D10.1

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.	D10.1
File Name	D10.1 Data Management Plan
Work Package	WP10
Dissemination Level	Public
Contractual Submission Date	August 2024 (M6)
Actual Submission Date	30th August 2024
Name responsible Institution	Masaryk University (MUNI)
Key Words	Data Management Plan (DPM), FAIR data, repository, data security.
Abstract	Plan for data management within the project. The update of DMP will be submitted twice during the project lifetime (M18, M36).

Co-funded by the European Union

Funded by the European Union under Grant Agreement No. 101136533.

History of changes

Version	Publication Date	Change
0.1	24.04.2024	Initial draft
0.2	11.6.2024	Correction and revision by MUNI Data Management Specialist
0.3	18.7.2024	Revision by partners (USN, UR)
1.0	29.8.2024	Final version

List of Contributors

Beneficiary	Name	Author/Contributor /reviewer
MUNI	Iva Sedláková	Author
MUNI	Michal Růžička	Reviewer
UNS	Mary Anderson-Glenna	Reviewer
UNICA	Erika Orrù	Reviewer

Abbreviations

EDUC – European Digital UniverCity
R&I- Research and Innovation
RI – Research Infrastructures
OSEG – Open Science Expert Group
EGRI – Expert Group on Research Infrastructures
RA – Research Assessment
WP – Work package
ZENODO – a general-purpose open repository
SOP – Standard Operating Procedure

List of Tables

Table of Contents

1. Introduction

The European Digital UniverCity (EDUC) counts among the pioneer alliances of the European Universities Initiative. Performance gaps between EDUC members from Widening and non-Widening countries persist, especially in R&I and the adoption of ERA policies. The EDUC-WIDE aims to reinforce this community by reducing the gaps between its members from “old” (Advanced) and “new” (Widening) EU Member States. EDUC-WIDE will facilitate the upgrading of the R&I culture of the EDUC University Alliance to match the current priorities of the European Research Area (ERA).

A Data Management Plan (DMP) is an essential document that outlines how data will be handled during and after a project. It ensures that data is properly organized, stored, preserved, and accessible, facilitating data sharing and reuse.

This document provides a clear and comprehensive approach to managing data, ensuring it is handled efficiently, securely, and ethically throughout the project's first months (approx. 12 months). The initial version of the document covers coordination and support measures – how we store and share analytical data, event participants, etc.

When Seed Projects (M16+ and, to a lesser extent, EDUC Fellowships and RI Access Projects) are selected, new elements will be added. MUNI will review and update DMP every six months.

2. Data Summary

Will you re-use any existing data?

EDUC-WIDE builds on the previous EDUC-SHARE project (02/2021–02/2024).
The EDUC-SHARE did not generate scientific data and results; it produced knowledge and experience-sharing tools developed to facilitate interactions among EDUC Alliance institutions.

However, we will re-use the EDUC-SHARE datasets that are available via https://www.educalliance.eu/. The dataset can be used in the provided format without any conversion needed.

What types and formats of data will the project generate or re-use?

In the first half of the project, we will be using the following data formats and types:

- DOCX, XLSX, PPTX, PDF
- PNG, JPEG, TIFF, HEIC, GIF, BMP
- SVG, AI, EPS, PSD, HTML, CSS, JS, JSON, XML

These are standardized formats suitable for long-term archiving.

Table 1. Overview of the collected or generated information per work package:

Work Package	Information collected/generated	Information published	Data formats
WP1+2 Open Science	• EDUC-SHARE strategy • Practice in support of Open Science in EDUC • Recommendations on Open Science support • Citizen Science Projects	• Reports • Meeting minutes	• docx, xlsx, pptx, pdf
WP3+4 Research Assessment	• Research assessment methods at EDUC partners • National/institutional practices in RA in Widening countries	• Reports • White Paper • Meeting minutes	• docx, xlsx, pptx, pdf
WP5+6 Career Diversity	• Research career advisory services methodology • GEP agenda at EDUC universities • EDUC Fellowhips call documentation	• Reports • Annotation of Training • Meeting minutes	• docx, xlsx, pptx, pdf
WP7+8 R&I Collaboration	• EDUC-SHARE Research infrastructure catalogue • Research Infrastructures catalog/OpenUp • Call documentation	• Report • Database of Research infrastructures • Publications • Meeting minutes	• docx, xlsx, pptx, pdf
WP9 Dissemination and Communication	• Online content: social media, posts on the website, and websites and social networks • Project propagation materials: newsletter, project logo	• Reports • Templates, logo • Communication and dissemination guidelines • Newsletters • Website	• docx, xlsx, pptx, pdf • jpeg, tiff, png, gif, bmp, heic • ai, svg, eps, psd • html, css, js, json, xml
WP10 Project Management	• Financial reporting, risks • Ethical issues	• Reports, Meeting minutes • Deliverables (as in the DoA)	• docx, xlsx, pptx, pdf

How data in our subprojects will be handled

In the framework of R&I cooperation, EDUC-WIDE will launch open calls (T7.2 Call for Seed Projects, T7.3 First Call for Access to Research Infrastructures, T8.3 Second Call for Access to Research Infrastructures) in which researchers will submit (sub-)projects. Recipients of Seed Projects (T8.2) will be obliged to deliver a simple Data Management Plan covering the results of their research (with support from their respective OSEG members). The resulting publications will have to be published in Open Access mode.

Data from the EDUC-WIDE research activities (Seed projects) will be covered in the Mid-term Data Management Plan (M18) and Final Data Management Plan (M36).

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

As EDUC-WIDE is funded by the Program "Widening participation and spreading excellence," the main focus is to align the performance gaps between EDUC members from non-Widening (Advanced/old) and Widening (new) European countries in EDUC-WIDE, specifically 1) Masaryk University, Czechia; 2) University of Pecs, Hungary and 3) Vasyl Stefanyk Precarpathian National University from Ukraine.

All data will be used to align the gaps between universities in focus areas of 1) Open Science, 2) Research Assessment, 3) Career Diversity, and 4) R&I Collaboration; among others, recommendations on Open Science, strategy for reforming Research assessment, or Report on Capacity building towards Professional Support for Career Diversification and Gender Equality.

What is the expected size of the data that you intend to generate or re-use?

We will have only a small amount of data stored (GiBs).

To whom might your data be useful ('data utility'), outside your project?

Data will be useful to researchers, administrative staff, and students at all partner universities of the Alliance as they will be used to create tools for internal evaluation or recommendations.

In WP1 (Open Science), we will start by analyzing the practice in EDUC and globally (T1.2) and proceed to the formulation of recommendations (T1.3). The recommendations will be directed not only to the EDUC Alliance but also to other university alliances and higher education institutions within Europe and globally.

Data collected in WP2 (Research assessment) will support the development of a strategy towards the reform of research assessment in EDUC Universities in compliance with the principles of CoARA.

Research assessment is firmly grounded in national systems and policies. An environment fostering excellence in science needs to be facilitated by national regulations, incentives, and financing. We will prepare a white paper firmly grounded in good practice gathered from across EDUC and will discuss the practice directly with policymakers and stakeholders in Widening countries. (T4.3)

3. FAIR Data

3.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

A persistent identifier will be added to the data (documents) when placed into the long-term repository (e.g., Zenodo). However, the project consortium will discuss the option of assigning the identifier to all deliverables directly by MU institution.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Standard metadata responds to the formats mentioned in Section 2 (Data Summary).

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

We will be keeping the relationships between data precise in the file names.

We made a SOP (Standard Operating Procedure) for file naming:

project acronym_file specification (deliverable no./name of the report)_responsible partner acronym

E.g., for example:

EDUC-WIDE_D10.1 Data Management Plan_MUNI

Will metadata be offered in such a way that it can be harvested and indexed?

Metadata will be available in a form that can be harvested and indexed (managed by the used repository/repositories).

3.2 Making data accessible

Work-in-Progress Storage

Project data and outcomes will be stored in institutional data storage (institutional Microsoft M365 cloud space of Masaryk University), a secured institutional cloud service with adequate legal agreements, professional technical maintenance and operations, and sufficient access control. This environment is primarily a platform by Masaryk University for MU staff for document management, which is the primary type of data in our project. Permission setting (for access to partners) is possible from the top level to the level of individual documents and lists.

Repository

Will the data be deposited in a trusted repository?

The long-term repository Zenodo (where data are assigned metadata and are publicly available) is planned to be utilized.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Assigning an identifier is not a standard procedure in Zenodo.

Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

We will work with the philosophy as *open as possible* for our data.

EDUC-WIDE will produce 23 deliverables. The data will be published as soon as possible after collecting it. All of them will be public and accessible on the project website or archived on Zenodo and linked via DOI from the project website.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No embargo will be used, and all data will be opened immediately.

Will the data be accessible through a free and standardized access protocol?

Data will be openly available on the project website. Usually, it can be accessed via a standard HTTPS protocol.

Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

All our metadata can become completely open over time (website, Zenodo). CC-BY licensing will be applied.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data will be available for at least five 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 longer as no significant data volumes will be preserved (there are small costs of preserving them).

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

No specific software is required to access the data.

3.3 Making Data Interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

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 (see Section 2 Data Summary).

3.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

The output data are mainly textual reports to be read by humans. For validation purposes, all the sources used will be properly cited, and the methodology used will be included in the documents.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

As stated in Section 3.2, all of our data can become completely open over time.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Data produced and used in the project will be useable by third parties.

Will the provenance of the data be thoroughly documented using the appropriate standards?

The versioning of the document and the post-implementation of the work will be recorded directly in the documents.

Describe all relevant data quality assurance processes.

The EDUC-WIDE Project Management Group will establish the process of internal reviews of outputs (documents). Prior to publication/submission as the official project output, the document will be approved by all partners/experts from each university (Approval document). These comments and modifications should be considered and implemented/rejected by the original authors. The Lead Project manager will do the final check.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

FAIR is a central part of our data management plan; it is considered in every decision. We use the FAIR data process to make our use of the data as efficient as possible. Making our data FAIR is, therefore, not a cost that can be separated from the rest of the project.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

None of the used repositories charge for their services.

Who will be responsible for data management in your project?

The Lead Project Manager from MUNI, Iva Sedláková (ORCID identifier 0009-0003-3515-0359), and work package leaders are responsible for the data management.

To execute the DMP, additional specialist expertise from MUNI was applied as we have such trained support staff available.

Recipients of Seed Projects (T8.2) will be obliged to deliver a simple Data Management Plan covering the results of their research (with support from their respective OSEGmembers), therefore Seed Projects researchers would be responsible for data.

How will long-term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

The minimum lifetime of the archive is five years.

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)?

Project members use their password-protected laptops.

Data are stored within the secured servers of each institution. All data centers where project data is stored carry sufficient certifications.

All project web services are addressed via encrypted HTTPS connections. The EDUC project website is hosted by FERLING Ltd. on a VPS rented at RakcForest Ltd.

The OpenUp platform will be used to implement internal Seed project calls (registration and submission). The platform is operated by the University of Pecs in compliance with the provisions of Regulation (EU) 2016/679 of the European Parliament and the Council on the protection of individuals. The data gathered during the registration (the username, first and last name, e-mail address, and type of user) are gathered with the subject's explicit and voluntary consent.

The University of Pecs ensures adequate security of the data subject's personal data, including protection against unauthorized or unlawful processing, accidental loss, destruction, or damage, by implementing appropriate technical and organizational measures.

We are running the project in collaboration with different groups and institutes. A collaboration agreement that describes who can have access to what data in the project is set.

Project members have been instructed about the project's generic and specific risks.

Will the data be safely stored in trusted repositories for long term preservation and curation?

We will be archiving data for long-term preservation during and beyond the project in suitable repositories (Zenodo, Czech National Data Repository, for example).

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The EDUC-WIDE Ethics Advisor will check abstracts and Ethical Issues in proposals for Seed Projects (M12), EDUC Fellowships (M22+), and RI Access (M8, M25) and point out potential risks.

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

We do not collect any research data connected to individuals, i.e., “personal data.”

Data gathered within the OpenUp platform will not be provided without the data subject’s explicit and voluntary consent. The name and e-mail address of the data subject for the purposes of sending information mailings based on the data subject’s explicit and voluntary consent as well. The processing of personal data will continue until the subject’s consent is withdrawn or until the extent and duration necessary for the purposes for which the data is processed.

During the project, each partner will process personal data by providing informed consent for data sharing and long-term preservation to interested persons. Each partner has a Data Protection Officer who is responsible for data processing. Partner institutions will collect and share data with the other project partners and store them on a secured platform.

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)?

We will be using the following policies and procedures for data management: **MU Open Science Strategy 2022–2028** (https://openscience.muni.cz/media/3477939/en-a-strategicka-cast-strategie-open-science-mu-2022-2028-uprprebal.pdf). Strategic framework for effective implementation of Open Science at Masaryk University (Coordinator).

We use the Data Stewardship Wizard with its *Common DSW Knowledge Model* (ID: dsw:root:2.6.3) knowledge model to make our DMP. More specifically, we use the https://dsw.muni.cz:443/wizard DSW instance where the project has a direct URL: https://dsw.muni.cz:443/wizard/projects/79e01e5f-2b8e-4022-8516-9ff659a9f8de.

8. Conclusion

The document introduces the project EDUC-WIDE data management approach. It provides an initial analysis of the data sources used or generated during the project's initial phase – approximately—the first 12 months.

The current document will be a living document throughout the project. The DMP will be updated regularly every six months, submitting the Mid-Term Data Management Plan (M18) and Final Data Management Plan (M36) as deliverables.

Within Seed Projects (and, to a lesser extent, Call for Access for Research Infrastructures and EDUC Fellowships), the Data Management will have to handle proper scientific data, metadata, ethical issues, and future publications that will arise from them.