Memorial Sloan Kettering Cancer Center

Data Catalogs, Metadata, and Data Discovery:

How data catalogs extend and enhance the IR ecosystem.

Anthony Dellureficio,
Associate Librarian for Data Management Services

Steering Committee Member, Data Discovery Collaboration

MIRL Symposium
November 17, 2021

Data Discovery Collaboration (DDC)

Community of Practice

- Multi-institutional consortium,
- Institutional and individual members,
- Platform-agnostic,
- Support FAIR data principles in the broader context of library and institutional systems,
- Focus on the data catalogs and other metadata expertise,
- Support research data discoverability through:
  - metadata,
  - outreach,
  - software development.

We welcome members without data catalogs!

What is a Data Catalog?

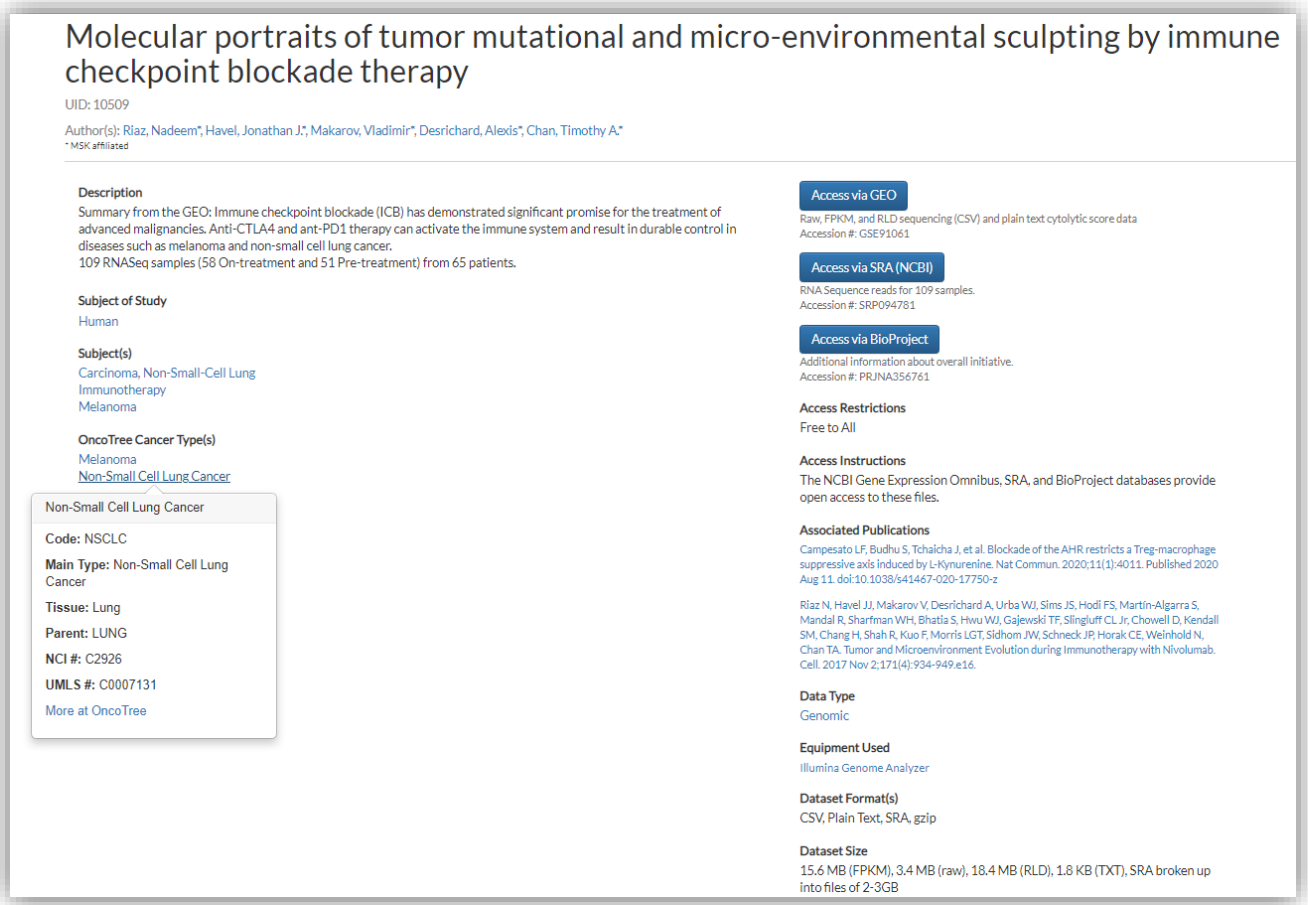
MSK Data Catalog - https://datacatalog.mskcc.org/

Includes datasets, code, analytical tools, or other research outputs not traditionally included in bibliographic catalogs.

Features:

- Metadata only - not repository,
- Curated, enhanced metadata,
- Wayfinder describing access and restrictions,
- Connect researchers,
- Standardized/normalized schema,
- Institution-specific - not aggregator,
- Identify relevant analytical tools,
- Track reuse, publications, and funding,
- Highlight otherwise undiscoverable datasets,
- Accommodate concerns over PHI exposure.

Flexibility to integrate with local infrastructure!

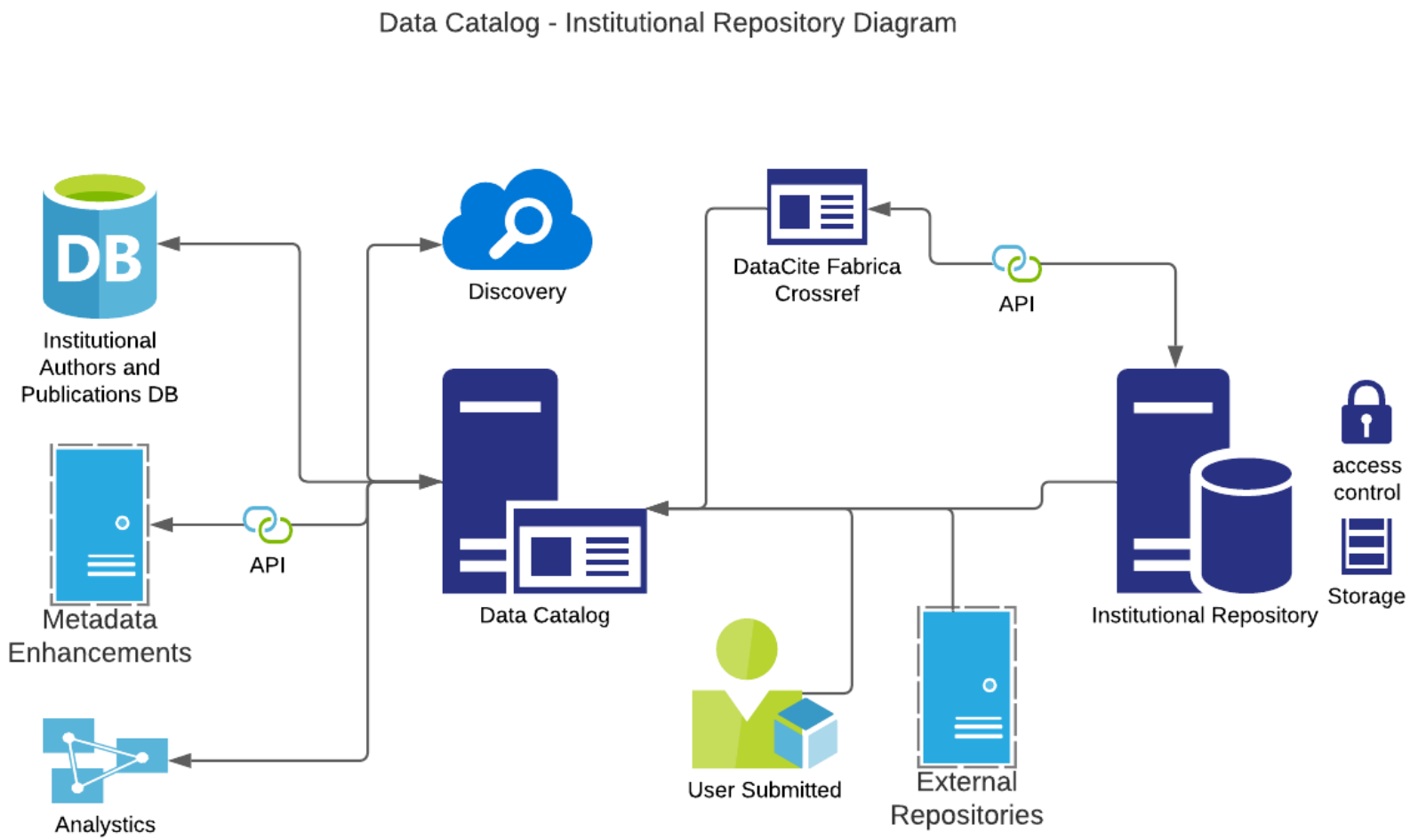


Data Catalogs and Institutional Repositories

Reinforce FAIR data principles

Metadata ecosystem includes:

- Data catalog,
- Institutional Repository,
- External repos,
- Institutional author/pub db,
- PID minting applications,
- Discovery platform,
- External taxonomy/ authority dbs
- Other library resources.



Thank you!

Resources and contact info

Data Discovery Collaboration (DDC) Steering Committee:

- **Nicole Contaxis** (NYU HSL)…….. Nicole.Contaxis@nyulangone.org
- **Jason Clark** (Montana State University)…….. jaclark@montana.edu
- **Anthony Dellureficio** (MSK)……..dellurea@mskcc.org
- **Melissa Ratajeski** (University of Pittsburgh)……... mar@pitt.edu
- **Terrie Wheeler** (Weill Cornell)…….. tew2004@med.cornell.edu

DDC Website: https://datadiscoverycollaboration.org/

MSK Data Catalog: https://datacatalog.mskcc.org/

Catalog code: https://github.com/msk-library/data-catalog

Recent DDC publication on curation through data catalogs:

Sheridan H, Dellureficio AJ, Ratajeski MA, Mannheimer S, Wheeler TR. Data Curation through Catalogs: A Repository-Independent Model for Data Discovery. *Journal of eScience Librarianship* 2021;10(3): e1203. https://doi.org/10.7191/jeslib.2021.1203.