DOIs in RDA

André Schaaff

Centre de Données astronomiques de Strasbourg

DCP Session - 10/11/2018

IVOA, College Park USA, 8-10/11/2018

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Remarks

- RDA is agnostic as much as possible
- You will find mainly the generic PID acronym in the documents

- A RDA feedback reported during the previous DCP sessions
- A brief overview (and links) concerning PIDs in RDA

Data Citation of evolving Data Recommendation

- Related in previous DCP sessions
- Implementors are listed on the RDA WG pages

The challenge

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-actionable manner
- allows us to cite and retrieve that data as it existed at a certain point in time, whether the database is static or highly dynamic
- is stable across different technologies and technological changes

Credits: RDA DC WG

Data Citation of evolving Data Recommendation

- Rules concerning PID

Which PID system should be used? Any PID system can be applied according to the institutional policy.

R8 – Query PID: Assign a new PID to the query if either the query is new or if the result set returned from an earlier identical query is different due to changes in the data. Otherwise, return the existing PID.

R9 – Store Query: Store query and metadata (e.g. PID, original and normalized query, query & result set checksum, timestamp, superset PID, data set description, and other) in the query store.

R10 – Automated Citation Texts: Generate citation texts in the format prevalent in the designated community for lowering the barrier for citing the data. Include the PID into the citation text snippet.

R11 – Landing Page: Make the PIDs resolve to a human readable landing page that provides the data (via query re-execution) and metadata, including a link to the superset (PID of the data source) and citation text snippet.

Solution: The WG recommends solving this challenge by:

- ensuring that data is stored in a versioned and timestamped manner.
- identifying data sets by storing and assigning persistent identifiers (PIDs) to timestamped queries that can be re-executed against the timestamped data store.

Credits: RDA DC WG

PID Information Types (PIT) Recommendations

- Recommendations (Endorsed)
  - https://rd-alliance.org/group/pid-information-types-wg/outcomes/pid-information-types
  - Essential types of information associated with PIDs
  - API proposal and demonstrator

PID Information Types(PIT) Recommendations (2)

Name	Range	Identifier	Flags
Type: Citation Information (EXAMPLE namespace) 11314.2/d5396a97c316a0eaca055846ba4233ac			
Title	STRING	11314.2/07841c3f84cbe0d4ff8687d0028c2622	
Creator	STRING	11314.2/31810b2c24913929bb5e0d4d949de9f7	
Publication date	DATE	11314.2/daed5901fbbe2570ee95c4009c739de2	
Language	STRING	11314.2/56211d62153b3500ce3b16cf86d6b403	optional
License	STRING	11314.2/2f305c8320611911a9926bb58dfad8c9	optional
Type: System level access information (EXAMPLE namespace) 11314.2/09d35f22e48b60284029ba51c17e2944			
Creation date	DATE	11314.2/6b3e1230d1b68965e290b16a43d2f46d	
Deletion date	DATE	11314.2/7e78be9736ad7f6bb5fb31218821eba5	optional
Permissions	STRING	11314.2/d057258f7b406fd9aad5a3893aba8208	optional
Checksum	STRING	11314.2/56bb4d16b75ae50015b3ed634bbb519f	
Object size (in bytes)	STRING	11314.2/0006e2b8e2f6e1ecce836e593bed38ae	
Type: Aggregation information (EXAMPLE namespace) 11314.2/699d487eff50c2e10982f4b85ed053a9			
Parent object identifier	IDENTIFIER	11314.2/f9e66e5f64ba3179d8f1e64138c69e04	optional
Child object identifier	IDENTIFIER	11314.2/f8db9e3b5f97aa8168fbd59788476375	optional
Type: Versioning information (EXAMPLE namespace) 11314.2/6b507d787dd06e4eb8f23b5bb56ae8bb			
Predecessor identifier	IDENTIFIER	11314.2/467d9ba30e2d9879fd9d483f319e462c	optional
Successor identifier	IDENTIFIER	11314.2/fc78024cb9dac0b0a80ed631ea650d4b	optional
Type: Preliminary example for EUDAT core information (EUDAT namespace) 11314.2/5f45666fc8689e3565728ca512c1b5e7			
Checksum	STRING	see above	
Format	STRING	11314.2/1a4f53a28b72d4bf4f8fdda7a2089595	
Data identifier	IDENTIFIER	11314.2/24dd85c4a3d39fb0d7e83a510a5041c6	
Metadata identifier	IDENTIFIER	11314.2/58a44100d2bcd1a34fb87eb87bc6f701	
Repository of Record	IDENTIFIER	11314.2/5546b0166091d9ae869f081f5548f3fc	
Mutability flag	BOOLEAN	11314.2/7c81e954eaead6a2f772abd83986d3e9	
Landing page address	URL	11314.2/66af2639d388977e81b85f6413df1e2c	
Date of deposition	DATE	11314.2/35837218f18dcc54a2d32e0fb30fa7fb	

Credits: RDA PIT WG

Scholarly Link Exchange WG

- It aims to “enable a comprehensive global view of the links between scholarly literature and data”.
- It takes into account Existing work + international initiatives (including heterogenous PID systems) to propose global information commons
- A result (including a schema): http://www.scholix.org/

Group of European Data Experts in RDA

- Interested (work) document “List of Assertions on PIDs”
  - https://www.rd-alliance.org/system/files/documents/GEDE%20PID%20Focus%20Area%20discussion%20document%20v2%20%28002%29.docx

… and also

- The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies
  - https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases-wg/outcomes/fairsharing
- RDA/WDS Publishing Data Workflows WG Recommendations
  - https://rd-alliance.org/system/files/Workflows_for_Research_Data_Publishing-_Models_and_Key_Components_submitted.pdf

… and also (2)

- Research Data Repository Interoperability WG Final Recommendations
  - https://www.rd-alliance.org/group/research-data-repository-interoperability-wg/outcomes/research-data-repository-0
- RDA Research Data Collections WG Recommendations
  - https://rd-alliance.org/group/research-data-collections-wg/outcomes/rda-research-data-collections-wg-recommendations

… and also (3)

- Data Description Registry Interoperability WG Recommendations (a reference to DOIs connections)
  - https://www.rd-alliance.org/system/files/DDRIOutputSpecification.pdf
- Data Fabric IG
- To follow: Software Source Code Identification WG, not yet endorsed

Conclusion

- Many done or ongoing actions in the frame of PIDs with a wide scope (papers, datasets, queries, software, …)
- We have to take this work into account