Using the IR as a Research Data Registry

Daryl Grenz, Eirini Mastoraki, Han Wang, Mohamed Ba-essa
University Library, King Abdullah University of Science and Technology (KAUST), Saudi Arabia

**Concept**

Just as an institution may benefit from knowing about all publications that are produced by researchers affiliated to the institution, there is also potential value in having a record of other forms of research outputs, such as datasets and software code, that are released by affiliated researchers. One avenue for identifying such outputs is on the basis of stated relationships between research data and publications. Aside from the text of the publication itself, there are also several API sources for such relationship information.

**Projects**

**Retrospective Link Harvesting**

Partial Automation Process

- Query DataCite API for relatedIdentifiers matching publication DOIs or PubMed IDs
- Generate repository records when:
  - relationship is IsSupplementTo
  - dataset creator is KAUST-affiliated on the related publication
- Update existing publication records with relationship statements pointing to the related datasets

**First Batch Outcome (June 2017)**

- 9299 Queries to DataCite by publication DOI
- 724 DataCite records retrieved
- 654 Metadata records added to repository
- 274 Publication records updated with relationships to datasets

**Data Repositories Represented**

- Cambridge Crystallographic Data Centre (CCDC)
- Figshare
- Dryad

**New Publication Processing**

Manual Workflow

1) Check for “Data Availability” or similar sections.
2) Capture included identifiers or links, and their type.

**Outcomes**

Data Availability Statements
(For 3528 publications between Feb. 2017 and May 2018)

- Without Data Availability Statements: 3259 (92%)
- Statements with Links or IDs: 269 (8%)
- Statements without Links or IDs: 93 (3%)

ID and Link Types Found
(For 176 publications with links or IDs)

- Accession Numbers: 47 (27%)
- Code URLs: 24 (14%)
- Dataset DOIs: 26 (15%)
- Other URLs: 46 (26%)
- Multiple Types of IDs or Links: 16 (9%)

**Future Work:**

Adopt Unified Approach to Registering and Representing Research Data in the IR

**Growing Potential of Scholix Services**

**DLI (beta)**
Queried May 2017

- Javascript Query UI + OAI-PMH
- REST API returns JSON records

- 669 relationships to 193 publications
- 211 related items with non-DOI ID types

**Scholexplorer**
Queried May 2018

- ENA accession numbers from Published
- PDB accession numbers from RCSB
- EMBL accession numbers from Published
- ENA accession numbers from Published
- PDB accession numbers from RCSB
- ENA query URLs from DNA

- 3794 relationships to 574 publications
- 565 related items with non-DOI ID types

**Contact Email:** daryl.grenz@kaust.edu.sa

Repository Record: [https://tinyurl.com/IR-as-data-registry](https://tinyurl.com/IR-as-data-registry)