NRC·CNRC

Research Data Alliance Outputs

David Schade

Canadian Astronomy Data Centre
Canadian Virtual Observatory
Canadian Advanced Network for Astronomy Research (CANFAR)

Co-Chair: Research Data Alliance Interest Group
Disciplinary Collaboration Framework

RESEARCH DATA ALLIANCE

O&A Members 56

Active Organisational & Affiliate members

MEMBERSHIP Members: 8284

Becoming a member of RDA is simple and open to both individuals and organizations

Register now

RDA Groups WG & IGs: 102

Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups

NEWS & EVENTS

RECENT BLOGS

FOLLOW US

Outputs and Recommendations

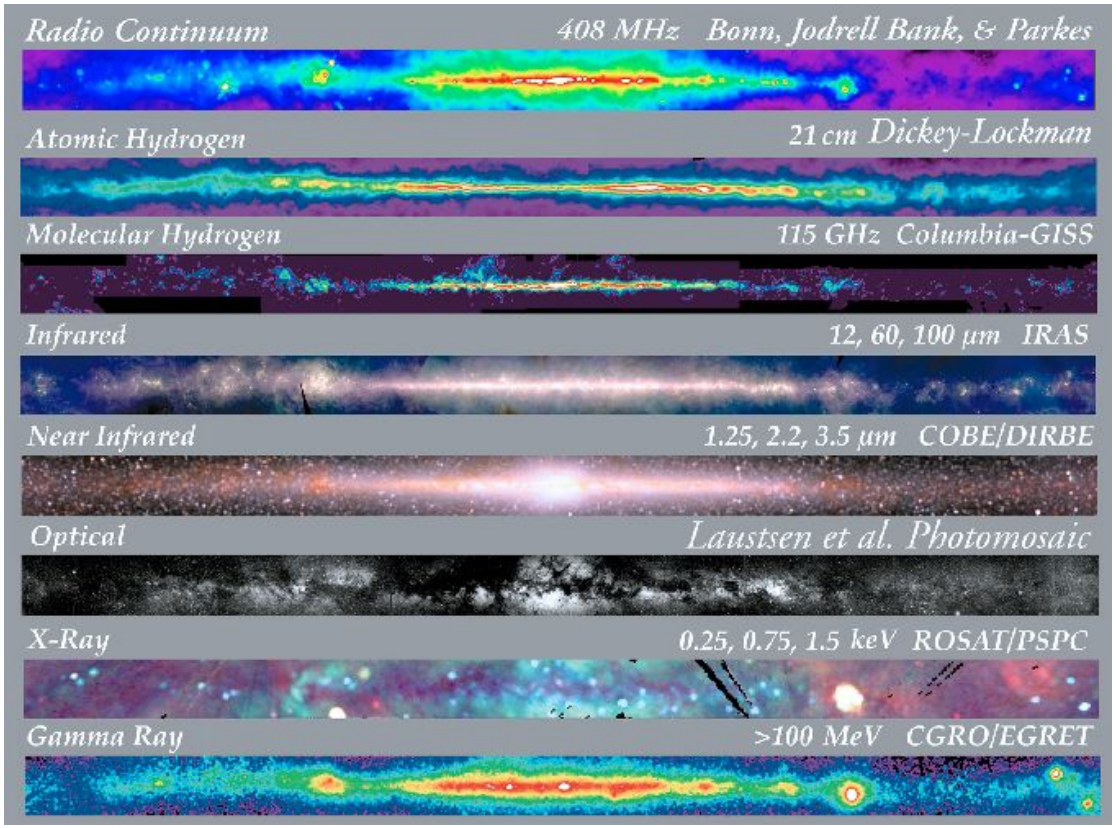
- Scalable Dynamic-data Citation Methodology
- Data Description Registry Interoperability Model
- Basic Vocabulary of Foundational Terminology Query Tool
- Data Type Model and Registry
- The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies
- Persistent Identifier Type Registry

RDA
RESEARCH DATA ALLIANCE

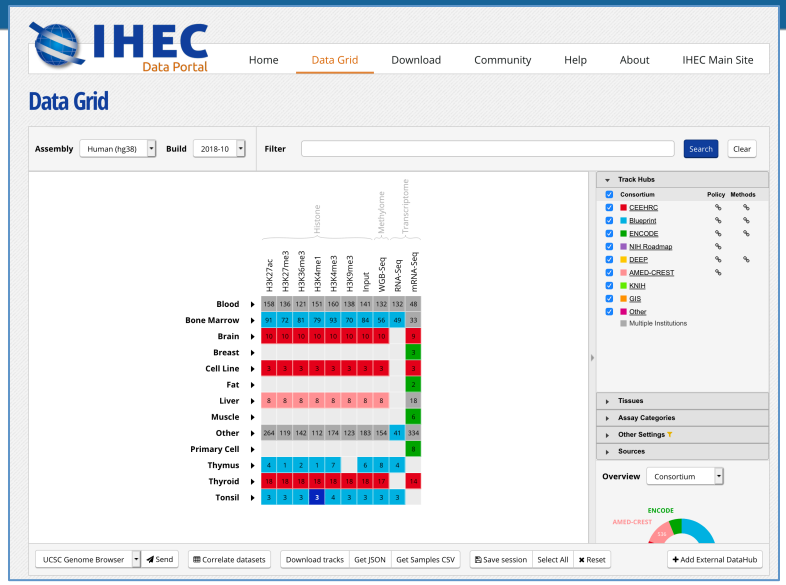
RDA is a place where you could look for relevant outputs that represent useful thinking on your subject

Astronomy is multi-wavelength & multi-facility

The universe cannot be understood without looking across a large range of (wavelength) energy of “light”

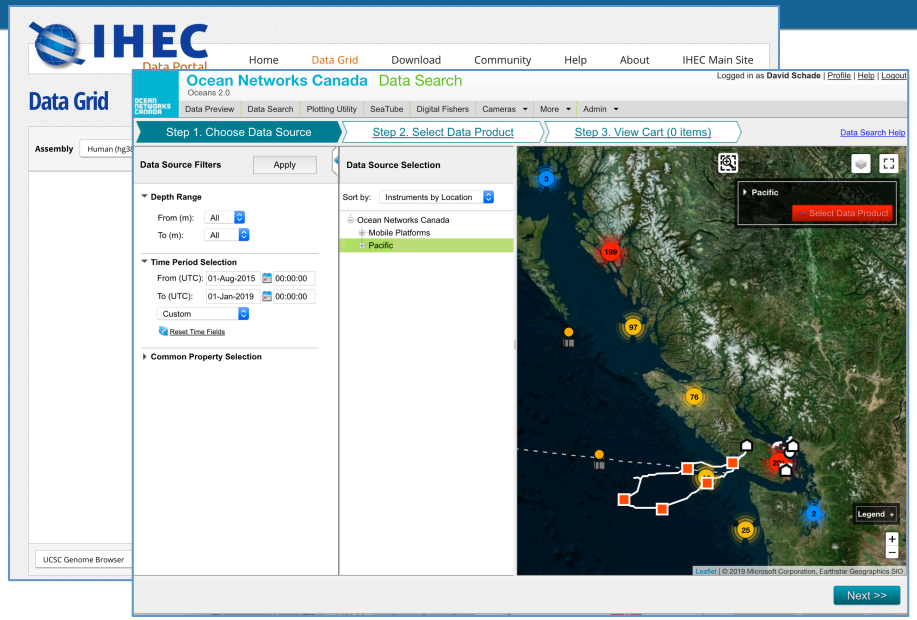


Integrative Research is everywhere



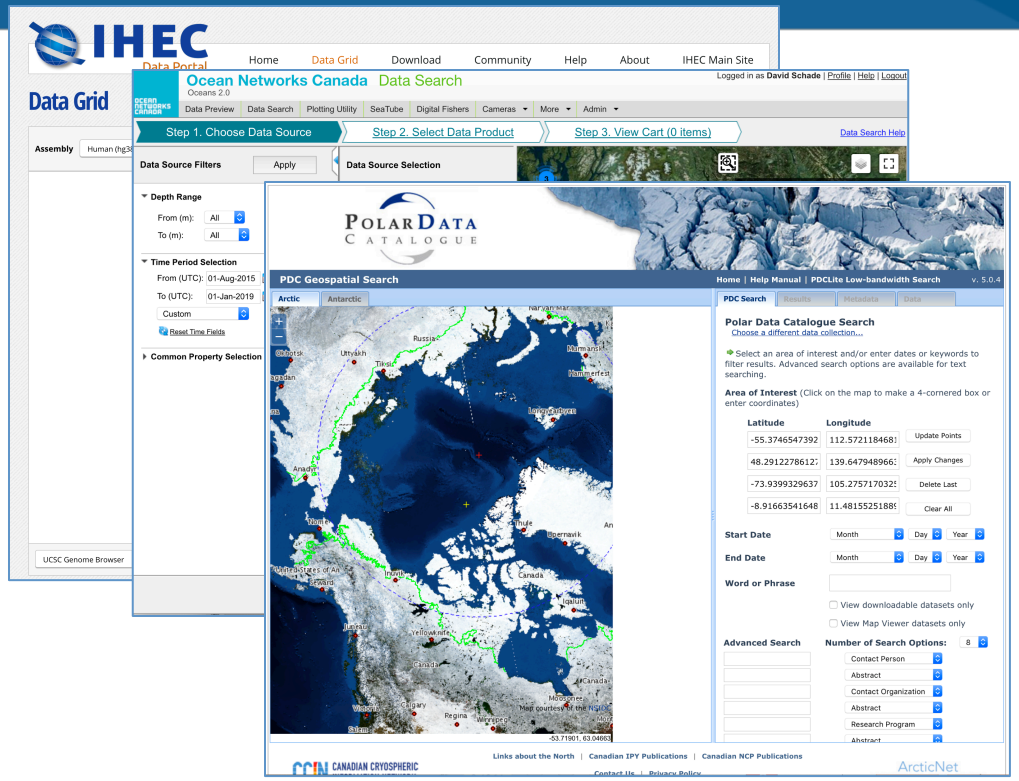
Standardization of metadata to support InterOperability is *everywhere*

Integrative Research is everywhere



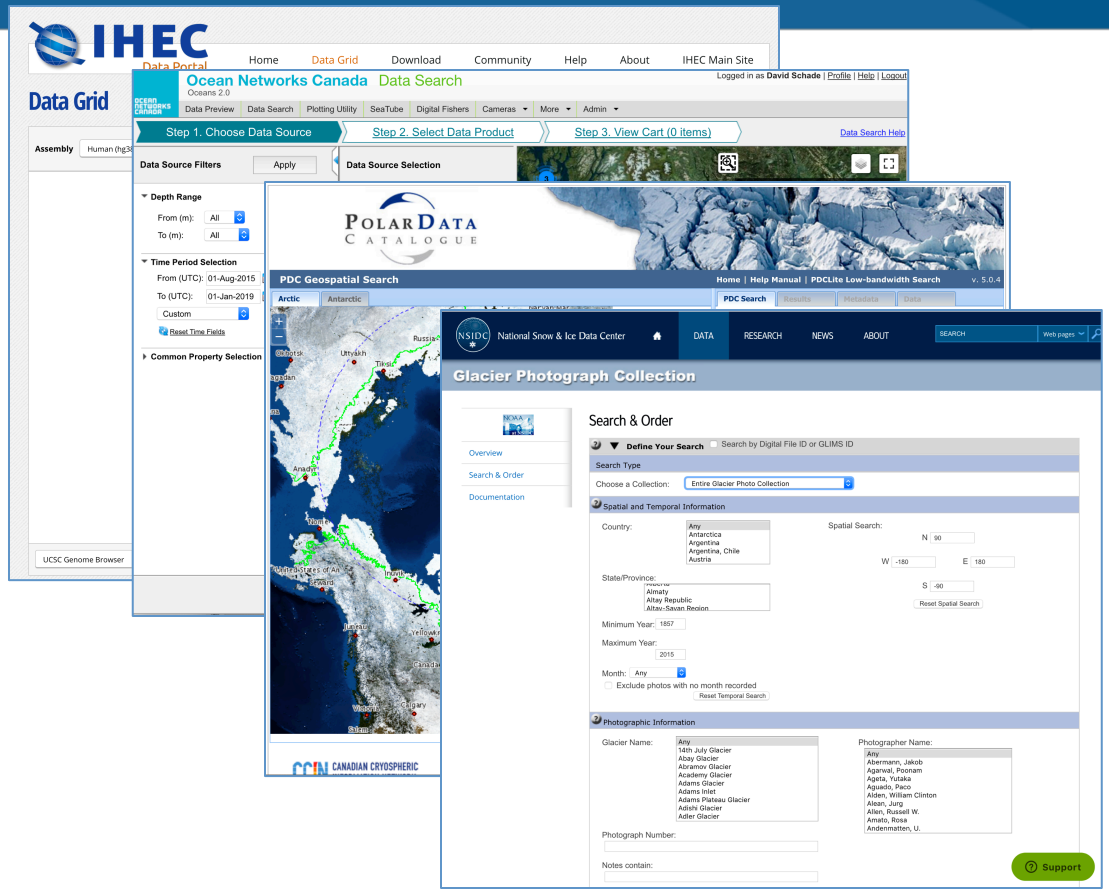
Standardization of metadata to support InterOperability is *everywhere*

Integrative Research is everywhere



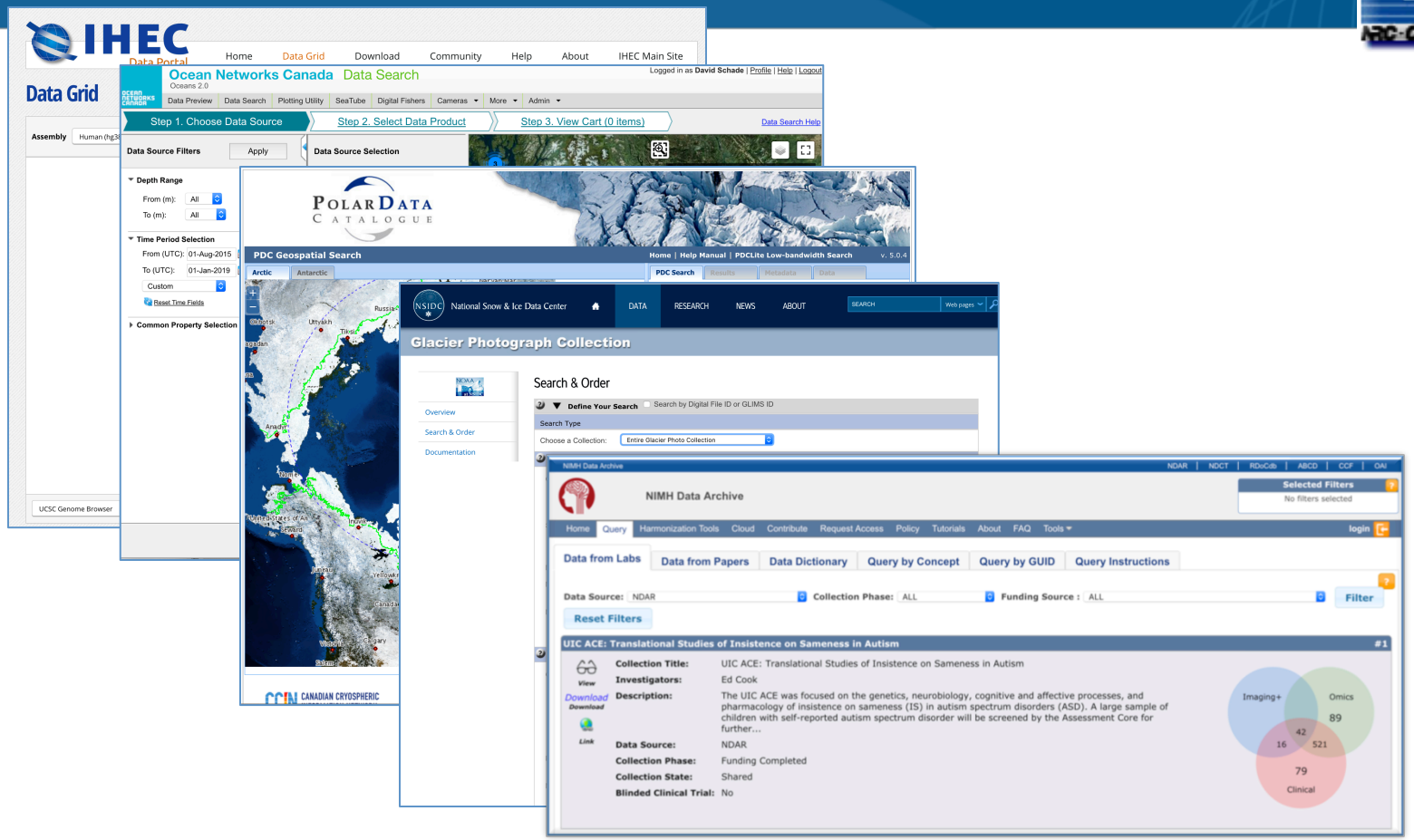
Standardization of metadata to support InterOperability is *everywhere*

Integrative Research is everywhere



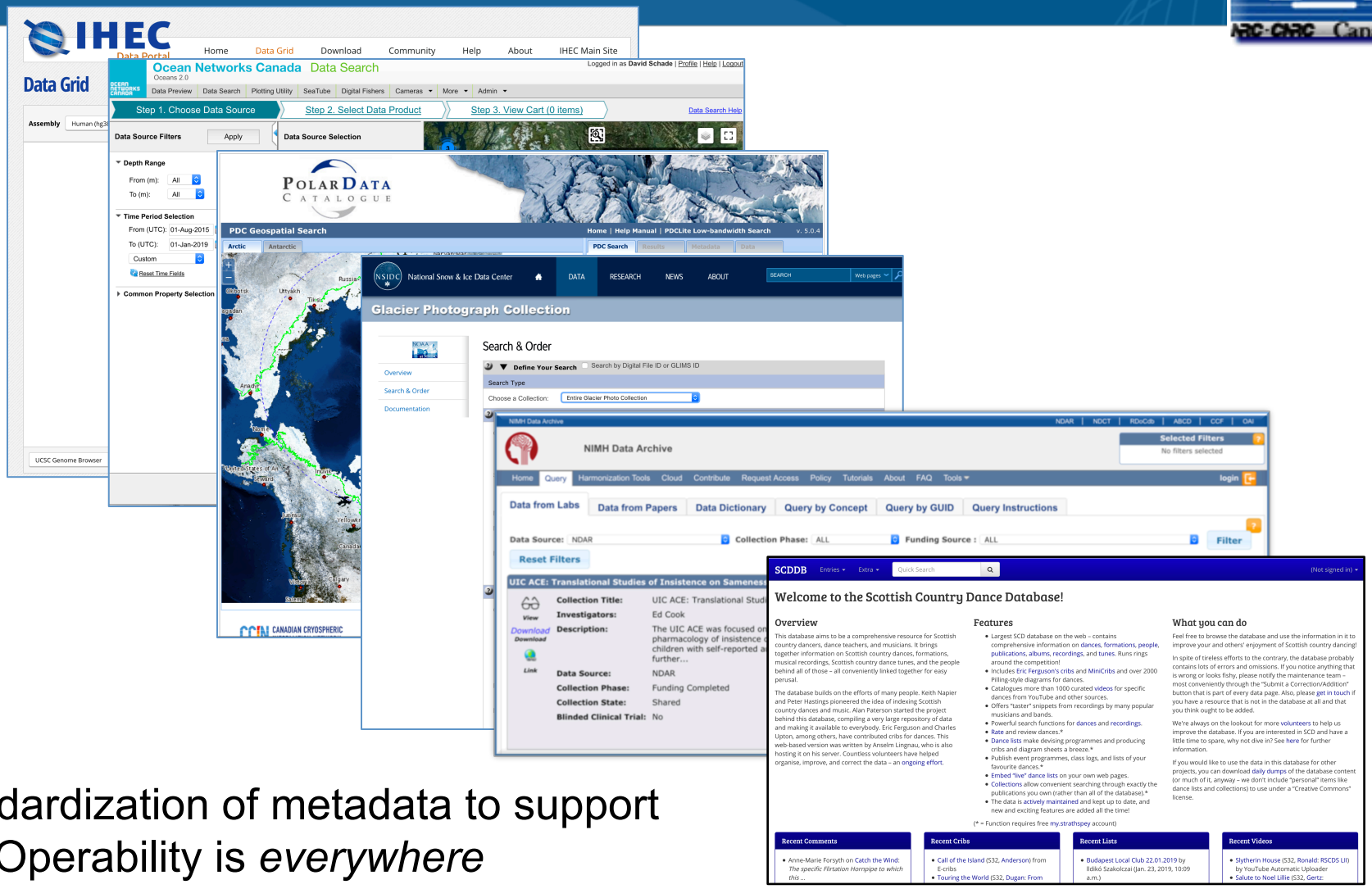
Standardization of metadata to support InterOperability is *everywhere*

Integrative Research is everywhere



Standardization of metadata to support InterOperability is *everywhere*

Integrative Research is everywhere



Standardization of metadata to support InterOperability is *everywhere*

For example:

Data Citation of Evolving Data

Recommendations of the Working Group on Data Citation (WGDC)

Andreas Rauber, Ari Asmi, Dieter van Uytvanck and Stefan Pröll

Revision of October 20th 2015

I. Making Data Citable

These WGDC recommendations enable researchers and data centers to identify and cite data used in experiments and studies. Instead of providing static data exports or textual descriptions of data subsets, we support a dynamic, query centric view of data sets. The proposed solution enables precise identification of the very subset and version of data used, supporting reproducibility of processes, sharing and reuse of data.

Goals of this WG are to create identification mechanisms that:

- allows us to identify and cite arbitrary views of data, from a single record to an entire data set in a precise machine-

A. Preparing the Data and the Query Store

Prepare existing data sources and provide the required infrastructure, which is needed for implementing the query based approach.

- **R1 – Data Versioning**: Apply versioning to ensure earlier states of data sets can be retrieved.
- **R2 – Timestamping**: Ensure that operations on data are timestamped, i.e. any additions, deletions are marked with a timestamp.
- **R3 – Query Store Facilities**: Provide means for storing queries and the associated metadata in order to re-execute them in the future.

B. Persistently Identify Specific Data Sets

When a data set should be persisted, the following steps need to be applied:

Recommendation on data citation

- Relies on executing a query that returns exactly the dataset to be cited
- Saves storage
- Versions, timestamps of data and queries

Recommendation on data citation

I view this as high-level work that presents obstacles to practical implementation

- The assumption is that all datasets that you want to publish/cite can be represented as a query (or set of queries)
  - Not all published data exists in queryable systems
  - Datasets are distributed across systems
  - Astronomers find data from queries (sometimes) and then apply arbitrary selections

RDA
RESEARCH DATA ALLIANCE

RDA is a place where you could look for relevant outputs that represent useful thinking on your subject

Disciplinary Collaboration Framework Interest Group

Ultimately, a major goal of the Research Data Alliance is increasing the power of ***cross-disciplinary research***

The universe cannot be understood by studying a single element

Metadata: The principal challenge

Data models: the metadata and metadata structures that characterize a data collection

Development of cross-discipline Data Models is the principal challenge

Generic Stream

- Very light metadata
  - limited “Findability”
- Little Integration or harmonization of metadata
  - limited accessibility
- Mix of content: data, literature, links, documents
- Not developed by researchers
- Provide useful service
- Support obligation to publish data

Research-centric Stream

- Metadata that has sufficient depth for the research
- Data Integration
- Management of data types and links to appropriate applications
- Developed by researchers themselves
- Integral part of the advancement of research

Disciplinary Collaboration Framework Interest Group

There are successful examples ***cross-disciplinary research*** and ***Integrative cross-disciplinary Research Data Management systems***

Let's learn from those successes

END