Is there a relationship between Data Collection, Data Curation and Data Sharing? Experiences from a Restoration Project in East Africa and the Sahel

Pietro Bartolini (ICARDA)
Asma Jeitani (ICARDA)
Christine Magaju (ICRAF)

RDA
RESEARCH DATA ALLIANCE

INTEREST GROUP ON AGRICULTURAL DATA (IGAD)

Virtual Meeting
May 2020

ICARDA
Science for resilient livelihoods in dry areas

MEL
monitoring evaluation and learning

World Agroforestry

Introduction

- Farmers in Ethiopia, Kenya, Mali and Niger implemented on-farm planned comparisons to test and innovate land management practices that restore agricultural productivity and ecosystem health

- Performance of each of the options was systematically monitored to produce rigorous evidence on the constraints and conditions for implementation and the variables of success for specific restoration options

Innovation at scale = harnessing technology to meet local needs within an appropriate enabling environment

Look at what is there, local knowledge

Engage with stakeholders to develop diverse and inclusive options and an appropriate enabling environment

Adapt options [technologies, market interventions, policies] to context

Refine what works where and for whom

Nurture innovation platforms

Data collection process

- Performance of options was monitored using electronic data collection.
- Data collectors used Open Data Kit (ODK) mounted on mobile phones.
- Electronic data capture was used to enable real-time data collection and analysis.
- Data collection activities produced a series of datasets surveying rural households in Kenya, Ethiopia, Mali and Niger.

Data Management

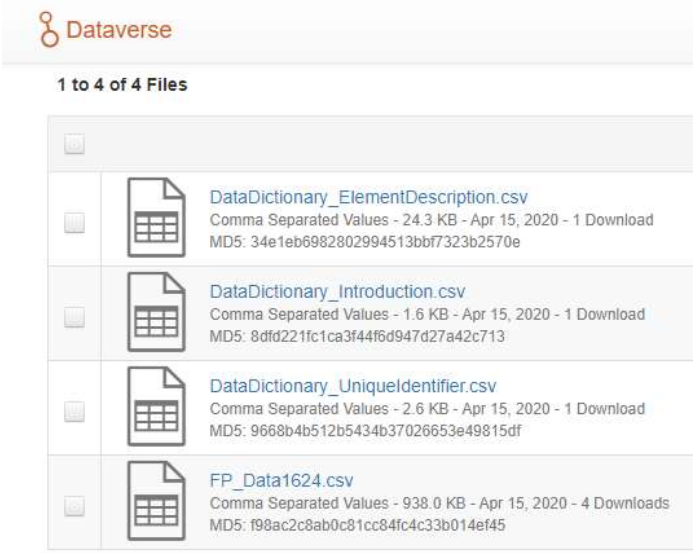
- In May 2019, ICARDA's M&E team presented the activities related to data curation and dissemination at the annual team meeting in Nairobi.

- An introduction on data curation techniques was provided, analyzing the dataset structure, suggesting improvements and developing a plan for the curation in the following months.

- The datasets structure was very similar, although the quality of the data varied depending on the collector proficiency with the instrument.

Data management cont.

- A close relationship with the ICRAF team allowed to complete the curation process, cleaning the data and standardizing the codes utilized.
- The result is a series of datasets containing high quality data, published on the institutional repository on Dataverse, and ready to be shared.
- The collaboration made the entire process easier and faster.



- At the same time it highlighted the need to establish the collaboration with the data collection team from the planning phase, in order to define the criteria by which the dataset has to be built.

Data Sharing

- The creation of a data paper related to the datasets is the last step of the entire process.
- Data papers provide more context and description than a simple dataset stored in a repository.
- It increases traffic towards associated research articles and data, leading to more citations and more collaborations.

Write-shop Organization

- Back in March, our team was going through the process of writing a data paper in order to share the results of the datasets.
- A write-shop in Tashkent was planned, then the global pandemic and the lockdown happened, changing our plans.
- The write-shop was re-structured to take place online using highly focused group meetings.

Write-shop cont.

- The new agenda adopted a mix of synchronous online meetings and asynchronous collaborations.

- Synchronous online meetings were held for commenting purposes, providing feedback, and discussing next steps.

- Asynchronous collaborations constituted the majority of the write-shop as it allowed schedule flexibility and increased productivity.

- This proved to be highly productive, resulting in the drafting of 2 data papers and 1 journal article.

Conclusion

Our experience shows that:

- Dealing with data collection, data curation, and data sharing activities as a whole produces better results than dealing with each one separately.
- Strong commitment of each team and a good communication are key in achieving desired results.
- Setting clear deadlines is extremely important.
- Publishing the datasets in a data paper can help convincing more scientists to collaborate in the data curation process.

Thanks for your attention