

Research Data Alliance 第7回総会参加報告

1. 概要

2016年3月1～3日に、Research Data Alliance（以下、RDA）の第7回総会が、東京の一橋大学一橋講堂を会場として開催された^{注1)}。今回の総会のテーマは「Making data sharing work in the era of Open Science」であった。RDAは、研究データのシェアリングを目的として、2012年9月に発足した団体である。技術的・社会的なインフラの構築をめざしているが、動きが極めて活発かつ速く、研究データにかんするディスカッションの中心的な場となっており、オープンサイエンスを進めていくにあたっては、その動きに注目していく必要がある。RDAでは、研究者やエンジニア、図書館職員など、様々な分野の専門家がボランティアベースで参加し、それぞれが興味のあるコミュニティに参加して、ディスカッションを重ね、検討を進めていく。コミュニティは、BoF、IG、WGに分かれている。BoFはBirds of a featherの略で、ICT系で良くもちいられるとのことであるが、同じ興味を持つ参加者のゆるやかな意見交換のコミュニティである。BoFで意見交換を重ねることからスタートし、課題が決まると、Interest Group（以下、IG）を立ち上げ、その課題についての具体的な検討が行われる。さらに、その課題は具体的な実行計画をもとに、Working Group（以下、WG）として18ヶ月間で解決され、提言や提案としてまとめられる^{注2)}。基本は個人参加であるが、様々な分野の中心的な人物が参加しており、その影響力は大きい。総会は半年に一回開催されるが、総会は全体セッションの他、コミュニティによって運営される様々なセッションにより構成されている。また複数のコミュニティによるジョイントのセッションも多い。

今回の総会開催に当たり、機関リポジトリ推進委員会からは、鈴鹿工業高等専門学校 青山、お茶の水大学 香川、国立極地研究所 南山、千葉大学 三角の4名が参加した。本報告では、それぞれ参加したセッションについて報告する。

注1)大会プログラムは下記

<https://rd-alliance.org/plenary-meetings/seventh-plenary/programme.html>

注2) RDAについては、下記が詳しい。

恒松他. 研究データ同盟 (Research Data Alliance) 第2回総会. 情報管理 Vol. 56(2013) No. 10 P724-727

https://www.jstage.jst.go.jp/article/johokanri/56/10/56_724/_pdf

2. 各セッションについて

以下、個々のセッションについて、参加報告を行う。

1) 3月1日 (火)

(1) 09:30 - 11:00 Opening Plenary - Hitotsubashi Hall

① 基調講演

- Mark Parsons, RDA Secretary General
- Yuko Harayama, Executive Member, Council for Science, Technology and Innovation Cabinet Office, Japan
- Yasuo Kishimoto, Deputy Director-General, Science and Technology Policy Bureau, Ministry of Education, Culture, Sports, Science and Technology Japan
- Satoru Ohtake, Principal Fellow, Japan Science and Technology Agency
- 【Video Message】 Michael Hager, Head of Commissioner Gunther Oettinger's Cabinet, European Commission

それぞれの立場からのメッセージが寄せられた。以下のようなコメントがあった。

- インターネット技術の発達によって、サイエンスがドラステックに変化しようとしている。
- 分野の違いを障壁とせず、世界規模で活発なディスカッションを行い、政策的な提案や推奨指針を策定していく必要がある。
- 国際協力の元、研究データ管理に必須であるシステムやプラットフォームといった基盤整備をオープンな環境で進めることにより、サイエンスのエコシステム構築を目指す。プライバシー等に配慮すべきデータも存在するが、オープンであることは社会にとって重要。
- 大量のデータからユーザが求めるものに素早くアクセスできるようにしなければならない。新たなディスカバリを確立し、社会に革新をもたらす必要がある。
- データの相互運用性の確保や処理フローの確立、十分な容量のストレージ、コミュニケーションツール等、議論すべき課題が多く、RDA 構成員には高いレベルのパフォーマンスが求められている。関連する研究教育機関や政府とも連携する必要がある。

② パネルディスカッション

「Towards a Global Internet of Data in the Open Science Era」というテーマの元、各パネリストより、各国におけるオープンサイエンスに関する取り組みの紹介があった。

- Jean-Claude Burgelman, European Commission
- Yasuhiro Murayama, National Institute of Information and Communications Technology, Japan
- Barend Mons, European Open Science Cloud High Level Expert Group
- Mark Parsons, Research Data Alliance
- Ross Wilkinson, Australian National Data Service
- Stephen Wolff, Internet2

以下のようなコメントがあった。

- 欧州の研究者にデータストレージや再利用環境を無償で提供している The European Open Science Cloud (EOSC) の取り組みとして、「A Digital Single Market Strategy for Europe」(<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1447773803386&uri=CELEX:52015DC0192>) の公開 (2015/5/6) や、ディスカッションを通して関係者へのサポートを強化している等の報告がなされた。
- 共通的な認識として、より広い意識喚起や共通的な標準の整備、ベストプラクティスの策定、最新のプライバシーやデータ保護、著作権に関する翻訳の必要性、ハードウェアの容量、異なる分野のコ

コミュニティへの働きかけ、長期間に渡って必要となるデータ管理職の整備等の課題が挙げられた。

・オープンサイエンスはオープンアクセスとは異なり、データがより多量（ビッグデータ）であることを意識する必要があるといった話や、信頼性のあるアクセスを提供するサービスおよびシステムの必要性、多様なソースのデータ管理や機関における研究データの集約等、システム的かつデータサイエンティストのような専門的なデータ管理スキルが要求されるという指摘もあった。

・データは研究のためだけではなく、産業界や社会に還元する必要があるとの認識も示された。

(2) 11:30 - 13:00 Breakout 1 - Working Meeting Session 1

IG Education and Training on handling of research data: Focus on Research -Data handling related competences and skills - Conference Room 4

<概要>

<https://rd-alliance.org/ig-education-and-training-handling-research-data.html-0>

<資料>

<https://rd-alliance.org/group/education-and-training-handling-research-data/wiki/ig-etrd-meeting-presentations-rda7-tokyo-1>

IG Education and Training on handling of research data: (The IRG-ETRD) は、研究データの取り扱いに関する教育や訓練に関連した情報交換を目的としたグループである。データサイエンティスト、データライブラリ、データキュレーターなどの新しい専門職の職能の定義や、その養成のための研修もターゲットとしている。

セッションでは、まず Laura Molloy 氏により、グループの活動についての説明があった。特にリサーチライブラリアンについては、下記サイトに必要なスキルがまとめられている。

<https://rd-alliance.org/node/971/all-wiki-index-by-group>

上記サイトで紹介されている、スキルは膨大なものであり、現時点では、情報収集のフェイズであるが

- ・ Data citation (CE) (Would suggest this as essential - AN)
- ・ Data literacy (CE)
- ・ Data visualisation (CE)
- ・ Data mining (CE)
- ・ Databases (CE)

などのスキルについて、情報源とあわせて紹介されている。

また EDISON プロジェクトについての紹介があった。EDISON は EU により助成を受けている 24 か月のプロジェクトで、2015 年 9 月にスタートした。data scientist を新しい職種として確立することを目指しており、技術的なトレーニングばかりでなく、ポリシー策定やガバナンス、ビジネス等の社会への実装も見据えた議論が行われている。

EDISON プロジェクト公式サイト) <http://www.edison-project.eu/>

全般的な印象として、人材育成について、すでに組織的・戦略的に議論が進められており、日本に比べて相当に進んでいるし、関わっている人間も多い。一方で現時点では、どのようなスキルが必要かを収集・整理している段階であり、育成プログラムとしてはまだ試行段階のようであり、必要なスキルについての整理はこれからである。日本として人材育成をどのように考えていくべきかの議論を早々にはじめるべきであろう。

(3) 11:30 - 13:00 Breakout 1 - Working Meeting Session 1

IG PID – Room 102

<概要>

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

<報告>

<https://rd-alliance.org/group/pid-ig/post/rda-pid-ig-tokyo-plenary.html>

研究データの相互運用性を向上させ、リンクによるデータ参照の基盤的な役割を果たす Persistent Identifiers (PID: 永続識別子) について、情報交換やディスカッションを行うために設定されたセッション。永続識別子の普及に取り組む関係者（外部プロジェクトや RDA の他の分科会）から、最新の活動状況について報告がなされた。

① PID Infrastructure: Update on the THOR Project

Project THOR (<http://project-thor.eu/>) は、全ての研究者が全てのキャリアにおいて、シームレスに永続識別子にアクセスできることを目指すプロジェクト。技術や人的基盤の整備、オープンサイエンスに関する調査等を実施しており、ands、British Library、CERN、DRYAD、DataCite、Elsevier、EMBL-EBI、ORCID、PANGAEA、PLOS の 10 機関がパートナーとしてプロジェクトに参加している (<http://project-thor.eu/the-thor-partners/>)。本セッションでは、各研究プロセスで異なるバージョンのデータが作成されるためバージョン管理が重要であるといった解説や、国際的な動向として 2015 年 6 月から CERN openlab に EMBL-EBI が参加した事例

(<http://news.embl.de/lab-matters/1506-cern-openlab/>) の紹介があった。

また、Project THOR の活動報告として、PID に関する情報を収集・発信する THOR knowledge hub (<https://project-thor.readme.io>) の開設、テクニカルワークショップやカンファレンスを開催して対面でのコミュニケーションやフォローアップを実施していること、DataCite や ORCID の API に関する情報収集を継続していること等も報告された。

② PID Interoperability in Action: Crossref, DataCite, and ORCID Auto-Update

CrossRef、DataCite、ORCID の連携機能について紹介があった。

日中韓では同じ苗字が多く、中でも日本人の人名はより複雑で人物の特定が困難な場合も多い。ORCID には 190 万人の研究者の情報が登録されており、CrossRef により ORCID と出版社のメタデータ (DOI を含む) が紐付けられている。DataCite profiles 上で自動更新を設定しておくと、出版社が自身の ORCID 入りのメタデータを登録すると自動的に自身のコンテンツとして紐付けるサービスや、その際のソース (出版社) の優先順位を研究者自身が設定できる機能について紹介があった。実際に DataCite の検索サービスでは、ORCID で検索する方が正確で不要なデータが表示されない事例のデモも行われた。また、出版社側からもデータ登録時に著者に ORCID の入力を求めるケースも出始めており、ORCID の普及に向けた取り組みが進んでいることが窺えた。

加えて、データと論文のリンクについては RDA/WDS Publishing Data IG (<https://rd-alliance.org/groups/rdawds-publishing-data-ig.html>) が検討していることや、データから機関へのリンクについては共通の識別子が、助成機関へのリンクについては共通のレジストリが必要という話もあった。

③ Organization IDs: Call for Community Action

CrossRef が提供する助成を受けた研究論文等を検索できるサービス Funding data Search (<http://search.crossref.org/fundref>) や、NISO が 2013 年に発表した機関 ID に関する推奨指針 [NISO RP-17-2013 Institutional Identification: Identifying Organizations in the Information Supply Chain]

(http://www.niso.org/apps/group_public/project/details.php?project_id=110) について紹介があった。

また、機関 ID に関連する取り組みやイベントとして、Jisc CASRAI-UK Pilot、CNI Spring 2016 Membership Meeting (April 4-5, 2016, San Antonio)、FORCE2016 (April 17, 2-19, 2016, Portland) に関する情報提供があった。

④ Open Microphone session: RDA and community PID initiatives

永続識別子に関連するプロジェクトやWGによる事例紹介があった。

Data Publishing WG、Data Services WGからは、持続性のあるインフラを既存システムの拡張で整備できるようにしなければならないといった話や、永続識別子は重複除去の解決に役立つといった解説があった。

欧州における研究データの保存・共有の基盤整備を推進する EUDAT からは永続識別子に関する政策や B2HANDLE (<http://eudat-b2safe.github.io/B2HANDLE/>) の紹介、長期間に渡る質の保証がユーザにとつての最大のメリットとなるといった話や、ベストプラクティスに沿った設計により証明可能な手法でデータは管理されるべきといった話があった。

SESAR (<http://www.geosamples.org/>) が提供する、地球環境科学系分野の標本に付与される識別子 IGSN (International Geo Sample Number) についての紹介では、DOI と同じ技術手法が用いられ、9桁のコード (例: IGSN:HRV003M16) を岩石標本等に付与して利用している。スキーマは Registration Schema と Description Schema (<http://schema.igsn.org>) から構成され、Elsevier が発行する雑誌の著者ガイドにも論文の投稿データに IGSN を含めるよう記載されている事例

(例: <https://www.elsevier.com/journals/lithos/0024-4937/guide-for-authors>)

の紹介があった。

欧州のパートナーで構成されるコンソーシアム The European Persistent Identifier Consortium (ePIC) からは、handle system に基づいて欧州の研究コミュニティに永続識別子を提供するサービスや API の紹介、DONA 財団による Multi-Primary Administrators (MPA) フレームワーク (<https://dona.net/mpa/>) の紹介があった。

研究基盤における認証や承認のインフラ (AAIs: Authentication and Authorization Infrastructures) に関する課題を解決するための IG である Federated Identity Management (FMI) IG からは、SAML2.0 対応や X.509 証明書、属性の活用や認証のシームレスな統合等、技術的な側面からの検討状況について報告があった。

永続識別子は検索における正確性の向上だけでなく、他のデータセットとのリンクや名寄せを行う場合の重要な要素となるため、最新の活用技術や先行サービスの動向を継続的に情報収集する必要があると感じた。

(4) 14:30 - 16:00 Breakout 2 - Working Meeting Session 2

Joint meeting of IG Education and Training on handling of research data, WG RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World: International Coordination of Research Data Education and Training Requirements -Conference Room2

<概要>

<https://rd-alliance.org/joint-meeting-ig-education-and-training-handling-research-data-wg-rdacodata-summer-schools-data.html>

IG on Education and Training on Handling of Research Data と WG RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World のジョイントミーティングであ

る。‘research data science’を進めるためには、様々な、データを扱うためのスキルを取得するための、トレーニングコースが必要である。その目的のために、いくつかの提案がなされている。

CODATA-RDA WG は、2016年8月に‘Research Data Science’に関する2週間のサマースクールを企画している。今回のコースは基本編として、様々な研究分野に共通の要素をとりあげたものになっているが、次の段階としては、分野にそって「香りづけ」したサマースクールに発展させていくとの事であった。こちらのスクールは、研究データをあつかう研究者ないし技術スタッフ向けのものであり、インフラ開発やソフトウェアや可視化、統計、機械学習など研究に直結したデータ処理技術が含まれており、図書館員向けではない。しかしデータサイエンスの全体像をとらえるためには参考になるであろう。サマースクールについては、作業用ファイルではあるが、下記が最新のものである。

スクールの趣旨

<https://drive.google.com/file/d/0B1DCCRXRF8ydjZMSUd4d1l6cEE/view>

カリキュラムのアウトライン

<https://docs.google.com/document/d/1tqMMVfeufPZqM5pwTOZAsQ8aGUhuF1Mo2yLIaeAa46A/edit>

CODATA-RDA Working のページ

<http://www.codata.org/working-groups/research-data-science-summer-schools>

その他、EDISON プロジェクト、デジタルキュレーションのプロジェクトの DigCurV についての紹介があった。

参考)

1. EDISON Project <http://www.edison-project.eu/>
2. DigCurv Project Curriculum Framework and Lenses
<http://www.digcurv.gla.ac.uk/>

(5) 14:30 - 16:00 Breakout 2 - Working Meeting Session 2

Joint meeting of IG Metadata, WG Metadata Standards Catalog, IG Data in Context, WG Metadata Standards Catalog: Domain Discussion on Initial Standard Canonical Set of Metadata Elements-
Hitotsubashi Hall Remote access available

<概要>

<https://rd-alliance.org/joint-meeting-ig-metadata-wg-metadata-standards-catalog-ig-data-context.html>

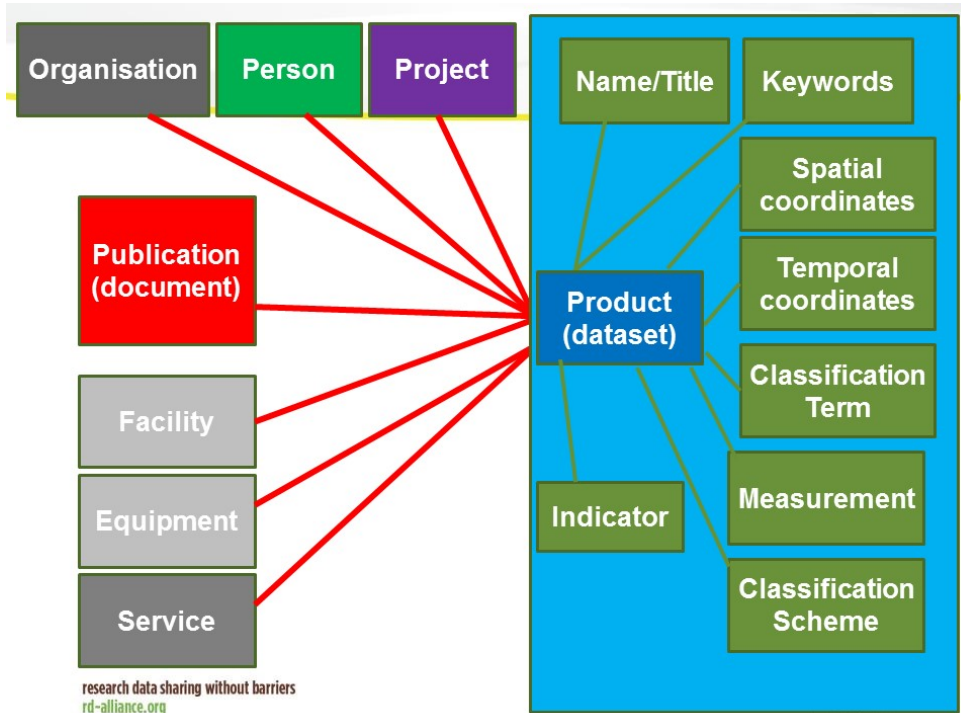
ドメイングループおよび技術グループの確認を踏まえ、パリでの前回総会においてメタデータ要素のドラフト版を提示した。要素のドラフト版は、Data in Context IG が担当したユースケース分析、Metadata Standards Directory WG が主に策定したメタデータ標準から、Metadata Interest Group (MIG) のコーディネートにより作成したもの。実際には、要素の内部や他の要素との間に関連性を持つ等、複雑な構造となる場合が多いため、新たな標準要素セットを策定するというよりは、各ドメインが使用および希望するメタデータ要素について理解を深めることを目的に開催された。

まずは、RDA のメタデータに関係する 4 グループが 2014 年 11 月に共同で策定した「Metadata Principles and their Use」

(<https://rd-alliance.org/metadata-principles-and-their-use.html>) について解説があった。メタデータは単なる記述や発見のためだけに用いられるものではなく、関連性や品質、権利等の情報を含むことで、データをユーザ、ソフトウェア、電子リソースに紐付ける機能を果たす、といった解説がなされていた。

また、数種の標準を比較して共通的な要素を抽出し、必要最小限の要素に集約するというユースケース分析を実施した結果、複数の値を持つ場合や複数言語で記述される場合が多いため、フラットな構造からセマンティック技術を活用し要素を相互リンクさせる構造化への転換が必要という結論に至った経緯が説明された。

各データセットのリンク概念図



続いて、要素のドラフト版についてオープンなディスカッションが行われ、主な意見として以下のものが挙げられた。

- ・多様な識別子が存在する要素（特に役割や時代）は、識別子の種類を明示して区別する。
- ・要素のリスト化はできたが、本当に難しいのは語彙とオントロジー。セマンティックへの対応が最大のチャレンジとなる。
- ・メタデータ作成は労力が大きいため、コミュニティでよく議論する必要がある。
- ・各リポジトリのシステムが採用しているメタデータには様々なバリエーションがあるため、マッピングして活用する必要がある。
- ・データの再利用を促進するためにはライセンスを適切に定義する必要がある。W3C 推奨の標準的なライセンスを使用するか、またはデータセットに特有のライセンスを策定するか、検討の余地がある。研究者が自身の目的に応じたライセンスを選択できるようにしなければならない。

Product (dataset)要素のドラフト版

Unique Identifier (for later use including citation) of the dataset (or software...)
Location (URL) of the dataset
Description
Keywords (terms)
Temporal coordinates

Spatial coordinates
Originator (organisation(s) / person(s))
Project
Facility / equipment
Quality
Availability (license, persistence)
Provenance
Citations
Related publications (white or grey)
Related software
Schema
Medium / format

次の段階として、各要素の定義を公開するとの予告があり、セッションに用いたプレゼン資料とともに後日 Web サイト上に掲載された。

<https://rd-alliance.org/metadata-p7-tokyo.html>。

日本でも課題として話題に上がることのある、フラットから構造化への転換について、世界的にも議論されていることが確認できた。また、「Metadata Principles and their Use」は基本理念として研究データに限らず、他の資料種別のメタデータ管理においても参考になる部分がある。今回提示された要素は共通的かつ相互運用性の確保に重点を置いたため、Dublin Core をアレンジしたような非常にシンプルな定義となっていた。要素定義よりも語彙とオントロジーの策定の方がより難易度が高いというコメントがあったとおり、実際にどのような技術で構造化するのかといったことや、書誌記述とリンク記述の書き分け、他のデータセットとのリンクの仕組み等、具体化・詳細化について引き続き情報収集を行う必要があると感じた。研究データに関するメタデータについては、オープンアクセスリポジトリ連合 COAR (

<https://www.coar-repositories.org/activities/repository-interoperability/ig-controlled-vocabularies-for-repository-assets/>)

や OpenAIRE (<https://guidelines.readthedocs.org/en/latest/>) 等でも検討やガイドラインの整備が進行しているため、各プロジェクトの相違点等を考慮して参考にする必要がある。

(6) 16:30 - 18:00 Breakout 3 - Working Meeting Session 3

IG long tail of research data: Managing diverse data sets: challenges and incentives - Hitotsubashi Hall

<概要>

<https://rd-alliance.org/ig-long-tail-research-data.html>

データ共有に関する調査分析やプロモーションを実施する IG で、このセッションでは今までの取り組みの報告や、今後の方向性に関するディスカッションが行われた。(アジェンダ: https://rd-alliance.org/sites/default/files/Longtail_agenda_March2016.pdf)

メタデータに関する調査結果では、Dublin Core と DataCite が最も一般的なスキーマであったことや、半数未満のリポジトリで DOI が活用されていたこと、メタデータはリポジトリ内での検索には役立っているものの、より広範囲な検索エンジンやディレクトリを用いたディスカバリには対応できていないという課題が紹介された。また、データ登録プロセスの簡素化に逆行してでもデータの発見や再利用のためにメタデータを詳細に記述すべきかについての議論や、データの発見性を向上させるた

めに DOI を付与してデータを論文に紐付けること、データ管理計画に関する戦略の策定について、ディスカッションを実施していることが報告された。

研究データ管理ツールに関する 2015 年の調査結果から、研究者はデータ管理のために様々なツールを利用していることが判明しており、リポジトリにそのような機能が実装されればリポジトリを研究データ管理ツールとして利用するインセンティブとなり得ることが示された。また、各分野でデータ管理に必要な機能等に関する小規模な調査を実施し、利便性、柔軟性、オープンソースであること、分析機能を有していること、の切り口による評価を実施したことが報告された。

また、関連する調査として「Belmont Forum Open Data Survey 2014」

(<http://dx.doi.org/10.5281/zenodo.16384>) についても紹介があった。

今後のトピックや方向性に関するディスカッションでは、ヨーロッパの助成団体（政府系ファンドを含む）や出版社等の関係する団体との連携をさらに進めること、IG としての組織的な戦略策定やミッションの集中化、データ共有に係るインセンティブの設定、効果的な管理ツールやサービス、ワークフローの提示によるデータ登録の促進、業界を超えた相互運用性の確保や発見性の向上を目指すこと等が議論された。

このグループは、機関リポジトリや図書館員が研究データ管理で実行可能な役割に焦点を当てて調査や広報活動を行っているため、図書館関係の事例や概要的な情報収集に役立つものと思われる。

2) 3月2日 (水)

(1) 9:00 - 10:30 Plenary Session - Hitotsubashi Hall

① RDA/WDS Publishing Data Bibliometrics Recommendations (09:10 - 09:20)

<プレゼンテーション資料>

<https://rd->

[alliance.org/sites/default/files/attachment/01-%20RDA%20Bibliometrics%20Outputs_FINAL.pptx](https://rd-alliance.org/sites/default/files/attachment/01-%20RDA%20Bibliometrics%20Outputs_FINAL.pptx)

データメトリクスはなぜ必要か?という問題提起がまずあった。funder、機関n、その他、すべてのステークホルダーがデータのインパクトと価値を知りたいがっている。研究者本人も自分のデータがどれだけ影響力を持つか知りたい。しかしインパクトと価値を知るためには、それには伝統的なジャーナルよりも多くの情報が必要であろう。

そのためには

- ・全体像の記述と集中すべきエリアの確定。

要素の洗い出し。citation count, download, social media (Altmetrics), peer-reviewed papers mention etc...

- ・より使われるために何が必要か。standards, data citation, consistent use of IDs/DOIs, culture change "a belief that they are valid"

- ・誰に対してインパクトがあるか。data producers, research funders, repository managers, data users.

等の検討が必要である。

なお

California digital library がデータカウント手法を作成しており、NSF も支援している。

<http://mdc.lagotto.io/>

② RDA/WDS Publishing Data Services Recommendations (09:20 - 09:30)

<プレゼンテーション資料>

https://rd-alliance.org/sites/default/files/attachment/02-RDA-P7-recommendation_and_adoption.pptx

目的は文献と研究データのリンクをすすめ、「見える化」「見つけやすさ」を向上し、再利用性を高

めることである。またクレジットを付与する仕組みについても検討している。現状は、文献とリンクしていないデータが多量に存在するが、これは技術的問題だけではなく慣習的な問題でもある。その対応のために集約型のソフトを作成し、全体をリンクさせたい。その目的のために、data / literature interlinking service を構築し、OpenAIRE、PANGAEA と連携しプロトタイプを作った。メタデータは一般的標準的なものをベースにした。このモデルは、複数のハブが互いにリンクしあうものである (multi-hub model)。新規 hub とし参加について歓迎したい。

③ RDA/WDS publishing data workflows recommendations (09:30 - 09:40)

<プレゼンテーション資料>

https://rd-alliance.org/sites/default/files/attachment/03-RDA-WDSPublishingDataWorkflows_P7_adoptionSessions_v04.pptx

データ流通の向上のために、データの publishing のためのワークフローを検討している

参考) <https://zenodo.org/record/34542#.VmVJqMrWlqc>

(2) 11:00 - 12:30 Breakout 4 - Working Meeting Session 4

IG Data Foundations and Terminology: expanding vocabulary coverage and services - Conference Room 4

<概要>

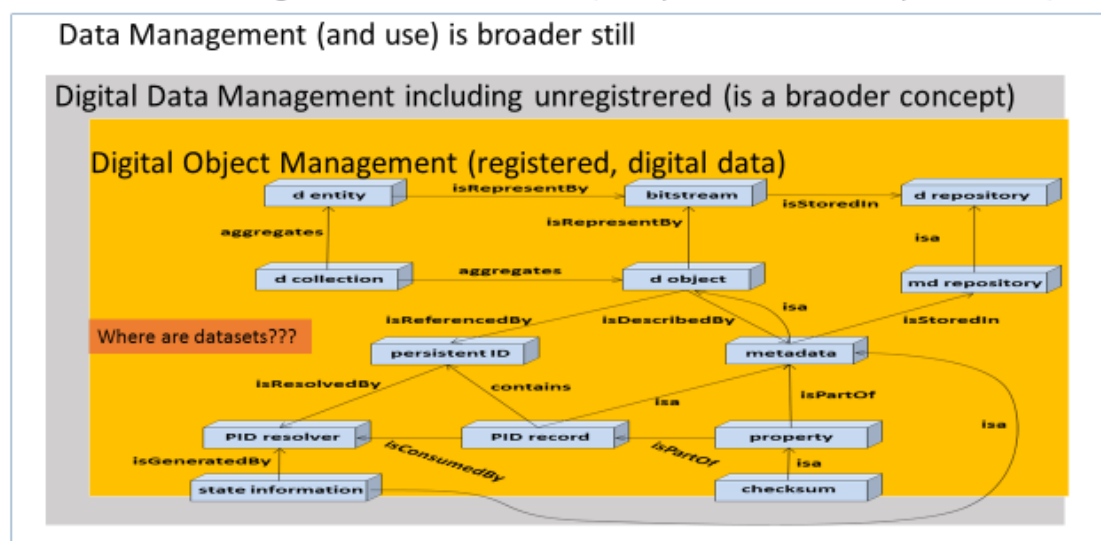
<https://rd-alliance.org/ig-data-foundations-and-terminology.html>

本 IG では、データ管理におけるフローや用語を標準化し、構造化して概念図を作成することを目標としている。RDA DFT WG の成果を引き継いでいるため、これまでの検討事項は下記ページを参照されたい：

<https://rd-alliance.org/ig-data-foundations-and-terminology.html>

図は、前回までの IG で作成された概念図のたたき台である：

Concept map overview of Core Terms Broadening the Discussion (Stepwise or Scope-wise)



Metadata group からの情報提供を受け、オープンデータにおける element の目的を 3 つに整理し直し

た上で、標準的な OA ポリシーに沿ったデータマネジメントの概念図を作成したい、との説明がなされた。なお、element の 3 つの目的は discovery、context、detailed だが、各々重なり合っているとのことである。

ディスカッションでは、主に概念図の relationship の妥当性、新たな用語の追加について活発な議論がなされたものの、時間の都合上引き続きコアメンバーによる電話会議に持ち越しとなった。

(3) 11:00 - 12:30 Breakout 4 - Working Meeting Session 4

Joint meeting of IG Data Rescue, IG Data Fabric, IG Preservation e-Infrastructure, IG Domain Repositories, IG Libraries for Research Data: Rescuing, Re-Using and Sharing Data At Risk - Conference Room 1

<概要>

<https://rd-alliance.org/joint-meeting-ig-data-rescue-ig-data-fabric-ig-preservation-e-infrastructure-ig-domain-repositories>

データレスキューをテーマとしたセッションである。最初に DataRescu のグループについての紹介があり、社会科学系のパンチカード等のデータのレスキューの事例が紹介された。

その他、2016 年 9 月にアメリカコロラド州で開催されるワークショップ (Workshop on "The Rescue of Data At Risk", Boulder (CO), Sept 8-10, 2016) についての紹介があった。

また第三回 international Elsevier Award for "Data Rescue" in the Geosciences についての紹介があった。

参考)

○2015 International Data Rescue Award in the Geosciences

<https://www.elsevier.com/physical-sciences/earth-and-planetary-sciences/the-2015-international-data-rescue-award-in-the-geosciences>

○CODATA による data rescue のページ

<http://codata.org/blog/2015/07/02/data-at-risk-and-data-rescue/>

(4) 15:00 - 16:30 Breakout 5 - Working Meeting Session 5

BoF on research data repository interoperability Conference Room 102

<概要>

<https://rd-alliance.org/bof-preparation-research-data-repository-interoperability-wg.html>

<要旨>

<https://rd-alliance.org/system/files/documents/2016-03-RepositoryInteroperabilityMeetingNotes.pdf>

このコミュニティの目的は、種々の研究データリポジトリの相互運用であり、そのためにリポジトリの標準的な API (Generic repository API) の開発を目指している。開発のコンセプトとしては removing barriers, easier collaboration, creating commonalities である。ディスカッションとしては以下のようなコメントがあった

- ・標準化する試みについて、利用者が簡単に使えることを指向するのは大切だが、技術発展のスピードが速すぎる。標準仕様を決められるか? (あるいは、決めることに意味があるのか?)
- ・11 月にある IEEE の big data ワークショップにプロポーザルを出したい、そのためには何が必要かを検討する必要がある。テレビ会議でディスカッションを続けたいが、追加の協力者募集したい。

3) 3月3日 (木)

(1) 09:00 - 10:30 Breakout 6 - Working Meeting Session 6

IG Libraries for Research Data: Applying Global Information-sharing and Collaboration in Libraries to Local Practice - Conference Room 3

<https://rd-alliance.org/ig-libraries-research-data.html>

<概要>

<https://rd-alliance.org/ig-libraries-research-data.html>

<配布資料>

<https://rd->

[alliance.org/system/files/documents/RDA%20P7%20Libraries%20for%20Research%20Data%20Interest%20Group%2020160303.pdf](https://rd-alliance.org/system/files/documents/RDA%20P7%20Libraries%20for%20Research%20Data%20Interest%20Group%2020160303.pdf)

<要旨>

<https://rd-alliance.org/system/files/documents/RDALibrariesforResearchDataP7-Notes.pdf>

図書館員向けに研究データ管理に関する情報提供やスキルの習得方法等の紹介を行っている Libraries for Research Data (L4RD)による、図書館関係者の国際的な情報共有の促進や各国の事例共有のために設定されたセッションである。

① Brief introduction to L4RD and overview of group activities to date

L4RD について概要説明があり、第2回総会からセッションを開始し、現在では188人が参加している等が紹介された。

② 23 Things: Libraries for Research Data

図書館員向けの研究データ管理に関する学習リソースをパンフレット形式でまとめた「23 Things: Libraries for Research Data」に関する紹介があった。英語のほか、日本語にも翻訳され、RDAのWebサイト上で掲載されていることや

(英語版)

https://rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_Management.pdf

(日本語版)

<https://rd->

[alliance.org/system/files/documents/23Things_Libraries_For_Data_RDA%EF%BC%BF.jp.pdf](https://rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_RDA%EF%BC%BF.jp.pdf))

データ管理計画やデータリテラシー等に関するオンラインリソースをThe Australian National Data Service (ANDS)のWebサイト上で公開する取り組み

(<http://ands.org.au/partners-and-communities/ands-communities/23-research-data-things>)の紹介があった。

③ Panel: “Asia-Pacific Trends in Libraries and Research Data”

東アジア地区から事例紹介がなされた。日本からは南山協力員が報告者として参加し、機関リポジトリを中心としたオープンアクセス活動の周辺状況や、機関リポジトリ推進委員会での研究データ管理に関する取組みの紹介を行った。質疑応答では、人事異動制度について、参加者からリポジトリ業務は必ず異動サイクルの一つとして位置づけられているのか等の質問があった。

韓国からは、データリポジトリ「Datanest」に関する紹介があった。DSpaceをベースにSamsung社が開発したシステムで、セントラルとローカルのシステムで構成され、データは国立データセンターに集約しているが、各機関のサーバ上に保持することも可能。また、既にDSpaceを導入している機関

は、既存システムを使い続けることもできるとの紹介があった。図書館員がメタデータの正規化や管理を行っており、図書館員が必要とされる部分はメタデータに関するスキルであるといった話もあった。今後の予定として、処理速度等のパフォーマンスの強化、ビッグデータの適切な蓄積方法についての検討が挙げられていた。

シンガポールからは、国内のほとんどの大学で機関リポジトリを構築しているという現状報告や、データ管理に関するガイドラインやベストプラクティス、オンラインコースウェア MANTRA (<http://datalib.edina.ac.uk/mantra/>) を参考にしており、データ共有ポリシーの策定を検討していること、研究分野によってデータの定義が異なることを意識しなければならない等の話があった。

④ Organizational roadmap for adopting RDM services: review and feedback of outline

COAR より、研究データ管理に関する機関向けのロードマップを作成中で、機関リポジトリ等の既存システムの拡張方法に焦点を当てた内容となっている旨の報告があった。

⑤ Organizational Models for Libraries Providing Data Services

イリノイ大学からの事例報告として、図書館員がどのように研究データ管理をサポートできるかを理解すること、研究データ管理のために必要な組織構造を検証すること、相互連携可能な構造やサービスを模索することをプロジェクトの目標としている旨の説明があった。今後の予定として、組織的なアプローチや機関意思決定方法等に関するより詳細な分析、普及活動等が挙げられていた。

研究データ管理は、欧米やオーストラリアにおいて先行して取り組まれてきたが、今回アジア地区で初めて総会が開催され、日本、韓国、シンガポール各国からの事例報告においても活発な意見交換がなされる等、アジア地区での普及も RDA の視野に入りつつあるように感じた。国の政策や普及しているシステム等の相違はあるものの、LARD や関連する先行事例は参考になる部分が多いため、引き続き情報収集を続ける必要があると感じた。

(2) 11:00 - 12:30 Breakout 7 - Working Meeting Session 7

BoF e-Infrastructure for Global Change Research - Room 102k

<概要>

<https://rd-alliance.org/bof-e-infrastructure-global-change-research.html>

グローバルな環境変化に対する調査を行い、戦略の提示や情報提供、リスク調査を担う団体である「The Belmont Forum」の取り組みについて紹介するためのセッションである。

初めに The Belmont Forum について、14 カ国以上の各分野の科学者によって構成される団体のコミュニティ戦略や実装プランに関するレポート「A Place to Stand: e-Infrastructures and Data Management for Global Change Research」(<http://www.bfe-inf.org/document/community-edition-place-stand-e-infrastructures-and-data-management-global-change-research>) を刊行しているという概要説明があった。

レポート中では5つの推奨指針(1. ポリシーの策定、2. 研究コミュニティにおけるコミュニケーション、コラボレーション、コーディネーションの推進、3. 効果的なデータ管理計画、管理ポスト設置の促進、4. 国際的かつコミュニティにとってのベストプラクティスを定めること、5. 産業を超えたトレーニングカリキュラムの構築支援)や、4つのアクションテーマ(1. Coordination Office、2. Data Planning、3. E-infrastructure、4. Human Dimensions)等に関する解説がなされている。また、各テーマは、The French National Research Agency (ANR)、Commonwealth Scientific and

Industrial Research Organisation (CSIRO)、Deutsche Forschungsgemeinschaft (DFG)、European Commission、Sao Paulo Research Foundation (FAPESP)、Japan Science and Technology Agency (JST)、Ministry of Science and Technology (MOST)、Natural Environment Research Council (NERC)、National Science Foundation (NSF)等のメンバー機関により実施されている等、プロジェクトの体制についての説明があった。

プロジェクトの専任のスタッフは数名で、メンバー機関からの助成や人的支援によりプロジェクトを進行している点や、こうした国際的なプロジェクトに日本からも参加している点が印象深かった。内容はポリシー策定や体制等、戦略的な側面に関することが多いため、意思決定権者向きの内容であると感じた。

(3) 12:30 - 13:00 Closing Session - Hitotsubashi Hall

第7回総会の参加者(合計357名)等について総括がなされるとともに、大竹暁科学技術振興機構上席フェローによる閉会挨拶があった。併せて、第8回総会および第9回総会について、各国担当者より紹介があった。

・ RDA 第8回総会

場所：アメリカ コロラド州 デンバー

期間：2016年9月11～17日

テーマ：「From Big Data to Open Data: Mobilizing the Data Revolution」

(International Data Week 2016 イベントの一環として開催)

詳細：<https://rd-alliance.org/plenary-meetings/rda-eighth-plenary-meeting.html>

・ RDA 第9回総会

場所：スペイン バルセロナ

期間：2017年4月5～7日

詳細：<https://rd-alliance.org/plenary-meetings/rda-ninth-plenary-meeting.html>

以 上