RDA-P13-Proposal

Gail Clement, Interest Group Co-Chair

1/11/2019

Group Name:

RDA/CODATA Legal Interoperability Interest Group

Meeting title:

Legal Interoperability and Intraoperability of Research Data: The Case of the Research Compendium

Please list the meeting objectives:

The proposed session continues the Interest Group's efforts to apply the Principles & Guidelines for Legal Interoperability of Research Data (released 2016; Legal Interoperability Interest Group' (2016)) to significant use cases arising from the research community. This process has enabled the group to assess fulfill three key objectives:

1. to test and affirm the applicability of the Guidelines document to real-world data sharing efforts arising in the research data community;

2. to identify possible areas for extending the Guidelines document to accommodate needs not presented in the foundational case studies;

3. to train and support research data stakeholders in their adoption and implementation of the IG's Legal Interoperability Principles and Guidelines (Legal Interoperability Interest Group' 2016).

In past plenaries, our sessions have focused on real world use cases involving datasets and databases intended for open sharing. For Philadelphia, we will focus on a different, although complementary, data object – the research compendium – which represents a multipart research output comprising not only data but also the textual narrative that presents the research findings; code and algorithms; documentation and any auxiliary resources allowing readers/re-users rerun the analysis and verify or build on the claims. (Gentleman and Lang 2007).

Research compendia are an emerging form of data-rich research output gaining increasing adoption as a replacement for the 'static' and incomplete genre of the published paper typically represented in a PDF or HMTL file with, at best, hyperlinks to disparate files and datasets that require a reader to retrieve, compile, and execute in their own environment, often without success because of hidden dependencies or unreliable linkages. Examples of today's research compendia include Jupyter notebooks containerized in Docker images; Rmarkdown or Latex computational narratives (integrated text and code) with GitHub repositories openly distributed as Binder emulation; or dynamic web publications integrated with interactive data dashboards using Shiny or other web technologies (Nüst 2018).

Early adopters of research compendium include large scale data producers such as:

- the LIGO Gravitational Wave Open Science Center (https://www.gw-openscience.org/about/);

- the NSF-sponsored Whole Tale initiative at NCSA (https://wholetale.org/);

- and the variously-published reproducible reports by research authors in the Open Science community eg Marwick, Boettiger, and Mullen (2018), Boettiger (2018).

This use case is significant for the IG's focused consideration at Plenary because these multi-part research outputs – each with its own distinct rights management needs – implicates not only concerns with legal interoperability of the compendium as a whole, but also legal INTRA-operability of its constituent parts.

Message boards, tweets, GitHub issues, and other communications within the Open Science Community reflect considerable confusion and misinformation regarding the legitimate rights management and licensing of these increasingly important, multi-part outputs. We anticipate that the proposed Philadelphia session exploring the Research Compendium use case will yield initial content worthy of redistribution in the form of an extension to the Guidelines in a next release, or a separate paper explaining how the IG's Guidelines apply to Research Compendia.

Meeting agenda:

- Brief Review of the IG's Legal interop Principles and Guidelines, as released in 2016 (15 minutes)
- Introduction to the Research Compendium as an emergent FAIR research output (30 minutes)
- Application of the Guidelines to a Research Compendia (30 minutes)
- Discuss, Q&A, and next steps (15 minutes)

Group chair serving as contact person:

Gail Clement

Contact person's email:

gperetsm@caltech.edu

Please give a brief introduction describing the activities and scope of the group:

The Legal Interoperability Interest Group Develop has researched, compiled, and published core principles and guidelines of best practices through which legal interoperability can be achieved, and link to related information resources online. We continue to work with key stakeholder groups to get the core principles and guidelines of best practices adopted. Additionally, we promote better understanding and greater use by the stakeholder groups in the research community of the agreed approaches to legal interoperability of research data, focused on highlighting and enabling better integration and reuse of such data. Mechanisms for promotion and implementation support include presentations at professional conferences; publication in the peer reviewed literature; and training of researchers through the CODATA-RDA Summer Schools, the Force11 Scholarly Communication Institute, and through local copyright education programs at research institutions worldwide.

Additional links to informative material:

- [IG Home page and background materials] (https://www.rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html)
- Released output via RDA and CODATA (Legal Interoperability Interest Group' 2016)

Short Group Status:

This group is active

Please specify who is your target audience and how they should prepare for the meeting:

Research data creators, providers, and publishers; Copyright specialists and Counsel at Research Institutions; Open Source software developers and platgorm providers of tools for reproducible reporting.

Participants are encouraged to read the papers referenced in this proposal to prepare for the presentation and discussion.

Estimate of the required room capacity:

100

Type of meeting:

Informative meeting

If available, would your group be interested in being assigned an early career / student to support meeting note taking?

Yes

Would you like to request support for remote participation in your group meeting?

Yes

Avoid conflict with the following groups

1. Data policy standardisation and implementation IG Avoid conflict with the following group
2. RDA/WDS Scholarly Link Exchange (Scholix) WG Avoid conflict with the following group
3. RDA/WDS Publishing Data IG Privacy Policy

References

Boettiger, Carl. 2018. "From Noise to Knowledge: How Randomness Generates Novel Phenomena and Reveals Information." *Ecology Letters* 21 (8): 1255–67. https://doi.org/10.1111/ele.13085.

Gentleman, Robert, and Duncan Temple Lang. 2007. "Statistical Analyses and Reproducible Research." *Journal of Computational and Graphical Statistics* 16 (1): 1–23. https://doi.org/10.1198/106186007X178663.

Legal Interoperability Interest Group'. 2016. "Legal Interoperability of Research Data: Principles and Implementation Guidelines." https://doi.org/10.5281/zenodo.162241.

Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. "Packaging Data Analytical Work Reproducibly Using R (and Friends)." *The American Statistician* 72 (1): 80–88. https://doi.org/10.1080/00031305.2017.1375986.

Nüst, Daniel. 2018. "Reproducibility Service for Executable Research Compendia: Technical Specifications and Reference Implementation." Zenodo. https://doi.org/10.5281/ZENODO.2203843.