Ref. Ares(2023)3328039 - 12/05/2023

Project number: 101082506

Project acronym: NIAGARA

Project title: Navigating the digital landscape: universities partnering for change

DATA MANAGEMENT PLAN

VERSION 1.0

Date: 12.05.2023

Prepared by:

Mahmood Khosrowjerdi (INN)
Kelefa Mwantimwa (UDSM)
Marek Deja (UJ)
Dorota Rak (UJ)
Andrew W. Malekani (SUA)

Co-funded by the European Union

About the project

The objectives of the NIAGARA project directly address digital transformation, inclusion, and diversity, and will impact participation in democratic life and civic engagement through a focus on digital competence. The connection between digital competence and participation in democratic life and civic engagement is well-established, particularly regarding issues such as misinformation, disinformation, and fake news. Indirectly and in the long term, the grant will likely contribute to addressing issues related to the environment and climate change by supporting education and learning at the tertiary level, as well as research. In particular, the NIAGARA project aims to enhance digital resources within university libraries across Tanzania and Zambia, ensuring equal access to library services for all students and staff. In addition, we will examine the essential digital skills required by both librarians and students to deliver efficient library services. To support their development, we will provide short courses tailored to enhance these digital capabilities. The NIAGARA project aligns with international, EU, national, and institutional priorities, making it a vital initiative for academic progress.

Project Profile

Project Title	Navigating the digital landscape: universities partnering for change
Project Acronym	NIAGARA
Funding Agency	Co-funded by European Union
Funding Scheme	Erasmus+ Grant
Project Number	101082506
Project Partners	1. Høgskolen I Innlandet (INN), Norway 2. Oslomet–Storbyuniversitetet (OSLOMET), Norway 3. University of Dar es Salaam (IMS), Tanzania 4. Sokoine University of Agriculture (SUA), Tanzania 5. University of Zambia (UNZA), Zambia 6. The University of Barotseland Limited (UBL), Zambia 7. Uniwersytet Jagiellonski (UJ), Poland
Start Date	02 January 2023
End Date	02 January 2026
Duration	36 Months
Project Manager	Elin Opheim
Project Coordinators	Jamie Johnston (OsloMet), Kelefa Mwantimwa (IMS), Wylustan P. Mtega (SUA), Christine W. Kanyengo (UNZA), Greg Kwalombota Mubita (UBL), Sabina Cisek (UJ)

Deliverable Information

Deliverable Number	D.2.2
Deliverable Title	DMP
Work Package (WP) Number	WP 2
WP Coordinator	UDSM
Contribution of Partners	The names/affiliations of contributors are printed in the title page.
Nature of Data	report/plan
Delivery Date	12.05.2023

1. Data Summary

A Data Management Plan (DMP) is one of the key elements to effectively manage project data. Generally, the DMP describes the Data Management Life Cycle for the data to be generated and processed by the project. The main objective of NIAGARA's Data Management Plan is to facilitate project data generation and management during and after the end of the project. DMP enhances effective data management including proper recording, documentation, reliable storage, and dissemination and sharing. Specifically, the DMP describes the purpose of data collection and its relationship to project objectives, types and formats of data, origin of data, data collection methods and instruments, data processing and analysis. Along with these, it entails data storage, preservation and archiving, software/database/repository, metadata, data publications and sharing, Data FAIR principles, data quality and security, and role and responsibilities of the partners. DMP is a living document and can be revised as needed during the course of a project.

During the lifetime of NIAGARA, various types of data would be gathered to contribute to the aims of this project. The data could be primary and secondary. The major sources of primary data in NIAGARA are surveys, interviews, and observations. The gathered data are expected to provide a basis for deeper insights on digital competencies and data management skills of librarians and LIS students. The gathered data would be the major building blocks of WP2, WP3, WP4, and WP7. Specifically, the data and information generated contribute to the development of digital literacy and data management curricula for in-service librarians from partner universities, recommending digital literacy and data management course curricula and contents for LIS degree and non-degree students in partner universities, developing interactive learning materials on digital literacy and data management for in-service librarians and LIS students.

NIAGARA will generate a variety of data (with various formats) via methods such as interviews, surveys, and observations over its three years (2023-2025). The data would be in various formats such as audio (interview audio files, in formats such as mp3), text (interview transcriptions, in formats such as .doc, .docx, .ppt, .pptx, .pdf), statistical data (tabular/textual data of surveys, in formats such as .sav, .csv, .xls), video (e.g., courses on digital competencies and data management skills, in formats such as .avi,), illustrations (.jpeg, .png), drawings (.dwg), and so on.

We could use publicly available data (secondary information/data) such as policy documents, curriculums, and online informational sources. We would open data licenses before using/incorporating them into our analyses/reports.

The origin of the secondary data that would be used in this project is not clear yet. However, if we use the openly available data, we will follow the good practice in academia and follow the licenses of the data for reuse. The gathered data via surveys would often result in small-sized files. The interview audio files would be around 500 megabytes per 60 minutes.

The data of the NIAGARA project would be useful for partner institutions, relevant institutional stakeholders (e.g., EU and HEIs) and groups of people such as LIS students, faculty librarians, and future researchers. In a narrow sense, data can be important in the context of generating new scientific publications and new knowledge. In a broader sense, 'data utility' can potentially contribute to supporting the decision-making processes of the management staff, improving the institutional development strategy, increasing access to digital sources and resources, or improving human life etc.

2. FAIR data

2. 1. Making data findable, including provisions for metadata

The data produced and/or used in the NIAGARA project will be deposited in DataverseNO and discoverable with metadata and identifiable Digital Object Identifiers (DOIs). All data files would get a unique DOI. DOIs provide a persistent and unique identifier for digital objects and enable reliable access to the data. The metadata associated with each dataset will include information such as the title, author(s), contact information, description of data, subject and keywords, and related publication (if the data is used in an academic publication or report), language, funding information, distributor, depositor, deposit date, time period, date of collection, data type, and software used to analyze the data.

In the NIAGARA project, we follow the naming conventions of DataverseNO (https://dataverse.no/). We would name the data files consistently. The data could be named using work packages numbers, deliverables code from grant agreement and the university or institution gathering data for the particular work package. This will help to ensure that the data is easily identifiable and associated with the appropriate project components. This will help to find the right data files among collaborators of NIAGARA too. Specially, we follow the recommendations of DataverseNO as follows:

- Using descriptive names for files, but short (less than 25 characters)
- Spaces are not used in the file names. We use instead the underscores (e.g., wp1_survey), hyphens (e.g., wp1-survey), or camel case (WP1Survey)
- We would not use special characters such as \ / ? : * ” > < | : # % ” { } | ^ [ ] ` ~ æÆ øØ åÅ äÄ öÖ in file naming.
- We would use the international date convention YYYY-MM-DD (e.g., 2017-10-25) in file naming.
- We would use identical file names for different formats of the same data.

For the spreadsheets, we follow the recommendations of DataverseNO as follows:

- One table = one file (one spreadsheet)
- One column = one variable

- One row = one observation / sample
- One cell = one value / piece of information
- The first row is the header including variable names.
- Variable names must not include special characters, spaces, or start with a number.
- Using the international date convention YYYY-MM-DD (e.g., 2017-10-25).

A general pattern for file naming of NIAGARA would be as follows: WorkpackageNumber_InstitutionAbbreviation_DeliverableNumber_VersionOfData_Date

For example, a data file (that is the "deliverable number 6" and) produced by JU (WP7) would be named as following: WP7_JU_D6_V01_2023-02-22.

In order to optimize possibilities for re-use of data, the keywords may be derived from the example publication and its associated datasets. The keywords would be added to the metadata associated with each dataset.

We provide clear version numbers for each dataset to ensure that final users can easily track and understand any changes made to the data over time. This will help to maintain data quality and ensure reproducibility of research findings. The version would be added to the file names as follows: v01 or V01 (i.e., the first version of the data file). For example, WP7_JU_D6_V01_2023-02-22.

Metadata standards provide structured descriptions of datasets, making them more easily discoverable and understandable by other researchers. We aim to publish the final publishable datasets in DataverseNO, and this data archive follows the Data Documentation Initiative (DDI) standard. This standard is popular in describing social science datasets, covering elements like study description, data file description, variable-level metadata, and question-level metadata.

2.2. Making data openly accessible

The data gathered via interviews (just the anonymized transcriptions) and online surveys (anonymized) would be published open access as default (but at proper time after the end of project in accordance with the EU regulations).

There are several platforms and data repositories available that facilitate the dissemination of research data. The NIAGARA data would be deposited in DataverseNO.

DataverseNO (https://dataverse.no/) is a national, generic repository for open research data, owned and operated by UiT (The Arctic University of Norway). DataverseNO is aligned with the FAIR Guiding Principles for scientific data management and stewardship. The technical infrastructure of the repository is based on the open-source application Dataverse, which is developed by an international developer and user community led by Harvard University. DataverseNO is CoreTrustSeal certified.

The data files of NIAGARA would be deposited in at least two versions: a use-friendly version (such as .xls, .sav, and so on) and a preferred file format, that is, non-proprietary, open (with documented international standards using standard character encoding, preferably Unicode (e.g., UTF-8), and uncompressed (space permitting). To access the user-friendly version the users will need software tools

that are capable of working with those data formats. For example, for survey data, users would need tools that can handle spreadsheet formats such as .CSV, .sav, and so on.
To access (open and analyze) most of the preferred file formats (such as R codes and raw data files) the users can use open access software to open and work with the data.

The information about the required software to access the data would be included in the metadata description of data. This is usually done during the depositing process.

The NIAGARA project will use DataverseNO to deposit the data. The data deposited in DataverseNO will be made available in compliance with the FAIR principles. By default, the deposited datasets will have a creative commons license (CC0 or CC BY 4.0) assigned allowing users to reuse the dataset without restrictions or obligations. The selected repository and the applied license will allow the NIAGARA project to ensure that its research data are well-documented, widely recognized, interoperable, and easily accessible for other researchers and institutions, facilitating collaboration and knowledge sharing.

The Inland Norway University of Applied Sciences (INN) has an institutional agreement with DataverseNO and depositing data in this archive would be possible without any additional costs to the project. The INN institutional curator of research data is part of the contributors of the NIAGARA project, and this could facilitate the process.

A data access committee may not be necessary for the NIAGARA project. The data will be anonymized and published under open licenses that only allow for identifying institutions, which means that there are no identifiable individuals or sensitive information that need to be protected. As such, the data can be made widely available without the need for a committee to manage or restrict access. However, it is important to ensure that the data is well-documented and properly formatted to ensure that it can be easily accessed and understood by other researchers and stakeholders.

2.3. Making data interoperable

The data produced in the NIAGARA project are designed to be interoperable, allowing for data exchange and re-use between researchers, institutions, organizations, and countries. The project adheres to standards for formats and aims to be compliant with available (open) software applications. This approach facilitates the re-combination of datasets from different origins, enhancing collaboration and knowledge sharing among various stakeholders involved in the project or interested in the project's findings.

To achieve this interoperability, the project follows the Data FAIR principles, which ensure that the data are Findable, Accessible, Interoperable, and Reusable. By adhering to these principles, the NIAGARA project ensures that its data are well-documented, stored in widely recognized and compatible formats, and shared through appropriate platforms, allowing other researchers and institutions to access and utilize the data efficiently.

To ensure that data related to surveys, recorded interviews, and interview transcripts are interoperable, it is essential to use widely recognized and compatible formats. We would follow the preferred file formats of DataverseNO as follows.

By using “preferred file formats”, we can improve the interoperability of data, making them easier for other researchers and organizations to access, analyze, and re-use the data in their own work.

DataverseNO utilizes Data Documentation Initiative (DDI) metadata schema, a widely recognized standard for describing social science datasets. This adherence to the DDI metadata schema allows Dataverse-stored datasets to be indexed by Google Dataset Search and other search engines. Additionally, Dataverse assigns DOIs to datasets for enhanced discoverability and long-term access.

Preferred file formats for long-term archiving in accordance with DataverseNO

File type	Preferred file formats (examples)
Audio	• Uncompressed and lossless Wav or AIFF (.wav/.aiff) • Compressed and lossless FLAC (.flac) • Compressed and lossy Mp3 (.mp3)
Container file	Container files are automatically unpacked when uploaded.
Image	• Uncompressed TIFF (.tif or .tiff) • Compressed and lossless PNG (.png) • Compressed and lossy JPEG (.jpg)
Text (slides, illustrations)	• PDF/A (.pdf) combined with original file
Text (tables)	• Tab separated Unicode plain text (.txt)
Text (text)	• Plain text (.txt, .md) If formatting needed: • XML, PDF/A (.pdf) combined with original file
Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es
Transcription	File format: • PDF/A (.pdf) combined with original file • PDF/A (.pdf) combined with Comma/Tab Separated Values (.csv/.txt) Font: • Unicode IPA (e.g. Charis SIL, Doulos SIL, Gentium Plus, Andika), ASCII SAMPA
Video	• MPEG-4 (.mp4)
Array data	• netCDF (.nc)
Statistical analysis	• R (.R, .RData) • SPSS (.dat/.sps) • STATA (.dat/.DO)
Qualitative data analysis	• Basic data in preferred file format, e.g., PDF/A, plain text in Unicode (.txt) • Analysis dump/package as REFI-QDA Project (.qdpx)
Workspace dump formats for mass spectrometry	• mzML (.mzML)

To make our data interoperable, it is essential to follow widely accepted data and metadata vocabularies, standards, and methodologies. These ensure that our data can be easily understood,

accessed, and used by other researchers, institutions, and organizations. For the NIAGARA project we would follow the DDI metadata standard for describing social science datasets, covering elements like study description, data file description, variable-level metadata, and question-level metadata.

Controlled Vocabularies and Ontologies: Using controlled vocabularies and ontologies in metadata ensures consistency and enhances discoverability. There are controlled vocabularies and ontologies more appropriate for our project's goals related to librarianship, social science, digital transformation, and information literacy. Some examples of controlled vocabularies that could be usable in describing the gathered data via NIAGARA include:

- Library of Congress Subject Headings (LCSH): LCSH is a widely used controlled vocabulary for subject indexing in libraries. It covers various subject areas, including social sciences, digital transformation, and information literacy. LCSH can help us describe and categorize data more effectively and consistently. URL: https://www.loc.gov/aba/cataloging/subject/
- ERIC Thesaurus: The ERIC (Education Resources Information Center) Thesaurus is a controlled vocabulary specifically designed for education-related research, including topics like information literacy, digital competence, and library science. It can help describe our data more precisely and make it more discoverable in education-focused search engines and repositories. URL: https://eric.ed.gov/?ti=all

We would use those parts of subjects/terms related to NIAGARA that are included in relevant controlled vocabularies such as LCSH or ERIC in the metadata field of the deposited data in DataverseNO (if applicable).

2.4. Increase data re-use (through clarifying licenses)

To maximize data re-use, we will license the data using permissive licenses that allow for the widest re-use possible. One common choice for this purpose is the Creative Commons licenses, such as CC BY 4.0 (Attribution), which requires users to attribute the source when using the data but allows for free sharing, adaptation, and use for any purpose, including commercial. Other Creative Commons licenses can be chosen based on the specific requirements and restrictions deemed necessary by the project team.

The data will be made available for re-use as soon as possible during and after the completion of the project or publication of the research results in scientific journals. If an embargo is necessary to give the project team enough time to publish results, we will specify the duration and justification for this embargo depending on funder requirements. It is usually after the end of a project (no later than 12 months). However, we aim to minimize the embargo period to promote timely data sharing and collaboration.

The data produced and/or used in the project will be usable by third parties, especially after the end of the project. We strive to ensure that our data is well-documented, stored in widely recognized and compatible formats, and shared through appropriate platforms to facilitate access and re-use. If the re-use of certain data is restricted due to confidentiality, ethical, or legal reasons, we will clearly explain these restrictions and make efforts to share aggregated, anonymized, or redacted versions of the data when possible.

The data's reusability will be maintained for as long as it remains relevant and useful to the research community. The duration may vary depending on the nature of the data related to the project's objectives, and the evolution of the research field. In the case of quantitative data, we will store the data in DataverseNO that ensures long-term preservation and accessibility for a minimum of 10 years. For qualitative data, we will adopt similar strategies, with a focus on preserving the context and meaning of the data, as well as ensuring that any necessary anonymization or privacy protection measures remain in place.

3. Allocation of resources

The depositing of data files of NIAGARA in DataverseNO is free of charge for the project. The data curation would be done via an expert curator before publishing the datasets. The researchers printed as contributors on the title page would be responsible for data management of the NIAGARA project.

The datasets published in DataverseNO are discoverable and openly available to anyone with access to the Internet. DataverseNO is responsible for secure archiving and data availability for at least 10 years after being assigned DOIs.

4. Data security

In the case of the NIAGARA project, the acquisition and processing of sensitive data for which, for example, the General Data Protection Regulation (GDPR) could apply, is not envisaged. For other areas of data handling, the following regulations and recommendations will be followed: Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L1024), UNESCO Recommendation on Open Science (https://tinyurl.com/em2zn5sw) and the Declaration on Access to Research Data from Public Funding (OECD Legal Instruments, https://legalinstruments.oecd.org/en/instruments/157).

The method of securing data deposited in the repository is described in detail in DataverseNO Preservation Policy (https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/). The guarantee of ensuring the security of data stored in this repository results from the adopted international legal regulations as well as standards and good practices, for example the FAIR Guiding Principles.

5. Ethical aspects

In the NIAGARA project, there are no plans to use, share, or preserve personal data. As such, there are no ethical or legal issues related to data sharing that could arise from the project. The project team will ensure that all data collected, stored, and shared are done so in accordance with relevant ethical and legal requirements. Individual cases of data collection will be registered with Norwegian Agency for Shared Services in Education and Research (Sikt) / Norwegian Centre for Research Data (NSD) in the event that any directly or indirectly personally identifying information is to be gathered in the process of doing interviews, conducting observations and/or administering questionnaires.

Another issue in the field of ethics is licenses under which data can be used and processed. In accordance with the open access policy, NIAGARA data which is going to be deposited in DataverseNO will be subject to Creative Commons licenses, in particular CC0 (public domain) or CC BY 4.0.

Finally, the questionnaires used to collect data will not include questions that could lead to the collection of personal data. Additionally, informed consent for data sharing and long-term preservation of aggregate data will be acquired. Participation in surveys is completely voluntary and the participants have the possibility to stop their participation at any time. The project team will ensure that all data collected and shared are done so in a manner that is respectful of the privacy and confidentiality of individuals. Any potential ethical or legal issues related to data sharing will be addressed on a case-by-case basis.

6. Other issues

We would follow the national/funder/sectorial/departmental procedures for data management such as:

European Commission, Data Management, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Norwegian national strategy on access to and sharing of research data, https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/?ch=1