

EXTERNAL_CATALOG_CONNECTIONS

External Catalog Connections

PRISM - Platform for Research Infrastructure Synergy Mapping

This document outlines potential connections and integration opportunities with external research tool catalogs and registries.

1. RDA IOI Infrafinder

Website: <https://www.infrafinder.io/>

Maintained by: RDA Infrastructure Opportunities & Integration (IOI) Interest Group

Purpose: Global registry of research infrastructures across all scientific domains

Connection Opportunities

- **Bidirectional lookup:** PRISM tools can link to Infrafinder entries for institutional context
- **Metadata enrichment:** Import governance, funding, and organizational data
- **Lifecycle alignment:** Map Infrafinder infrastructure stages to MaLDReTH lifecycle stages
- **API integration:** Potential REST API for automated data synchronization

Data Fields Alignment

PRISM Field	Infrafinder Equivalent
Tool Name	Infrastructure Name
Description	Description / Purpose
Organization	Host Institution
URL	Website URL
Category	Research Domain

Implementation Notes

- Infrafinder focuses on **physical** research infrastructures (labs, facilities, instruments)
 - PRISM focuses on **digital** research tools and software
 - Complementary rather than overlapping - potential for cross-referencing
 - Could create "Tool → Infrastructure" relationships (e.g., LIMS → Lab Facility)
-

2. OpenAIRE Research Graph

Website: <https://graph.openaire.eu/>

Maintained by: OpenAIRE Consortium

Purpose: Open scholarly communication graph linking publications, data, software, and organizations

Connection Opportunities

- **Software identification:** Link PRISM tools to their OpenAIRE software records
- **Publication tracking:** Associate tools with research outputs that used them
- **Impact metrics:** Import citation counts and usage statistics
- **Funding data:** Connect tools to grant funding information

Data Fields Alignment

PRISM Field	OpenAIRE Equivalent
Tool Name	Software Title
Description	Software Description
URL	Code Repository / Landing Page
Organization	Creator Organization
is_open_source	License Type (Open vs Proprietary)

API Endpoints

```
# Search for software by name
GET https://api.openaire.eu/search/software?q=<tool_name>

# Get software details
GET https://api.openaire.eu/search/software/<software_id>
```

3. Research Organization Registry (ROR)

Website: <https://ror.org/>

Maintained by: ROR Community

Purpose: Persistent identifiers for research organizations

Connection Opportunities

- **Organization linking:** Associate tools with their developing institutions via ROR IDs
- **Provenance tracking:** Link tool creators to verified organizations
- **Collaboration mapping:** Identify multi-institutional tool development
- **Metadata standardization:** Use ROR IDs as canonical organization identifiers

Data Fields Alignment

PRISM Field	ROR Equivalent
Organization	Organization Name
-	ROR ID (e.g., https://ror.org/abcd1234)
-	Organization Type
-	Country / Location

Implementation Pattern

```
# Example: Store ROR ID alongside organization name
tool.organization = "University of California"
tool.ror_id = "https://ror.org/04yrm1557" # UC Berkeley
```

4. bio.tools (ELIXIR Tools Registry)

Website: <https://bio.tools/>

Maintained by: ELIXIR Infrastructure

Purpose: Registry of life sciences software tools and databases

Connection Opportunities

- Domain-specific enrichment:** Import detailed metadata for bioinformatics tools
- EDAM ontology:** Standardized operations, topics, and data types
- Publication links:** Associate tools with their primary citations
- Technical specifications:** Import supported data formats, platforms, languages

Data Fields Alignment

PRISM Field	bio.tools Equivalent
Tool Name	Tool Name
Description	Description
URL	Homepage
is_open_source	License
Category	EDAM Topic
-	EDAM Operation
-	Supported Input/Output Formats

API Integration

```
# Search bio.tools
GET https://bio.tools/api/tool?q=<tool_name>

# Get tool details
GET https://bio.tools/api/tool/<biotoolsID>
```

5. FAIRsharing.org

Website: <https://fairsharing.org/>

Maintained by: FAIRsharing Community

Purpose: Registry of standards, databases, and policies for research data

Connection Opportunities

- **Standards compliance:** Link tools to the standards they implement
- **Database connections:** Identify which databases tools can access
- **FAIR assessment:** Import FAIRsharing FAIR scores and compliance indicators
- **Policy alignment:** Connect tools to institutional/funder data policies

Data Fields Alignment

PRISM Field	FAIRsharing Equivalent
Tool Name	Resource Name
Description	Description
URL	Homepage
Category	Subject Domains
-	Standards Implemented
-	Related Databases
-	FAIR Indicators

Use Cases

- Identify tools that implement specific metadata standards (e.g., Dublin Core, DataCite)
 - Find tools compatible with specific repository platforms
 - Track FAIR-enabling tools across lifecycle stages
-

6. GitHub / GitLab / Software Heritage

Platforms: github.com, gitlab.com, softwareheritage.org

Purpose: Source code repositories and software preservation

Connection Opportunities

- **Code repository linking:** Direct links to tool source code
- **Version tracking:** Monitor tool releases and updates
- **Contributor data:** Identify development teams and communities
- **License verification:** Confirm open source status and license types
- **Activity metrics:** Stars, forks, commits, issues as quality indicators

Data Fields Alignment

PRISM Field	Repository Equivalent
Tool Name	Repository Name
Description	Repository Description
URL	Repository URL
is_open_source	License (parsed from LICENSE file)
-	Stars / Forks Count
-	Last Commit Date
-	Primary Language
-	Contributors Count

GitHub API Example

```
# Get repository information
GET https://api.github.com/repos/{owner}/{repo}

# Response includes:
{
  "name": "tool-name",
  "description": "Tool description",
  "html_url": "https://github.com/owner/repo",
  "license": { "name": "MIT" },
  "stargazers_count": 1234,
  "open_issues_count": 42,
  "updated_at": "2024-10-06T12:00:00Z"
}
```

7. EOSC Marketplace

Website: <https://marketplace.eosc-portal.eu/>

Maintained by: European Open Science Cloud (EOSC)

Purpose: Catalog of research services, tools, and training materials

Connection Opportunities

- **Service integration:** Link PRISM tools to their EOSC service offerings
- **Access points:** Connect to cloud-deployed tool instances
- **Training materials:** Associate tools with learning resources
- **Usage policies:** Import terms of use and access conditions

Data Fields Alignment

PRISM Field	EOSC Marketplace Equivalent
Tool Name	Service Name
Description	Service Description
URL	Service URL
Organization	Service Provider
Category	Scientific Domain
-	Service Category
-	Access Type (Open/Restricted)
-	TRL (Technology Readiness Level)

8. Wikidata

Website: <https://www.wikidata.org/>

Purpose: Free collaborative knowledge base with structured data about everything

Connection Opportunities

- **Unique identifiers:** Q-numbers for tools, organizations, concepts
- **Multilingual data:** Tool names and descriptions in multiple languages
- **Relationship mapping:** Rich graph of "part of", "uses", "developed by" connections
- **Authority control:** Link to other identifier systems (ORCID, DOI, etc.)

Query Example (SPARQL)

```

# Find research software tools
SELECT ?tool ?toolLabel ?website WHERE {
  ?tool wdt:P31 wd:Q7397. # instance of: software
  ?tool wdt:P366 wd:Q42240. # use: research
  OPTIONAL { ?tool wdt:P856 ?website }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}

```

9. re3data (Registry of Research Data Repositories)

Website: <https://www.re3data.org/>

Maintained by: re3data.org consortium

Purpose: Global registry of research data repositories

Connection Opportunities

- **Repository tools:** Identify software that powers data repositories
- **Preservation connections:** Link tools to long-term preservation repositories
- **Data deposit:** Map tools that enable deposit to specific repositories
- **Subject domains:** Align repository subjects with tool categories

Data Fields Alignment

PRISM Field	re3data Equivalent
-	Repository Name
-	Repository Software Platform
Tool → Repository	Data Submission Tools
Lifecycle Stage	Data Lifecycle Support

Use Case

Map PRISM "PRESERVE" and "STORE" stage tools to re3data repositories they support:

- Zenodo → Zenodo repository entry
 - Dataverse → Harvard Dataverse entry
 - DSpace → Multiple institutional DSpace instances
-

10. Research Data Alliance (RDA) Outputs

Website: <https://www.rd-alliance.org/>

Purpose: RDA working group outputs, recommendations, and tools

Connection Opportunities

- **RDA tool registry:** Potential future RDA-maintained tool catalog
- **Recommendation compliance:** Link tools to RDA recommendations they implement
- **Working group outputs:** Associate tools with relevant RDA WG deliverables
- **MaLDReTH alignment:** Direct connection to parent MaLDReTH II initiative

Integration Points

- RDA Metadata Standards Catalog
 - RDA Data Type Registries
 - RDA Persistent Identifier Tools
 - RDA Data Fabric recommendations
-

Implementation Priorities

Phase 1: High Priority (Immediate Value)

1. **GitHub API:** Enrich open source tools with repository data
2. **ROR:** Standardize organization identifiers
3. **bio.tools:** Import bioinformatics tool metadata (if applicable to domain)

Phase 2: Medium Priority (Enhanced Metadata)

1. **OpenAIRE:** Link to publications and software records
2. **FAIRsharing:** Connect tools to standards and databases
3. **Wikidata:** Add unique identifiers and multilingual support

Phase 3: Future Integration (Advanced Features)

1. **EOSC Marketplace:** Service deployment and access
 2. **IOI Infranfinder:** Physical infrastructure connections
 3. **re3data:** Repository software mapping
 4. **RDA Outputs:** Recommendation compliance tracking
-

Technical Implementation Patterns

1. External Identifier Storage

Add fields to `ExemplarTool` model:

```
class ExemplarTool(db.Model):
    # ... existing fields ...

    # External identifiers
    github_url = db.Column(db.String(500))
    biotools_id = db.Column(db.String(100))
    ror_id = db.Column(db.String(100))
    wikidata_qid = db.Column(db.String(50))
    eosc_id = db.Column(db.String(100))
    openaire_id = db.Column(db.String(100))
```

2. API Integration Example

```
import requests

def enrich_from_github(tool_name):
    """Fetch metadata from GitHub API."""
    url = f"https://api.github.com/search/repositories?q={tool_name}"
    response = requests.get(url, headers={'Accept': 'application/vnd.github.v3+json'})

    if response.status_code == 200:
        data = response.json()
        if data['total_count'] > 0:
            repo = data['items'][0]
            return {
                'description': repo['description'],
                'url': repo['html_url'],
                'is_open_source': repo['license'] is not None,
                'stars': repo['stargazers_count'],
                'language': repo['language']
            }
    return None
```

3. Bulk Enrichment Strategy

```
# Pseudocode for enrichment workflow
for tool in tools_without_metadata:
    # Try GitHub first (if appears to be software)
    github_data = enrich_from_github(tool.name)

    # Try bio.tools (if in life sciences domain)
    biotools_data = enrich_from_biotools(tool.name)

    # Try Wikidata (general fallback)
    wikidata_data = enrich_from_wikidata(tool.name)

    # Merge data with conflict resolution
    tool.merge_external_data([github_data, biotools_data, wikidata_data])
```

```
# Validate and commit
db.session.commit()
```

Data Quality Considerations

Matching Challenges

- **Name variations:** "R Project" vs "R" vs "R Statistical Software"
- **Disambiguation:** Multiple tools with similar names
- **Scope differences:** PRISM digital tools vs Infrafinder physical infrastructure

Validation Requirements

- **Manual review:** Flag automated matches for human verification
- **Confidence scores:** Track match quality (exact, probable, uncertain)
- **Update frequency:** Schedule periodic re-enrichment from external sources
- **Provenance:** Record data source for each field

Governance

- **License compatibility:** Ensure external data can be used and redistributed
 - **Attribution:** Credit data sources appropriately
 - **Privacy:** Respect contributor data protection requirements
-

Contributing

To propose additional external catalog connections:

1. Create an issue describing the catalog and connection value
 2. Document API availability and data licensing
 3. Provide mapping between PRISM and external catalog fields
 4. Submit pull request with documentation updates
-

Last Updated: October 2025

Maintainer: MaLDReTH II RDA Working Group

Contact: <https://www.rd-alliance.org/groups/mapping-the-landscape-of-digital-research-tools-ii-maldreth-ii>