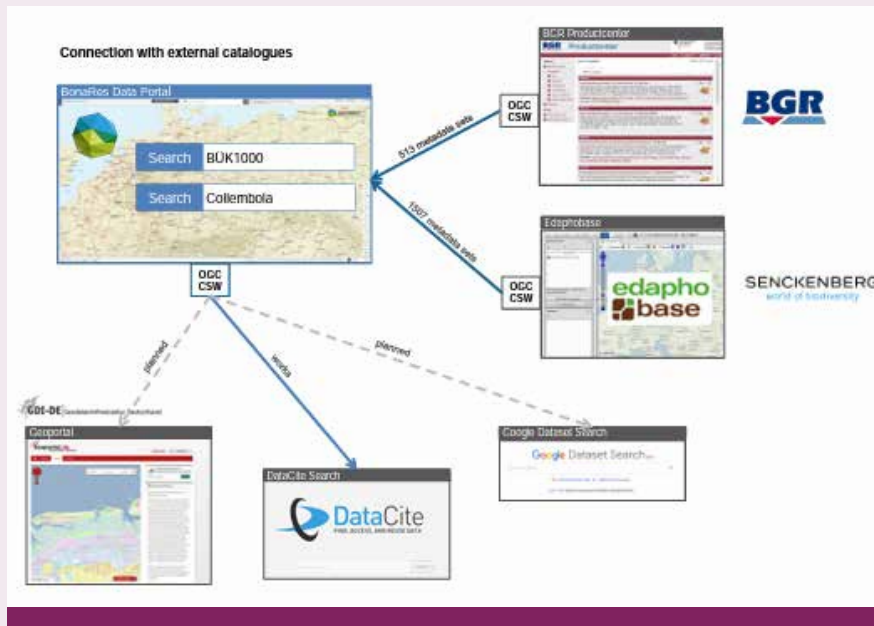


Module 1

**Making soil data findable,
accessible, interoperable,
reusable and open**

Case Study



Making soil data findable, accessible, interoperable, reusable and open

How developing a community-driven soil database can support farmers in decision making

Summary

Healthy soils are the foundation of a sustainable bioeconomy and face challenges such as oil sealing, soil consolidation, biodiversity loss, soil erosion, climate change and a growing world population. **BonaRes** – also known as ‘Soil as a Sustainable Resource for the Bioeconomy’ – is a

funding measure of the German Federal Ministry of Education and Research (BMBF). Their goal is to extend the scientific understanding of soil ecosystems and improve the productivity of soils and other **soil functions**, and develop new strategies for a sustainable use and management of soils through BonaRes Centre Knowledge Portal.

The BonaRes Data Centre collaborates with over 50 research institutions across 10 interdisciplinary collaborative BonaRes Funded projects to acquire and provide soil-related research data for scientists, decision makers and farmers. Their Knowledge Portal consists of structured schema data repositories supporting complex interactions of physical, chemical and biological soil processes.

Key learnings

- Making data about soils findable, accessible, interoperable and reusable (FAIR) can help protect and enhance soils as a natural resource.
- Pooling data from multiple research initiatives improves the knowledge base available to others to inform decision making.
- Long Term Field Experiments (LTFE) represent a significant portion of data that it had not been possible to fully utilise. By standardising metadata it is possible to increase historical and future data utility for a significant level of insight.
- It is important to establish well written data guidelines that respect the interests of both data collectors and users that are supported by proper data licences.

The challenge

Soils are an important natural resource, they filter harmful substances, prevent them getting into the water supply, and hold on to nutrients, carbon and water which are critical for agriculture and food production. Healthy soils are the foundation of the bioeconomy, but represent a global challenge due to the impact of growing populations on land use, natural draining and biodiversity, which in turn increases soil erosion and compaction.

¹ BonaRes, 'Map of Long term field experiments', <https://datenzentrum.bonares.de/map-ltfe.php>. Accessed December 2020.

Long-Term Field Experiments (LTFE) are field experiments with the minimum duration of 20 years and a static design.¹ More than 200 LTFE with running times of more than 20 years were identified in Germany. Because of the range and span of the time periods covered, LTFE data is highly diverse, decentralized and difficult to access and compare.

LTFE data is also inconsistently described, and important information about each dataset is not captured in a standard way, making it hard to find. For example, search results may not return all the data from different sources due to inconsistent metadata.

The solution

BonaRes brings together soil experts from over 50 institutions to research ways of balancing efficient use of soils with protecting the environment in order to support improved decision making by farmers. The network draws on the international community to garner support and encourage access to and use of the knowledge base it is creating.

The BonaRes Data Centre merges research data from BonaRes funded collaborative projects with special emphasis on LTFE datasets.² Doing so significantly helps in analyzing long term changes seen in soil characteristics and also aggregates geospatial soil data derived from different sources like public authorities, private agencies and research institutes. BonaRes attaches unique identifiers to datasets to facilitate better search and identification, and to increase users' confidence they are referring to the right data.

² BonaRes : "Long Term Field Experiment Map"
<https://lfe-map.bonares.de/>

The team ensured the data was findable, accessible, interoperable and reusable within project design:

- **Findable** – The team publishes an open inventory of data they hold. The inventory contains a number of fields to help users understand and use the data, for example a description and information on accessibility and quality. The database is designed to recognise standard **global search terms for agriculture**, thus increasing the ease of data exchange between different databases. To aid data contributors the team devised a rating-based system to recommend the suitability of each standard for soil data. However, because these recommendations are not mandatory, curating and preparing inconsistent data for publishing can require increased time and resources.
- **Accessible** – The team publishes the data in a consistent manner by providing guidance on suitable open formats and a support team to coach researchers through curating and providing data.

- **Interoperable** – The team conducted a research study to identify suitable standards to use for soil data and established a standardized metadata schema for researchers to contribute their data under.³
- **Reusable** – Maximum reuse of data is achieved through the use of a **standard open licence** permitting anyone to access, use and share data, and BonaRes use this option by default, following any embargo period. In instances where an open licence is not appropriate the team uses **globally recognised standard licences** which includes standard permissions that are easy to understand.

³ BonaRes' Metadata Schema for Soil Research Data', <https://datenzentrum.bonares.de/metadata-schema.php>

The impact

BonaRes data sets can now be discovered from different search engines including DataCite Search and Google Dataset Search, and datasets from linked external catalogues can be found via the BonaRes Data Portal, meaning the data is more findable.

Researchers are now comfortable in following global standards as recommended under BonaRes metadata schema which has become part of their publishing processes. Researchers are more confident in sharing their data and do not need to worry how it will be maintained for future use. The BonaRes platform takes care of all the technical aspects and ensures data is made widely available.

Researchers are now able to share data easily using links to a single repository, as compared to initial sharing of individual datasets by email allowing for improved access and version control through the BonaRes platform. Standardised processes also make acknowledgement and citation of their work simpler.

Lessons learned

The team faced a number of challenges in making data FAIR and open:

- **Protecting people from harm.** The majority of data held within the database is soil data, however occasionally data is submitted that is about private land, and therefore there is a risk of individuals being identified. To safeguard people the data is about, the team developed a set of guidelines and advice for publishers for anonymising personal data before providing it to the data center.
- **Fear others would unfairly benefit.** Data contributors expressed a concern that research would be picked up and used by others and careers would suffer as a result. To manage these fears BonaRes require the record of the data to be added to the open inventory, so people know it exists, but then allow an embargo of up to two years to ensure first right of use of the data.
- **Fear of misuse.** Traditional approaches to using research data included researchers keeping track of who is using their data, requiring access via a log in, and providing a name and email address. BonaRes engagement with the open science community has helped to change attitudes and the requirement to log in has been removed – the data is open access.
- **Quality of metadata.** The metadata schema contains 43 fields to add important information about the data and researchers initially found this cumbersome to complete. The team made a decision that only half of these would be mandatory, thus balancing the provision of key information to ensure FAIR data with the likelihood of researchers being able to take part.

- **Legacy data.** Legacy data or data from longer term research often needs to be digitised to ensure it is accessible. This can be a complex and time consuming process, often taking one employee up to a year to ensure it is in the right format. Following globally accepted standardised metadata and vocabularies has helped in overcoming the challenges for present and future datasets.
- **Standards for data.** Whilst the team encouraged data contributors to use specific standards for soil data, they did not make this mandatory initially. Therefore those datasets without the necessary standardised fields or values required additional time to curate prior to publishing. Having clearly defined mandatory and optional parameters under the metadata schema reduces the amount of data cleaning and curation required.

By taking a community-driven approach and engaging data contributors and users when considering how to make data FAIR, it was possible to understand and address fears and concerns, as well as ensure data was provided in a way that worked for everyone.

Future plans

To date the portal holds only primary data, generated by the research within the network. In future the team will include data from other sources. BonaRes are currently working to streamline their approach and consider licensing and intellectual property before they are able to make data from third parties accessible.

To embed FAIR data principles further BonaRes could consider how users could interact with the catalogue in different ways, for example by downloading the metadata related to all datasets.

The catalogues interface, although rich, relies on search being the starting point. Providing alternative ways to browse or download the catalogue could help users to navigate and find useful resources, such as allowing direct access to the hierarchical search that appears after a search has started.

Further resources

If you are planning to make FAIR data a core component of a project in a similar way to this case study there are resources in the Data Sharing Toolkit to help you do this and manage similar challenges:

- Module 2 – Assessing in-country potential for data sharing
- Module 3 – Reusing data from third-party sources
- Module 4 – Protecting individual's rights when sharing data
- Module 5 – Sharing data through data licensing
- Module 6 – Minimising harmful impacts from data sharing
- Module 7 – Ensuring sustainable access to data

Data Sharing Toolkit



ACKNOWLEDGEMENTS

This document was authored by the Open Data Institute and CABI as part of a Bill & Melinda Gates Foundation funded investment.

The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or CABI.

datasharingtoolkit.org

DOI: 10.21955/gatesopenres.1116742.1

cabi.org | theodi.org | gatesfoundation.org

 **CABI** Data Sharing Toolkit



BILL & MELINDA
GATES *foundation*



Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.