




ICSU
WORLD DATA SYSTEM

国際シンポジウム「ビッグデータ時代の図書館の挑戦—研究データの保存と共有」
2014年2月5日

研究データをめぐる世界的な動向： WDS,RDA,データパブリケーション



村山泰啓

(独)情報通信研究機構 統合データシステム研究開発室長
京大大学生存圏研究所・客員教授
Ex Officio Member, ICSU-WDS Scientific Committee
日本学術会議 WDS小委員会幹事

International Programme Office Hosted by 
Based in Tokyo, Japan

目次

- **オープン・データ：現状の議論**
 - 政府の動き、アカデミーの意識
- **学術データ事業 ICSU-WDS**
 - アカデミー側、科学情報資産の長期保持・利用
- **データ・コンソーシアムRDA**
 - G8+O6国政府の議論など。科学者 & IT研究者
- **いま進められていること**
 - 例：データパブリケーション、データサイテーション



2

“Big Data” and US President Obama



Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE: ANNOUNCES \$200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some of the Nation’s most pressing challenges.

To launch the initiative, six Federal departments and agencies today announced more than \$200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.

データを めぐる動向

G8首脳会議で議論されるオープンデータと透明性 オープンガバメントと日本の動向

東京大学政策ビジョン研究センター 特任研究員
佐々木 一

Facebook 45 | Twitter 7 | 40

G8サミットにおける「オープンデータ憲章」

2013年6月に英国で開催されたG8サミットにおいて、首脳宣言が明記され、具体的な取り組み内容などについて「オープンデータ憲章」が発表された。

原則1：原則としてのオープンデータ
データによっては、公表出来ないという合理的な理由があることを認識しつつ、府のデータすべてが、原則として公表されるという期待を醸成する。

原則2：質と量
時宜を得た、包括的かつ正確な質の高いオープンデータを公表する。データの情報は、多言語に訳される必要はないが、平易かつ明確な言語で記述；データが、読みや解析の限界など、その特性がわかるように説明される可能な限り早急に公表する。

原則3：すべての者が利用できる
幅広い用途のために、誰もが入手可能なオープンな形式でデータを公表する。可能な限り多くのデータを公表する。

原則4：ガバナンス改善のためのデータの公表
オープンデータの恩恵を世界中の誰もが享受出来るように、技術的専門性や経データの収集、基盤及び公表プロセスに関して透明性を確保する。

原則5：イノベーションのためのデータの公表
オープンデータ・リテラシーを高め、オープンデータに携わる人々を育成する。将来世代のデータイノベーターの能力を強化する。

スケジュール（共同アクション）

2013.10 国別行動計画を作成し公表



(林 雅之、オープンデータ概論 ver.2、2013、
<http://www.slideshare.net/mhayasi/20130805?ref=http://blogs.itmedia.co.jp/business20/2013/08/slideshare192-8e85.html>)



Open (Research) Data

G8 Science Ministers Statement London UK, 12 June 2013

Introduction

We, the G8 Science Ministers met in London on Wednesday 12 June with Presidents of our respective national science academies, as part of the UK's G8 Presidency. At this unique meeting we discussed how our nations could lead efforts to improve the transparency, coherence and coordination of the global scientific research enterprise in order to address global challenges and maximise the social and economic benefits of research.

1. Global Challenges
2. Global Research Infrastructure
3. Open Scientific Research Data
Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of
4. Expanding Access to Scientific Research Results

Report:
www.royalsociety.org

Copies on table outside

英国王立協会がWGをつくり、
調査・出版：
「公開事業としての科学」



Science as an
open enterprise

June 2012

THE
ROYAL
SOCIETY

[Geoffrey Boulton, CODATA conf., 2012]

6

科学的方法論と科学

- 科学とは。
科学的発見・理解(原著論文)の根拠とは。
 - 実験方法の記録、第3者による再現実験
 - コミュニティ(学会等)による自由な相互批判、検証
- 「科学」という「制度」:
 - 情報がオープンで共有されることが必須だった。
→ 検証、再現性の担保(「科学」という方法論の信頼性)
- 科学的発見(原著論文)
 - 論文の固定、評価、公表、保存、引用、再利用
 - 「論文出版」⇒研究コミュニティ、社会で知を共有。

7

A crisis of replicability?

NATURE | VOL 483 | 29 MARCH 2012

REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

*Source of citations: Google Scholar, May 2011.

科学研究成果の「再現性」の危機？
重要な研究は、2次的な論文を多く生産する。
←論文の結果の再現性があってもなくても、根拠が疑われない事例。(！)



[Geoffrey Boulton, CODATA conf., 2012]

THE ROYAL SOCIETY

科学研究成果の 再現性担保における問題の例

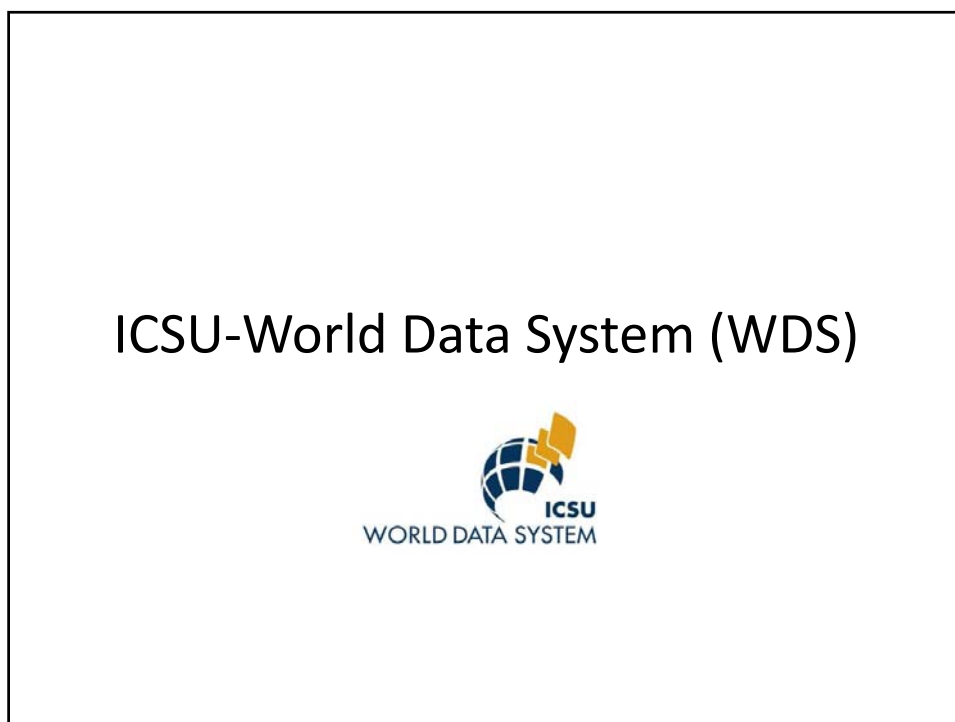
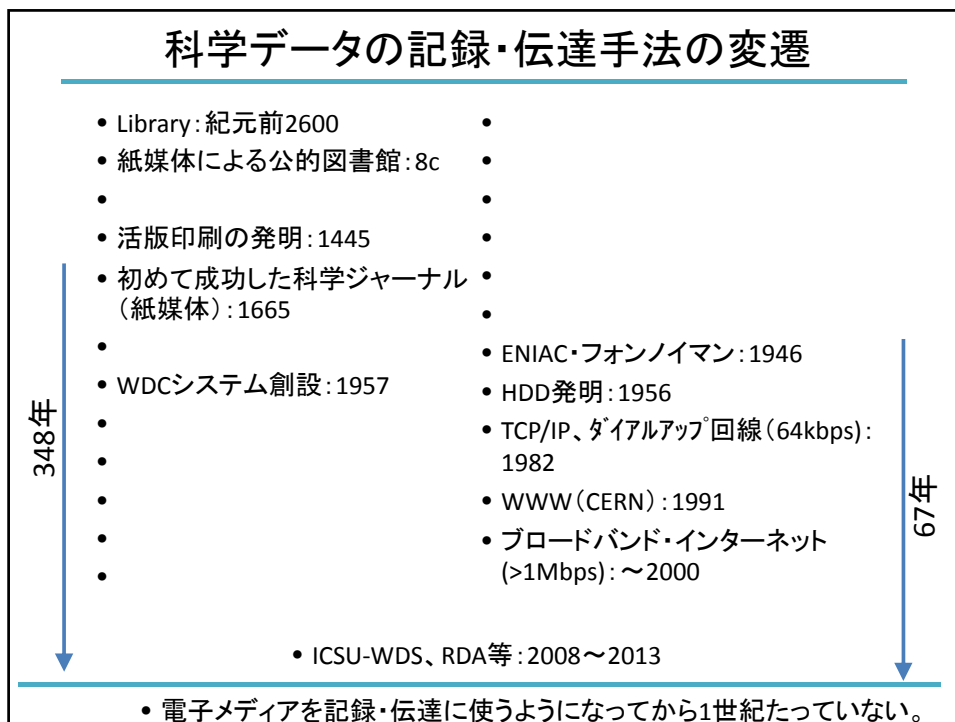
- 論文から得られる情報が不十分
 - 記述が不十分
 - 研究基盤としての情報が不十分
- 再現できない事象の検証
 - 例: 気候変動、巨大地震...

9

科学的方法論とデータの問題






- 科学的発見・理解(原著論文)の根拠の考察例
 - 再現性高: 実験方法の記録、第3者による再現実験
 - 再現性低: 一過性の現象記録(データ)の再確認(地球科学[気候変動・地震etc.]、生命科学etc.)
- 根拠となるデータの固定、評価、公表、保存、引用、再利用
 - 科学知の基礎として共有しなくてよいのか。
 - ⇔論文の固定、評価、公表、保存、引用、再利用
 - 「データ・パブリケーション」という概念が成立するか？
- 電子的な文献・データの保存・公開・配信の課題

10



ICSU-World Data System事業の創設

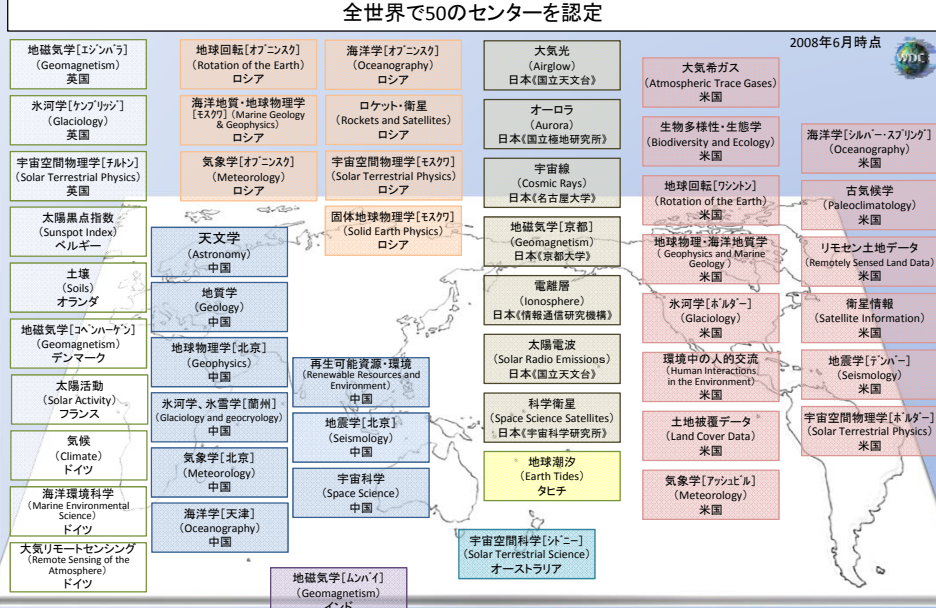
(参考: ICSU(International Council for Science; 国際科学会議)。1931年創設。科学界の国連、世界代表とも。)

<p>これまで (1950年代～)</p>	 	<p>WDC (World Data Center) ・(当時は紙・フィルム記録が主体) 各国の機関が科学データを保管する態勢</p> <p>FAGS (Federation of Astronomical and Geophysical Data Analysis Services) ・天文、地球物理データの解析サービス組織</p> <p>WDC(全世界で50センター、日本では7センターが認定)</p>	
<p>2006年頃～</p>	<p>再編</p>	<ul style="list-style-type: none"> ●最新のIT技術への対応の遅れ ●分野横断型のデータ利用への対応の遅れ ●分野の偏り ●等の問題点を検討 	
<p>現在 (2008～)</p>		<p>2008年10月28日 ICSU 第29回総会(※)にて WDS創設</p> <p>※モザンビーク共和国 首都マプトで開催</p>	

WDSに至る経緯 国際認定されたWDC(2006年時点)

全世界で50のセンターを認定

2008年6月時点



<p>地磁気学[エンババ] (Geomagnetism) 英国</p> <p>氷河学[ランブッジン] (Glaciology) 英国</p> <p>宇宙空間物理学[オホムン] (Solar Terrestrial Physics) 英国</p> <p>太陽黒点指数 (Sunspot Index) ベルギー</p> <p>土壌 (Soils) オランダ</p> <p>地磁気学[ノルウェー] (Geomagnetism) デンマーク</p> <p>太陽活動 (Solar Activity) フランス</p> <p>気候 (Climate) ドイツ</p> <p>海洋環境科学 (Marine Environmental Science) ドイツ</p> <p>大気リモートセンシング (Remote Sensing of the Atmosphere) ドイツ</p>	<p>地球回転[オプニスク] (Rotation of the Earth) ロシア</p> <p>海洋地質・地球物理学 [モスクワ] (Marine Geology & Geophysics) ロシア</p> <p>気象学[オプニスク] (Meteorology) ロシア</p> <p>天文学 (Astronomy) 中国</p> <p>地質学 (Geology) 中国</p> <p>地球物理学[北京] (Geophysics) 中国</p> <p>氷河学、氷雪学[蘭州] (Glaciology and geocryology) 中国</p> <p>気象学[北京] (Meteorology) 中国</p> <p>海洋学[天津] (Oceanography) 中国</p> <p>地磁気学[ムンバイ] (Geomagnetism) インド</p>	<p>海洋学[オプニスク] (Oceanography) ロシア</p> <p>ロケット・衛星 (Rockets and Satellites) ロシア</p> <p>宇宙空間物理学[モスクワ] (Solar Terrestrial Physics) ロシア</p> <p>固体地球物理学[モスクワ] (Solid Earth Physics) ロシア</p> <p>再生可能資源・環境 (Renewable Resources and Environment) 中国</p> <p>地震学[北京] (Seismology) 中国</p> <p>宇宙科学 (Space Science) 中国</p>	<p>大気光 (Airglow) 日本(国立天文台)</p> <p>オーロラ (Aurora) 日本(国立極地研究所)</p> <p>宇宙線 (Cosmic Rays) 日本(名古屋大学)</p> <p>地磁気学[京都] (Geomagnetism) 日本(京都大学)</p> <p>電離層 (Ionosphere) 日本(情報通信研究機構)</p> <p>太陽電波 (Solar Radio Emissions) 日本(国立天文台)</p> <p>科学衛星 (Space Science Satellites) 日本(宇宙科学研究所)</p> <p>地球潮汐 (Earth Tides) タヒチ</p> <p>宇宙空間科学[ドニー] (Solar Terrestrial Science) オーストラリア</p>	<p>大気希ガス (Atmospheric Trace Gases) 米国</p> <p>生物多様性・生態学 (Biodiversity and Ecology) 米国</p> <p>地球回転[ワシントン] (Rotation of the Earth) 米国</p> <p>地球物理・海洋地質学 (Geophysics and Marine Geology) 米国</p> <p>氷河学[ボルダー] (Glaciology) 米国</p> <p>環境中の人的交流 (Human Interactions in the Environment) 米国</p> <p>土地被覆データ (Land Cover Data) 米国</p> <p>気象学[アッシュビル] (Meteorology) 米国</p>	<p>海洋学[シルバースプリング] (Oceanography) 米国</p> <p>古気候学 (Paleoclimatology) 米国</p> <p>リモセン土地データ (Remotely Sensed Land Data) 米国</p> <p>衛星情報 (Satellite Information) 米国</p> <p>地震学[デンバー] (Seismology) 米国</p> <p>宇宙空間物理学[ボルダー] (Solar Terrestrial Physics) 米国</p>
---	---	--	---	---	---

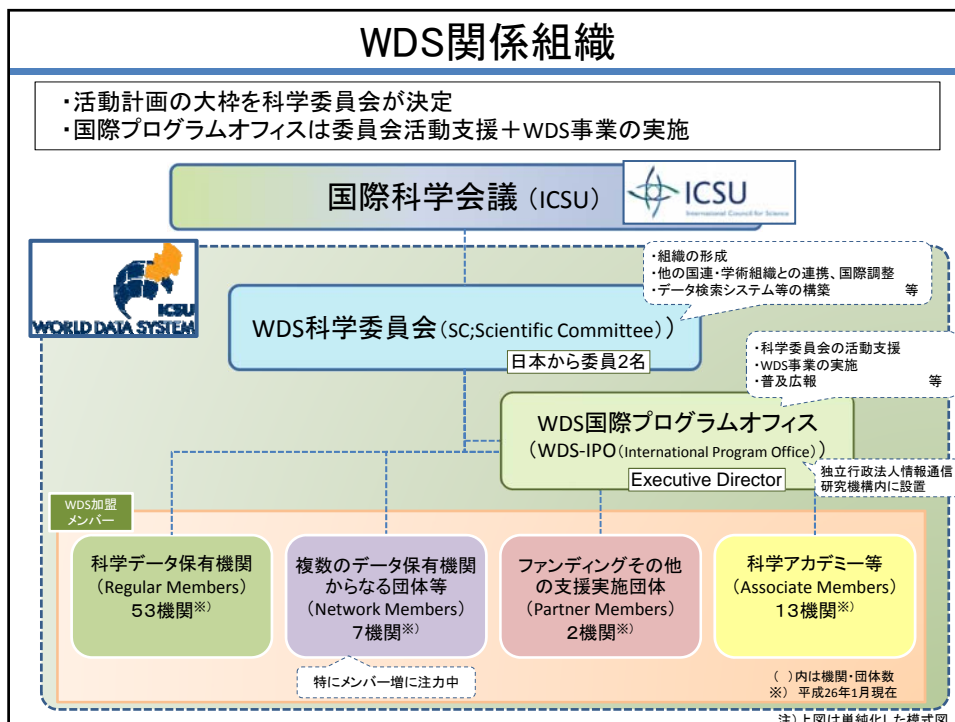
Strengthening international science for the benefit of society

WDS Objectives

30th ICSU Rome Sept 24 - Oct 01 2011
General Assembly
Associated Meetings and Related Events

- Enable **universal and equitable access** to quality-assured scientific data, data services, products and information;
(一般公開するポリシー；品質評価されたデータ)
- Ensure **long term data stewardship**;
(長期的なデータ保全)
- Foster compliance to agreed-upon **data standards and conventions**;
(標準的、共通形式の利用を促進)
- Provide **mechanisms to facilitate and improve access** to data and data products.
(データのアクセシビリティの改善)

[B. Minster, 2011] 15



WDS-IPO Opening Ceremony at Tokyo

Venue: Tokyo, Japan Date: 9 May 2012



Speeches also given by: ICSU Executive Director Dr. Steven Wilson, WDS-IPO Executive Director Dr. Mustapha Mokrane, WDS-SC chair Prof. J-B. Minster, NICT Dr. Y. Murayama

17

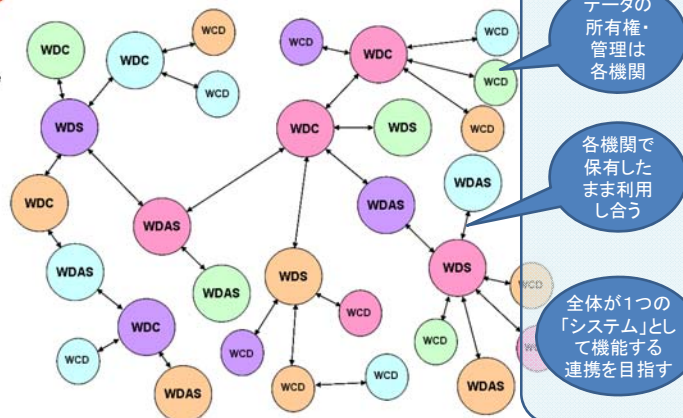
17

“System of data systems”

Parallel concept to GEOSS

WDC = World Data Center
 WDS = World Data Service
 WDAS = World Data Analysis Service
 WCD = World Center for Data

- = International
- = National
- = Regional
- = Disciplinary
- = Interdisciplinary



Illustrative example of the interoperability between the components of a GDSS

[WDS-scientific committee, 2011]

Practical Steps to Go 今後の重点課題

Working Groups has been Established.
WGを設置して、下記課題を議論

- WDS-WG** { 1. Data Publication, Data Citation
Standards, Practices: データ出版・参照(サイテーション)、
管理運用の枠組を議論。
- 学術出版社、ライブラリとの連携:
Elsevier, Thomson-Reuters, Wiley ,Springer,...
- Data Citation TG(後述), DataCite(※¹), RDA, IODE(※²)等
他のグループとの連携
※1 科学データへのDOI(Digital Object Identifier)付与など、データサイテーションを推進する国際コンソーシアム
※2 International Oceanographic Data and Information Exchange: 国連下の海洋観測データの国際交換組織
- WDS-WG** { 2. Open Metadata Catalogue
WDS加盟機関のグローバル検索を可能に。
- RDA-WG化を検討** { 3. Scalable Knowledge Network
将来の国際データネットワークのための最適な
メタデータ利用や参照モデルを議論。

Research Data Alliance (RDA)



RDA (Research Data Alliance) について

- 研究データ共有を加速し、実験・プラクティスを実現していくコンソーシアム。
 - 2013年3月正式発足。
 - 米、欧、豪がそれぞれにローカルオフィスを設置
 - いずれ総事務局を設置したい。
 - G8・GSO (Group of Senior Officials) 下のデータWG議論が契機とのこと
- 組織モデル：IETF (Internet Engineering Task Force)の手法を、科学データに適用。
 - 実質的な国際標準・国際相互結合体制の形成を目指す。
 - 研究者・技術者による自由な参加。
 - ボラティア・合意ベースでの最善の解決策を作っていくたい。
 - ICSU、CODATA、WDS、社会科学分野などとも協力・連携（国際的に活躍する専門家はだいたい限られているが...）

[村山、恒松、2013]

RDA Plenary 2 -- September 16-18, Washington D.C. -- 3 days of Peace, Love and Data

14

■ RDA Plenary 2

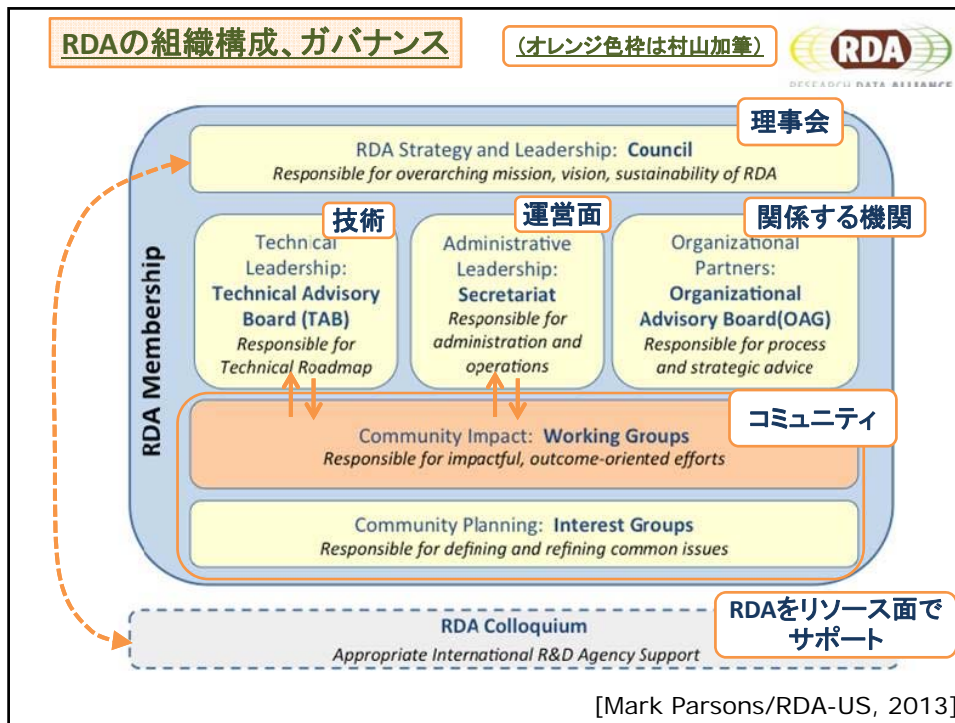
- 368 participants from 22 countries and all sectors
- All-hands stakeholder talks and RDA working meeting
- Data Citation Summit convened by DataCite, CODATA/ICSTI, Force11, ESIP, DCC, etc. to create a common agenda
- ~5000 tweets over 3 days



- しかし、日本4名、中国・台湾各4~5名
- オンライン登録者>800名
- 他国際活動との整合性に留意

Fran Berman

[Mark Parsons/RDA-US, 2013]



Groups that Met at the RDA Plenary

テーマ提案・検討中

- Birds-of-a-Feather**
 - Linked Data
 - Chemical Safety Data
 - Education and Skills Development in Data Intensive Science
 - Libraries and Research Data
 - Cloud Computing and Data Analysis Training for the Developing World
- Working Groups**
 - Data Type Registries
 - Metadata Standards
 - Practical Policy
 - Persistent Identifier Types
 - Data Foundations and Terminology
 - Data Categories and Codes
- Interest Groups**
 - Agricultural Data
 - Big Data Analytics
- Data Brokering**
- Certification of Trusted Repositories (joint with ICSU-WDS)**
- Long tail of Research Data**
- Marine Data Harmonization**
- Community Capability Model**
- Data Publishing (joint with WDS)**
- Toxicogenomics Interoperability**
- Research Data Provenance**
- Data Citation**
- Metadata**
- Economic Models and Infrastructure for Federated Materials Data Management**
- Engagement**
- Preservation e-Infrastructure**
- Legal Interoperability (joint with CODATA)**
- Global Registry of Trusted Data Repositories and Services**
- Digital Practices in History and Ethnography**
- Data Citation Harmonization Summit**
 - DataCite, FORCE 11, CODATA/ICSTI, ESIP, DCC, etc.

MISSION 遂行 (18か月)

WG検討中

BOLD = new since last Plenary

[Mark Parsons/RDA-US, 2013]

Info:
enquiries@rd-alliance.org
[@resdatall](https://twitter.com/resdatall)

Fran Berman

RDA 2nd Plenary全体の印象

•おもな参加者層:

- 各分野でデータを使う専門研究者(自然科学・医学・社会科学等)
- IT研究者、計算機科学
- 図書館やデータサービス機関の専門家
(「ライブラリ・サイエンス」が非常に重要・大きくコミット)

•IT研究・計算機科学者が(役職者に)目立つ印象。

•会議資金

- 米NSF・NIST等ファンディング機関
- マイクロソフトリサーチ(副社長トニー・ヘイ氏も出席)。



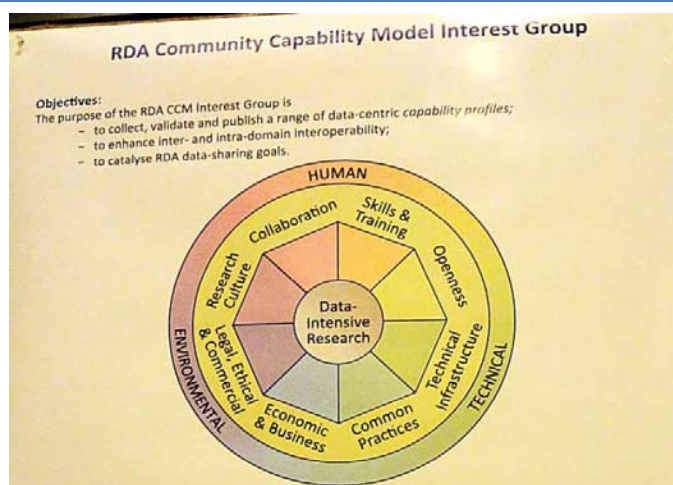
<RDA役職・関係者からの聞き取り>

•オープンデータの動向

- EC組織内: オープン・ガバメントとオープンデータ(研究データのオープン化)は異なる
- そのため、オープンガバメントとは別にRDAが別途立ち上がった理解。
- ITシステム開発でない議論の成熟が世界中の研究データのオープン化に必須という認識らしい。
 - データ形式、法・ルール・標準などの議論が先行

→参考資料: 次ページのポスター講演(マイクロソフト+バス大学)

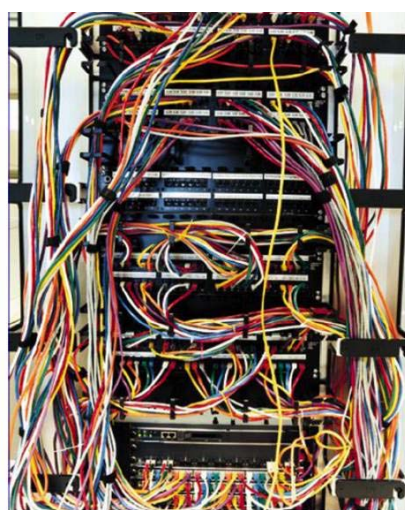
RDAでのポスターセッションから：



(マイクロソフトリサーチ、英バス大学)

- RDA (Research Data Alliance) was established March 2013.
- Promoted in G8 science minister meeting June 2013.
- Big data science or data intensive science would come with solution of the human, environmental, and technical challenges.

インターネットはつながっても、
データは...



[Mark Parsons/RDA-US, 2013]

オープンデータへの道

データ共有・公開組織の例:

- IODE(国際海洋学データ交換機構)

– 国連下。海洋データ。WDS加盟。



- IVOA(国際バーチャルオブザバトリ連盟)

– 天文データ



- ILRS(国際衛星レーザー測距データサービス)

- IGS(国際全地球測位データサービス)

– いずれもNASAがホスト。WDS加盟。



- IERS(国際地球回転・基準系事業)



- IUGONET(超高層大気長期変動の全球地上ネットワーク観測・研究)

– 超高層大気観測データ





などなど...

(和名は仮訳。)

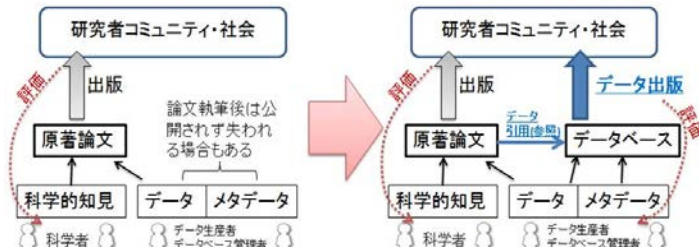
いま進められていること

“data publication”、“data citation”とは何か？

- データパブリケーション
 - データを「出版」する仕組み：
 - 課題：データの「査読」「固定」「公表」等をどうするか。
 - 課題：ID標準化、引用ルール確立、評価手法など国際団体等で模索中
- データサイテーション
 - データを文献のように「引用」「参照」する仕組み
 - 課題：ID標準化、引用ルール確立、評価手法など国際団体等で模索中
- データを出版・引用・参照すると
 - 論文・書籍と同様、知的生産力の基準に。→ 研究職・教育職の業績評価。
 - 信頼できるデータ生成・提供は現代では科学者の仕事。← 評価

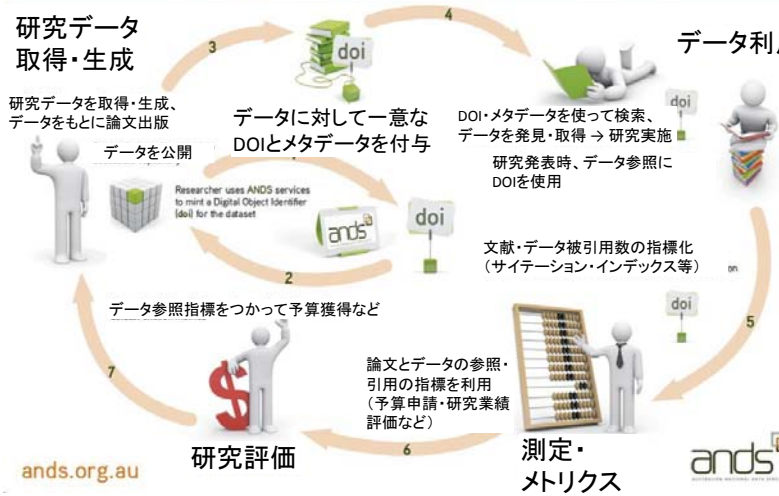
従来の仕組み ← データ出版を含む新しい仕組み



[地球電磁気・地球惑星圏学会, 2013] 31

国立機関によるデータサイテーション事業推進(豪)

Building a Culture of Data Citation



[和訳は暫定版]

Australian National Data Services
<http://www.ands.org.au/cite-data/index.html>

32

まとめにあたって

- オープンデータ(の科学面)とは
 - 研究データ・研究データの共有と、社会の利益・信頼
 - 研究者間のオープンな相互批判・検証も「科学」の一部。
 - データの保持、利用態勢は、今後どうするのか／どうなるのか。
⇔文献の保持・利用態勢(図書館、出版):
 - オープンにしやすいデータ・しにくいデータ
- ICSU-WDS (World Data System)
 - 国際的なアカデミーが主体となり、科学データの相互利用、長期保存を推進。
 - 過去50年以上の世界的なデータ保存体制の基盤
- RDA (Research Data Alliance)
 - データを相互利用・活用するための国際コンソーシアム
 - 米・欧・豪がアクティブ。
 - 情報・データの技術開発・インフラの視点
- 動いている・動きつつある活動
 - データサイテーション、データパブリケーション
→オープンなデータ利用の「しかけ」。大手出版社の動き。

ご清聴ありがとうございました。

