RDA 11 Berlin

feedback

André Schaaff

Centre de Données astronomiques de Strasbourg

DCP Session

IVOA, Victoria, 28/05-01/06/2018

Purpose

- A summary of selected sessions of RDA Berlin, Plenary 11, March 2018

WG Provenance Patterns

- Nick Car, Geoscience Australia
  - A database of elementary use cases
  - General enough to be recognized in many situations
  - Various Provenance patterns have been designed for them
  - W3C based
  - Implemented in Prov-RDF in TripleStore
- Interest for IVOA provenance use cases and how we have re-used the W3C concepts and serialisation formats
- https://www.rd-alliance.org/groups/provenance-patterns-wg

☐ WG Storage Service Definitions

- Works on a common vocabulary for the description of data storage services
  - machine readable
  - enough to describe both in-house storage services and various cloud storage
- We had "real world" examples
- Modern data model types (UML, DB Schema, Data exchange Schema, Semantic Web)
- Ontology for a Storage System Model + vocabularies for terms as needed
- http://purl.org/storagesys

BoF EOSC-related European Projects

- EOSC
  - Build a European Open Science Cloud
  - Offer a virtual environment for storage, management, analysis and re-use of research data
  - Important item for new European projects
  - Not build from scratch, EUDAT (B2Find service / IVOA Registry related in Sesto), EGI are core participants
  - First deadline, 2021

BoF EOSC-related European Projects (2)

- Aim of the session was to find new topics for RDA
  - Links with Data Usage Metrics WG (user statistics for data)
  - Link with Software Source Code IG as source code is also a product of Science, also PID IG
  - Link with Disciplinary Collaboration Framework IG to involve the community in the discussion
- Françoise attended the collocated event after RDA

IG Software Source Code

- Widely described during the Santiago DCP Session
- IG is interested in source code (and not applications / tools) linked to data production and processing
- Well organised with surveys to capture how the participants are acting in this domain
  - https://www.rd-alliance.org/sites/default/files/attachment/questionnaire.pdf
  - https://docs.google.com/spreadsheets/d/1gZXZcmL3E3B5mVbWwgpvr2z1dPyFpHF5L9JvOrGbsyI/edit#gid=0

⬜ IG Software Source Code (2)

- Important topic: Software Citation
  - Review of Force11 Software Citation WG results (Arfon Smith (STSI) was one of the group leader of this work, ADASS XXV BoF about this topic)
  - Different Identifier for each version of a source code ?

IG Repository Platforms for Research Data

- Repository interfaces for data analytics
  - Move the analysis to the data rather than moving the data to the analysis
  - If Data spread in multiple locations => split the analysis among multiple locations
  - Various solutions, but need a common understanding of the data, such as a data model
  - Providing functionality through a common API (how if multiple data types ?). Involves discovery and use of discovered data
  - Enable data visualization capabilities to look at data stored at one repository.

IG Repository Platforms for Research Data (2)

- Relationship to DMPs, integration with repositories

- Policies can be additional metadata relevant to data

IG Repository Platforms for Research Data (3)

- Can we develop a data architecture that is implementable ?

- Extend OAIS to an Archive interaction layer

- Common interoperable protocols and interface specifications for repositories

- Enabling access to any object within an OAIS repository

- A use case is needed.

IG Virtual Research Environments

- VRE and Science Gateways, Virtual Laboratories

- Aim: common architecture or best practices to enable the interoperability, avoid duplication of effort

- Key skills Issues in Europe (VRE), North American (SG), Australia (VL)

- https://docs.google.com/presentation/d/1f5rC3Z B-AKZpw3XFSRelDV42H1jOEmRr08zJu-1CR_w/edit?usp=sharing

☐ WG Data Usage Metrics

- Standardized metrics for data like for articles
- Overview of existing initiatives in this field (Make Data Count, DataOne, Elsevier, etc.)
- Some Data is loaded but not used
- Citation and usage (some people are using but not citing)
- Work with many datasets, cite them all ?
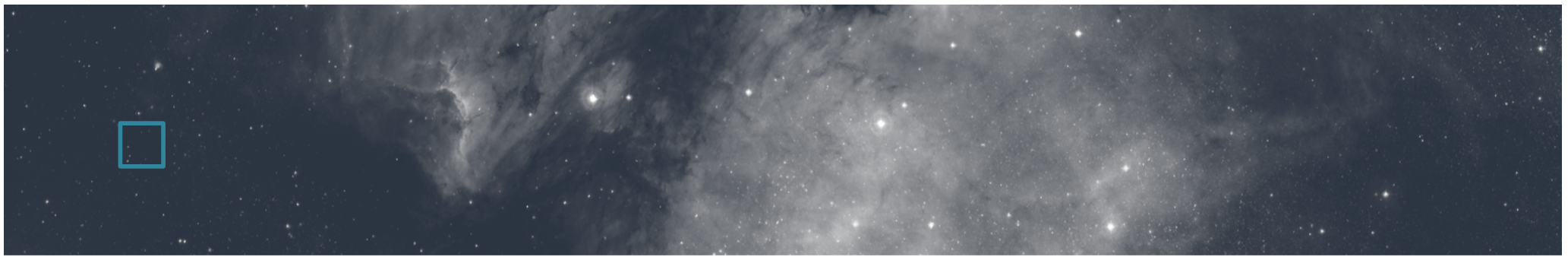- Realistic to put a DOI on each request (example: a database with 1 million queries per day...) ?

IG Research Data Architectures in Research Institutions

- Interoperability between institutional research data infrastructures and international or discipline based infrastructures

- New IG => presentation of best practices at Helsinky University and University College London

- Several differences but in any case a good place to share, to adapt, etc.

Conclusion

- Still no dedicated IG / WG to our domain
  - Has been discussed during Santiago DCP session
  - To small participation
- But, still many opportunities to participate and to get involved

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