparison • Paired-end Illumina sequencing data in fastq format. • 1TB • The paleogenomic data allow to estimate population-specific features that can potentially discriminate between a natural or human-based origin of the elephants. Evolutionary biologists & bioinformaticians interested in elephant evolution & megafauna extinction
3	• No existing data is being reused • Data will be generated in the following formats: Graphics/images: JPEG, PNG, PDF, PPTX; Tables: CSV, OPJU, XLSX; Text: DOCX, PDF, TXT • The more significant data sizes will stem from SEM and TEM images, because they are stored in high-resolution bitmap format. Approximately, the total size of the dataset for a single sample characterization is expected to be less than 20 MB • The goal of the project is to characterize the surface of the materials/catalysts using microscopy (SEM and TEM) and spectroscopy (XPS) techniques. This data will help us to understand the morphology of the samples. To do this, we will obtain images from TEM and SEM, in addition to the data set from XPS, using the corresponding equipment. The data will be useful for research groups working on nanomaterials, supported catalysts, heterogeneous catalysis, and surface characterizations
4	• No existing data is being reused. • The raw data includes several thousand of micrographic movies in (.tiff) format as well as meta files. These will be processed yielding maps of the protein in .mrc format for which atomic models will be build (.pdb, .cif files) • The raw data will sum up to several TB and the model and map files will be a few GB • The cryo-EM micrographs collected in the scope of this project should enable the determination of protein structures. The map and the model files can be used for the visualization of the results and might help researchers drawing their own conclusions or developing new hypotheses
5	• No existing data is being reused. • The questionnaires were scanned into PDF format, and the summarized data are stored in an MS Excel file. Raw sequencing data are in .fastq file format. Summaries are generated in .csv files. Further data analysis will be carried out in MS Excel. The

	publication will be written in MS Word, and figures will be generated using MS PowerPoint. - At this point, we guess 10 GB - Microbiome samples were collected from participants' hands, and participants completed a questionnaire addressing key factors that may influence the composition of the microbiome (e.g., whether they have pets, handle raw meat, bake with yeast, etc.).
6	- No existing data is being reused. - JPG 500 megabytes – 1 gigabytes - The primary objective of data collection is to further confirm the formation of nano- and/or microparticles and to investigate their morphology and structure. These characteristics are critical for assessing the physicochemical properties of the systems under study, as they may significantly influence stability and efficacy.
7	- No existing data is being reused - Transcriptome-wide gene expression data in the form of raw RNA sequencing files (FASTQ format), followed by processed data files such as read count tables (CSV/TSV format) and differential gene expression analysis results (e.g., Excel or CSV files). - The expected size of the raw data is approximately 100–150 GB, processed data will be about 10 Mb - The data will be generated as a part of our research project “Analysis of post-traumatic stress disorder gene expression profiles in mice models of neuroinflammationThe data collection supports the project’s objective by enabling the identification of gene expression changes linked to posttraumatic stress disorder (PTSD) development, thereby uncovering potential genetic risk markers of the disorder
8	- No existing data are being reused. All data are newly generated specifically for the purpose of this project - Transcriptome-wide gene expression data in the form of raw RNA sequencing files (FASTQ format), followed by processed data files such as read count tables (CSV/TSV format) and differential gene expression analysis results (e.g., Excel or CSV files). - The expected size of the raw data is approximately 100–150 GB, processed data will be about 10 Mb - The data will be generated as a part of our research project “Analysis of post-traumatic stress disorder gene expression profiles in mice models of neuroinflammation. The data collection supports the project’s objective by enabling the identification of gene expression changes linked to posttraumatic stress disorder (PTSD) development, thereby uncovering potential genetic risk markers of the disorder.
9	- No preexisting data will be re-used for this project. - mrc format is open source and a standard in the field. These files can be easily converted into .tif files. - single images (approximately 30 MB each), tilt-series (approximately 1 GB each), or tomograms (ranging from 4 GB to 10 GB with all reconstruction files). - The data collected in this project aims to characterize, at high resolution and in situ, the internal structural organization of spermatid manchette microtubules. Specifically, the study focuses on identifying and analyzing novel internal structures within these microtubules. These internal structures are hypothesized to play a crucial role in regulating manchette function, which is essential for sperm head shaping and flagellum formation during spermiogenesis. By acquiring cryo-electron tomography (cryo-ET) data from lamellae prepared by cryo-FIB milling, the project seeks to provide molecular insights into the organization, spatial distribution, and functional significance of these internal structures in male reproductive biology.
10	No existing data is being reused.

	• JPEG format, Text file (MDOC). • The expected size for all the data will be around 13 GB • A key objective of the project is to gain deep insights into the relationship between vesicle structure, functional components and drug incorporation. So, thanks to the access to CEITEC at Masaryk University, we will be able to gain a complete picture of the developed phospholipid-based vesicles, unlocking structural insights essential to describe a nanosystem
11	• No existing data is being reused. • CSV • 6 TB • To understand the mechanisms that create resistance to conventional therapies (in this case, chemotherapy), it is important to know the genes that are expressed before and after applying it
12	• No existing data is being reused. DOCX, XLSX, PPTX, PDF, PNG, JPEG, TIFF, GIF, HTML, XML • The structure of GWD will be visualized by cryo-electron microscopy – several GBMB • The phosphorylation process is crucial for the physico-chemical properties of starch and at the same time it is the only natural modification of starch that occurs, so that the findings are not only decisive for the understanding of metabolic processes in plants but also occupy an essential position in the starch industry. Data will be useful to researchers, administrative staff, and students at all partner universities of the Alliance. Moreover, it will be published in a high-ranking journal and it is an important part of the PhD thesis.

3. FAIR Data

3.1 Making data findable, including provisions for metadata

- *Do data have a persistent identifier? Naming conventions used?*
- *What metadata will be created?*
- *What disciplinary or general standards will be followed?*

1	• Researcher's ORCID: https://orcid.org/0000-0002-0606-5636 • Standardized metadata formats will be used by: 1. Employing DataCite, Dublin Core, or the Metadata Object Description Schema (MODS). 2. Providing comprehensive descriptions, including detailed information about the dataset, its purpose, methodology, location, temporal coverage, and any special conditions or restrictions. 3. Using persistent identifiers by assigning unique and persistent identifiers (e.g., DOI, ORCID) to the dataset and associated publications • The RAW data will be available as supplemental material in future scientific publication(s).
2	• Project, sample and sequencing run identifier are automatically assigned to ongoing numbers by the NCBI SRA. • Project, sample and sequencing run identifier as well as species name (including higher level taxonomy), fossil specimen collection ID, sequencing specifications and author from the meta data will all be searchable in the NCBI SRA. Species name (including higher level taxonomy), fossil specimen collection ID, sequencing specifications and author will be associated with the raw data. Naming and metadata will follow NCBI SRA conventions, which are well-known in the field. The record will also have a unique URL.
3	• The data will be deposited in the institutional repository of the Universitat Jaume I, which assigns a DOI and handle-type persistent identifier to each record, as well as standardized metadata, digital preservation and exchange protocols. • Metadata are based on the Dublin Core schema (ISO 15836), a flexible standard widely used in all disciplines and completed with the OpenAire specifications • Metadata used to describe the data are entered into the Institutional Repository manually using the template provided by the Institutional Repository.
4	• Metadata will receive a DOI. • Data acquisition parameters and sample properties will be submitted to the PDB, EMDB and EMPIAR. • The submission of map and model files to the EMDB and PDB and of raw micrograph to EMPIAR is a standard procedure in the field of structural biology.
5	• We plan to publish a paper; it will have a DOI number (dataverse issues PID). Supplementary data will be available in the USN repository (Dataverse), it will also receive PID. The corresponding sequencing data are available in the European Nucleotide Archive (ENA). • Title, authors, keywords, description, Doi • Dataverse uses DataCite Metadata Schema 3.1. European Nucleotide Archive (ENA) repository, the use community-developed reporting standards in the form of sample checklists. Sample checklists are a defined set of minimum information required and validated during ENA sample registration
6	• The article will have Digital Object Identifiers (DOI). In imaging studies, naming conventions typically include keywords such as "imaging," "cryo-TEM," "SEM," and

	“morphology,” among others. Additionally, a clear and consistent versioning scheme will be implemented. - Each image will be accompanied by detailed metadata, including information on acquisition parameters (e.g., microscope type, magnification, resolution, contrast settings), sample preparation details, and experimental conditions. Additionally, metadata will include timestamps, file formats, and relevant annotations to facilitate data interpretation and interoperability. - Standardized terminology will help align data annotations, reduce ambiguities, and enhance compatibility across different systems and analytical tools
7	- The data will be published alongside a scientific article, which will be assigned a Digital Object Identifier (DOI). The key components of the naming structure will include: - Sample ID: A unique identifier for each mouse sample (e.g., MouseID_TreatmentCondition_Date). - Treatment Information: Indicating the type of treatment or condition (e.g., PTSD, Control, Neuroinflammation). - Sequencing Run: A reference to the sequencing batch or run (e.g., Run1, Run2). - Data Type: Indicating the type of data (e.g., raw_RNAseq, processed_data). - Version Number: to track data updates (e.g., v1, v2). For example, a sample name might look like: M123_PTSDFootShock_2025_Run1_raw_RNAseq_v1 - Descriptive Metadata (Title, Author, Abstract, Keywords), Data Provenance (Source, Creation date, Version History), Technical Metadata (File Format, File Size, Software/Tools Used), Data Quality and Integrity (Data Validation, Error Report) - The article will include supplementary materials with the original data, ensuring accessibility and allowing other researchers to cite and utilize the data.
8	- The dataset will be renamed with the name of the project. Version numbers or dates in the naming convention will be used to establish a consistent naming system that differentiates between major and minor updates. The new version with major changes and minor updates represented by incremental numbers. Dataset will be published in a repository e.g. Zenodo. - The metadata will include key information such as authors, description of data, sample origin, sequencing platform, data formats, and licensing terms.
9	- A structured versioning system will be implemented to ensure data traceability. File-based versioning: Raw data files (e.g., .mrc images and tilt-series) will be stored in an immutable format with timestamps. Processed data (e.g., sub-tomogram averages) will follow a structured naming convention indicating the processing step and dated version (e.g., SampleID_TiltSeries_v20250312.mrc, SampleID_TomoReconstruction_v20250313.rec). Metadata documentation: Each data acquisition session will be logged in the electronic lab notebook (eLabFTW), documenting key experimental details, processing parameters, and version updates. Repository version control: Data submitted to EMPIAR and EMDB will include version history in accordance with their data deposition policies. For all published data, tilt-series and tomograms are deposited on EMPIAR (https://www.ebi.ac.uk/empiar/) and sub tomogram averages on the EMDB (https://www.ebi.ac.uk/emdb/) with all required metadata. Both EMPIAR and EMDB provide DOI to their datasets.
10	- Files will be named according to a pre-agreed convention between the users. Microscope - Microscope images generate a range of metadata (field size, magnification, lens phase, zoom etc.) with each image, so the data will be accompanied by a text file which will describe the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information.

	• UNICA IRIS provides a Handle to their datasets. • common metadata standards in the field are used.
11	• Author ORCID: https://orcid.org/0000-0001-5725-5774, Data will have a DOI once they are deposited (e.g. Zenodo). • Descriptive metadata. • DataCite Metadata Schema.
12	• Since the aim is to publish the GWD structure in an internationally recognized journal, it is usually assigned a DOI. A persistent identifier will be added to the data (documents) when placed into the long-term repository. • Key words such as cryo-electron microscopy, structure elucidation are given. • Standard metadata responds to the formats mentioned in Section 2.

3.2 Making data accessible

- *Which of your data and metadata will be opened to others, and which not?*
- *What will be needed to reach the data?*
- *Embargo conditions and expected embargo termination?*
- *Which repository will be used to share data openly?*
- *How long will the data remain available and findable?*

1	• The metadata will be available on the database shared with UNICA and in supplemental material in the future scientific publication(s). • Through the DOI or the link related to the publication/s. • The embargo will be mandatory until the publication is accessible online. • Repositories of the open access publications, UniCA IRIS and UniCA EPrints. • UniCA EPrints and IRIS do not have expiry dates.
2	• All technical metadata as well as information on the sample of origin for the sequencing data will be made publicly available upon publication • Data can be accessed via the web-based NCBI SRA without further software needed. Specific software for faster, command line-based download of the data is freely available from the website (sra-tools), together with detailed manual and tutorial information • No embargo. • Raw sequencing data uploaded to and deposited in the NCBI SRA are curated by the database • Long-term curation of the data is done by the NCBI SRA.
3	• The data of images will be made openly available. However, industrial designs of microscopes and reference images for assembly may be closed by decision of the project consortium • To access them, it is necessary to request authorization by email. • No. • Data will be deposited in the University's Institutional Repository. The project website will provide links to the data in the Repository. • The Repository guarantees its reusability for at least 10 years.
4	• The processed maps and model files and the raw micrographs will be made publicly available in the EMDB/PDB/EMPIAR. • The intermediate data generated during data processing will not be made available as they are of little use and can take up to several TB. • They can be downloaded from the corresponding homepages without any charge • The data will be accessible upon publishing the results in a peer-reviewed journal.

	- The data will be submitted to the PDB, EMDB and/or EMPIAR - The models and maps submitted to the EMDB and PDB and the raw data to the EMPIAR are publicly available to any use for an unlimited amount of time.
5	- All data will be openly available - Access to the data will be provided through standard procedures, as is the case with other metadata in such repositories. Read more: https://ena-docs.readthedocs.io/en/latest/retrieval/general-guide.html#viewing-and-exploring-ena-records; https://ena-docs.readthedocs.io/en/latest/retrieval/file-download.html - The data will be under embargo until they are published in the journal - The metagenomic data will be uploaded to a public repository, most likely the European Nucleotide Archive (ENA). Supplementary data will be available in the USN repository, which (from next year) will be Dataverse. The selected journal will also have a repository. - ENA doesn't have an end date. USN RDA operates with a minimum of ten years.
6	- The data published in a scientific article will be opened to others. - The http to the article will be needed to reach the data. - No embargo. - The data will be made findable online by being included as an appendix to a journal article.
7	- The following data will be made openly available: - Raw and Processed RNA Sequencing Data - Experimental Metadata - Behavioral Test Results Closed data includes: Preliminary or Incomplete Data: Data that is still being analyzed or that has not yet been fully validated may be temporarily withheld until the analysis is complete to prevent the dissemination of potentially misleading or incomplete results. Confidentiality Agreements or Collaboration Concerns: In cases where data is collected as part of collaborative projects or under confidentiality agreements with other research institutions, certain data may be kept closed until publication or until permission for release is granted. - Access to the data, in case of any restrictions, will be provided upon request by contacting the corresponding author of the publication. Interested parties may submit a formal request outlining the purpose of data use, and access will be granted in accordance with ethical guidelines and data sharing policies. - The data will be fully accessible for reuse under the CC BY 4.0 license. - The scientific article will be published in a peer-reviewed journal and will include supplementary materials containing: a) The processed data; b) Metadata related to the experimental design; c) Documentation and analysis scripts used for data processing and statistical analysis - The data and associated metadata will remain publicly available and findable indefinitely, with long-term preservation ensured through a trusted repository
8	- The data obtained from this study will be published in an open-access journal relevant to the topic, specifically focusing on thyroid cancer or the molecular and cellular mechanisms involved in cancer. - Data will be open after publication. - Distributed Databases and a Central Repository - a suitable field-specific repository will be used to ensure that data is accessible from different sources.
9	- When electron microscopy data are submitted for publication, they are deposited in public databases (EMPIAR, EMDB) with an embargo that expires upon acceptance for publication. - To open them, only classical, open source, freely downloadable software can be used (e.g., etomo, UCSF Chimera, Fiji).

	• After the article's publication date is open. • They are deposited in public databases (EMPIAR, EMDDB). • The data will be stored for the maximum duration of the repository, at least 10 years.
10	• The data sets will be made publicly available after their publication • Access to the data can be requested by sending a simple email to those responsible for data management • Data will also be available in an institutional repository, UNICA IRIS, designed to collect both the bibliographic metadata and the digital version of research outputs. • The published data will be available online for the maximum duration of the repository, at least 10 years.
11	• All data will be open • via http://zenoddo.org • Data will be embargoed until publication. • Open access journals or the UJI repository. • The data will be stored for the maximum duration of the repository, at least 10 years.
12	• The entire data, as well as the methodology used, will be included in the documents and will be available for public access. • No specific software is required to access the data. • The data are part of the doctoral thesis and are also published in an open access journal. Only standard software is required to use them. • Data will be available for at least ten years after the end of the project. However, as these will be put into the repositories; it is expected they will be preserved there much.

3.3 Making Data Interoperable

- *Will the metadata be provided in relevant standards?*
- *Where possible, controlled vocabularies, keywords, thesauri, and ontologies are used for description:*

1	• Employing DataCite, Dublin Core, or the Metadata Object Description Schema (MODS). This ensures that the metadata can be understood and utilized globally • Using common vocabularies and taxonomies like ICD for medical data or Dublin Core for metadata.
2	• Data deposition and availability will follow strict and internationally confirmed rules of sequencing read deposition (institutionalized in the NCBI SRA). • Where possible, controlled vocabularies, keywords, thesauri, and ontologies are used for description: The datasets will follow the NCBI SRA recommendations.
3	• They provide information about the title and date of the data, authorship, keywords, versions, project. • Relevant keywords are used to support data discovery and reuse, including terms such as "MXene", "hybrid catalyst", "surface characterization", "XPS", "TEM", and "SEM". While we do not currently use a formal controlled vocabulary or ontology, the keywords are selected based on standard terminology commonly used in catalysis and materials science literature. Data quality is assured with the standardized procedures established in the quality assurance document available on the project portal.
4	• <i>YES, they are made accessible to the community by submitting to the established repositories of the PDB and EMPIAR.</i> • They can be searched by the protein's name, the experimental conditions and the provided metadata and will get a DOI.

5	• Dataverse: DataCite Metadata Schema 3.1. ENA uses community-developed standards (see link under point 2.1) • Subject-specific keywords will be provided to make the data findable.
6	• Standardized terminology will help align data annotations, reduce ambiguities, and enhance compatibility across different systems and analytical tools. • Additionally, metadata will include timestamps, file formats, and relevant annotations.
7	• Metadata will be provided in accordance with the DataCite Metadata Schema. • We will use standard vocabularies and ontologies for all data types, including Gene Ontology (GO) for gene functions, Experimental Factor Ontology (EFO) for experimental conditions, and Uberon for anatomical terms, ensuring inter-disciplinary interoperability.
8	• Each dataset will undergo rigorous validation checks to ensure accuracy and consistency. Data will be collected and processed using standardized methods/protocols to reduce errors and ensure consistency across the dataset. • We will use standard vocabularies and ontologies for all data types, including Gene Ontology (GO) for gene functions, Experimental Factor Ontology (EFO) for experimental conditions, and Uberon for anatomical terms.
9	• Experimental metadata are recorded on a standardized form describing, for each "cryo storage box," the sample preparation and freezing conditions and are linked to the images named as "box number – grid number." • Controlled vocabularies and ontologies are used where applicable to ensure consistent and accurate data description and facilitate interoperability. When submitting data to public cryo-EM databases (EMPIAR and EMDDB), we follow their recommended metadata schemas and controlled vocabularies, including standardized terminology for imaging parameters, specimen types, and processing steps. In addition, keywords describing the biological context (e.g., spermatid manchette, SPACA9, microtubules, cryo-electron tomography) are selected from established biomedical ontologies such as MeSH (Medical Subject Headings) and the Gene Ontology (GO), where relevant. We use consistent terminology across our electronic lab notebook (eLabFTW), file naming conventions, and data repositories.
10	• The images will be accompanied by contextual documentation, according to standard practice for synthetic methodology projects: Text files (MDOC) which detail the experimental procedures and compound characterization. • Controlled vocabularies, keywords, thesauri, and ontologies are not applied in the description of the data. Instead, the data are documented using free-text descriptions.
11	• Metadata will be in industry-standard formats. • Standard vocabulary. Keywords: quiescence, neural stem cells, glioblastoma, temozolomide.
12	• The output data are mainly textual reports to be read by humans. • No specific measures are needed for interoperability beyond using English and commonly used data formats. For validation purposes, all the sources used will be properly cited, and the methodology used will be included in the documents.

3.4 Increase data re-use

- *The data are clearly licensed*
- *It is clear how, when and by whom the data were created and processed*
- *The data and metadata meet the relevant standards of the field*

1	• (CC0), Open Data Commons (ODC): Open Database License (ODbL) which ensures users can share and modify data • No, this information will be included in the open access publications • Yes.
2	• CC-BY license. • Since the data will be in NCBI SRA, it will be in standard form and well described.
3	• All metadata will also be openly available under a CC0 public domain license. The data may be reused by anyone after the project has ended, according to the conditions of the Creative Commons 4.0 International Attribution (CC BY) license. • A Readme.txt file will be added to the data files with additional information to facilitate interpretation and reuse of the data • Dataset from the project will adjust to normalized format: ASCII, TXT, CSV, XML, ... Keywords will be selected from controlled vocabularies that are suitable for the specific type of data. The OAI-PMH protocol will be used for harvesting.
4	• The data can be used without limitation • The data will be associated with the name of the depositor – probably Jakob Ruickoldt – and all others of the resulting papers • Yes.
5	• YES, with general CC licenses • This information will be provided in the description field of the datasets, and in the ReadMe File provided with the datasets • Yes, see above about the metadata standard
6	• The data will be made available for reuse in line with the journal policy. • Yes, information on how, when, and by whom the data were created and processed is tracked and documented by the software used, which automatically logs all operations. • Yes
7	• The data will remain reusable indefinitely, as it will be deposited in a public repository with an open-access license (CC BY 4.0). • The key components of the naming structure will include: Sample ID: A unique identifier for each mouse sample (e.g., MouseID_TreatmentCondition_Date). Treatment Information: Indicating the type of treatment or condition (e.g., PTSD, Control, Neuroinflammation). Sequencing Run: A reference to the sequencing batch or run (e.g., Run1, Run2). Data Type: Indicating the type of data (e.g., raw_RNAseq, processed_data). Version Number: to track data updates (e.g., v1, v2). For example, a sample name might look like: M123_PTSDFootShock_2025_Run1_raw_RNAseq_v1
8	• An open license CC-BY. • The data and metadata meet the standards in their field.
9	• All published datasets are deposited in public repositories such as EMPIAR and EMDB, which require the use of clear open-access licenses. These datasets are made available under the Creative Commons CC BY 4.0 license, allowing reuse, redistribution, and adaptation with proper attribution. This ensures the widest possible dissemination and reuse of the data

	- Each dataset is associated with detailed metadata that are systematically recorded in the electronic lab notebook (eLabFTW) and also included in the metadata submitted to public repositories - Data formats (.mrc, .rec, .tif) comply with the community standards for cryo-electron microscopy, and metadata follow the requirements set by EMPIAR/EMDB, including imaging parameters, sample descriptions, and processing workflows.
10	- The data sets will be made publicly available. - The data and metadata meet the standards in their field, and the provenance of the data will also be clear (The data were created and processed by a CEITEC technician using cryo-TEM).
11	- CCBY - In Zenodo, you're asked for authors, dates etc - Neuro-oncology and stem cells
12	- As stated, all of our data can become completely open over time. The dataset can be used in the provided format without any conversion needed. - Data produced and used in the project will be usable by third parties.

4. Allocation of resources

- *Who will pay for data storage for wide usage by other researchers?*
- *Who will be responsible for data management in your project?*

1	• The University of Cagliari will, if feasible, handle the data allocation. • Dr. Antonio L. Manai and Prof. Cristina Cocco will oversee the data
2	• NCBI SRA does not require payment for data storage. • The project leader is responsible for data management and the sequencing data have to be deposited before submission of the associated research article.
3	• The data will be deposited in the Institutional Repository and the University therefore assumes the costs of storage, security and preservation. • Data management responsibilities are primarily handled by Judith Medina, José Mata and the university (UJI). This includes data collection, processing, analysis, documentation, and secure storage.
4	• The submission to the EMDB, PDB and EMPIAR is free of charge and funded by the European commission, NIH, UKRI-BBSRC, Wellcome Trust and EMBL-EBI. • The backup of the raw data on magnetic tape will require ca. 250 €/12 TB per year of storage paid by the group of Prof. Wendler at the University of Potsdam.
5	• Data will be stored on the institution's servers. There is no extra fee for this project. • Prof. Dr. Jörn Klein is the project leader, responsible for data management.
6	• The data are stored in the Institutional repository, with no additional costs. • Data will be managed by me.
7	• We will utilize free and open-source software tools for data processing, documentation, and sharing. • The responsibility for data management in the project lies with the Principal Investigator (PI), Olha Strilbytska. She will oversee all aspects of data collection, documentation, storage, sharing, and long-term preservation.
8	• The estimated costs for making data FAIR are likely €10,000. • To cover these costs, funding will come from the other project budget allocated for data management, with additional support sought from grants or institutional funds dedicated to open research data initiatives.
9	• The cost for short-term storage is allocated to the team's research budget, and the cost for long-term storage is covered by the institute and university. We recently increased the long-term storage space on the institute's servers, giving our team access to 300 TB for long-term data storage. • The team leader is responsible for data sharing and implementing the Data Management Plan (DMP). Long-term storage on the institute's servers is under the responsibility of the system administrator.
10	• The data will be stored on an HDD, and this does not require payment • Simone Pani and his PhD supervisor
11	• I will pay with my UJI2023-29 project. • Maria Angeles Marques Torrejón
12	• None of the used repositories charges for their services. • Paola Caria.

5. Data security

- *What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?*

1	• Data will be classified and categorized based on its sensitivity (e.g., public, confidential, restricted). This classification will help in determining the level of security needed for each type of data. Then Role-Based Access Control (RBAC) will be used to ensure that only authorized individuals can access specific types of research data. Access permissions based on the role of the individual in the research project will be performed. In addition, the use of strong passwords will be used for systems that store or manage research data. Password managers can be used to keep track of complex passwords. Sensitive data are not present since the samples are exclusively from an animal model. • ONE drive UNICA with password then after publication the data will be available in the open access manuscript.
2	• All data deposited in the NCBI SRA are cross-deposited to the European Nucleotide Archive and the DNA Data Bank of Japan. Secure storage is guaranteed all each of their server. • Project members use their password-protected laptops. Data are stored within the secure servers of the institution.
3	• The data deposited in the Institutional Repository of the Universitat Jaume I will be preserved following the policy and security mechanisms of this Repository. • Non-public data will be stored on a secure University web server and backed up to national research cloud services.
4	• Data are stored on our high-performance computing cluster. The data can only be accessed by members of the AG Wendler. Raw data will be saved on external hard drives and tapes – they are furthermore deposited to EMPIAR. • Maps and models will be deposited in a publicly available repository EMDb and (PDB) certified by CoreTrust Seal.
5	• Data are stored on the encrypted institutional server. The university IT department carries out regular backups to ensure data security. Only researchers working on the project have access
6	• Data will be regularly backed up and stored in secure, encrypted systems, both on-site and in cloud-based repositories, adhering to institutional security protocols. • During data transfer, encryption will be applied to prevent unauthorized access. • Lastly, a data recovery plan will be in place to address any potential data loss, ensuring quick restoration from backup systems if needed. The data are stored in a hard disk.
7	• During the research phase, data will be accessible only to authorized team members and stored on secure institutional servers with encrypted backups. Upon publication, processed datasets and metadata will be deposited in a trusted open-access repository. • To ensure data recovery, regular backups of all collected and processed data will be performed and securely stored in two physical locations: one copy at Vasyl Stefanyk Precarpathian National University (PNU) and one at Masaryk University.
8	• Data files will be stored on a dedicated, password-protected hard drive to ensure data security. • The data will be regularly backed up to multiple secure locations. These backups will include both raw data and associated metadata.
9	• All computers used to process the data are encrypted, as required by the CNRS, and have local backups. In case of an incident, machines can be backed up, and data can be

	retrieved from Drobo racks. For long-term storage, data are stored on the institute's servers located on campus. - Backups and archival storage: Older versions of datasets will be preserved on the institute's servers for long-term reference while ensuring the latest validated version is clearly marked for analysis. Raw data are stored on Drobo hard drive racks configured as RAID to prevent data loss in case of hard drive failure. - Additionally, all data are copied onto hard drives located in different rooms.
10	- Data will be collected in local HDD, where they will be accessible in local only to my supervisor and I. - Data access will be granted by the project coordinator to researchers involved in the project. For these experiments, data will be stored in local hard disk in simple format as JPEG/TIFF. The raw data will be uploaded to the external device delivered or they will be available for secured (encrypted and authenticated) data transfer over the wide-area network, with a specific procedure provided by CEITEC itself.
11	Data will be kept in a drive where only the research team can access it (Google Drive).
12	- Project members use their password-protected laptops. Data are stored within the secure servers of the institution. This includes backup and restore of data on the data storage for the servers provided. 7 versions of the backup are created. Deleted versions of backups are kept for 90 days until final removal. Depending on the agreement, backups are made daily or weekly to the connected tape library with automated logs of these processes for independent review.

6. Ethics

- Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?*

1	The project has received approval for animal model treatments from the technical committee (MSMT-20810/2023-3).
2	Not relevant
3	Not relevant
4	Not relevant.
5	Not relevant
6	Not relevant.
7	The research will be conducted under the approval of relevant ethical review boards to ensure that animal welfare standards are met
8	Ethical approval for the study was obtained from our local committees (Prot. NP/2022/1373).
9	Not relevant.
10	Not relevant
11	The data is checked by the ethics committee at the university.
12	Not relevant

7. Other issues

- *Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?*

1	• AgID, a guideline for Open Data, Italian Open Data License (IODL)
2	• Open Science Guidelines of the University of Potsdam: publish.UP Open Science Guidelines of the University of Potsdam
3	• The Code of Good Practice in Research and Doctoral Studies of the Universitat Jaume I, which ensures a legal and ethical use of data, is followed. • The policies and procedures followed are those corresponding to the funding of this project and those of the Universitat Jaume I.
4	• We follow the guidelines set up by the DFG • Open Science Guidelines of the University of Potsdam: publish.UP Open Science Guidelines of the University of Potsdam
5	• https://www.usn.no/english/about/strategy-and-focus-areas/open-science-policy
6	• UNICA's Open Science Strategy https://en.unica.it/en/research/research-evaluation/open-science
7	• The project will follow the institutional research data management policy and national open science guidelines. PNU's Open Science Strategy will be applied where relevant to ensure compliance with FAIR principles.
8	• UNICA's Open Science Strategy https://en.unica.it/en/research/research-evaluation/open-science
9	• The CNRS has an Open Science strategy that focuses on four main themes: open access to publications, sharing research results (including data), developing tools, and promoting data reusability and interoperability. A formal Research Data Plan moreover commits to implementing FAIR data principles (Findable, Accessible, Interoperable, Reusable), see science-ouverte.cnrs.fr. CNRS's HAL platform is broadly utilized across French universities (including the University of Rennes) for open archiving of publications (https://hal.science/). Additionally, the University of Rennes has launched a dedicated Science Ouverte (Open Science) portal, embedding Open Science principles (scienceouverte.univ-rennes.fr).
10	• Open access goals will be pursued through the establishment and maintenance of the institutional repository for research outputs, UNICA IRIS.
11	• The policies and procedures followed are those corresponding to the funding of this project and those of the Universitat Jaume I.
12	• Open Science Guidelines of the University of Potsdam: publish.UP Open Science Guidelines of the University of Potsdam.

8. Conclusion

The document provides a comprehensive summary of the data collection and measurement approaches undertaken as part of the 1st Call for Access to Research Infrastructures, covering 12 distinct research topics.

The next update is scheduled for Month 36 and will encompass research activities conducted in the second half of the EDUC-WIDE project.