WP2 - Training on organic cultivar testing Module 9 – Data management in on-farm cultivar testing

Unit 9.3 Practical examples - data collection, data repositories & data documentation

Julia Cooper & Charlotte Bickler (Organic Research Centre)

Frederik Rey & Ambrogio Costanzo (ITAB)

LiveSeeding

Co-funded by the European Union

UK Research and Innovation

2. Data Collection & Documentation

- **Collection methods**: protocols, instruments, apps used (e.g. Open Data Kit, tablets, GPS).
- **Data quality assurance**: calibration, training of enumerators, double entry, validation rules.

A quick introduction to data collection protocols

Top ten tips on data collection methods

1. **Clarity of Purpose –**
   - Collect only the data you need to answer the research question.
   - Avoid "collect everything" → it wastes time and creates GDPR/ethics issues.

   See Unit 9.2 on GDPR

2. **Standardisation**
   - Use standard methods and protocols (e.g. ICASA variables, FAO methods).
   - Same units, same measurement technique, same timing → ensures comparability.

   Important for interoperability!

3. **Consistency & Training**
   - Everyone collecting data must be trained to do it the same way.
   - Farmer cooperators, technicians, researchers → consistency avoids bias.

4. **Accuracy & Precision**
   - Use calibrated instruments, correct measurement techniques.
   - Record exact values (don't round in the field unless protocol says so).

   Makes your data robust and valuable

5. **Documentation & Context**
   - Record not only the measurement but also the who, when, where, how.
   - Context (weather, soil conditions, crop stage) often explains variability.

   Important for reusability!

LiveSeeding

Top ten tips on data collection methods

6. **Data Recording Systems**
   - Use robust tools: notebooks, field sheets, or digital apps.
   - Build in checks (drop-down lists, range checks, validation rules).
   - Always date and label records clearly.

   Builds confidence in your dataset/robust data for statistical analysis

7. **Backups and Copies**
   - Digital entries → sync or back up daily.
   - Paper notes → photograph/scan regularly.
   - Don't rely on memory or scraps of paper.

   Minimises risk of losing data!

8. **Ethics and Permissions**
   - Collect only data farmers consented to provide.
   - Respect privacy (no unnecessary identifiers).

   See Unit 9.2 on GDPR

9. **Pilot / Pre-test**
   - Try out your data collection tool on a small scale.
   - Fix ambiguities and errors before rolling out widely.

10. **Quality Checks**
    - Build in regular checks for missing values, outliers, or inconsistencies.
    - Better to catch errors early in the field than later in analysis.

LiveSeeding

Data recording – some methods

- Old-fashioned pen and paper/clipboard

- Toughpad?

- Mobile phone app?

Self-directed learning: Fieldbook phone app for data collection

Deep-dive into open data repositories

Why store data in an open repository?

- Meets the principles of FAIR data, specifically "accessible"

- Required by many funding bodies e.g. HEurope

- Good practice when publishing in peer-reviewed journals

- Allows "reuse" of data so increases impact of your project

- Can you think of other reasons to use an open repository?

LiveSeeding
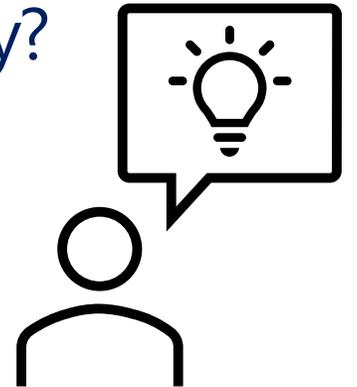
Choices for a repository

Domain-specific Repositories

Generalist Repositories

Institutional Repositories

e.g. for soil data
https://www.bonares.de/service-portal/data-repository#data%20reuse

Often using figshare...or dataverse...

The Dataverse® Project

DRYAD

figshare

MENDELEY DATA

OSF

Vivli
CENTER FOR GLOBAL CLINICAL RESEARCH DATA

zenodo

LiveSeeding

Find a research data repository

https://www.re3data.org/

Take a few minutes to search this registry of data repositories to find repositories that might be suitable for storing on-farm cultivar evaluation data. Be prepared to explain what you have found to the class.

Useful open data repositories

https://dataverse.org/

The **Dataverse**® Project

Open source research data repository software

Researchers — Enjoy full control over your data. Receive *web visibility, academic credit*, and *increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal Dataverse collection?

Journals — Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. Want to find out more about journal Dataverse collections?

Institutions — Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

Developers — Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools*, or other research and data archival systems with the Dataverse Project. Want to contribute?

How Dataverse Supports FAIR

- FINDABLE: Persistent Identifiers (DOIs, Handles) for datasets (and files), and enabling metadata indexing by search engines

- ACCESSIBLE: Open access to (meta)data and ensures data can be downloaded in machine-readable formats

- INTEROPERABLE: Provides standardised metadata schemas and enabling the integration of of data with research tools and platforms

- REUSABLE: Licenses that clearly state how data can be used and ensuring that data are well documented and preserved for long term use

Datasets - Metadata

Metadata Fields

Choose the metadata fields to use in dataset templates and when adding a dataset to this

☑ Use metadata fields from Harvard Dataverse

☐ Citation Metadata (Required)   [+] View fields
☐ Geospatial Metadata   [+] View fields
☐ Social Science and Humanities Metadata   [+] View fields
☐ Astronomy and Astrophysics Metadata   [+] View fields
☐ Life Sciences Metadata   [+] View fields
☐ Journal Metadata   [+] View fields

Files | Metadata | Terms | Versions

⬆ Export Metadata ▾

Export options
Dublin Core
DDI
DataCite
DDI HTML Codebook
JSON
OAI_ORE
OpenAIRE
Schema.org JSON-LD

Citation Metadata ∧

Dataset Persistent ID ❔ doi:10.7910/DVN/HADLCK

Publication Date ❔ 2021-05-26

Title ❔ Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania

Author ❔ International Institute of Tropical Agriculture (IITA)
University of Bonn, Germany

Contact ❔ Use email button above to contact.

Muthoni, Francis (International Institute of Tropical Agriculture (IITA))

Description ❔ A sub-national field assessment of land degradation was conducted in the Kongwa districts of Tanzania in December 2019. 34 sampling plots were selected using a stratified sampling method based on a land cover map. One site that hosts Africa RISING technologies on land rehabilitation was purposely selected to act as a control. The primary sampling plots measured 100*100m and were subdivided into replicate sub-plots measuring 30x30m. A sub-sample of 3 sub-plots was selected in each primary plot for assessment of land degradation. The observations from 3 subplots (30*30) were averaged to obtain an aggregate value for the larger plot (100x100m). The antecedent biophysical conditions in the sampling plot were recorded i.e. the land use, degree of slope, topographical position, soil color, crop types grown and land tenure. The percentage of area that was undegraded in each plot was also estimated visually.

A questionnaire for mapping land degradation and sustainable land management was applied for visual assessment of the type, extent, degree, and direct causes of land degradation. The different types of land degradation, for example erosion by water, were scored whether they are present or not, in addition to their extent and degree (intensity). The extent represented the proportion of a sub-plot covered by different types of land degradation, the degree was divided into 4 classes in ascending order of intensity of land degradation (0 = Light, 1 = Moderate, 2 = Strong, 3 = Severe). Moreover, the type, purpose, extent, and effectiveness of sustainable land management practices (SLM) were visually assessed in every subplot. Data were recorded using the mobile-based KoboCollect toolbox and transmitted to a cloud database for storage and descriptive analysis.

Subject ❔ Agricultural Sciences

Keyword ❔ land degradation (AGROVOC) http://aims.fao.org/aos/agrovoc/c_34823
sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561
soil erosion (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2651
TANZANIA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_7608
EAST AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2442
AFRICA SOUTH OF SAHARA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_166
AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_165

Topic Classification ❔ Natural resources (AGROVOC) http://aims.fao.org/aos/agrovoc/c_5091
Sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561

Notes ❔ If you are interested in accessing restricted datasets, kindly fill out this data use agreement form.

Language ❔ English

Producer ❔ International Institute of Tropical Agriculture (IITA) https://www.iita.org/ IITA

Useful open data repositories

Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit research papers, data sets, research software, reports, and any other research related digital artefacts.
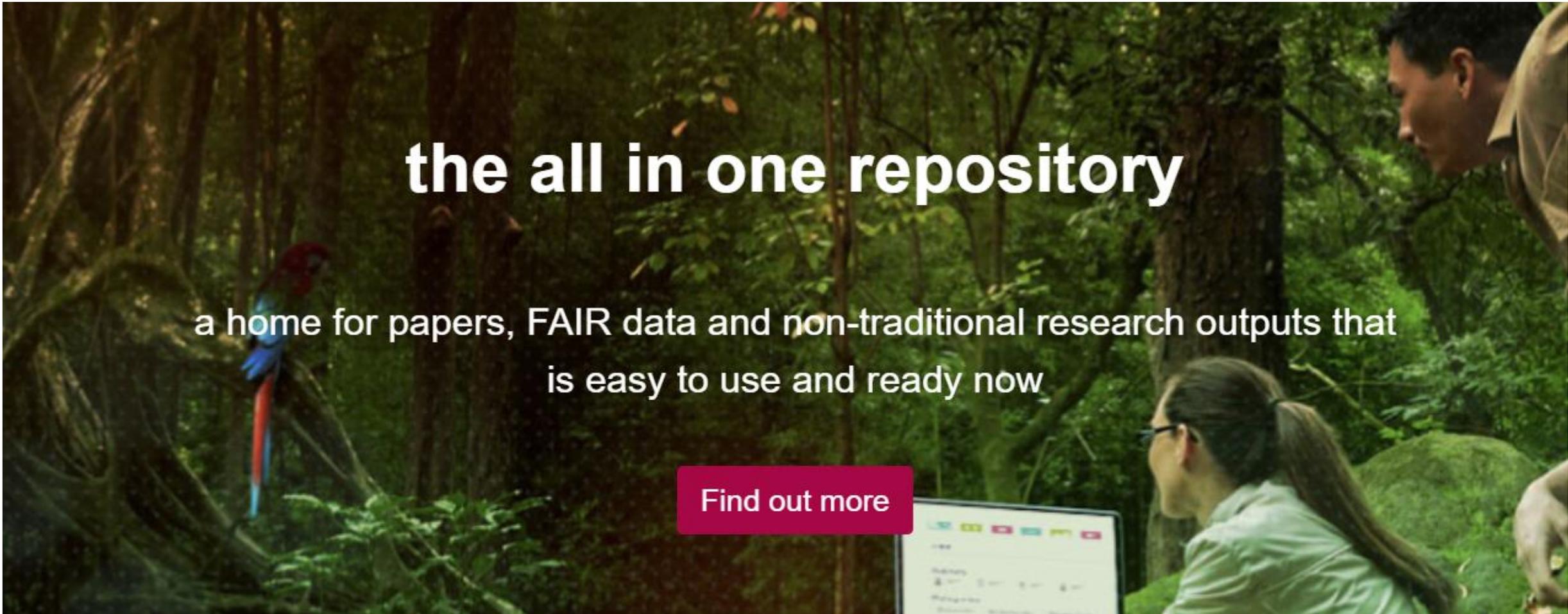

zenodo

Featured communities

EU Open Research Repository

Browse

Open repository for EU-funded research outputs from Horizon Europe, Euratom, and earlier Framework Programmes.

LiveSeeding

Useful open data repositories

And figshare: www.figshare.com   Commercial product for building repositories

the all in one repository

a home for papers, FAIR data and non-traditional research outputs that is easy to use and ready now

Find out more

LiveSeeding

Self-directed learning: exploring open data repositories

Finally, a word on licenses...

License	What it means	Use case
CC0 (Public Domain)	No restrictions. Anyone can use, share, or modify without permission or credit.	Great for maximum data sharing (e.g. open government datasets).
CC-BY (Attribution)	Free use, share, adapt — but must **credit** the creator.	Most common for research data (e.g. Zenodo uploads).
CC-BY-SA (Attribution + ShareAlike)	Same as CC-BY, but if you adapt, you must license your new dataset under **the same terms**.	Keeps derivative data equally open.
CC-BY-NC (Attribution + NonCommercial)	Must **credit**, but can't be used for commercial purposes.	Suitable for teaching/research only.

LiveSeeding

LiveSeeding