DataCite Use Cases Around the World: IITA

Olatunbosun Obileye, **Olayemi Oluwasoga**

Blog post published July 15, 2024 in *DataCite Blog - DataCite*
https://doi.org/10.5438/asx2-v395

Abstract

DataCite use cases around the world: Welcome to our blog series shining a spotlight on how research institutions integrate DataCite infrastructure into their systems and workflows.

Keywords

Information Systems (Subfield), Global Access Program, Guest Blog Post

Recommended citation

Obileye, O., & Oluwasoga, O. (2024, July 15). DataCite Use Cases Around the World: IITA. *DataCite Blog - DataCite*. https://doi.org/10.5438/asx2-v395

Copyright

Copyright © Olatunbosun Obileye et al. 2024. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DataCite use cases around the world: Welcome to our blog series shining a spotlight on how research institutions integrate DataCite infrastructure into their systems and workflows. Join us as we explore real-world examples, best practices, and the impact of the DataCite community.

Olayemi Oluwasoga is the Open Data Engagement Manager at the International Institute of Tropical Agriculture (IITA), an entity within CGIAR, with a passion for organizing, securing, and making information accessible. With over 20 years of experience in statistical analysis and data management, she has a proven track record of developing and implementing data management plans, ensuring data quality and compliance.

A headshot of Bosun Obileye

Olatunbosun Obileye:
What were the reasons at your institution for joining DataCite as a member?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
The International Institute of Tropical Agriculture (IITA) is an agricultural research institution that creates research data from different thematic areas. We have a need to maintain a repository for aggregating the research outputs from various projects undertaken by the institution, the need to have a central location to access data for reuse, the need to ensure FAIR compliance, donors' requirements as well as the institutional policy on open access, which stimulated IITA to join DataCite. DataCite is seen as a strategic organization that can provide Digital Object Identifiers (DOIs) for all kinds of research outputs with appropriate metadata that could promote the visibility of the research data from IITA.

IITA joined DataCite to assign DOIs to our datasets and other research outputs. IITA's institutional repository warehouses all research data, excluding genomic data. Social science, plant nutrition, plant health, natural resource management, and project reports are some of the resources in IITA's institutional repository. All these require DOIs to provide a persistent identity for each unique research output per project.

The interoperability of DataCite's metadata with other related identifiers like ORCID and ROR also attracted the institution to DataCite's metadata and the various APIs which could be used for other integrations within the organization. These attributes help improve the visibility of IITA's scientists as well as the institution.

A headshot of Bosun Obileye

Olatunbosun Obileye:
How are DataCite services currently used at your institution?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
DataCite resources are utilized in various ways beyond the creation and assignment of DOIs to research data. First, DataCite Fabrica is used to register DOIs for individual datasets. The DataCite API is used to register bulk datasets from various projects and creators. IITA has various specialized research databases like Cassavabase, Musabase, Yambase, and other critical research databases for specific crops. These are integrated into the institutional repository, built on CKAN. CKAN fetches the metadata from each database, unlocking the repository through some scripts (because the databases require authentication while CKAN is an open data repository) while DOIs are assigned to the datasets coming from the research databases automatically through DataCite's API.

We also implemented multi-styled citations through DataCite metadata schema to ensure secondary users of IITA's dataset provide the appropriate credits to data downloaded or reused from the institutional repository.

Diagram showing multi-styled citations in IITA's institutional repository .

Diagram showing multi-styled citations in IITA's institutional repository.

Additionally, the institution generates reports on the metrics, usage, and impact of various project data shared through the repository. These metrics are accessed via DataCite Fabrica and DataCite Commons. These help the Data Management team to create appropriate reports and make recommendations to scientists on how to improve the quality of their work outputs.

Data visualization of 3,335 works, including graphs on publication year, work types, top depositors, and fields of science. The majority of works are Kulakow and Asfaw from the field of agriculture. Major contributors include Kulakow and Asfaw.
Diagram showing some reports generated from DataCite's Commons for the International Institute of Tropical Agriculture's repository

A headshot of Bosun Obileye

Olatunbosun Obileye:
What were the most important steps in integrating DataCite infrastructure into your systems/ workflows?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
The following steps were important in integrating DataCite's infrastructure into IITA's system:

1. Ensuring DataCite's URL is registered as a trusted URL in IITA's perimeter firewalls as well as within the end-point firewall.
2. Allowing URL redirection for all traffic (2-ways) from IITA to DataCite and DataCite to IITA to ensure DOI resolutions are not blocked. Additionally, CGcore metadata schema v2 is mapped to DataCite's metadata schema.
3. Related Identifiers are made mandatory for all data uploads along with the mandatory and recommended fields.
4. When a new version of CGcore metadata schema is released, it is mapped accordingly with the DataCite Metadata Schema

Bar chart illustrating the cumulative sum of counts for different connection metadata

in IITA's Institutional Repository.
Diagram showing connection metadata in IITA's institutional repository.

A headshot of Bosun Obileye

Olatunbosun Obileye:
What challenges did you face during the implementation and how did you address them?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
There were a number of challenges faced during the implementation. A skill gap was identified for new hires, which was overcome with continuous training and support provided by DataCite's support team.

Additionally, the location of the repository was another concern. The need to comply with national laws on data residency was a concern. To ensure this was not compromised, the repository was hosted on premise (locally).

Some metadata schema mapping problems were encountered with the old Fabrica and the metadata schema from DataCite because IITA uses a unique metadata schema called CGcore Metadata Schema v2. A new version of the CGcore metadata properly addressed this challenge. The metadata being populated from the repository could then be mapped to the DataCite Metadata Schema.

A headshot of Bosun Obileye

Olatunbosun Obileye:
What kind of outreach, communication, and training did you do for staff and researchers at your organization before, during, and after adopting DataCite services? What worked, what didn't?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
Before the implementation, the management drafted a policy to back the adoption of open science and mandated all students, researchers and visitors to abide by the policy. A department was created to be responsible for data management and open science by the senior management of the institution. Additionally, skilled staff were hired to implement and maintain the system. Also, awareness-raising activities were carried out across the various stations of IITA. Seminars were organized and materials were created to sensitize users ahead of the new systems to be implemented.

During the implementation, the institution was inundated with flyers and news bulletins on the stages of the work while researchers' inputs were continuously sought to gain the buy-in of the main users. Agile methodology was used to ensure recommendations from scientists were implemented in the development (system). During the user acceptance test, the test instance of Fabrica was used to simulate how DOIs relate to datasets deposited in the institutional data repository.

After the implementation, a launch date was announced for the new repository which is fully integrated with DataCite's infrastructure. A lot of outreach and activities were organized for the launch. During the launch, the management of the institution was present while scientists were invited to talk about how the new platform could support various research outputs of the institution. The Deputy Director General in charge of R4D (Research for Development) emphasized the importance of ensuring all research data is uploaded to the new platform. Subsequently, seminars were organized across IITA hubs (consisting of countries in clusters). Further sensitization on the need to have DOIs for datasets and the linking of published papers with datasets were also carried out.

A headshot of Bosun Obileye

Olatunbosun Obileye:
What impact has DataCite adoption had internally?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
The adoption of DataCite had an impact in several ways internally.

1. **Increased institutional visibility.**
   IITA is a member of CGIAR. The CGIAR has a big data platform where FAIR metrics were implemented to provide means of measuring the FAIR compliance of research data from various institutions within the CGIAR. Due to the integration of DataCite's Metadata Schema with CGcore Metadata Schema v2, it was observed that IITA's FAIR score within the CGIAR was among the centers ranked first in the FAIR ranking within the CGIAR big data platform. This visibility also helped with the image of the institution among the CGIAR centers.
2. **Trust from funders.**
   Internally, researchers could access DOIs from past research projects and share those with funders when requesting project follow-up grants for scaling up the project. The impact of the data made available with DataCite DOIs could also be seen by funders.
3. **Verified publication datasets.**
   Writers, researchers, and students simply share DOIs of their research data with publishers when reviewing their works in peer-reviewed journals. This saves a lot of time and makes their work verifiable among reviewers.
4. **Improved MEAL (monitoring, evaluation, accountability, and learning).**
   Based on reports from data metrics in DataCite Commons, data that were not making enough impact could easily be reviewed to understand the reasons. It helped researchers to improve their research topic.
5. **Metrics for research, researchers, and institutional impact measurement.**
   The impact of outputs is measured and rewarded by the institution. DataCite provides the tools to view the impact of works by researchers within the institution.
6. **Integration of gray databases, discoveries of standalone databases, and standardization of research data output.**
   IITA has been able to scout and gather various research databases existing in silos and integrate them into the institutional data repository to provide the appropriate visibility for the datasets. Due to the standard workflow for uploading research data into the repository, which requires the completion of the metadata records, many works were made to follow global best practices for exposing data with appropriate licenses.

7. **Improved processes.**
   Assignment of DOIs became part of the data curation workflow. The registration and assignment of DOIs from DataCite to research results is one of the most important prerequisites for the quality control workflows in the institutional data repository.

A headshot of Bosun Obileye

Olatunbosun Obileye:
What advice do you have for other institutions planning to adopt DataCite infrastructure?

A headshot of Olayemi Oluwasoga.

Olayemi Oluwasoga:
DataCite services are great. We like the fact that there is a test account to try out innovations and services and the mapping of metadata is quite easy.

Diagram showing the number of datasets per year in IITA's institutional repository.

Diagram showing the number of datasets per year in IITA's institutional repository.