Building Better DOI Metadata: Lessons from the PID Network Germany Pilot at TU Dortmund

Norbert Gövert and Paul Vierkant Kathrin Höhner

Blog post published July 20, 2026 in *DataCite Blog - DataCite*
https://doi.org/10.5438/2ek0-a312

Abstract

The quality and interoperability of metadata are essential for connecting research outputs, people, organizations, and infrastructures across the scholarly record.

Keywords

Information Systems (Subfield), Metadata, PID Network Germany

Recommended citation

Kathrin Höhner, N. G. and P. V. (2026, July 20). Building Better DOI Metadata: Lessons from the PID Network Germany Pilot at TU Dortmund. *DataCite Blog - DataCite*. https://doi.org/10.5438/2ek0-a312

Copyright

Copyright © Norbert Gövert and Paul Vierkant Kathrin Höhner 2026. 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.

The quality and interoperability of metadata are essential for connecting research outputs, people, organizations, and infrastructures across the scholarly record. As a project partner in the PID Network Germany project, DataCite addresses this need by bringing together stakeholders from research and cultural heritage sectors to strengthen the adoption, standardization, and practical implementation of persistent identifier (PID) systems in Germany. As part of this work, pilot implementations with LMU Munich and Technical University Dortmund explore how community-informed metadata guidance can be translated into concrete improvements in local repository and DOI workflows. These experiences closely align with DataCite's vision of metadata as a shared global asset that is open, connected, reusable, and maintained through community collaboration. This interview explores how the pilot of TU Dortmund contributes to DataCite's overall goal of improving metadata. In this interview, Kathrin Höhner from TU Dortmund University Library responds to our questions about the pilot, the resulting improvements to DSpace, and the lessons learned during implementation.

1. Which metadata or workflow challenge did you set out to address through the pilot?

Our overarching goal is to increase the visibility of TU Dortmund University's research outputs. Metadata should therefore be reusable and interoperable across different systems and contexts. To achieve this, we consider high-quality, well-structured metadata essential, not only within our local repository, but also when metadata is submitted to DataCite as part of DOI registration.

TU Dortmund University Library has been operating a DSpace-based repository since the 2000s. The repository hosts a wide range of publications, including dissertations, project reports, and Open Educational Resources. Until now, the repository has used a non-hierarchical Dublin Core metadata format, with metadata made available via OAI-PMH for various dissemination purposes. A central challenge was that authors were recorded only as text strings. This made it difficult to disambiguate people reliably, which is essential for the long-term reuse and interoperability of metadata.

Through the pilot, we set out to improve the way DSpace handles persistent identifiers (PIDs), particularly ORCID iDs for people and ROR IDs for organizations. The aim was to strengthen the DOI metadata workflow and ensure that richer, more accurate metadata can be transmitted to DataCite.

2. How does your DSpace–DataCite workflow currently work, and where did you see opportunities for improvement?

Our strategy is based on mapping concepts rather than simply providing text strings. Wherever possible, we aim to use PIDs to make metadata more reliable, machine-actionable, and reusable. In the existing workflow, DSpace metadata is exposed via OAI-PMH and used for different distribution and registration purposes. However, the previous metadata model did not sufficiently

support the registration of PIDs for people and organizations. This limited the quality of the metadata that could be passed on to DataCite and reused in other contexts.

The PID Network pilot gave us the opportunity to improve this workflow by extending DSpace so that identifiers such as ORCID iDs and ROR IDs can be transmitted more effectively to DataCite. Since we also reuse DataCite metadata in other contexts, these improvements benefit us directly as well. Better metadata in DataCite means better metadata can flow back into discovery, reporting, and other institutional workflows.

3. Which PID or metadata fields were most important for your implementation work, and why?

The most important fields for our implementation work were those that allow people, organizations, and DOI relationships to be described more precisely.

A key part of the project focused on the placement of DOIs within the DSpace metadata. This included storing DOIs assigned by the repository itself under *dc.identifier.doi* and storing external DOIs under *dc.relation.hasversion*. For the registration of metadata with DataCite, ORCID iDs for individuals and ROR IDs for organizations were particularly important. Only by using such PIDs is it possible, in the medium and long term, to automatically disambiguate person and organization entities in metadata management. This is crucial for improving the quality, reliability, and reusability of metadata.

Additional implementation work addressed the inclusion of ROR IDs for organizations and organizational units, the use of the *nameType* attribute for people and organizations through configurable entities, the provision of item version information when versioning is used in DSpace, the provision of metadata in the DataCite Metadata Schema via OAI-PMH, and the use of ROR IDs for publishers.

4. How did the pilot connect local institutional needs with broader open-source development in DSpace?

From the outset, it was a requirement of the project that the results of the development work should be incorporated into the DSpace core. This was important because TU Dortmund's local needs are shared by many other institutions using DSpace: improving DOI metadata, supporting PIDs, and making metadata more interoperable are not isolated local concerns.

The implementation was carried out by The Library Code GmbH on behalf of TU Dortmund University Library, with the goal of making the resulting developments available to the entire DSpace community. This means that the pilot resulted in more thandid not result only in a local customization. Instead, the work has become part of the shared open-source development process for DSpace. Once these changes are included in a DSpace release used by institutions, the wider community can benefit from the improvements without having to reproduce the development work locally. This approach also benefits TU Dortmund directly. By contributing the

features to the DSpace core rather than maintaining them as local patches, the university can receive the extensions through regular DSpace upgrades and reduce the long-term maintenance effort associated with custom code.

5. What were the main technical or organizational lessons learned during the pilot?

We had excellent collaboration with The Library Code throughout the project. We were also pleased to find that many parts within DSpace were already designed in a way that made it possible to implement the project objectives within the established time and budget constraints.

At the same time, the pilot showed that implementing improvements in an open-source core codebase requires careful planning. Because the changes were intended for the DSpace core, the development process does not end with local implementation. It also involves preparing the code for community review, submitting pull requests, and responding to feedback from the DSpace community.

The pilot also confirmed that robust metadata models are essential. In the context of the project, we had to conclude that DSpace's entity concept does not fully meet our expectations and differs from other conceptual approaches. This was an important lesson for us, particularly because our approach depends on mapping metadata concepts and identifiers rather than relying on strings alone. The pilot resulted in several improvements that were submitted to the DSpace community as pull requests. These changes have since been merged into the main DSpace branch and included in DSpace 10.0, released at the end of May 2026.

- Store local DOIs in DSpace under *dc.identifier.doi* and external DOIs under *dc.relation.hasversion*, including a migration script for existing data and a change from http://dx.doi.org to https://doi.org.
- Add support for ROR IDs for organizations and organizational units.
- Transmit the *nameType* attribute for people and organizations using configurable entities.
- Add item version information to metadata when versioning is used in DSpace.
- Enable DataCite metadata export through OAI-PMH.
- Add support of ROR IDs for publishers.

6. What would you recommend to other institutions using repository software that want to improve the quality and completeness of their DOI metadata?

It is difficult to offer general recommendations that apply equally to all institutions, because repository systems, metadata practices, and local workflows differ widely. However, our experience has confirmed the value of focusing on metadata concepts and robust data models rather than simply improving individual metadata strings. For institutions aiming to improve DOI metadata quality, we would recommend looking closely at where PIDs can be integrated into existing

workflows. ORCID iDs for people, ROR IDs for organizations, and clear DOI relationship types are especially important because they support disambiguation, interoperability, and long-term reuse.

The pilot also showed us the value of connecting local implementation work with broader open-source development. By developing the features directly for the DSpace core, the results can be reused by the wider community and maintained through regular software updates. This makes the work more sustainable than a purely local customization.

The post Building Better DOI Metadata: Lessons from the PID Network Germany Pilot at TU Dortmund appeared first on DataCite.