
Plan Overview

A Data Management Plan created using DMPTuuli

Title: Using digital micro certificates in the recognition and validation of skills of the informal sector in Senegal

Creator: Satu Järvinen

Affiliation: University of Eastern Finland

Template: General Finnish DMP template

Project abstract:

The objective of my research project is to develop an AI-driven system utilizing digital micro-certificates to identify and validate the skills of workers in Senegal's informal sector and enhance their employment opportunities. This project integrates innovative technology with local expertise through collaboration with Senegalese stakeholders and research institutions. The initiative leverages SkillSafari's AI technology, specifically designed to recognize and validate skills acquired in the informal economy.

The research will be conducted in Senegal, with a particular focus on youth and women. The project will develop a digital platform that enables the documentation, assessment, and certification of skills acquired in the informal sector. The ultimate goal is to establish a scalable model that can improve employment prospects for individuals in vulnerable positions.

Funding is critical for the development of this innovative technology and the investigation of its practical impact. This project holds significant potential to advance skills recognition and promote social equity in Senegal and farther in the African continent.

ID: 19345

Start date: 01-01-2022

End date: 31-12-2026

Last modified: 05-02-2025

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Using digital micro certificates in the recognition and validation of skills of the informal sector in Senegal

1. General description of the data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

The research overall is based on existing research and literature in addition to specific research carried out during research. The existing literature is made available through referencing and naturally identifiers (Url, DOI) when available. Also statistical data made available by i.e. government officials may be used, though access is most likely entirely limited to results and not the raw data itself.

Data collection in this research is collected through surveys, interviews and observations. The interview and observation data is transformed into data sets for further examination. During this all data allowing for identification of the people interviewed and/or observed will be removed.

All data will be stored using common file types with focus on open data formats (i.e. csv, odt, mp4). The transcriptions of the video files will be checked and anonymised to make them ready for archiving.

Due to the complicated nature of data collection, complications due to limited access to the target population and language barriers, the amount of data collected will be limited to apr. 20 people from each country (Senegal and Rwanda). Therefore the size of data will not be extensive, the recorded videos will require the most space but still remain small enough to not cause concern for proper storage.

1.2 How will the consistency and quality of data be controlled?

For the surveys, interviews and observations the consistency of data is controlled through having a pre-set data collection method (asking the same questions from each participant) and recording responses under the given headings. That data is then handled and analysed through a selected method that ensured all data is analysed with the same method. Since interviews will be carried out most likely in local languages, employing qualified interpreters and ensures that they 1) understand the the research questions and 2) produce a reliable translation of the interviews.

Quality of the data is approached in some ways: naming of files, structure of repository where files are stored including explanations of data, version control of data, documentation of analysis and partially sharing the data as much as possible for other researchers to review. In this research, all data except video and audio files containing personal information of the research participants will be shared under a creative commons license.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

The most important ethical issue related to my data and it's management relate to personal information of research participants. The surveys are anonymous but interviews and observations are not. Therefore information collected has to be GDPR compliant. Senegal and Rwanda being African countries do not require GDPR registries, but the researcher being European and research being done in Europe makes GDPR compliance a requirement. More largely GDPR ensures the privacy of the research participants is ensured.

To ensure compliance, all participants will be asked to fill a consent form and only people understanding and signing the form will be allowed participation in the research. The form should follow the guidelines of a GDPR registry form. In addition to explaining the general principles of how data is stored and manages, the form will explain that any information containing personal information (video and audio files) will not be shared and only stored in a password protected space for 5 years. The data used for analysis that will be shared does not include personal data, therefore we will ask for consent to use and share it. Also participants will have the right to demand for and access the general research data if they so want to ensure none of their personal data is misused. The consent forms will be prepared in French and English and ideally also in Kinyrwanza and Wolof.

2.2 How will you manage the rights of the data you use, produce and share?

The existing literature and research data is managed through referencing the work and through that ensuring access to it to any interested party.

The data produced will be shared using a creative commons license asking for attribution, denying commercial use and asking to share the used data in the same format. The selection of this license is based on an idea of free data that serves common purposes with attribution to the origin of the data.

By following existing common research practices the research will not break any intellectual property rights. The final work will be the intellectual property of the author and when reused should be referenced to properly. However, apart from a clear case of plagiarism misuses of the data produced and shared and hard to find and contest. Luckily no real harm will be done in such occasions and the final work does not contain patents or any other information that would require special attention be put in the protection of intellectual property. Rather the reuse of the data should ideally help develop more sustainable learning solution for Africa, where too little research data currently exist.

3. Documentation and metadata

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

The formatting of data and the provision of metadata will conform to UEF standards and guidelines (unless some project financier requests something different). This includes clear data description, annotation, contextual information and documentation, e.g. unique identifier for each transcript, cover sheets with details such as date, place and interviewee details. Noting that no personal information is shared, only anonymised data.

The the storage, management, identification and publication of data the Fairdata.fi service will be used.

The data will be organised under specific articles and topics with references to different versions identified in the name (V1, V2...). The name structure will be consistent, i.e. interview datas from Senegal and Rwanda will stored using same logic to that the file name and place in the repository clearly indicated which interview was conducted and where the interviewee comes from. (i.e. "Article_1_Inerview_1_SEN" and "Article_1_Inerview_1_RWA")

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will the data be backed up?

Data will be stored using the Fairdata.fi service which also does regular backups of data. This with the exception of the video files containing personal data and falling under GDPR that will be stored in a UEF provided secure, password protected cloud space and deleted after 5 years.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

I control access to data. Access is granted to i.e fellow researchers working on an article together. The version management systems of the online repository combined with naming different versions of the data ensures it stays compliant.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

All data except of data containing personal information of research participants (interview and observation videos / audios) will be made available and published. All data and metadata are made available in the fairdata.fi service.

5.2 Where will data with long-term value be preserved, and for how long?

The fairdata.fi service has separate tools to support data management, metadata description, safe storage as well as digital preservation over decades, technology cycles and research projects

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management?

Satu Järvinen, Doctoral Researcher, University of Eastern Finland

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

The fairdata services used are offered by the Ministry of Education and Culture in Finland and produced by CSC – IT Center for Science Ltd. Other required are offered by UEF. These resources do not require monetary investments.