

# 第8回RDA総会報告

in Denver, CO, US

込山悠介 (Yusuke Komiyama)

国立情報学研究所 (National Institute of Informatics)

[komiyama@nii.ac.jp](mailto:komiyama@nii.ac.jp)

# 自己紹介

- 辻山悠介 (こみやま ゆうすけ) / Yusuke Komiyama
  - 国立情報学研究所 コンテンツ科学研究系 助教
  - 博士(農学)
  - 専門分野
    - バイオインフォマティクス
      - データ解析、データベース、機械学習
    - 人工知能
      - セマンティックWeb、Linked Data、オントロジー、メタデータ
    - オープンサイエンス
      - 学術情報基盤(HPC、ネットワーク、仮想マシン、クラウド、認証)
      - 研究データ管理基盤(レポジトリ、Webサービス、データベース)
  - 連絡先
    - [komiyama@nii.ac.jp](mailto:komiyama@nii.ac.jp)
    - <https://orcid.org/0000-0001-6468-3718>

# 第8回RDA総会

(Research Data Alliance 8<sup>th</sup> plenary)

- 開催地 : Denver, US
- 期間 : 15-17<sup>th</sup>, September 2016
  - International Data Week (IDW)2016 の一部として開催
- セッション数 : 70
- 私の視点
  - RDA入門。(Newcomer ガイダンスから出席)
  - 研究データ管理のためのレポジトリ技術、メタデータの互換性、仮想研究環境の情報収集。



# Research Data Alliance: Philosophy, Structure, and Governance

---

Mark A. Parsons  
0000-0002-7723-0950  
Secretary General

RDA for Newcomers  
Denver, USA  
15 September 2016



## RDA: State of the Art WG, IG & BoF

RDA Plenary 8, 2016  
Denver

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)



RESEARCH DATA ALLIANCE

## RDA for Newcomers

15 September 2016  
Denver, CO

## Website Tutorial

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

Presentation adapted from Yolan  
Meleco (2015).

# 参加したセッション

BoF: birds of a feather  
IG: interest group  
WG: working group

- DAY 1
  - (1) IG New Paradigms for Data Discovery: Kickoff
  - (2) IG Brokering: Matching Interoperability Requests and Offers
  - (3) IG Repository Platforms for Research Data
- DAY2
  - (4) IG Long tail of research data: 10 ways to support data diversity and the long tail of research data
  - (5) WG Research Data Repository Interoperability
  - (6) IG RDA/CODATA Legal Interoperability: Final Principles and Implementation Guidelines for Legal Interoperability of Research Data
- DAY3
  - (7) IG Virtual Research Environment: VREs/Virtual laboratories/science gateways: opportunities for developing a more coordinated approach to support interoperability across different systems
  - (8) Joint meeting of IG Active Data Management Plans, IG Preservation e-Infrastructure, IG Reproducibility: Tools Convergence: Integrating Data Management Plan and Preservation Tools

BoF: birds of a feather (ミートアップ、同好会)

IG: interest group (研究会、検討会)

WG: working group (作業部会、委員会)

# 参加したセッション

## • 1日目

- (1) IG データ発見のための新たなパラダイムのキックオフ
- (2) IG 互換性の要求と提供をマッチングさせる仲介交渉
- (3) IG 研究データのためのレポジトリプラットフォーム

## • 2日目

- (4) IG 研究データのロングテール: データ多様性と研究データのロングテイルを支援する10(+ $\alpha$ )事例
- (5) WG 研究データレポジトリの互換性
- (6) IG RDA/CODATAの法的互換性: 研究データの法的互換性のための最終的な原則と実装の指針

## • 3日目

- (7) IG 仮想研究環境: 仮想研究環境/仮想ラボ/サイエンスゲートウェイ: 非同一システムの横断的な互換性支援に向けた、更に調整した手法を開発するための機会。
- (8) “IG 能動的データ管理計画”, “IG e-Infrastructure維持”, および “IG 再現性: ツールの集中: データ管理計画と保存ツールの統合” の合同ミーティング。



# Data Discovery Interest Group Kickoff Session *September 15, 2016* *RDA 8th Plenary meeting, Denver CO*

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

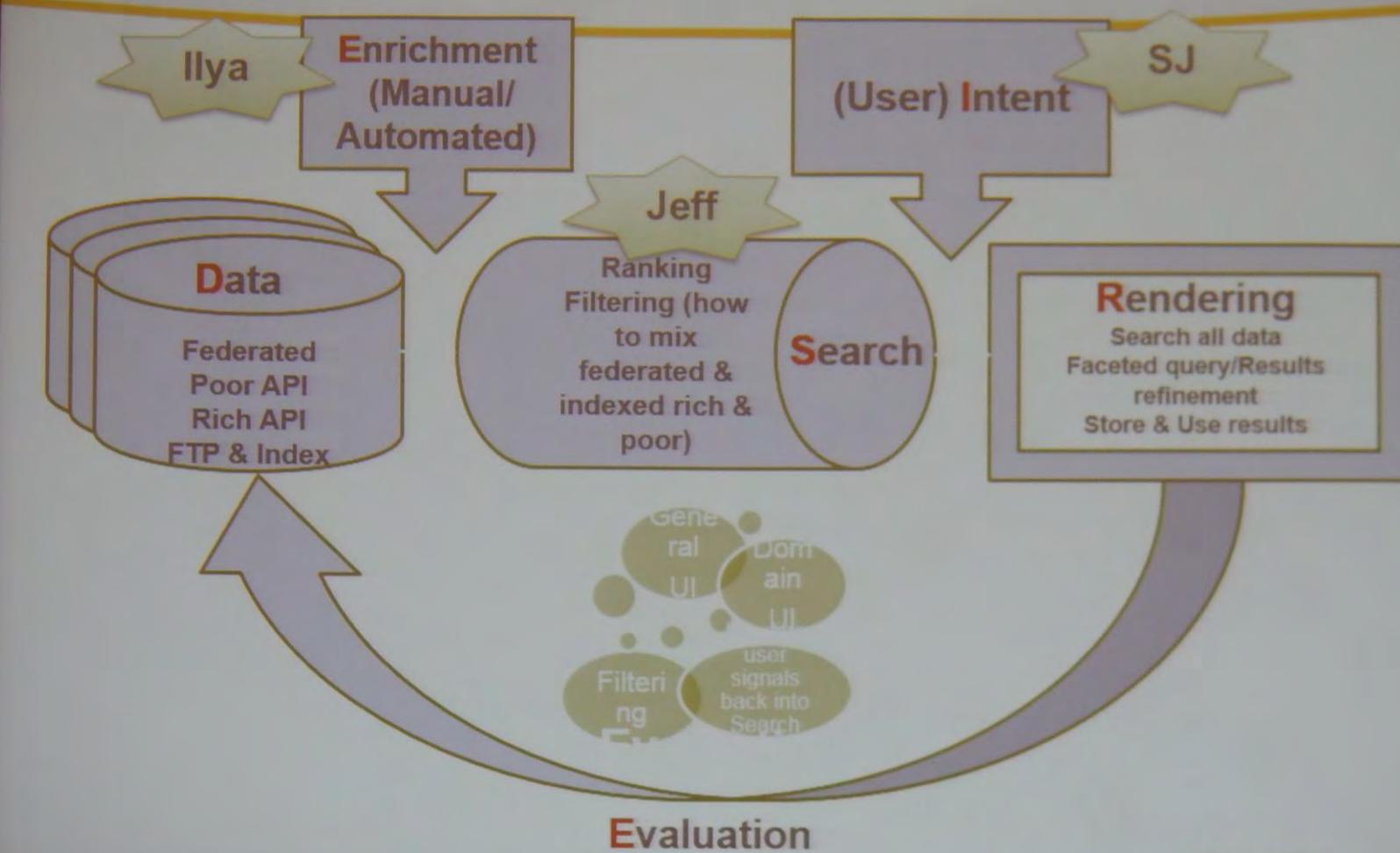
Anita de Waard,  
Siri Jodha Singh Khalsa

## • 概要

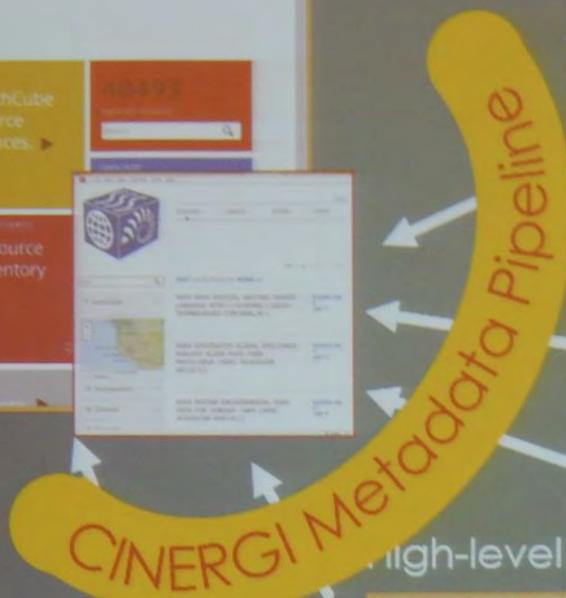
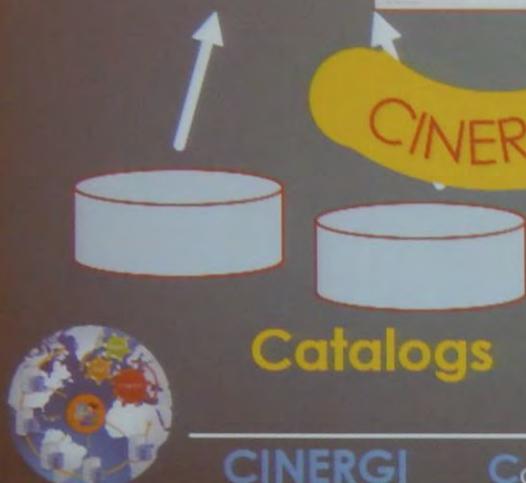
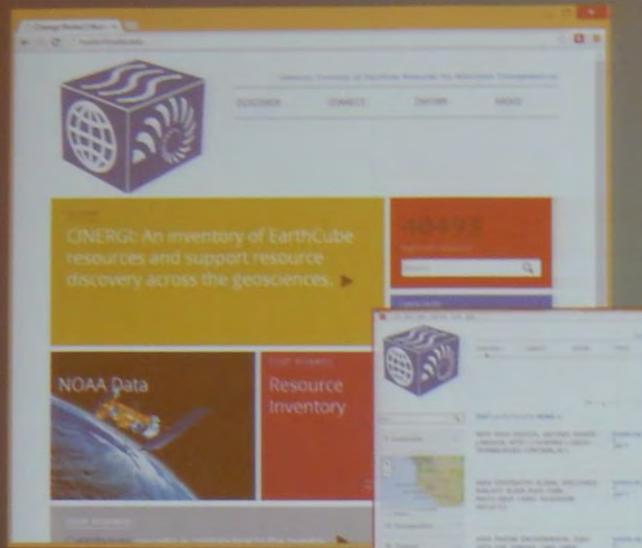
<https://rd-alliance.org/groups/data-discovery-paradigms-ig>

- 地質学者 M. Lee Allison博士の追悼。
- データディスカバリーの望まれるモデル
  - Data, Enrichment, Search, Intent, Rendering (DESIRE)の解説。
  - データ、質的向上、検索、意図、表現。
- 3件の話題提供
  - Community Inventory of Earth Cube Resource for Geoscience Interoperability (CINERGI),
  - DataMed,
  - User Profile を使った Search Personalizationの発表。
- 質疑応答では
  - データ発見のための公共API
  - データ発見のためのベンチマー킹検索結果
  - 他のIG/WGとの重複と交流
  - コンテンツの質的向上と側面的な検索の方法
  - 妥当なランキングのためのアルゴリズム
  - Search Personalizationにおけるユーザープロファイルのための文書とガイドライン(の策定)
  - などに付いて会場を交えたパネルディスカッションが行われ、
  - Possible focus topicsとして23件にまとめられた。

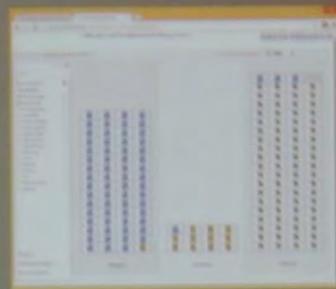
# The DESIRE Model of Data Discovery



# Metadata aggregation in CINERGI



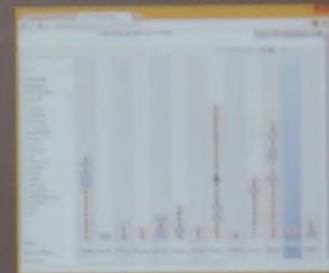
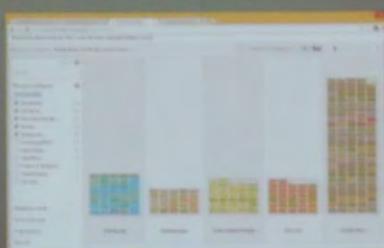
## Domain Inventories



RCN  
(Research Coordination Networks)



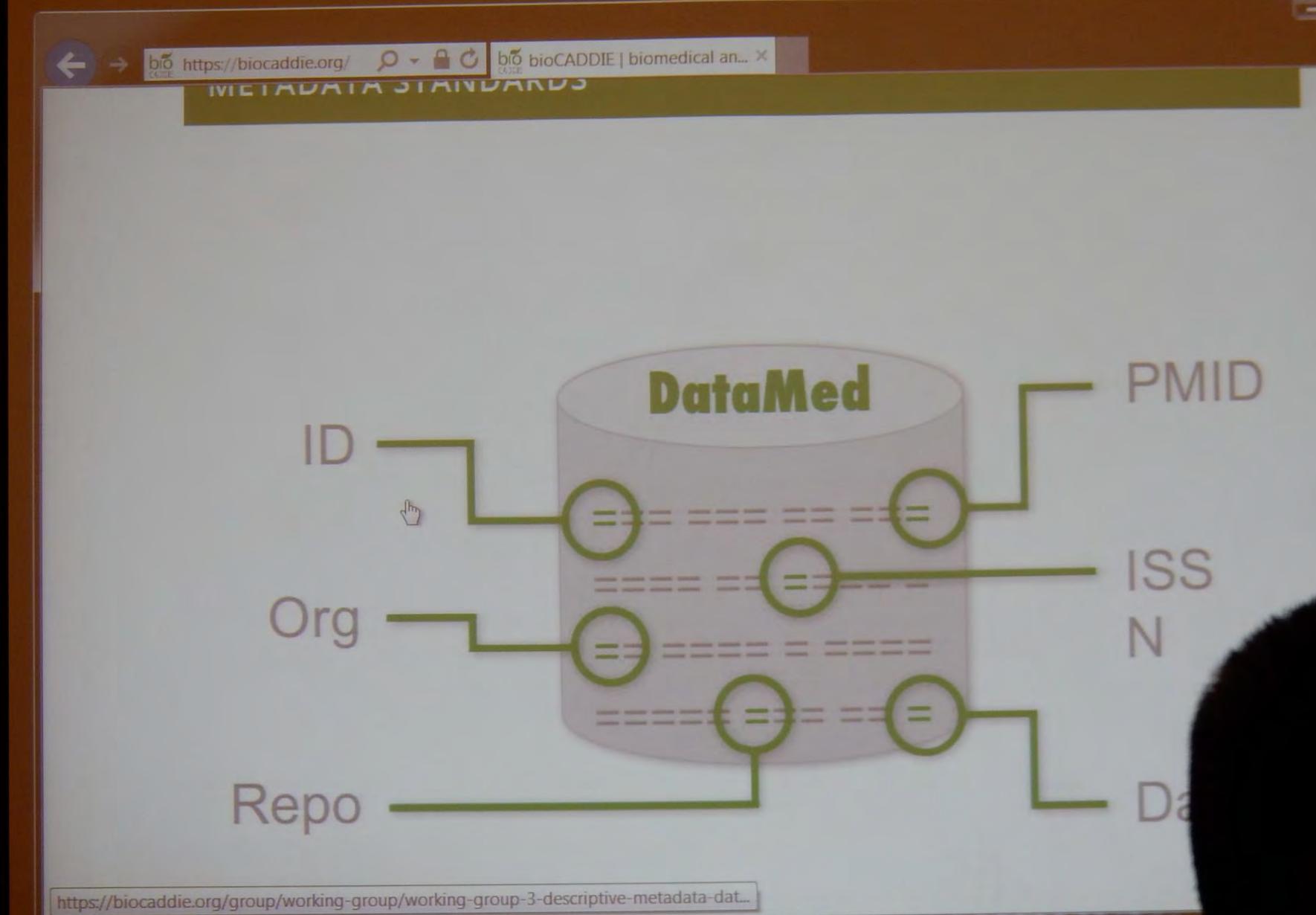
## Domain workshops



CINERGI

Community Inventory of Earthcube Resources for Geoscience Interoperability

## 個別事例2: DataMed (NIH) におけるメタデータ集約と統合



# Search Personalization

- Keyword-based search is imperfect
  - Ambiguity in meaning
  - Varies with context
  - What meaning does the user assign to the term?
- Personalization in information retrieval
  - Long history ("ask a Librarian")
  - Area of intense research since advent of search engines
  - Commercial interest in tailoring information presented to individuals

ORCIDや研究者ID等から明示的な情報と、  
Webサーバー・クライアントのブックマークやログからユーザー行動分析した潜在的情報で、  
Search Personalizationを実現するために、ユーザープロファイルを作りましょう。



## 本セッションでされた 総合討論(フロア)の議論内容

- データ発見のための公共API
- データ発見のためのベンチマー킹検索結果
- 他のIG/WGとの重複と交流
- コンテンツの質的向上と側面的な検索の方法
- 妥当なランキングのためのアルゴリズム
- Search Personalizationにおけるユーザープロファイルのための文書とガイドライン(の策定)

## Possible focus topics

- 1. Deduplication and cross-repository issues
- 2. Identifiers and how they help in search
- 3. Data citation: how do we access/use?
- 4. Relevancy ranking for structured data?
- 5. Enrichment tools for faceting and ranking
- 6. Domain-specific vs. generic issues: interfaces and enrichment
- 7. Different discovery platforms for Open Search, science-focused OS profile?
- 8. Metadata standards to enhance data discovery, e.g. schema.org and such
- 9. Models and methods of personalisation
- 10. Identify core elements of Findability
- 11. Automated integration of records; granularity and findability
- 12. Common APIs
- 13. Upper-level ontologies for search
- 14. Creating test collections for search evaluation and methods of evaluation
- 15. Collections and granules: build tool that enables guidance for data submitters on how data is organized
- 16. Guidelines for making your data findable! Best practices based on experiences.
- 17. Identify collections of use cases for users: e.g. browsing vs search
- 18. Measures of data quality: and impact of findability
- 19. Define series of reference datasets – can be used to do these metrics
- 20. Identify list of prototyping tools, use by WG!
- 21. Cross over between domains: how to enable cross-walk between domains
- 22. “Return to the semantic”: schema has been popularized by crowdsourcing rather than 1 researcher.
- 23. Implementing schema.org as it exists! How does it apply to science?



8<sup>th</sup> RDA Plenary  
Denver, Sep 2016

# Brokering IG session: matching Interoperability requests and offers

W. HUGO, S. NATIVI, JAY PEARLMAN, S.J. KHALSA



## • 概要

<https://rd-alliance.org/groups/brokering-ig.html>

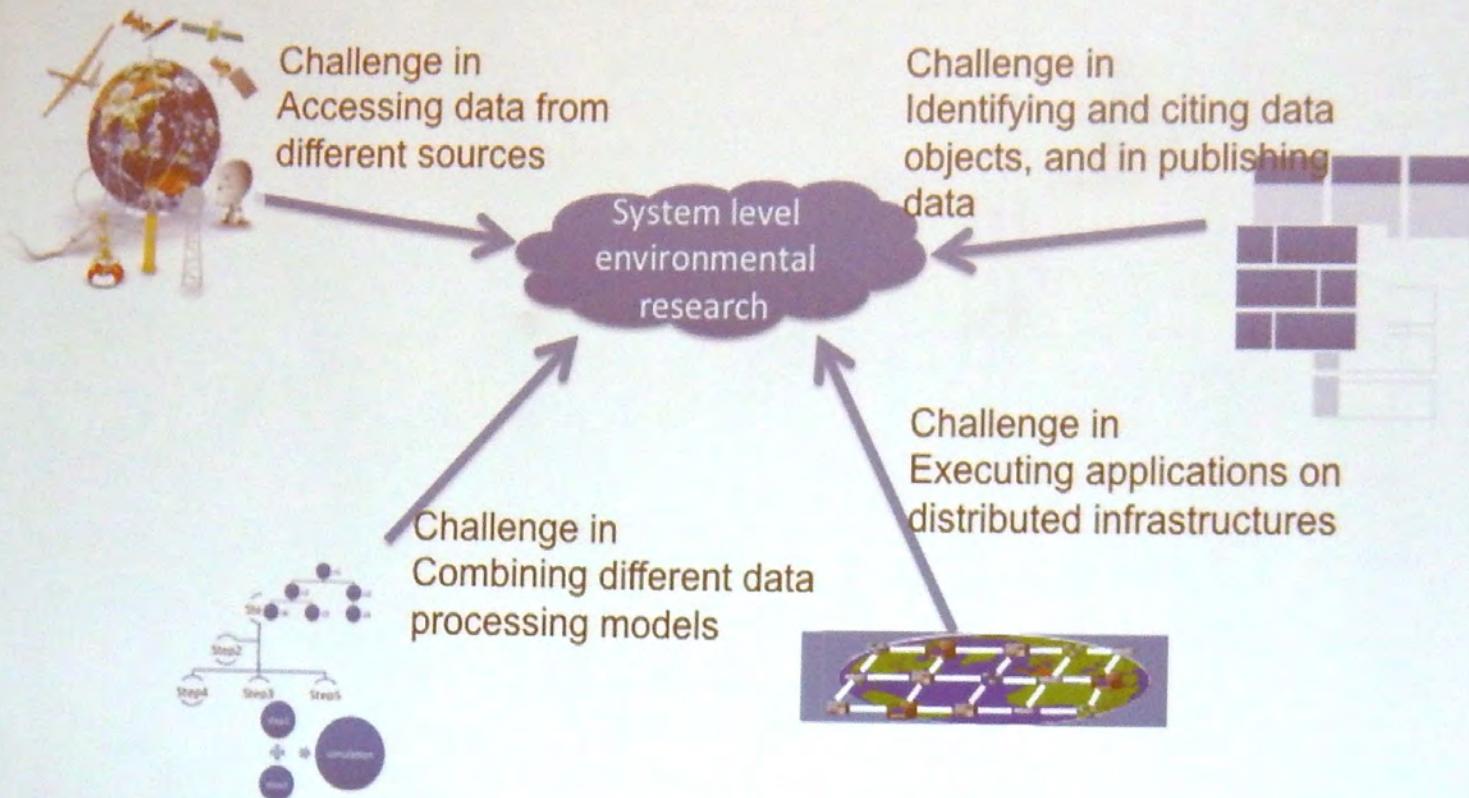
- 会場の参加者全員の簡単な自己紹介。
- Brokering Frameworkと可能性。
- 他のRDAグループとID関連の議論。
- New WG case statement,
- データコミュニティによるブローカリングアプローチの採用のための認識されたチャレンジに関する議論。
- 結果のまとめ。

# Broker Traits

The Broker that (effectively) works in a heterogeneous environment containing multiple remote objects that interact synchronously or asynchronously typically demonstrates the ability to:

1. **Finalize requests on behalf of its clients** against a vast supply system –e.g. by transforming different interoperability protocols;
2. Support **many clients at the same time** in a dynamic way;
3. Access **large, distributed, and heterogeneous supply systems** in a dynamic way;
4. Be **fully autonomous** from its clients and accessed supply systems;
5. Be **flexible and configurable** (even at run-time);
6. Be **extensible**.

# Rationale



Credit: Data Fabric IG  
Envri+ use Case

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)





RESEARCH DATA ALLIANCE

# The RESEARCH DATA ALLIANCE

## WG: Brokering Framework

Wim Hugo – ICSU-WDS/ SAEON

Michael Diepenbroek, Pangaea

Stefano Nativi, CNR

Jay Pearlman, IEEE

research data sharing without barriers

[rd-alliance.org](http://rd-alliance.org)

## Vision and Mission

標準化方法における出版・発見とブローカリングの開始とコンポーネント調停を容認する  
ブローカリングフレームワークでの記述・開発・試行および実装

### ■ Vision

- “To describe, develop, test, and implement a Brokering Framework that allows publication, discovery, and invocation of brokering and mediation components in a standardised manner.”

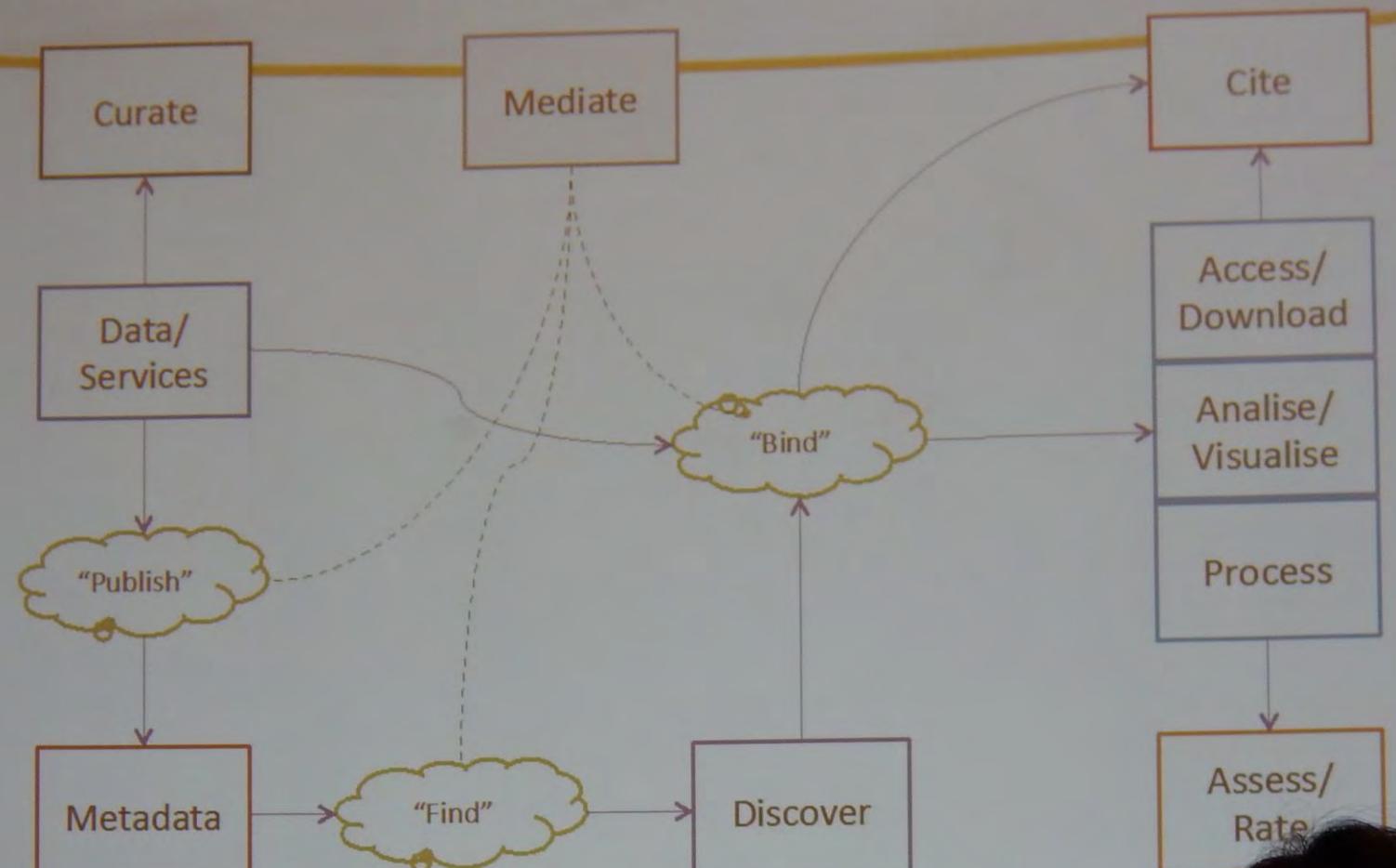
### ■ Mission

- Community consensus on the way in which brokering and mediation components are described, discovered, and invoked, based on real use cases – “Brokering Framework” – leading to a formal service and content standard. *The formalisation of the standard is an external process and may involve participants in the working group, but is not a task for the working group;*
- Demonstrating the value of the framework by developing, testing, and commissioning a reference implementation of a brokering and mediation registry.

## Deliverables and Milestones

| Task or Deliverable  | RDA Plenary | Expected Timelines |
|--|-------------|--------------------|
| Use cases and classification of brokering and mediation components | 8           | 3 months           |
| Develop content and service standards                              | 9           | 9 months           |
| Create a registry of mediation options                             | 9           | 9 months           |
| Test environment - initial/third party implementation              | 10          | 12 months          |
| Develop protocols for the test environment                         | 10          | 12 months          |
| Contingency and slippage provision for the above: final delivery   | 10          | 15 months          |
| Governance of registry and test environment                        | 10          | 15 months          |
| Recommendations and report   | 11          |                    |

## Value Proposition: Common Implementation Architecture



“Predictable Assembly from Reliable Components”



## Repository Platforms for Research Data

Co-Chairs: Stefan Kramer, Ralph Müller Pfefferkorn,  
David Wilcox

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

## • 概要

- 会場に新規参加者が多かったため、研究データのためのレポジトリプラットフォームIG一般の紹介。
- レポジトリプラットフォーム必要要件のコンテキストにおける研究データユースケースの集約と分析を実施しようとしているグループ。
- 事例として
  - Dissertation Data,
  - National Climate Repository Platform,
  - IRODS Data Grids,
  - Nanoscropy Open Research Data Repository,
  - Vector-bom Disease Network,
  - OpARA - Institutional Repository,
  - CurateND,
  - CUAC,
  - TAILwang,
  - SURF sara TDR,
  - Institutional Life Cycle RDM 等が報告された。
- IGの作業テーブルについての短い文書 "The development of the matrix of use cases and functional requirements for research data repository platforms" の紹介、
- Appendixとして各ユースケースがまとめられている。<http://doi.org/10.17606/M67P43>。
- 2日目に関連したResearch Data Repository Interoperability WGが開催されることの告知。
- グループとして次のステップについて議論。

## The development of the matrix of use cases and functional requirements for research data repository platforms

Last updated: 2016-09-26 by Stefan Kramer

### Introduction

The [Repository Platforms for Research Data Interest Group](#) (hereafter: RPRD IG) of the Research Data Alliance (RDA) was formally launched at the [5th RDA Plenary Meeting](#) in San Diego, California, in March, 2015. Originally proposed as a Working Group (WG), but changed to an Interest Group following feedback from the RDA Technical Advisory Board, the RPRD IG had the charter, within the 18-month timeframe expected of an RDA WG, to "gather and analyze research data use cases in the context of repository platform requirements," with the "primary deliverable [being] a matrix relating use cases with functional requirements for repository platforms" for research data. The primary target audience for the matrix was envisioned to be developers and service providers of repository software, with secondary audiences including repository implementers and managers, funding agencies, grant applicants, and researchers generating and using scientific data.

### Methodology

Members of the RPRD IG identified other RDA groups with expected outcomes or activities, as well as documents, relevant to its own work; these can be found in its wiki under [RPRD Related Work](#). RPRD IG members developed a Microsoft Word template for the collection of use cases, and identified communities of practice - and their online groups or mailing lists - from which to solicit use case submissions. The template, and the resulting collected use cases, can be found under [Collection of use cases](#); a listing of which communities were contacted when and by whom, under [RPRD Use Case Dissemination](#); and an example of a use case solicitation message in Appendix 1 of this document.

Specific RPRD IG members then volunteered to transfer the stated functional requirements and their respective importances<sup>1</sup> from the use case documents to a matrix in the form of a Google Sheet. That format/platform had been chosen because it lends itself well to convenient group collaboration and was familiar to IG members. The RPRD IG co-chairs then de-duplicated and combined the collected functional requirements, since many of them overlapped across use cases, and computed a "functional requirement score" for each, based on the importance values from the use cases, but inverted so that a *higher* value indicates *more* (combined) importance. At the top of the "Functional requirement score" column are summary statistics for that measure.

---

<sup>1</sup> which were expressed on a scale from 1 = very important to 5 = not at all important

ホーム 挿入 印刷レイアウト 数式 データ 校閲 表示

標準 ページ レイアウト ユーザー 設定のビュー ✓ 種類 ✓ 見出し ズーム (100%) ウィンドウ 先頭行の固定 先頭列の固定 分割 マクロの表示 マクロの記録

A1

Median: 8 / 1st quartile: 4 / 3rd quartile: 12 / Mean: 8.704515454545456

**Use Case**

| Category                         | Requirement  | Description  | Functional requirement score | Number of Mentions | Use Case     |         |              |        |        |       |           |      |          |          |                  |  |
|----------------------------------|--|--|------------------------------|--------------------|--------------|---------|--------------|--------|--------|-------|-----------|------|----------|----------|------------------|--|
|                                  |  |  |                              |                    | Dissertation | ICRCCDS | DKRZ Climate | NORDIR | VeeNet | OpARA | CuratedID | CIAC | TAIL wag | SURFteam | Inst. Life-cycle |  |
| Metadata                         | Support for different metadata (schemas), including domain-specificity and interoperability  | Any form of metadata (schemas) should be allowed. This includes metadata such as author, owner, license, source publication, library, date and time stamps, and domain specific metadata.  | 28                           | 7                  | 1            | 1       |              | 1      | 1      |       | 1         | 1    | 1        |          |                  |  |
| Persistent identifiers           | Assignment of PID / DOI  | All data ingest must "provide publication" or even before (e.g. when a paper was submitted but the data are not final yet) a PID has to be assigned to data and the collections it belongs to. It resolves to the "landing page" concept.  | 26                           | 7                  | 2            |         | 1            | 2      |        | 1     |           | 1    | 1        | 1        |                  |  |
| Authentication                   | Fine grained authentication and authorization. Allow the integration or the import from external authentication/authorization systems.                   | Provide different access rights for groups and individuals (roles, permissions) and allow the import of such objects (e.g. from identity management systems). In the case of confidential or proprietary data control over access to data is in the hands of those who provide the data  | 16                           | 4                  | 1            |         |              |        | 1      |       |           | 1    | 1        |          |                  |  |
| Data Access                      | Allow data providers to choose the level of access to data (e.g. Open Access).   | Control over access to data is in the hands of those who provide the data  | 16                           | 4                  |              |         |              | 1      |        |       | 1         | 1    | 1        |          |                  |  |
| Data Access                      | Provide state-of-the-art user interfaces and clients over the life time of a repository platform   | (Easy to use) User interface, features, and functionality should be updated over time to match the requirements and expectations of current researchers. They include web based user interfaces and others Data providers have to define what happens when to which dataset. E.g. for licensing and quality control regularly enforced policies are helpful. | 16                           | 4                  | 1            |         |              |        | 1      | 1     |           |      | 1        |          |                  |  |
| Policy Support                   | Allow the (automated) use of data policies   |  | 15                           | 4                  | 1            |         | 2            |        |        |       | 1         | 1    | 1        |          |                  |  |
| Publication                      | Provide data access statistics either by the use of external analytics services or internal monitoring of user activity.                                 | Provide the data publisher (or other stakeholders) with data access statistics to track and share usage and activity related to profiles, collections, datasets, and other metrics.  | 13                           | 4                  |              |         |              | 2      | 3      | 1     |           |      | 1        |          |                  |  |
| Submission / Ingest / Management | Provides interfaces (APIs) for the automated execution of tasks, e.g. to ingest data or to integrate data analysis tools and other external applications | API for automated execution of standard repository tasks and to interoperate with external tools useful to the stakeholders  | 13                           | 4                  | 1            |         |              | 3      | 2      |       | 1         |      |          |          |                  |  |
| Data Access                      | Provide (authorized) users access to versions of data (e.g. different simulation runs)   | Provide different versions of a data set   | 12                           | 3                  |              | 1       |              |        | 1      | 1     |           |      |          | 1        |                  |  |
| Data Access                      | Embargo date selection for data-depositing user  | The user can select a specific date (e.g. 2016-05-15) or time span (e.g. 1 year) for when the research data shall become available. The selectable default values can be set by the repository administrator to match the needs of the institution.  | 12                           | 3                  | 1            |         |              |        |        |       |           |      | 1        | 1        |                  |  |
| Data organization                | Collection virtualization / logical naming   | Managing data collections and their properties (like file names) independently of the storage system and storage resource naming.  | 12                           | 3                  | 1            |         |              |        |        |       |           |      | 1        | 1        |                  |  |
| Metadata                         | Allow data annotation by the data owner, by other (authorized) people or by automatic (metadata) extraction tools  | Additional annotation of data provides additional valuable insights of the data and aids in more comprehensive records over time.  | 12                           | 3                  |              |         | 1            | 1      |        |       |           |      | 1        |          |                  |  |
| Policy Support                   | Policy enforcement points  | Control all operations with administrator defined rules  | 12                           | 3                  | 1            |         |              |        |        |       |           |      | 1        | 1        |                  |  |
| Submission / Ingest / Management | Record audit trails  | Track changes to resource metadata and information relationships. Maintain a log of all events / operations. These logs should be easily readable, flow on and out of the repository and chart the connections that are.   | 12                           | 3                  | 1            |         |              | 1      |        |       |           |      | 1        |          |                  |  |
| Data Access                      | Sophisticated search capabilities for metadata and data both for humans and computers  | Allow search capabilities like full text search, schema specific search. Metadata and data should be indexable by search machines and indexable by external search engines. The search results should be presented in a meaningful way.  | 11                           | 3                  |              |         |              | 1      | 2      | 1     |           |      |          |          |                  |  |
| Submission / Ingest / Management | Easy to use ingest process with few barriers to participation  | Opening up many barriers to entry into the repository allows for the curation of unpublished materials and supports linking to published works in other places.  | 11                           | 3                  |              |         |              |        |        | 2     | 1         |      |          | 1        |                  |  |
| Location                         | Tight integration with (near) data processing.   | The large data volumes require data near processing (and data reduction) facilities like High Performance Computing. These should provide and maintain provenance information. Processing of large datasets  | 9                            | 3                  |              | 2       | 2            |        | 2      |       |           |      |          |          |                  |  |
| Authentication                   | Provide Single-Sign-On and/or support for different authentication methods   | Combined support for methods like Shibboleth or LDAP   | 8                            | 2                  | 1            |         |              |        | 1      |       |           |      |          |          |                  |  |
| Data Access                      | Allow local download of a selected set of information  | The ability to download content to a local device when the information conforms to proper handling and is marked as accessible   | 8                            | 2                  |              |         |              | 1      |        |       |           |      | 1        |          |                  |  |
| Integration                      | Federation   | Enable interoperability with other existing data management systems  | 8                            | 2                  | 1            |         |              |        |        |       |           |      | 1        |          |                  |  |
|                                  | Integration of Pico-ino data management  | All data management activities should be integrated  |                              |                    |              |         |              |        |        |       |           |      |          |          |                  |  |

コマンド

(4) IG 研究データのロングテール: データ多様性と  
研究データのロングテイルを支援する10(+α)事例  
(4) IG Long tail of research data: 10 ways to support  
data diversity and the long tail of research data



[http://cdn.zmescience.com/wp-content/uploads/2009/08/jellyfish\\_nom.jpg](http://cdn.zmescience.com/wp-content/uploads/2009/08/jellyfish_nom.jpg)

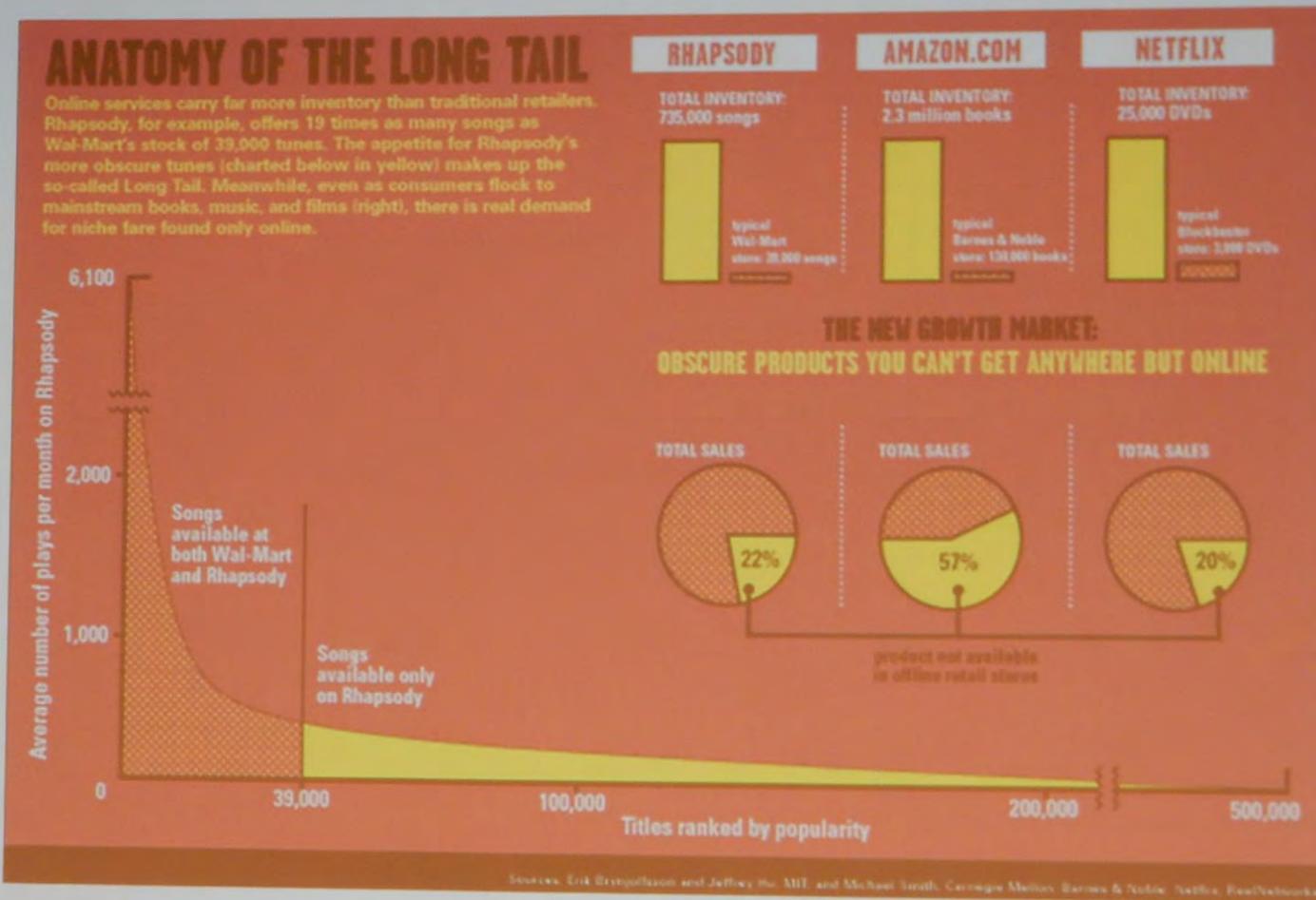
# • 概要

- - Motivation of the IG
- - Beyond Big Data
- - Long-Tail as in Economics
- - Long-Tail as in Research Data
- - Larger parts of research use small data
- - Big Data, Long-Tail Data
- - Institutional, domain or no repositories
- - Long-Tail of Research Data Interest Group
- - Long-Tail of Research Data Interest Group
- - Summary Findings of the IG
- - Objectives of the paper

(4) IG 研究データのロングテール: データ多様性と  
研究データのロングテイルを支援する10(+ $\alpha$ )事例  
(4) IG Long tail of research data: 10 ways to support  
data diversity and the long tail of research data

# “Long-Tail” as in Economics

Chris Anderson (Editor in Chief), Wired, Issue 12.10, October 2004



<http://www.wired.com/wired/images.html?issue=12.10&topic=tail&img=2>

# “Long-Tail” as in Research Data

P. Bryan Heidorn (LIS U Arizona) in Library Trends 57/2, Fall 2008

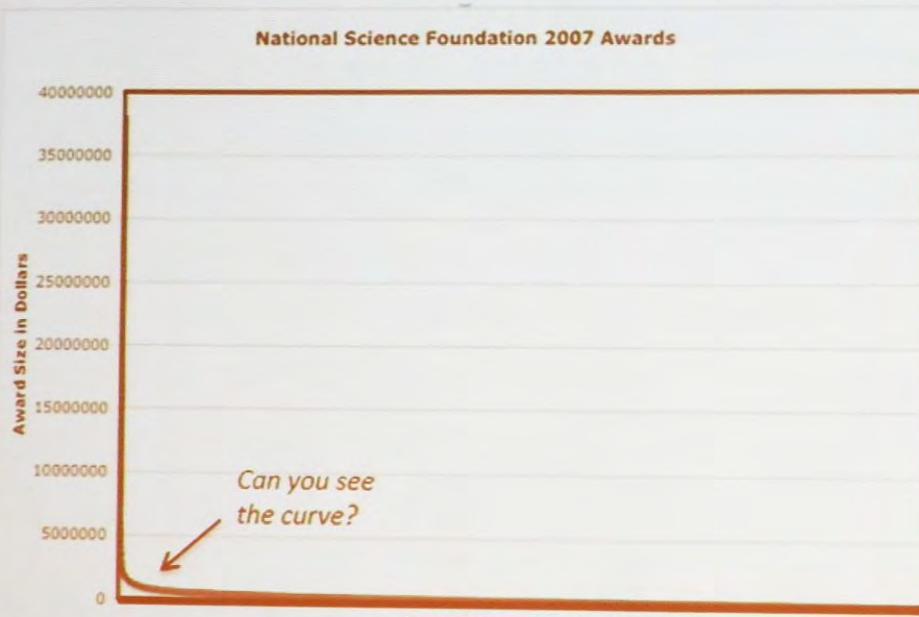
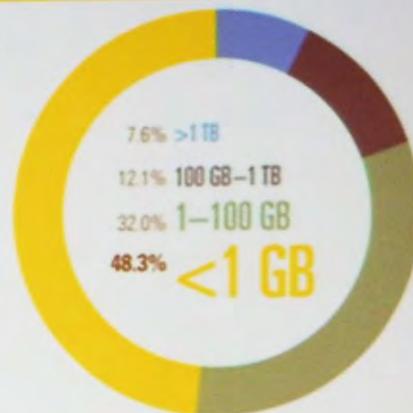
- ... While great care is frequently devoted to the collection, preservation and reuse of data on very large projects, relatively little attention is given to the data that is being generated by the majority of scientists.
- ... There may only be a few scientists worldwide that would want to see a particular boutique data set but there are many thousands of these data sets.
- ... The long tail is a breeding ground for new ideas and never before attempted science.
- ... The challenge for science policy is to develop institutions and practices such as institutional repositories, which make this data useful for society.

# Larger parts of research use small data

The 2011 survey by *Science*, found that 48.3% of respondents were working with datasets that were less than 1GB in size and over half of those polled store their data only in their laboratories.

*Science* 11 February 2011; Vol. 331 no. 6018 pp. 692-693 DOI: 10.1126/science.331.6018.692

What is the size of the largest data set that you have used or generated in your research?



Because there is only a tiny fraction of large projects and a loooooooooooooong tail of small projects

# Big Data, Long-Tail Data

| No. | Head                      | Tail                                      |
|-----|---------------------------|---|
| 1   | Homogeneous               | Heterogeneous                             |
| 2   | Large                     | Small                                     |
| 3   | Common standards          | Unique standards                          |
| 4   | Regulated                 | Not Regulated                             |
| 5   | Central curation          | Individual curation                       |
| 6   | Disciplinary repositories | Institutional, general or no repositories |

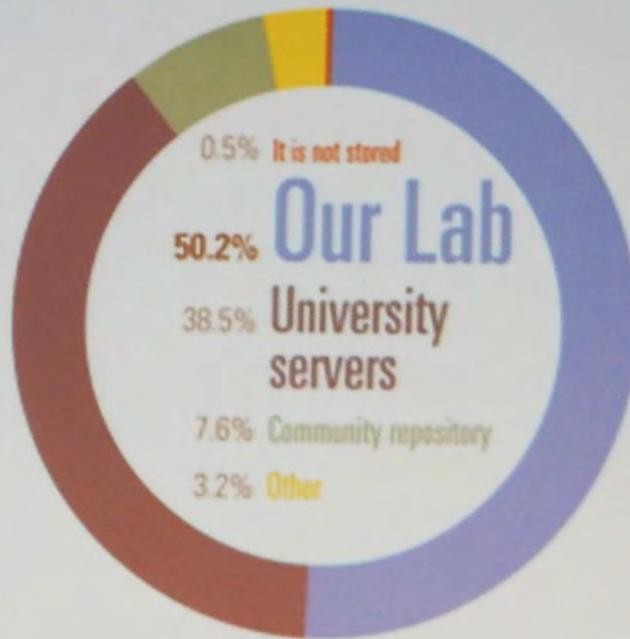
Adapted from: *Shedding Light on the Dark Data in the Long Tail of Science* by P. Bryan Heidorn. 2008

- “Disks in your drawer; server in lab basement”
- Long Tail Data exist across all disciplines

# Institutional, domain or no repositories

Where do you archive most of the data generated in your lab or for your research?

“ Even within a single institution there are no standards for storing data, so each lab, or often each fellow, uses ad hoc approaches. ”



Science 11 February 2011: Vol. 331 no. 6018 pp. 692-693 DOI: 10.1126/science.331.6018.692

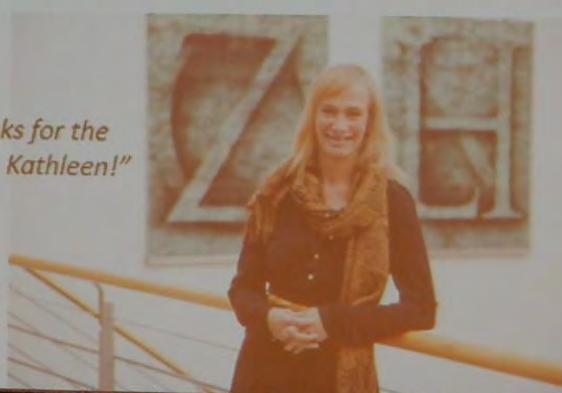
# Long Tail of Research Data Interest Group

- Accepted as an RDA Interest Group in Summer 2013
- 190 members from around the world

## Objectives

- To better understand the long tail
- To address challenges involved in managing diverse datasets
- To share and develop practices for managing diverse data
- To work towards greater interoperability across repositories

*"Thanks for the  
slides, Kathleen!"*



Kathleen Sheerer,  
COAR Executive  
Director and Co-  
Chair of the RDA IG

# Long Tail of Research Data Interest Group

## Activities-to-date

- Survey of discovery metadata
- Discussion of strategies for improving discoverability of datasets
- make it easier for researchers to deposit their data
- sharing practices about discovery
- interoperability across repositories (WG!)

- - Speakers Corner

- - DataOne
- - Portage
- - Figshare
- - OpenAIRE
- - eIRG
- - European Open Science Cloud
- - EarthCube
- - Mendele, Scholix
- - DataCite
- - Zenodo
- (実際には飛び入りも有り10名以上いた)

(4) IG 研究データのロングテール: データ多様性と  
研究データのロングテイルを支援する10(+ $\alpha$ )事例  
(4) IG Long tail of research data: 10 ways to support  
data diversity and the long tail of research data



# Confederation of Open Access Repositories – Repository Interoperability

Thomas Jejkal

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

# • 概要

- - Presentation of the Case Statement review results
- - Short talks on state of the art technologies
  - - Outcome of the COAR interoperability Project
  - - Refreshing the SWORD Protocol
  - - Import/Export Standards for Repository Resources
- - Discussion about suitability, gaps and other interoperability technologies
- - Summary of results and next steps



# Confederation of Open Access Repositories – Repository Interoperability

Thomas Jejkal

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

# Confederation of Open Access Repositories (COAR)

- - Repository Interoperability
- WG2: Repository Interoperability
  - - Started with COAR initiative in 2009
  - - Produced three documents until 2014
- The Case for Interoperability
  - - Focus on connecting Open Access repositories
  - - Distinguish between system level and semantic interoperability
  - - Identified technical and administrative challenges
  - - Goals: Analyze stakeholder input, environmental assessment, roadmap
- Current State of Interoperability
  - - Environmental assessment of 19 interoperability initiatives
- COAR Roadmap: Future Directions
- Conclusions & Current Work

# Confederation of Open Access Repositories (COAR)

- International association with > 100 members
- Build capacity, align policies and practices, act as global voice
- In RDA via Long tail of/Libraries for research data IG
- Four strategic directions:
  1. Promote development of global network of OA repositories
  2. Provide community support and build local capacity
  3. Define and promote interoperability, commons standards and best practices
  4. Stimulate development of value-added services

## WG 2: Repository Interoperability

- Started with COAR initiative in 2009
- Produced three documents until 2015
  - The Case for Interoperability for Open Access Repositories (2011)
  - Current State of Open Access Repository Interoperability (2012)
  - COAR Roadmap: Future Directions for Repository Interoperability (2015)

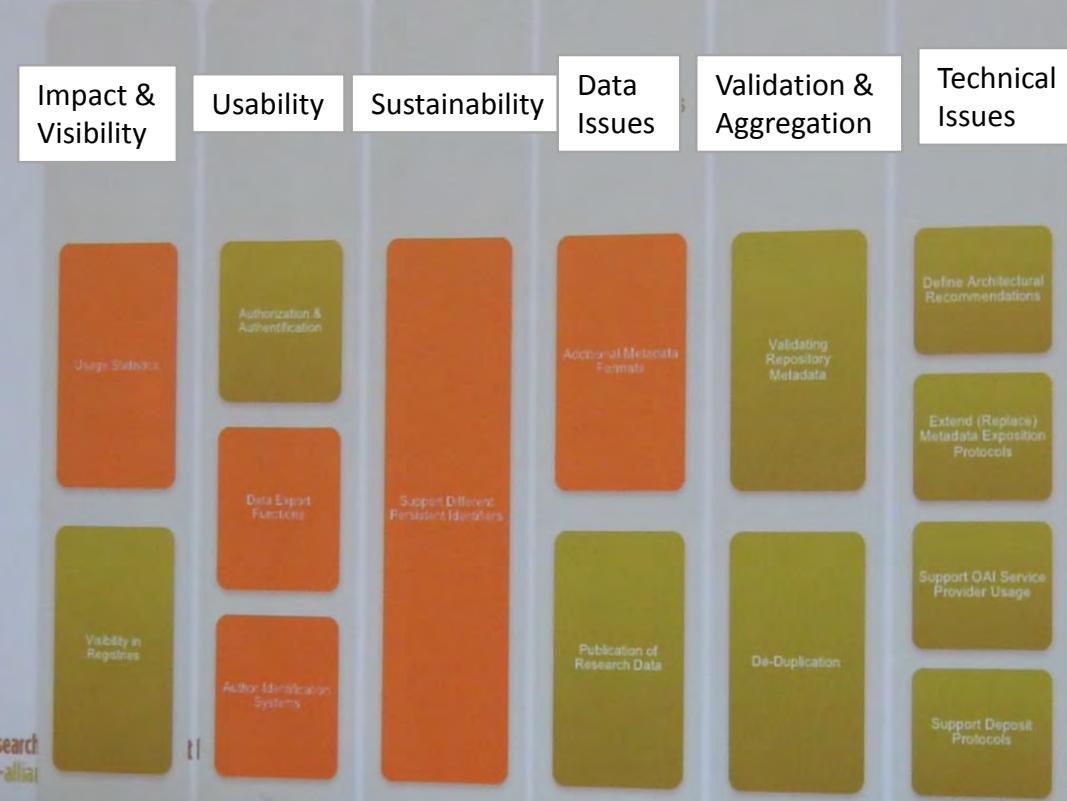
# Current State of Interoperability

- Environmental assessment of 19 interoperability initiatives



# COAR Roadmap: Future Directions

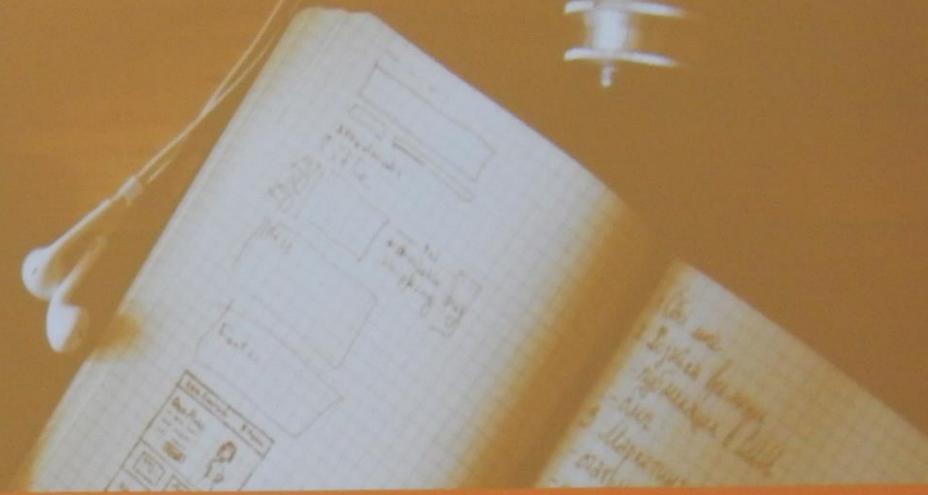
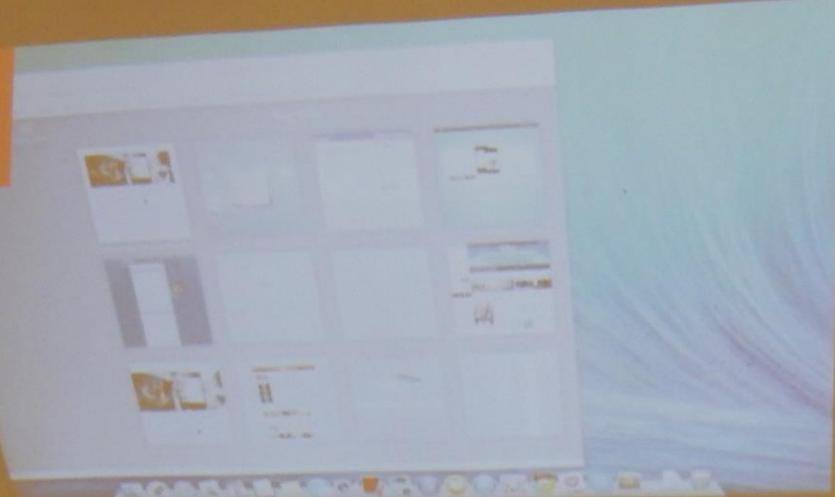
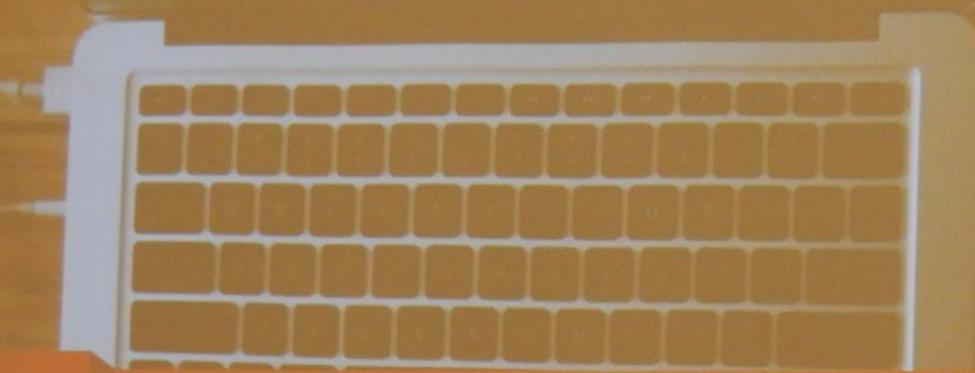
- 9 main issues identified depending on relevance and timing



# SWORD & SWORD for Data

- - Background
- - Summary of planned works
- - Technical development
- - User community
- - Sustainability
- - Archivematica case studies on SWORD implementation
- - Your feedback

SC



# SWORD & SWORD for Data

Dom Fripp (Senior Curation Metadata Developer)



2016

- » SWORD is a lightweight protocol for depositing content from one location to another.
- » Pushes of data and content. Opposite of the OAI-PMH protocol which harvests and pulls metadata.
- » SWORD1 – 2007 (<http://swordapp.org/>)
- » Jisc had supported the development of the SWORD2 deposit protocol with UKOLN and Cottage Labs (2008-10)
- » #dataspring project (2014) – push to publisher template

## Summary of planned works

- » Analysis, requirements and recommendations for updating SWORD and SWORD for data.
- » Based on recommendations update SWORD and SWORD for data.
- » Plan implementation.
- » Community engagement for development (esp. RDM community)
- » Plan curation and sustainability for SWORD.

## Technical development

- » Next generation repository requirements.
- » Taking into account data models e.g. Portland Common Data Model (PCDM) for data sets.
- » Taking into account RIOXX (OA compliance tool)
- » Taking into account other relevant metadata requirements/initiatives at Jisc (e.g. Research Data Shared Service)
- » Pathways for adoption of ResourceSync.

- » Advocacy:

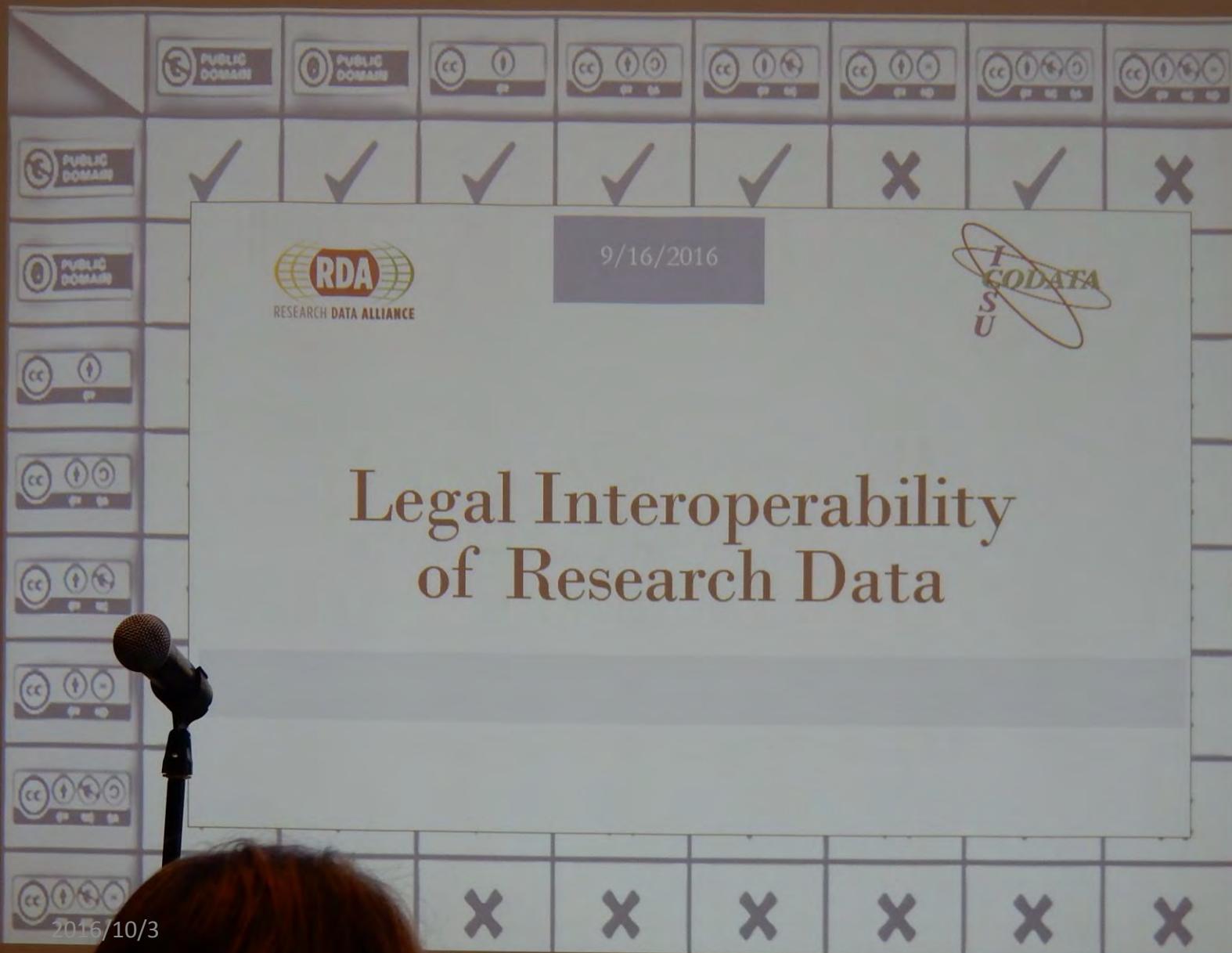
Explaining what SWORD is and how it can make a positive difference to their scholarly systems

- » Usable Implementations:

We already have some good SWORD software but we need more and we need it more widely used.

- » Community Space:

Develop materials already online into stronger focal point for SWORD.



## • 概要

- Legal Interoperability of Research Data
- - Interest Group Charter
- - IG Activities & Discussion
- - The Principles & Guidelines:Who?
- - Legal Interoperability Definition
- - Legal Interoperability Definition
- - Scope Note
- - Early Evidence of Interest/Uptake
- - Early Evidence of Interest/Impact
- - Legal Interoperability Principles and Implementation Guidelines: Dissemination Plan
- - Dissemination Plan
- - Questions
- - Data Citation (DC1)



RESEARCH DATA ALLIANCE

# Virtual Research Environments IG

Helen Glaves, Lesley Wyborn (& Keith Jeffery)

research data sharing without barriers  
[rd-alliance.org](http://rd-alliance.org)

# • 概要

- Virtual Research Environments IG
- - Introduction: rationale for the VRE IG and aims for the session
- - Case studies:
  - - Virtual laboratories - the Australian perspective on VREs (NCI)
  - - Science Gateways (USA), (Purdue University)
  - - Positioning Virtual Research Environments in the European Open Science Cloud Dimitris Koureas (Natural History Museum, UK)
- - Discussion
- - Conclusions and Next Steps

- What is a VRE?
- What is a science gateway?
- Consensus view
- Approaches
- Landscape
- 
- VRE Interest Group
- 
- - Nector
- - Science Gateways:
- - Positioning Virtual Research Environments in the European Open Science Cloud Dimitris Koureas (Natural History Museum, UK)

- (8) Joint meeting of IG Active Data Management Plans,  
IG Preservation e-Infrastructure, IG Reproducibility  
(8) “IG 能動的データ管理計画”, “IG e-Infrastructure維持”,  
および “IG 再現性”の合同ミーティング

## EAGER: Collaborative Research: Supporting Public Access to Supplemental Scholarly Products Generated from Grant Funded Research

Helen Berman, Kerstin Lehnert, Victoria Stodden

Research Data Alliance  
Denver, CO

September 17, 2016

# ● 概要

(8) Joint meeting of IG Active Data Management Plans,  
IG Preservation e-Infrastructure, IG Reproducibility  
(8) “IG 能動的データ管理計画”, “IG e-Infrastructure維持”,  
および “IG 再現性”の合同ミーティング

- EAGER: Collaborative Research: Supporting Public Access to Supplemental Scholarly Products Generated from Grant Funded Research
- NSF requires...
- - NSF DATA SHARING POLICY
- - NSF DMP Requirements
- - GEO Data Management Plan (DMP) Analysis
- - Characteristics of DSE
- - Findings
- - Divisional Data Policies
- - NSF GEO AC Study
- - DMP II
- - Idea: Build on Existing Tools
- - Research

# • 概要

(8) Joint meeting of IG Active Data Management Plans,  
IG Preservation e-Infrastructure, IG Reproducibility  
(8) “IG 能動的データ管理計画”, “IG e-Infrastructure維持”,  
および “IG 再現性”の合同ミーティング

- - Making DMPs actionable and public
- - Defining machine-actionable DMPs
- - Stakeholders and benefits
- - What do we want machine-actionable DMPs to do?
- - DMPonline themes
- - Repository use cases
- - Persistent identifiers (PIDs)
- - API to create a plan
- - Prioritization & next steps
- - Summary
- Information integration through Actionable Data Management Plans

# まとめ

- 一度参加したぐらいでは全貌は見えないが、入門ガイドンスもあり開かれた国際会議。ドメインのデータベース事情やレポジトリ関連技術の規格などについて最新の動向を調査することができる。詳細な技術情報はポスター発表等から収集することもできる。
- ただ技術レイヤーの話題は少ない、枠組みや国際協調のコンセプトを形成するためのきっかけ作りと中期的なワーキンググループの報告の場。分科会で1時間程度の総合討論による意見集約、グループ毎のWebページで意見交換、文書策定を進める作業方式。
- 各機関における意思決定者である事業現場のリーダーや管理職が参加するべきだろう。
- 初参加だとBoF, IG, WGの進捗、成熟度合いが分かりづらいため、興味があるセッションについて定点観測していく必要がある。
- 日本の行政機関、資金提供機関、研究機関からの参加がまだ少ないように感じた。国内でのネットワーク形成と国際連携を同時に進める必要があるだろう。