



## Deliverable 5.2: Initial Data Management Plan

Deliverable title: Initial Data Management Plan

Deliverable number: D5.2

Planned delivery date: 31/07/2020

Nature of deliverable: Open Research Data Pilot

Lead partner: UPS-UT3

Dissemination level: Public

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement n° 891744





|   |   |
|---|---|
| Acronyms .....  | 3 |
| Description of the deliverable .....  | 4 |
| Introduction .....  | 4 |
| 1) Surveys in rural areas. Participants involved: UPS .....   | 4 |
| 1.1) Data set description .....   | 4 |
| 1.2) Standards and metadata .....   | 5 |
| 1.3) Data sharing .....   | 5 |
| 1.4) Archiving and preservation (including storage and backup) .....                                  | 6 |
| 2) Reportages in schools. Participants involved: APDOC .....  | 6 |
| 2.1) Data set description .....   | 6 |
| 2.2) Standards and meta data .....  | 6 |
| 2.3) Data sharing .....   | 6 |
| 2.4) Archiving and preservation (including storage and backup) .....                                  | 6 |
| 3) Interviews and reportages during workshops. Participants involved: APDOC .....                     | 7 |
| 3.1) Data set description .....   | 7 |
| 3.2) Standards and meta data .....  | 7 |
| 3.3) Data sharing .....   | 7 |
| 3.4) Archiving and preservation (including storage and backup) .....                                  | 7 |
| 4) Other data: pre existing data, data from communication and dissemination activities, reports ..... | 7 |
| 4.1) Data set description .....   | 7 |
| 4.2) Data sharing .....   | 8 |
| 5) Intellectual property rights .....   | 8 |
| Conclusion .....  | 8 |

## Acronyms

APDOC: Association Les Petits Débrouillards d'Occitanie

CA : Consortium Agreement

DDE : Direct explanatory data

DDP: Direct personal data

DDR: Direct results data

DIP: Indirect personal data

DMP : Data Management Plan

DoA: Description of Action

DPO: Data Protection Officer

GA: Grant Agreement

GDPR: General Data Protection Register

INSA: Integrated Nitrogen Studies in Africa

MSCA: Marie Skłodowska Curie Actions

NANI: Net Anthropogenic Nitrogen Inputs

RISE: Research and Innovation Staff Exchange

SOP: Standard Operating Procedure

WP : Work Package

## Description of the deliverable

The INSA project Data Management Plan (DMP) specifies the data management life cycle for all datasets to be generated and collected, the workflow to make the data accessible for use by beneficiaries and the strategy for preservation of the data. This DMP describes therefore the life cycle of all data collected and processed in INSA. It gives also an overview of available research data, standards, access and terms of use of the datasets.

The present report describes the initial INSA DMP, which will be regularly updated over the course of the project whenever changes arise with regard to new research data, methodology, standards, access and terms of use of the data or other external factors.

## Introduction

We present in this section an overview of the datasets that will be collected in the project.

This mainly concerns personal data collection and/or processing, given that the project is mainly based on existing data bases and published data (WPs 2, 3, 4). Personal data refer to videos, reportages and interviews performed during the time span (WP3, 5). Additionally, this will also include a survey on agricultural practices on a village scale and link with social practices in rural areas (WP3).

The document will detail data on (1) surveys in rural areas, (2) reportages in schools, (3) reportages and interviews during workshops. Section (4) deals with other data (pre existing and related to communication and dissemination activities) and section (5) specifies intellectual property rights, background and access rights that need to be considered.

Dataset contents are briefly detailed in sections 1, 2 and 3. Exhaustive details are given in the Annex 1 (Description of Action - DoA) of the Grant Agreement and were declared as a register form by the Data Protection Officer (DPO) of UPS-UT3, to validate the compliance of the project to the General Data Protection Register (GDPR). The INSA project was also approved by the ethics committee of UPS-UT3 (through its Institutional Review Board - IRB).

### 1) Surveys in rural areas. Participants involved: UPS-UT3

#### 1.1) Data set description

Data collection during field investigations in Côte d'Ivoire, Benin and Senegal:

- a. Criteria identification / Sampling / recruitment: selection of survey villages, then at the village level, identification of quarters and families for the most possible farm diversity, related to the use of nitrogen in crops.
- b. Interviews with villagers without any discrimination regarding gender, age or other factors such as ethnicity. Every interviewee will receive full information on the status and objectives of the interviewers and will be duly informed of their rights regarding the data collected (access, retraction). Any person may refuse to be interviewed.

Informed consent will be established for the entire survey conducted in rural areas with farmers.

Sometimes, consent during agricultural surveys cannot be given in writing, because of illiteracy, or because of cultural reasons, such as refusing to sign a document that is considered as a document connected with the police or justice system, which could be interpreted as a threat for people and their goods. Non written consent will be formally proposed and independently witnessed by members of the community. Surveys will always be conducted by two researchers.

### 1.2) Standards and metadata

**Anonymization :** To avoid identification of the persons, concerning social data (age, lineage), geographic data (village, district), individuals cannot be identified, and the scale of indication must be greater than 50 households. Names of individuals will be replaced by numbers/codes and the correspondence between names and numbers/codes will be kept on two encrypted professional computers. DIP are disconnected from DDE and DDR in files.

**Pseudonymization:** this mainly concerns verbatim. If the quotation concerns a non specific record/account, pseudonymization consists in listing initials shifted in one direction (for example, Mr C.D. will become Mr A. B.) or the other (Mr A.B. to Mr C. D.). If the quotation is important only because It is linked to a person of reference (chief of village, representative for water and forest, herald), it cannot be pseudonymized and a written authorization for quotation right will be asked for.

Agricultural surveys will be conducted at the village scale and will link to social practices in Cote d'Ivoire, Senegal and Benin. As regards to socioeconomic data, the related surveys to be carried out will consist of interviews with GPS coordinates locating the interviewees' homes. We will collect / produce:

- a. Direct personal data (DDP): name, age, sex...;
- b. Direct explanatory data (DDE) socio-economic characteristics of the family of farmers: manpower, productions, techniques, credit, land, livestock...;
- c. Indirect personal data (DIP): GPS positioning, lineages;
- d. Direct results data (DDR): fertilizers, manuring, legumes and all other nitrogen-related practices;

### 1.3) Data sharing

a. For project members, DDP and DIP will not be accessible outside the two researchers in charge. Only DDE and DDR will be accessible and will be "pseudonymized" on interoperable platform (T5.1, metadata and data with secured password access).

b. For the scientific community, directly or through publications, and for the general public audience, only anonymized and "averaged" data will be available.

c. For the interviewees, points a and b will be explained to them. They will be able to exert their right of access to their own data and their right to delete it without justification. As the mobile network is the most reliable and widespread for accessing data in Africa, a permanent telephone number will be available for each country for transmitting and assessing these rights.

#### 1.4) Archiving and preservation (including storage and backup)

Data storage and management: raw survey data will be stored on only two encrypted professional computers and not on a shared server for ethics reasons. Data will be preserved for 5 years after the end of the project (31 Jan 2024). Paper forms (consent forms) will be kept for duration of 10 years after the end of the project in a specific closed envelope with the following text written on it: "I certify that this envelope contains X (number) of consent forms collected during the H2020-MSCA-RISE INSA-GA 891744 project", followed by the name of the coordinator.

## 2) Reportages in schools in Côte d'Ivoire. Participants involved: APDOC

### 2.1) Data set description

APDOC will bring pedagogic material in primary schools (for children over 6 years old) to explain the N cycle and its environmental impact, in accordance with teachers in Cote d'Ivoire (T5.3) and directors of schools. A formal authorization is being asked at "Division de la Vie Scolaire" in Abidjan to allow APDOC's intervention in schools.

APDOC activities in schools in Côte d'Ivoire will be filmed. These activities will be presented as a game and will have an educational role in order to raise awareness of children on the role of nitrogen in the environment. The pictures or videos taken during these activities will be used for the promotion of the project.

Consent of each pupil and their parent/guardian will be sought concerning their participation in the activity. Then, approval from the legally authorized representative and the agreement of the children or other minors will be asked regarding image rights. The participation agreement and image right forms together with information on the purpose of the activities will be provided to the families before activities are started. Children will not be interviewed; they will only be filmed (if agreement by legal representative is provided) while participating in the activities.

### 2.2) Standards and meta data

Consent forms were written based on standards provided on line by the French Ministry of National Education and supervised by the Data Protection Officer of UPS-UT3.

No meta data will be collected concerning reportages in schools.

### 2.3) Data sharing

Edited films and pictures will be freely accessible through the INSA project website and on social media channels (Twitter) of the project and beneficiaries for the promotion of the project.

### 2.4) Archiving and preservation (including storage and backup)

Data storage concerns archiving of files, videos and paper forms. Raw data from the APDOC activities in schools (reportages, videos) will be stored only on two computers belonging to APDOC technicians during the project, and on an encrypted hard disc specific to the project for a period of 5 years after the end of the project. The INSA website will be maintained for duration of 5 years after the end of the project. Paper forms (consent forms) will be kept for duration of 10 years after the end of the project

in a specific closed envelope with the following text written on it: "I certify that this envelope contains X (number) of consent forms and X information forms collected during the H2020-MSCA-RISE INSA-GA 891744 project", followed by the name of the coordinator.

### 3) Interviews and reportages during workshops. Participants involved: APDOC

#### 3.1) Data set description

Workshop participants will be informed that APDOC will film and take photos during the time of the workshops, both in Cote d'Ivoire (2021) and Kenya (2022). Participants will also be informed that photos and videos may be used for the promotion and communication of the project. They may accept or not to be filmed and/or interviewed.

#### 3.2) Standards and meta data

No meta data will be collected during workshops.

Consent forms were written based on standards provided on line by the French Ministry of National Education and supervised by the Data Protection Officer of UPS-UT3.

#### 3.3) Data sharing

Edited videos, photos and interviews will be freely accessible through the INSA website and on social media channels (Twitter, Facebook) of the project for the promotion of the project.

#### 3.4) Archiving and preservation (including storage and backup)

Data storage concerns files, videos and paper forms. Raw data (reportages, videos, interviews) will be stored only on two computers belonging to APDOC technicians during the project, and on an encrypted hard disc specific to the project for a period of 5 years after the end of the project. The website will be maintained for duration of 5 years after the end of the project. Paper forms (consent forms) will be kept for duration of 10 years after the end of the project in a specific closed envelope with the following text written on it: "I certify that this envelope contains X (number) of consent forms collected during the H2020-MSCA-RISE INSA-GA 891744 project", followed by the name of the coordinator.

### 4) Other data: pre existing data, data from communication and dissemination activities, reports, articles

#### 4.1) Data set description

Most of the work in the frame of INSA will be based on literature review, with already published data, and/or pre-existing unpublished data already collected by the participants of the project. No experimental survey will be performed in the framework of the INSA project, except those described in sections 1, 2 and 3 above. However, chemical analysis of available collected samples and interpretation of unpublished existing data may provide processed results, publications and reports.

Communication and dissemination activities will create data such as videos, reportages (as mentioned above), presentations, policy briefs, lectures and educational support material, reports, presentations, project documents, and publications.

Reports and scientific publications will be written during the course of the project.

#### 4.2) Data sharing

Videos, reportages, presentations, policy briefs, lectures, and educational material will be freely available without restricted access through the public part of INSA website.

Technical reports, secondment reports, reports on Net Anthropogenic Nitrogen Inputs (NANI) approach, Standard Operating Procedure (SOP) reports, as defined in the WPs description, presentations, project documents, publications related to nitrogen in Africa, and processed results from available data bases will be freely available for all members of the consortium through the interoperable platform via the INSA website. Access will be granted to members via a password.

We will ensure that data is discoverable, accessible, assessable and intelligible, usable beyond the original purpose for which it was collected, interoperable to specific quality standards. Moreover, all scientific articles in peer reviewed journals will be published with either green or gold Open Access, favoring anytime where possible gold Open Access journals, and made available through Open Access depositories (e.g. zenodo).

Provision of data sets through the interoperable platform will be dependent on the data policy defined inside each project related to INSA.

Internationally available repositories such as figshare repository will be used when possible for data and R codes (for data analyses).

### 5) Intellectual property rights

Knowledge shall be the property of the participating organization carrying out the work leading to that knowledge. Intellectual property shall be owned by the participating organization or the participating organizations generating it. Ownership and dissemination of results are defined in section 8 of the Consortium Agreement. Access rights (background included) are defined in section 9 of the Consortium Agreement.

### Conclusion

This first version of the DMP gives an overview of which data will be collected, processed and/or generated within the INSA project, which methodology and standards will be applied, which data will be shared publicly or with restricted access to the consortium only, and how data will be preserved after the end of the project. This information is subject to evolve during the life of the project, and the DMP will be updated accordingly.