2018-03-21 Trip Report -- 11th RDA Plenary, Berlin, Germany

About the Research Data Alliance

RDA plenary meetings are held every six months. The meetings have few plenary sessions — an opening session with keynote speakers and reports of work products, a closing, and a business session — and other than they keynotes, no invited speakers. RDA has 7000 members, and is world-wide with 50% of its membership from Europe. RDA is organized into birds of a feather sessions, which may lead to permanently standing interest groups, which may lead to scoped and temporary workgroups that produce work products for ratification by RDA. RDA currently has over 100 workgroups. The conference has 30 minute breaks, which include small poster sessions, gatherings at breakfast and lunch, a conference dinner, and a reception. These provide ample opportunities for conversation. In a typical workgroup meeting of 30 people, 75% are new people. The six month schedule, and the focus on work products and working groups creates a very vibrant, action-oriented environment.

Over the course of the three days I had productive conversations (in no particular order) with: Maryann Martone, Fiona Murphy, Amir Aryani, Gary Berg, David Wilcox, Tom Cramer, Michele Minielli, Josh Greenberg, Anne Theissen, Sören Auer, Corey Harper, David Dubin, Mark Leggott, Mark Musen, Leslie McIntosh Borelli, Nick Weber, Kathleen Shearer, Wolfram Horstmann, Raphael Ritz, Nick Weber, Peter Fox, Anne Theissen, Martin Fenner, Amanda Lawrence, Erik Schultes, Erik Car

Wednesday

Opening Session

Keynotes. SAP. Neuroscience. Both talks emphasized the role of data as a central nervous system for an organization.

Panel. Brennan, NLM

Bof towards a global open science commons

juan B the 7 c's of open science communication, collection, curation, computation, capacity, creation, collaboration

Simon Hodson. Codata. Re3data.org. Investigators collaborate on hiv, but data resides locally, isolated. Icsu world data system.

Ross ?? Australia. ANDS. Investments in capacity and collaboration. Need Trusted research data repository services.

Mark Leggott. Canada refocusing on Science. Now Canadians say open science. National science library of Canada folded when funds cut 70%

Brian Matthews. Eoscpilot.eu culture of disincentives, fragmentation between infrastructures, access to big data resources, barriers to interoperability. Pan-European governance. Science, technical, cultural challenges. OpenAire-Advance

Nick Weber. NIH Data Commons. Building across their institutes. Date, software, workflows, docs, articles. Nick.weber@nih.gov

IG. Kathleen Shearer and Wolfram Horstmann. Long tail of Research data

IG is interested in issues that arise as a result of the diversity of academic activities, and those not generally served by big data.

The focus was on repository activities. That's a shame. The problems occur much earlier in the research cycle.

Having spent fours years helping the researchers from a position in the deans office in the the college of liberal arts and sciences, I became well aware of problems facing these disciplines. That was 25 years ago and the issues have not changed— lack of resources, lack of data science skills, lack of colleagues who can participate in the various disciplines. A small investment goes a long way in the "long tail."

IG Gary Berg-Cross. Data Foundations and Terminology

I participate in this group helping to develop the TEDT

See slides. Raphael Ritz. Dft-tool.rd-alliance.org http://hdl.handle.net/21.11109/dft-2.0@#Active_Collection_2 Reptor turns a web server into a repository. Not a query tool.

Amir Aryani. Privacy.

Conference dinner at the Berlin Museum of Natural History

Dinosaurs! And more conversation about the future of research data.

Thursday

Plenary Session. Recommendations and Outputs

Dtr-1 data typing. Adopted by Deep Carbon ISO 11179

David Wilcox bagit for repository interchange. Include basic metadata data cite 4.0 schema github.com/RDADataRepositoryInteropWG

Anne Theissen. See data model regarding attribution based on prov-o. Enhances Darwin Core standard.

International materials registry wg Laura Bartolo. Discovery problem for material science, but general problem. GitHub.com/usnistgov/

Develop metadata standards at the ISO level. SC32 WG2. International community of metadata experts.

ISO 21838 Barry Smith. Top level ontologies.

AGU advances data practices for sharing. Enable Culture Change through Credit.

Shaik Meera. Rice Data interoperability.

Scholix. Elsevier. Link data with literature. Schema adoptable by hubs. Www.scholix.org

PID kernel WG. Wouter Haak. Need a tiny amount of metadata into PID records.

Wg. Data quality. Fai => R?

Remarkable session. Repositories recognizing they can not assess whether data in their repository is FAIR. Reusable is not an attribute of the repository. It is an attribute of the data in the context of a use case. Only domain specialists can determine if the data has appropriate specification to be re-used, and for what use case.

Wg. Provenance Patterns

First meeting. Http://patterns.promsns.org. GitHub RDAProvPatWG.

Excellent presentation by David Dubin contrasting VIVO-ISF and PROV with respect to provenance. Perhaps VIVO does not need PROV. VIVO already has reification and BFO based roles that can represent how data was created through work processes. PROV is not based on BFO and it's roles are not BFO roles. VIVO-ISF could assert that a PROV role was something in BFO, but it would not be a BFO role. For VIVO, PROV introduces duplicate classes (entity, time, role) already found in VIVO-ISF. A more detailed analysis and recommendation (both for VIVO and RDA) is warranted starting from Dubin's work

WG chair, Erik Car response was "we're using PROV" It will be up to VIVO-ISF to decide if and how to deal with PROV.

Presentation by Erik Car regarding his use case tool.

I made a comment regarding lumping and splitting the use cases to get to a set that represent reusable patterns. This is the expected activity as use cases are collected.

Most people were more interested in use cases for provenance than in provenance patterns.

BoF How expensive is FAIR?

Did not attend this session. Was in conversations, see above.

Plenary business session

Was not in this session. Was in conversations, see above.

Networking Reception

At the conference venue. Light fare. More conversations.

Friday

IG Vocabulary services

Task groups could lead to work groups. 150 members. Self-organizing. Very disorganized. Tackling and re-tackling middle level ontology questions. Several questions from participants indicating confusion.

Would like to participate in this group, but perhaps it needs to settle down a bit.

WG DDRI

Amir Aryani, ANDS. Research Graph. 50 million entities. 250 million vertices. Questions about data quality. Use PID. Met Anna. Martin Fenner. Mark Leggott.

National Computtional Infrastructure (Austalia). Amir presented on augmentation of their graph.

Fellow from GESIS presented on graph augmentation of the GESISS repository by Research Graph

My talk. Described the Research Graph VIVO Cloud Pilot https://doi.org/10.6084/m9.figshare.6022226.v2

Amanda Lawrence from APO

Closing Plenary

661 registrants. 41 countries. 179 from Germany.

P12 Will be in Gabarone Botswana, November 2018. With International Data Week

P13 Leslie McIntosh Borelli. Jane Greenberg. Philadelphia March/April 2019 Drexel University.