



Contribution ID: 150

Type: **Presentation**

Bridging Metadata Standards: Implementing the CDIF Framework for Enhanced Interoperability in Data Observatory catalog.

Tuesday 14 October 2025 12:47 (11 minutes)

The exponential growth of data has intensified challenges in achieving cross-disciplinary and cross-repository interoperability. Metadata standards—such as DataCite, Dublin Core, and OpenAIRE—play a pivotal role in data discovery and reuse, yet their heterogeneity creates fragmentation. This proposal presents the implementation of the CODATA-CDIF Conceptual Domain Interoperability Framework (CDIF) within the Chilean National Data Observatory catalog, a platform aligned with the Chilean National Research and Development Agency (ANID). Our work addresses critical interoperability challenges by harmonizing metadata practices across repositories, enabling federated search, and enhancing compliance with national and international funding mandates.

Key Challenges

1. **Compatibility Between Standards:** Mapping metadata elements across standards (e.g., OpenAIRE's "Embargo Period" vs. DataCite's "Date" fields) often involves complex many-to-one relationships, complicating automated workflows.
2. **Semantic Ambiguity:** Disparate naming conventions (e.g., "language" vs. "linguisticCoverage") hinder federated searches. Locating datasets in Spanish, for instance, requires querying variants like "es-CL", "Spanish", or "es".
3. **Value Standardization:** Free-text fields introduce ambiguity (e.g., "spatialCoverage" entries like "Chile" vs. "CL" vs. geographic coordinates). While controlled vocabularies (e.g., GeoNames, Wikidata) resolve this, their integration demands meticulous curation.
4. **Automation Barriers:** Schema mismatches impede scalable integration, necessitating flexible infrastructure to reconcile differences.
5. **Metadata Loss:** Crosswalks between standards risk losing properties. For example, OpenAIRE's "Citation properties" lack equivalents in DataCite.

Implementation Approach

Our CDIF-based framework harmonizes metadata across ANID-aligned repositories, DataCite, Dublin Core and OpenAIRE through:

- **Crosswalk Development:** Semantic mappings resolve ambiguities, such as distinguishing DataCite's "Date:dateType='Issued'" from Dublin Core's "dateIssued". These mappings are validated against ANID's grant reporting requirements, ensuring coverage of critical fields like funding attribution.
- **Controlled Vocabularies:** FAIR-aligned authorities (e.g., ISO 639-3 for languages, UNESCO Thesaurus for disciplines) standardize values, reducing ambiguity in searches.
- **Middleware Infrastructure:** A RESTful API dynamically translates metadata queries across standards, leveraging AWS services (OpenSearch for indexing, S3 \+ Athena for querying) to deliver fast, scalable federated search.
- **Modular Ontology Extensions:** Gaps in discipline-specific metadata (e.g., geospatial granularity in OpenAIRE) are addressed by extending CDIF's ontology, ensuring compatibility with domain-specific needs.

Case Study: Searching for "Cactaceae Species" Models

To illustrate interoperability challenges, consider searching for "Cactaceae species" models across repositories:

- **Zenodo:** Filters allow limiting results to “resource type != model” and “access != open”, yielding relevant datasets.
- **DataCite Commons:** No “model” filter exists; available filters (e.g., “Work type”, “License”) fail to narrow results effectively.
- **Harvard Dataverse:** Similar limitations—filters like “Dataverse Category” or “File type” do not align with the query.

Outcome: Only Zenodo returned targeted results. The lack of standardized filters in other repositories forces users into manual, time-consuming searches with no guarantee of success.

CDIF-Driven Solution:

Our implementation standardizes metadata properties and values across repositories. For example, the “resource type” field is mapped to equivalent terms (e.g., “Model”, “Simulation”) in DataCite and OpenAIRE, while controlled vocabularies enforce consistency (e.g., “language:es-CL” instead of free-text variants). An example spreadsheet example demonstrates mappings for 20+ properties, including:

- **Resource Type:** Mapped to “Model” (DataCite’s “Software”), “Dataset” (OpenAIRE’s “Research Product”).
- **Temporal Coverage:** Aligned with ISO 8601 across standards.
- **Discipline:** Normalized using UNESCO and OECD taxonomies.

While still in progress, we believe this approach holds the potential to enable efficient searches across repositories and platforms in the future.

Conclusion and Relevance to SciDataCon 2025

Our CDIF implementation underscores that interoperability requires both technical solutions (crosswalks, APIs) and governance frameworks to align stakeholders. This work aligns with SciDataCon’s focus on actionable strategies for data integration, offering insights for:

- **Policymakers:** Balancing flexibility with standardization in national metadata mandates.
- **Repository Managers:** Adopting modular frameworks to avoid schema lock-in.
- **Global Communities:** Addressing multilingual and multidisciplinary challenges through shared vocabularies.

We invite collaboration to scale this approach, particularly for repositories in underrepresented regions. By bridging the metadata gap, we empower researchers to focus on discovery—not data hunting.

Audience: data librarians, repository managers, metadata specialists, policymakers involved in research data management, and researchers interested in data discovery, interoperability, and the application of metadata standards.

Keywords: Metadata interoperability, CDIF, crosswalks, controlled vocabularies, FAIR data, federated search.

Primary authors: Mr ANTILAO, Alejandro (Data Observatory Foundation, Chile); Mr PAREDES LIZAMA, Álvaro (Data Observatory Foundation, Chile)

Presenter: Mr PAREDES LIZAMA, Álvaro (Data Observatory Foundation, Chile)

Session Classification: Presentations Session 4: Data Stewardship

Track Classification: SciDataCon Persistent Themes: Data Stewardship