Link to RDA13 session page

Attendees:

1. Andrew Treloar, ARDC, Australia
2. Ari Asmi, U. Helsinki
3. Birger Jerlehag, SND Sweden
4. Christina Ore, PYT
5. Danie Kinkade, BCO-DMO
6. Dawei Lin, NIH/NIAID
7. Dieter Van Uytvanck, CLARIN
8. Garry Baker, BGS, UK
9. Hideaki Takeda, NII Japan
10. Ingrid Dillo, DANS
11. John Watkins, UKRI-CEH, UK
12. Jonathan Crabtree, UNC Odum Institute
13. Jonathan Petters, Virginia Tech
14. Kelsey Druken , NCI Australia
15. Kevin Ashley, UKDCC
16. Kheeran Dharmawardena, Atlas of Living Australia
17. Koenraad De Smedt, UIB, Norway
18. Leighton Christiansen. US Dept of Transportation
19. Marina Soares e Silva, Mendeley Data
20. Mike Brown, UKRI-CEH, UK
21. Mustapha Mokrane, DANS
22. Nigel Robinson, Clarivate Analytics UK
23. Nikoloai Svoboda, ZALF, Germany
24. Pascal Suppers, DataHub Maastricht, NL,
25. Philipp Conzett, UiT The Arctic University of Norway, and DataverseNO
26. Robert Chen, SEDAC, Columbia Univ.
27. Rorie Edmunds, WDS
28. Sarah Callaghan, STFC, UK
29. Shelley Stall, AGU
30. Shigeru Yatsuzuka, NBDC, Japan
31. Stefan Reichmann, TU Graz, Austria
32. Stefanie Kethers, ARDC, Australia
33. Steve Diggs, Scripps/UCSD
34. Tom Morrell, Caltech
35. Wang Shuai, University of Tokyo, Japan
36. Will Dean, Temple Univ, USA
37. Wim Hugo, SAEON, South Africa

Notes:

A number of folks new to the IG, but most in the room have heard of the CoreTrustSeal

1st half - discussing case studies of CoreTrustSeal certification for repositories from a few disciplines (J. Westbrook, G. Baker, D. Jaquette)

2nd half - discussion of the TRUST principles. The White Paper link: **https://bit.ly/2Ih7g8F**

John Westbrook, PDB (Protein Data Bank)

Founded 1971 with 7 X-ray structures. Today has >150K structures

210 New drugs approved 2010–2016 with a target in PDB. > $100 b of funding of NIH contributed to those approvals to those targets (95%)

Several incentives for CTS certification process for PDB

- The increasing focus of funders on supporting FAIR data mgmt practices
- The cert documents the resource investment required for good mgmt of the full life cycle of data
- Relatively low barrier to gain certification, but there's a good balance between rigor and cert effort

Certification benefited funding renewal application

Challenges in obtaining certification

- Supporting CTS requires diverse expertise in data sciences
- Long time horizon of CTS objectives challenging to support with 3-5 year competitive funding cycles
- Resource burdens for robust CTS support not fully appreciated by funding agencies or grant reviewers (infrastructure not given the limelight in proposals, new features are)

Garry Baker, National Geoscience Data Centre

NGDC has a mission to preserve data/physical specimens from geologic research, and provide for its reuse – 900 TB + 500 TB on tape archive.

Drivers to get CoreTrustSeal certification

- Demonstrate their professionalism
- Consider the full lifecycle of the repository
- Benchmark their processes and services
- Provide confidence to their funders (i.e., NERC)
- Alignment to scientific publishers who are looking for this sort of trusted status

Benefits in obtaining CoreTrustSeal certification

- Getting repository team together was invaluable
- got implicit knowledge “out of their heads”
- Identified many areas they could improve
- provided a vehicle to explain the business need of repositories to managers and funders
- encouraged them to chance expectations for data donors
- to look more seriously at future stakeholder engagement activities

Donators, funders and journal publishers like this certification - ex. NERC considering specifying CoreTrustSeal for their Data Service as part of future commissioning processes

Daniel Jaquette, Linguistic Data Consortium

Hosted at U. Penn, 80 holdings in more than 90 languages

Motivation - Provides an international, recognized standard, and to do a sanity check of catalog’s quality - started work under the Data Seal of Approval, received CoreTrustSeal certification in the end

During the internal assessment process, some of our internal assumptions were off (learned who was doing what internally with respect to data management

Investigation into broader persistent identifiers

Discussion

Q: As to who your “organization” is when you’re assessing your Repository - how can we accommodate into these certification responses? There are several different organizational arrangements (autonomous Repository, semi-autonomous Repository, some repository actions handled by different parties)

Q: J. Westbrook – how do we define ‘organizations”? (CoreTrustSeal defines it as “Repositories”, not organizations)

A: M. Mokrane – Extended guidance exists, this and the initial CoreTrustSeal guidance is meant to be broad enough to encompass a variety of organization structures

A: K. Ashley – Audit processes deal with this distributed responsibilities within system administration; Doing this and failing an audit can sometimes unlock resources to address issues

Q: About the evolution of the certification of Requirements, are they being strengthened, etc.

A: R. Edmunds – A review of the CoreTrustSeal Requirements and processes is ongoing until the end of April, strong encouragement for the community to provide feedback

Dawei Lin – TRUST principles

Don't just consider making data FAIR, need to have homes for FAIR data (i.e. repositories)

FAIR defines data and metadata properties
TRUST describes the characteristics of data repositories that are responsible for managing and disseminating data over a long period of time
Want FAIR data in repositories we TRUST
TRUST – Transparency, Responsibility, User Community, Sustainability, Technology

Advantages of TRUST principles – A framework for trustworthiness, an umbrella for standards, an aspiration for operations, an aid for understanding

TRUST white paper: **https://bit.ly/2Ih7g8F**

FACT is the **acronym** for **FAIR**, Accurate, Confidential and Transparent … watch out for overlap!

Question about turning TRUST principles into actions (how to take principles into practice) – D. Lin says they welcome all comments, including this and other gap analysis

How do these TRUST principles overlap with the CoreTrustSeal review?
I. Dillo says the Board sees the two are separate and complementary – CoreTrustSeal may be a starting point but the TRUST principles are larger than this

Question about tiers of curation provided by a repository – how might we evaluate such TRUSTfulness?

Could be fine to have different levels of FAIRness within a repository...this would be more about the repository being Transparent about the levels of curation it applies to different datasets

Can also have different levels of TRUST

Maybe make a one-page memo about these TRUST principles after Helsinki, when comments are addressed