Connecting the Dots with DataCite DOI Metadata

August 20th, 2024   Metadata   Metadata Schema   6 min Read

https://doi.org/10.5438/k81t-zq43

One of the key factors in making research outputs discoverable is assigning unique, persistent identifiers (PIDs) along with rich metadata. DataCite provides services to register Digital Object Identifiers (DOIs), which provide a persistent way of identifying and accessing a particular resource. DataCite DOIs have metadata descriptions based on a comprehensive schema that includes 20 properties divided into mandatory, recommended, and optional elements, essential for ensuring that research outputs are easily found, accurately understood, and effectively linked within the research community. In this blog, we'll explore the value of metadata collected during DOI registration and the importance of specific properties within the DataCite Metadata Schema.

Enabling Discovery

Rich metadata allows researchers to provide detailed descriptions, significantly enhancing the discoverability of their outputs. DataCite DOIs and metadata are fully searchable and accessible through DataCite Commons, which provides a public web search interface to the PID Graph. The metadata can also be retrieved via our APIs. These services ensure that research outputs and resources are easily found by various metadata aggregators, including data portals and indexes.

Key properties for enabling discovery include:

- **Identifier:** Ensures the resource is uniquely and persistently identified, making it easy to locate.
- **Subject:** Improved findability through keywords and phrases, which is essential for findability.
- **Publisher:** Provides important context about the resources's origin, facilitating its discovery.

Enriching Understanding

DataCite DOI metadata provides rich descriptive information that helps users better understand and contextualize research outputs and resources. Detailed descriptions, keywords, and contextual information ensure thorough documentation, making it easier for users to grasp the content and use it correctly.

Key properties for better understanding and contextualization include:

- **Description**: Provides detailed information about the resource, explaining what it is and what it's for. It can save researchers a lot of time by quickly conveying the resource's scope and usefulness.
- **Creator**: Identifies the authors or organizations responsible for the resource, providing important context about its origin, offering context about its origins and ensuring proper attribution.
- **Title**: Provides a concise and descriptive name for the data, offering a quick understanding of its content.
- **Publication Year**: Indicates when the resource was published, providing temporal context. Knowing the publication year helps users assess the resource's relevance and currency, which is particularly important in fast-moving research fields.
- **Resource Type**: Describes the type of resource, such as a dataset, software, or image, helping users understand its nature and potential uses. The DataCite Metadata Schema is flexible and supports many different resource types.

- **Funding Reference**: Includes the funder's name and identifier along with the award number, URI, and title. This feature enhances transparency by allowing stakeholders to track the financial support behind research outputs and directly link them to their funders. Additionally, it provides important context about the origins and development of the research, highlighting potential biases.
- **Rights**: Is essential for clarifying how the data can be used, making sure users know about any restrictions or permissions.

A word cloud showcasing the most commonly registered licenses.

Connecting Research

Metadata does more than describe resources; it connects various pieces of information, creating a cohesive and comprehensive research ecosystem. Enhanced connectivity through metadata supports complex relationships like citations, funding references, and project associations. This interconnectedness is essential for a complete understanding of research activities and their impacts, allowing researchers to gain deeper insights, uncover new relationships, and enhance their understanding of the field; see examples from three DataCite projects highlighted on our blog.

Connection metadata is a key feature of the DataCite Metadata Schema. This includes properties that facilitate links between works and other entities, such as people and organizations, ensuring that DOIs and related identifiers are clearly documented. This capability supports complex relationships and fosters a more integrated and cohesive research ecosystem.

For example, a publication can be connected to its underlying dataset, software, and other related research through their respective identifiers. This interconnectedness supports comprehensive research workflows, facilitates interoperability, and enhances the overall research process.

There are other key aspects of connection metadata to consider—learn more about them in our blog post.

Key properties for connecting to people and organizations include:

- **Name Identifier:** Used to record a person's unique identifier, such as an ORCID iD, ensuring their work is accurately attributed to them.
- **Affiliation Identifier**: Uniquely identifies the organization a creator or contributor is affiliated with.
- **Funder Identifier:** Specifies the unique identifier of the funding entity.
- **Publisher Identifier:** Identifies the publisher with a unique identifier.

Building Research Narratives:

Connection metadata helps trace research stories and understand relationships between different works.

A key property for connecting to related works:

- **Related Identifiers**: Links to related resources like publications, datasets, software, and other research outputs. By establishing these links, metadata provides broader context and additional sources of information, enriching the research narrative.

For example, linking datasets to the publications that cite them allows researchers to track the impact and reuse of data, offering insights into how data contributes to further research and knowledge creation. Similarly, linking an instrument to publications and institutions illustrates its impact and the network of knowledge it belongs to. This type of metadata reveals the relationships the instrument has with various entities, including the publications that mention or describe it, the institute that originally created it, and any contributing institutes. It also captures references from other articles that discuss the instrument, providing a comprehensive view of its role in the research landscape.
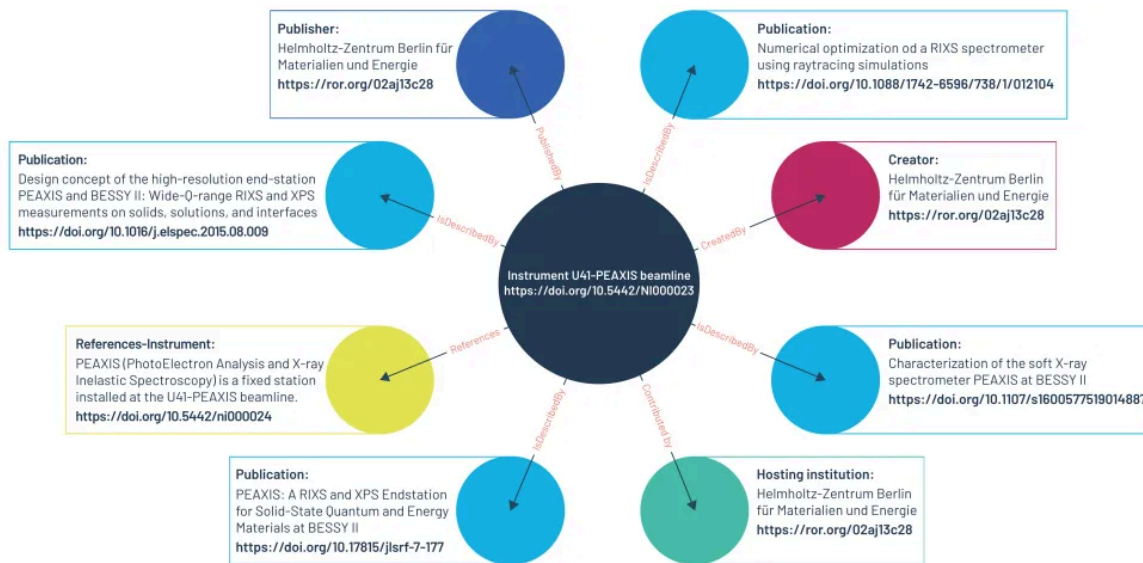
An example of a [DataCite DOI](#) for an instrument connected to other resources and research outputs.

By mapping these connections, we can see how the instrument fits into a larger research context, highlighting its impact. This detailed web of connections helps researchers understand the broader implications of their work and how it integrates into the existing body of knowledge.

Using connection metadata, researchers can answer questions like "Which samples were used in this study?" or "How are these datasets related?" This interconnectedness is crucial for constructing a comprehensive research narrative and contributing to the bigger picture of research.

Beyond the Basics: The True Impact of Metadata

The benefits of metadata are extensive, encompassing increased discoverability, better understanding, proper citation, and enhanced connectedness. DataCite DOI metadata exemplifies these advantages by providing a robust framework for capturing metadata. This framework ensures that research outputs and their metadata are not only accessible but also easily searchable and properly cited, enhancing their value and utility in the research community.

However, to fully realize these benefits, it's crucial to commit to expanding and updating metadata by prioritizing completeness and accuracy beyond the mandatory properties, and actively including recommended and optional properties and sub-properties.

By understanding and implementing rich, complete, and high-quality metadata, we can ensure that our research outputs remain valuable, accessible, and interconnected. This, in turn, supports the ongoing advancement of knowledge and innovation, driving progress.

Metadata is a fundamental component of modern scholarly information infrastructure. DataCite DOI metadata highlights the many benefits of robust metadata practices, from improving discoverability and understanding to ensuring proper citation and fostering connectedness.

Sara El-Gebali