

ジャパンリンクセンター活用の為の対話・共創の場(第2回)
～研究データへのDOI登録～

Research Data Allianceの活動状況

平成27年2月27日

独立行政法人 科学技術振興機構

情報企画部 上席主任調査員

恒松 直幸

-
1. Research Data Allianceとは何か？
 2. Data Citation Working Groupの活動状況

1-1 概要



- RDA(研究データ連盟)は、研究者とイノベーターが技術、分野、国を超えてデータをオープンに共有し、社会の大きな課題を解決するため、データのオープンな共有を可能にする社会的、技術的架け橋を築くことを目標とした国際的組織。

(<https://rd-alliance.org/about.html>)

- 2012年8月に創設。米国NSF、欧州委員会FP7、オーストラリア産業・革新・気候変動・科学・研究・高等教育省(DIICCS RTE)のAustralian National Data Service(ANDS)が支援している

(<http://www.cs.rpi.edu/~bermaf/Research%20Data%20Alliance.pdf>)

- **メンバー数:73カ国から1,694名 (2014年3月時点)**

エリア:EU(823) AU(62) US(620) 他(189)、

組織の種類:学術・研究(1,096) 政府・公的サービス(273) ITコンサル(58)

大企業(29) 他(238)

1-2 組織目標



- 「研究データ」の共有、交換、利用、二次利用、規格調和、見つけやすさ（アクセス）を促進させていくことを通じて、「国際的なデータ駆動型イノベーション（data-driven innovation）」と「科学的発見」を加速化させていくこと。
- **この実現のために、インフラ、ポリシー、プラクティス、規格等の開発と採択を行っていく。**

(<https://rd-alliance.org/get-involved.html>)

■ On the Web

【RDA (Research Data Alliance) 公式ウェブサイト】 <http://rd-alliance.org>

【RDA Europe (旧名称iCORDI) 公式ウェブサイト】 <http://europe.rd-alliance.org>

【Research Data Canada公式ウェブサイト】 <http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/about/index.html>

1-3 組織構造



(出典: RDA Presentation to G8資料 スライド14参照 (2013年12月18日ウェブアップ)
<http://www.slideshare.net/ResearchDataAlliance/rd-apresentation-forg8o6>
 Presentation by several RDA Council members to G8 Ministers)



1-4 単位組織（Interest GroupとWorking Group）



■ Working Group

- 限定された目的のために、創設から18ヶ月以内にDeliverableを完成する

■ Interest Group

- 関心を共有する参加者の集まり
- ここでの議論からWorking Groupを創設する場合もあり
- 18ヶ月よりも長いスパンで活動する

1. Data Citation (データサイテーション)
2. Data Description Registry Interoperability (データ記述のレジストリ相互運用性)
3. **Data Foundation and Terminology (データの定義)**
* 2014年夏終了予定
4. Data Type Registries (データ型レジストリ)
* 2014年夏終了予定
5. **Metadata Standards Directory (メタデータ規格ディレクトリ)**
* 2014年秋終了予定

6. PID Information Types (永続的識別子の情報タイプ)

* 2014年夏終了予定

7. Practical Policy (実用的な自動化ポリシー)

*2014年夏終了予定

8. Standardisation of data categories and codes (データ・カテゴリーとコードの標準化)

9. Wheat Data Interoperability (コムギ・データの相互運用性)

1-6 Interest Group



1. Agricultural Data Interoperability (農業データの相互運用性)
 2. Big Data Analytics (ビッグデータ分析)
 3. **Biodiversity Data Integration (生物多様性データの集成)**
 4. Brokering (ブローカリング)
 5. **Certification of Digital Repositories (デジタル・リポジトリの認証)**
 6. Community Capability Model
 7. **Data in Context (文脈におけるデータ)**
 8. Defining Urban Data Exchange for Science
 9. Development of cloud computing capacity and education for developing world research
 10. Digital Practices in History and Ethnography
 11. Domain Repositories (分野別リポジトリ)
 12. Engagement Group
 13. **Federated Identity Management (ID連携管理)**
 14. Legal Interoperability
-

1-6 Interest Group { 続き }



-
15. Ethics and Social Aspects of Data
 16. Long tail of research data
 17. Marine Data Harmonization (海洋データ形式の共通化)
 18. Materials Data Management (物質・材料データ・マネジメント)
 19. Metadata (メタデータ)
 20. Photon and Neutron Science (光量子科学・中性子科学)
 21. Preservation e-Infrastructure (eインフラの保存・維持)
 22. Publishing Data (データパブリケーション RDA/WDS 合同IG)
 23. Research Data Provenance (研究データの典拠管理)
 24. Service Management
 25. Structural Biology (構造生物学)
 26. Toxicogenomics Interoperability (毒性ゲノム学の相互運用性)
 27. Geospatial
 28. Materials Data, Infrastructure &
-

Data Citation Working Groupの活動状況

RDA 総会 第四回（2014年9月）における報告

https://rd-alliance.org/sites/default/files/140923_rda_wg_dc.pdf

Status: Working group *endorsed* in March 2014

Goals

- **Concentrating on the problems of dynamic (changing) datasets**
 - ✓ (But: should work also for non-dynamic data)
- **Assigning PIDs flexibly to arbitrary subsets of data**

***PID: Persistent Identifier**

Entire dataset, arbitrary subsets of rows/columns, subgraphs,...

SQL, XML, RDF, LOD, CSV, ...

■ スコープ

Does NOT deal with metadata, landing page design, bibliometrics, ...

■ アプローチ

- Creating concepts, recommendations and pilots/demonstrators
- Starting with conceptual evaluation of the approach, studying fitness, impact, scalability, changes required, ...
- Followed by actual pilot implementation

■ 要求事項

1. Ensure data is time-stamped and versioned
2. Assign PID to time-stamped query/selection expression

■ 成果 Deliverables:

1. Requirements: minimal set of functionality for data and service
2. Pilot implementations for selected data types
3. Reference architecture and guidelines for implementation
4. Pilots

<https://rd-alliance.org/groups/data-citation-wg/wiki/use-cases-pilots-precisecitation-data.html>

-
1. **LNEC: Portugese Civil Engineering Lab (SQL)**
 2. **NERC: UK Natural Environment Research Council data centres (SQL)**
 3. **CLARIN: XML**
 4. **MSD: Million Song Database (CSV, SQL)**
 5. **VAMDC: Virtual Atomic and Molecular Data Centre**

- Which timestamp to assign?
- (query, last update, last change to result set)
- Hashing for verification
- Annotating compulsory deletion
- How to handle distributed datasets
- Migrating data and queries