Booklet

Collection, entry and verification of data I

Description of tasks for collection, digitization and verification

Summary

About AfricaRice and Afrique-learning

AfricaRice:

AfricaRice is a leading pan-African rice research organization committed to improving livelihoods in Africa through solid science and effective partnerships. AfricaRice is a research center of CGIAR, which is part of a global research partnership on future food security. It is also an intergovernmental association of African member countries. Today, it has 30 member countries. The mission of AfricaRice is to contribute to poverty reduction and food security in Africa through research, development and partnership activities, aimed at increasing the productivity and profitability of the rice sector so as to guarantee the sustainability of the agricultural environment.

Afrique-Learning:

Afrique-learning is a Beninese cooperative which creates and manages vocational e-learning courses specially designed for African youth. Courses are tailor-made in collaboration with experts in the field with the aim of producing interactive, illustrated, interesting and easy-to-study courses that provide the student with important information in simple and appropriate language. Learning is done independently at the student's own pace, it is assessed and a course certificate is attained following a final test. Courses are available on computer, smartphone or android tablet. They only require a very modest bandwidth and are therefore within the reach of students. Registration and classes are free.

Acknowledgements

Development of the toolkit was supported by AfricaRice project: "Fostering the Impact of Rice Technologies for Better Livelihoods in Sub Saharan Africa (FIRITEL)" funded by the Belgian Development Cooperation.

Photos / grafics credits

[1] Photo of Tyck, on www.iwara.com

[2] Illustration produced by EUDOX BÉATITUDES

[3] Grafic created by the pedagogical team

Context of the booklet

This is the first part of the booklet that deals with data collection for a Smart-Valleys development project. Data is collected during the different stages of the inland valley development. It is used for project management, providing information on the progress of the project and making it possible to prepare project reports. At the same time, this data is also used as proof that tasks have been successfully completed by the service providers. The introduction to the concept of services can be found in booklet 5. A description of services can be found in guide 7.

Data and its collection

This booklet deals with data and its collection as part of an inland valley development project using the Smart-Valleys approach. It shows the important steps in collecting and the verification of data.

Why collect data?

- To monitor the progress of the project
- To determine the impact of the project.
- To check the different stages of the project.
- To enable the preparation of project reports.
- To create a database for further analysis.

Design and plan data collection

The project coordination is responsible for determining the objectives and procedures for data collection.

The type of data to be collected depends directly on its use and the objective to be achieved.

The procedures for collecting data are based on the tasks of the inland valley development stages, as well as the manpower required and the systems available.

Main stages of design and planning

1. Defining the project objectives.
2. Defining the data necessary to document the progress and impact of the project, and therefore monitor the achievement of the project objectives.
3. Developing the process of data collection and verification:
    a. workflow
    b. human resources required
    c. technical resources required

Data collection, entry and verification

There are three main areas of activity. Data collection, digitization and data verification. The three working stages follow one another and are interdependent. Collecting data in a conscientious and complete manner at the beginning is therefore crucial. After it is digitized, the data can be checked. At this stage, both the data itself, but also its format, is checked.

In traditional practice, the process follows these steps: the technician collects the data on printed sheets of paper which are then digitized, either by the technician or by someone else. Digitization is done manually by transferring the data from the completed data sheet to a spreadsheet (eg Excel).

If you use the ODK app or an equivalent system, the data is entered directly on electronic sheets, on a smartphone and is therefore already in digital form. This approach is faster and reduces the risk of errors during transcription.

The data must then be checked, according to well-defined criteria: technical criteria (eg “Has the soil cultivation process been properly documented?”), orthography, required unit (“ha” instead of “ m^2 ”), etc.

The task of the verification of data is almost as crucial as that of collecting the data itself. Wrong data gives a false picture of reality. This can lead to incorrect assessments of the project’s development or falsely influence decisions about subsequent interventions.

Both the data collector and the data controller therefore have an important responsibility.

Data collection and entry in the field [1]

I. Definition of the data to be collected and the collection procedures

Definition of the data to be collected

In an agricultural project of inland valley development using the Smart-Valleys approach, the main objective of the project is already defined. It is development of inland valleys for the water verification in rice field.

But there can be further objectives. For example, stopping the "rural exodus". In this case, young farmers in particular should be supported so that they can empower themselves. Another project objective may be to strengthen the role of women in rice cultivation. In this case, particular attention must be paid the booklet "Gender and rice production".

The technical sheets presented in the toolbox (technical booklet "Field sheets for data collection") concern the stages of selection and management of inland valleys.

This data makes it possible to assess and compare inland valleys. In the course of the project, a catalog of potential inland valleys will be created which might be developed during the project or in a subsequent project. The data collected also makes it possible to determine the economic structures and therefore to carry out an analysis of needs, which can benefit the project itself but also the follow-up projects.

On the basis of the available data, it should be possible to assess the success of project objectives and measure the progress of the project. If the project aims to increase the rice harvest, data from the harvest concerned should be compared with data from rice harvests in previous years. You can then measure the improvement and thereby the success of the project objective.

The following steps must be followed when planning data collection:

1. List the individual project objectives and its criteria
2. Identify what data will allow you to evaluate the success of that objective
3. Define the method by which the desired data can be collected in the easiest and cheapest way
    - you plan to survey only certain population groups
    - or that you survey different groups, in order to have a valid comparison

Definition of collection procedures

Basic procedures for data collection and verification are described in this booklet. This process must of course be adapted to the needs of the project. In cases where data is collected electronically, digitization is not necessary. You can also configure the datasheets so that additional data verification is completely unnecessary.

Use the booklets "The concept of "service" to implement the *Smart-Valleys* approach or any other project" and "The organization of an inland valley development project" to adapt the description of the service and the organizational structures to the needs of the project.

Follow the basic steps:

1. Adapt the technical sheets to the data it is necessary to collect in order to pursue the project objectives. ("Field sheets for data collection")

2. Adapt existing services and design new ones if necessary. ("Presentation of services for the selection and development of an inland valley")

3. Now adapt the data collection and verification processes to the relevant criteria - ie take the existing processes and remove steps or add new ones ("Data collection, entry and verification")

II. Training service providers

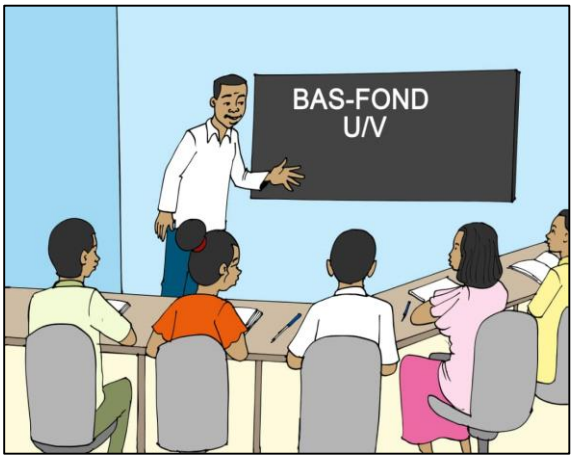
To ensure maximum data quality, service providers must be well trained and well equipped. The more these points are neglected, the more questionable the quality of the data will become.

It is always preferable to mix classroom training and online training. The online training serves as preparation before the classroom training, and as a follow-up after it. Online training also makes it possible to update trainers who will intervene in the classroom. Supervisors are typically trained first, and then deliver training to the technicians they supervise.

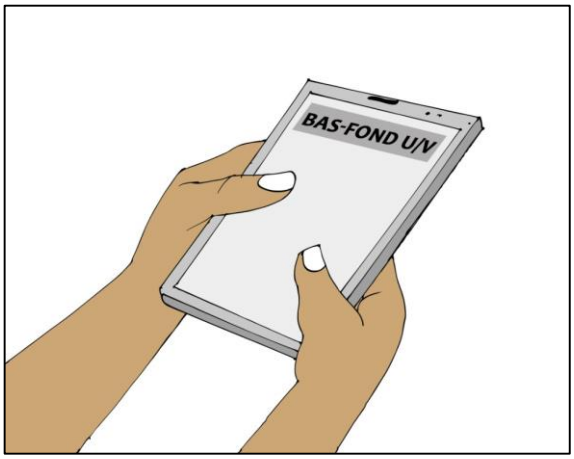
The training must be given by an experienced person, able to impart knowledge as concretely as possible. It must include as many exercises and field visits as possible, and must be done in advance of field operations. A good education is the basis of a good work in the field.

Key points for training

- The data collector and data controller must be familiar with agricultural development of inland valleys using the *Smart-Valleys* approach. This will allow them to identify and improve the quality of incorrect or erroneous data.
- The data collector must be familiar with survey methods and must understand how to interview people properly. For example, (s)he must not unconsciously give the answer to the respondents or influence them in any other way.
- Both the data collector and data controller must know the format in which the data must be transmitted. For example, numeric values can be written as numbers or words.
- Each service provider must master the work tools and applications necessary for the different stages of his work:
    - The data collector must be able to complete the data sheets or master using the ODK application to enter field data.
    - He must be able to use a camera and know which parts of an object can be photographed with which camera setting (close-up for a plant leaf or panorama for the whole field, for example).
    - For digitization, he must know and be able to use the relevant applications (Excel, Google Drive, etc.).
    - He must also be familiar with the cultural and traditional aspects likely to influence the surveys. The data collector must speak the respondent's language well. In this way, nuances can be perceived, interpreted correctly and, if in doubt, questioned. If not, a neutral translator must be employed.



Classroom training [2]



Online training [2]

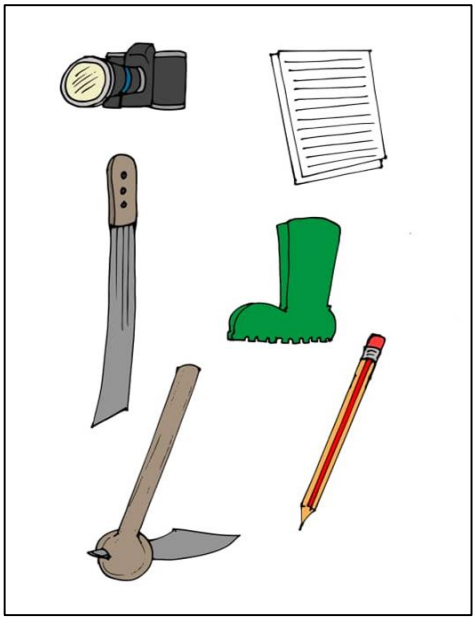
III. Data collection process

Before starting work

The data collector must first study the data to be collected in order to understand its characteristics and the reason for collecting the data. Only then can (s)he assess whether the responses and the data collected meet the criteria that were set when defining the data to be collected.

What working material does (s)he need?

Next, (s)he must ensure that (s)he has the necessary working tools. Firstly there are the tools for collecting the data itself, such as the data sheets to be filled out or the mobile device such as a phone or tablet to enter the data directly. Then, depending on the circumstances and requirements, (s)he may need a motorcycle, pen, rubber boots, a camera, test strips to measure soil pH, etc.



Data Collector work Tools [2]

The real work begins

Now the data collector can actually start the job. It is essential to always keep in mind the requirements of the data to be collected: what values the data can take and in what form it is presented.

Data collectors interviewing producers [2]

For example, the production area must be recorded in a certain unit. In some regions, area is still given in local units of measure. However, in order to be able to analyze the data, a universal unit, such as hectare (ha), must always be used. In this instance, the data collector will need to convert the local unit of measure into hectares. This means that the template of the data sheet to be completed must be adapted accordingly.

Here is an example of a data sheet to fill out, where the units to be used are explicitly specified:

3. Areas / Other crops

Instructions :

- Instructions : Give the areas in hectares (to be estimated with the rope or the GPS)
- Select the other crops cultivated in the lowland

a. Cultivable area : ___ ha | b. Total area cultivated : ___ ha

c. Area with rice crop : _____ ha

☐	Maize	☐	Sorghum / Millet
☐	Market gardening	☐	Sweet potatoes
☐	Cowpea / Soybean	☐	Others :

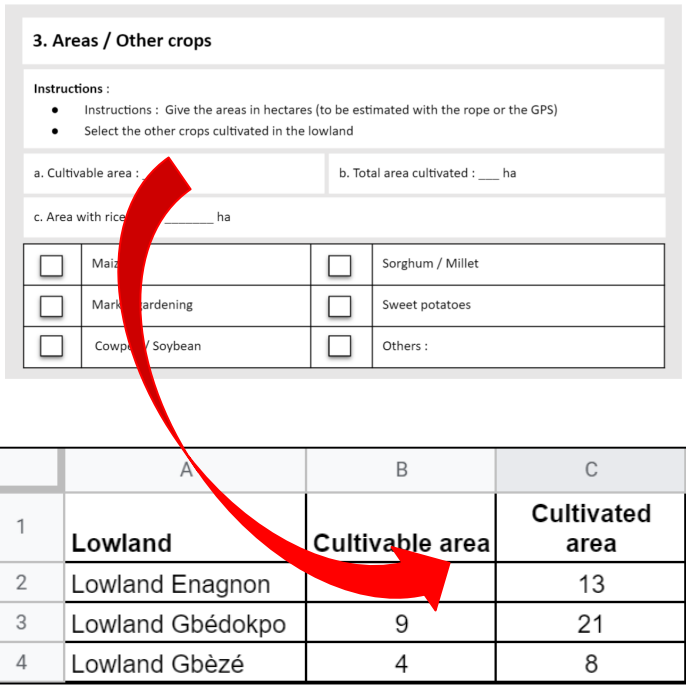
Extract of the data sheet: inland valley exploration [3]

After collecting the data, it is digitized. At this stage, it is strongly recommended that you check the data again before the service provider responsible for the verification does the final check. Any errors discovered during digitization can be corrected immediately by the data collector.

Design and definition | Training | Collection | **Digitization** | Verification

IV. Data digitization

[3]



	A	B	C
1	**Lowland**	**Cultivable area**	**Cultivated area**
2	Lowland Enagnon		13
3	Lowland Gbédokpo	9	21
4	Lowland Gbèzé	4	8

During digitization, the data is transferred to a MS Excel ™ type spreadsheet. In this form, the data can be easily verified and analyzed. The easiest method is manual transfer. Each value in the datasheet has its own column in the Excel table. It is important that each cell in the spreadsheet contains a unique value: a single cell must not contain data for corn and soybeans grown in the inland valley, for example. Such merging of data can create problems during further analysis. Several functions and options in the spreadsheet can help identify input errors, for example, the filter function.

V. Data verification

Data verification is the last of the three stages of data gathering during the project. At this stage, the data is checked to see if it is within the expected range of values. If the data collector indicates that an inland valley has 10,000 hectares, there is a high probability it is an error and that (s)he probably forgot to convert meters to hectares. Likewise, a harvest of 4000 tonnes of rice per hectare is not possible. The person in charge of the verification must therefore have a good knowledge of the data to be collected in order to discover such errors.

Secondly, (s)he checks for errors in how the data is written. This could be a misspelling or a number given as a word instead of a number. Such errors do not necessarily prevent the data from being understood , but the analysis will not be complete because this data will not be taken into account in its respective category and can thus distort the result of the analysis.

The table below shows examples of this type of error in the form of the data: the 5 was entered as a five, the inland valley shape was entered with a W instead of a V, and the word "Clay" is misspelled:

#	Name / inland valley indication	Accessibility	Cultivable area (ha) (number)	Culture / development intensity	Number of farmers (number)	inland valley shape (V / U)	Slope	Type of soil	Admitted to Phase 2
1	Ifèdé inland valley	No	**Five**	Rice - low	7	**W**	Steepy	**Clays**	Yes
2	Ifèdun inland valley	Yes	30	Rice - high	20	U	Slight	Sandy- clay	Yes
3	Gbénonkpo inland valley	Yes	22	Maize - high	**Te**	U	Slight	Silty-clay	Yes

Examples of data entry errors [3]

Such errors, relating only to the entry of the data in the form, can easily be corrected by the service provider responsible for the verification after communication with the data collector. But if the data is fundamentally flawed or even missing, the data collector will have to go back into the field to collect that data again.

This is of course a big effort for the data collector, as (s)he has to travel again and, if in doubt, arrange another meeting with the interviewee. This means more work for the supervisor as well because (s)he has to check the relevant data several times in case of doubt.