

科学技術分野のオープンデータの 国際動向

村山泰啓

（独）情報通信研究機構

ICSU-World Data System Scientific Committee (ex officio)

WDS国内委員会（日本学術会議WDS小委員会）幹事

科学技術オープンデータの背景

Background of open research data

- 科学(技術)と社会
 - 社会と科学者の関わりが問われている
- データ
 - 科学技術活動の重要な成果物
 - “1st class citizen” of scientific products
 - 公的資金研究によるデータ
 - 原則としての公開が求められている／Open as default
 - 公開・共有によるさらなる研究の加速
 - Eサイエンス、データ中心科学など

科学的方法論と科学

- 科学的方法論（「科学」という「制度」）
 - 研究の方法・過程・論理・結論等の記録、相互批評
 - 情報がオープンに共有されることが必須
 - 第3者による再検証の担保
 - 研究者コミュニティでのコンセンサス形成⇒社会との知識共有
 - これまで：ジャーナル、学界（講演会・シンポジウムetc.）
- 科学的発見（原著論文）と知の共有
 - 論文の固定、評価、公表、保存、引用、再利用

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.

結論の再検証できない論文が多数ある。
科学における「再検証可能性」の危機？ (村山)

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をどうするか
 - 例：環境、地球・宇宙、生命・生体...

科学的方法論とデータの問題

- 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」の概念・システムは成立するか？⇒「実証実験」(エルセビア、ワイリー、シュプリンガー、トムソンロイター...)
- 研究者、研究機関の活動を減退させては本末転倒
 - 公開データの範囲、猶予期間、利用条件...

科学技術情報の保存と共有

History: scientific record & communication

Public library (paper media) : 8c



Printing press/Gutenberg: 1445

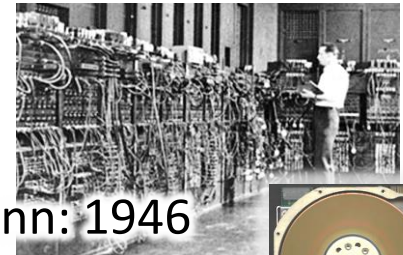


First scientific journal: 1665

Intl. Assoc. Academies: 1899

ICSU established: 1931

ENIAC, von Neumann: 1946



World Data Center system : 1957



Hard Disk Drive: 1956

TCP/IP, dial-up (64kbps): 1982

WWW (CERN): 1991

Broadband internet (>1Mbps): ~2000



349 years

印刷媒体

68 years

電子媒体



WORLD DATA SYSTEM



New global data initiatives: ICSU-WDS, RDA etc. : 2008~2013

ICSU-World Data System事業の創設

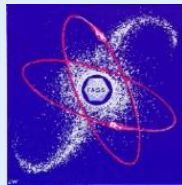
参考: ICSU (International Council for Science; 国際科学会議)。
1931年創設。科学界の国連、世界代表とも言われる。

これまで
(1950年代～)



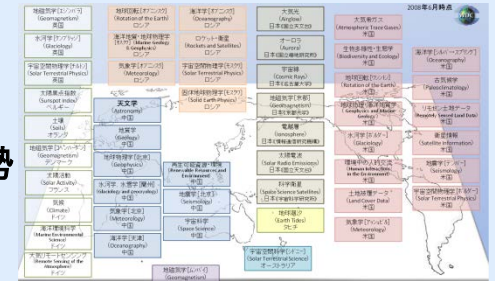
WDC (World Data Center)

- ・(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



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



WDC(全世界で50センター、
日本では7センターが認定)

2006年頃～

再編

- 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

現在
(2008～)



2008年10月28日
ICSU 第29回総会(※)にて
WDS創設

※モザンビーク共和国
首都マプトで開催



WDS (World Data System: 世界科学データシステム)

科学界の国連と呼ばれる国際科学会議(ICSU)が実施する、科学データ(ベース)に関する国際的取組の高度化を目指す国際学術事業／委員会

WDSの理念と目標

- 「品質管理された」データの「原則無償」での提供
 - － 研究に利用可能なクオリティ(精度や信頼性)の確保
- 長期的展望に立ったデータ管理体制の確保
 - － 人類資産としての貴重な科学データの長期保存
- 分散的な管理態勢の下での、共通性・相互運用性の向上
 - － 各国・各研究機関の取組をベースにした、分散型システムが基本。ただし共通性や相互運用性は追求
- 多分野横断型研究への対応
 - － より多分野に渡る横断型の研究へのニーズを反映
- 世界の特定地域に偏らないデータ活動
 - － 発展途上国等も含め、データの偏在を解消

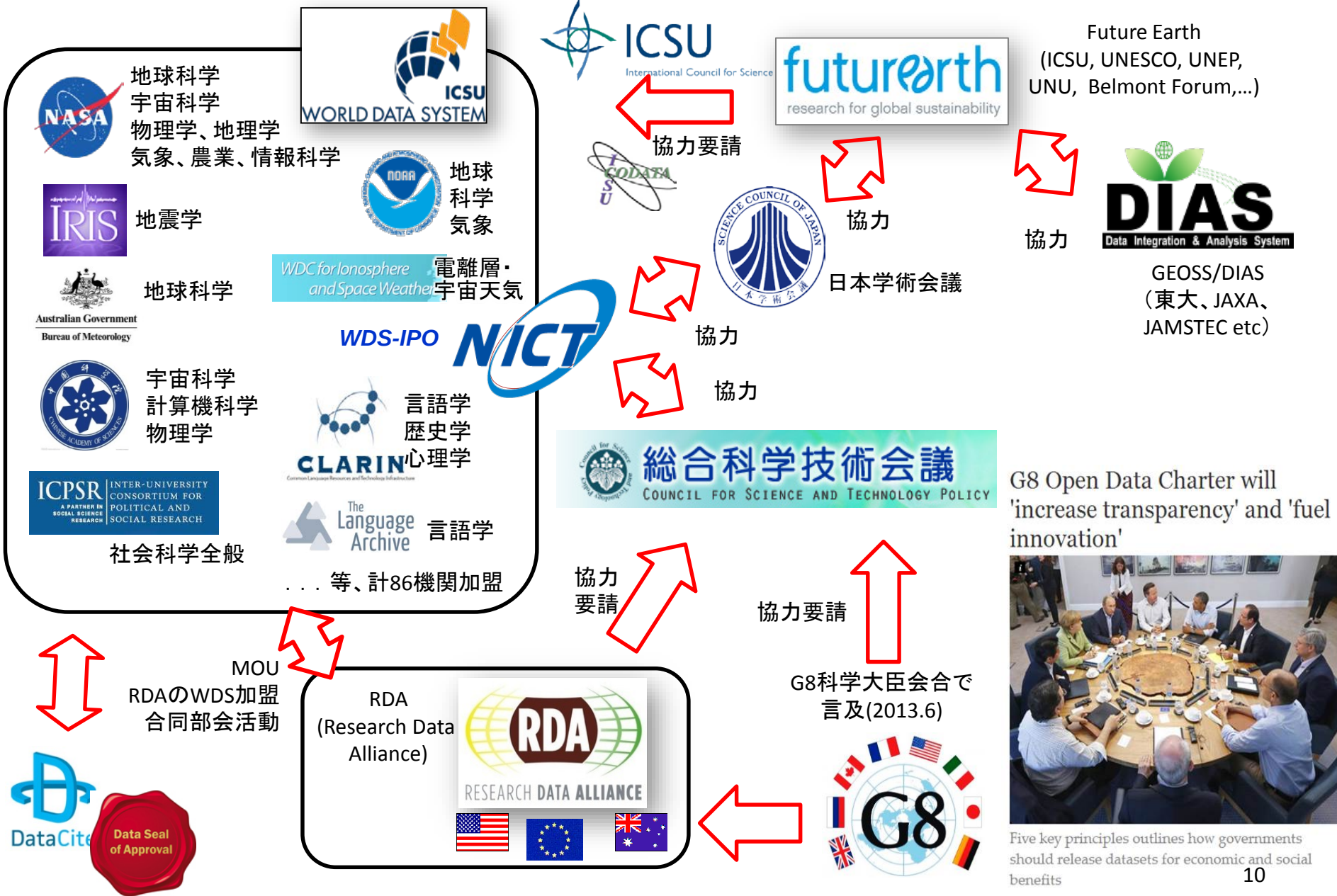


(「WDS Constitution(規約)」より)

➡ これらの理念を具体化する各国・関係研究機関の連携態勢の構築を目指す

「WDS」は、実現を目指す各国・関係研究機関の連携態勢や、同プログラムを実施するICSU内組織の名称としても使用される

日本から見た関係国際機関の概観

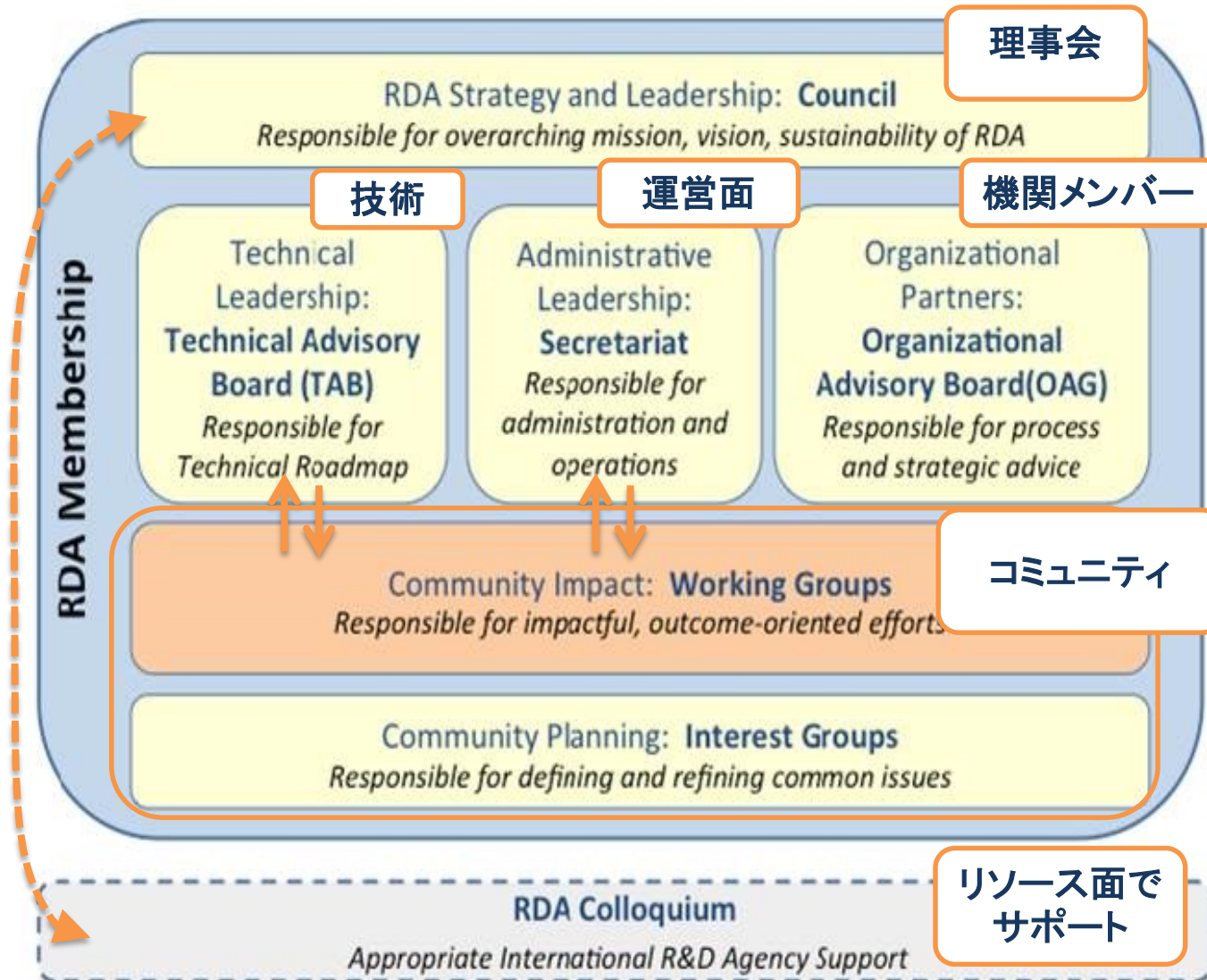


Research Data Allianceについて

- 研究データの共有を加速し、技術・プラクティス等を実現していくコンソーシアム。
 - 2013年3月発足。
 - 米、欧、豪が少額ながら資金を出しているとのこと。
 - G8・GSO (Group of Senior Officials) 下のデータWG議論が契機
 - 研究のオープン・データと、オープン・ガバメントは枠組みが異なるとの理解(ECのWG担当者による)。
- IETF* の組織モデルを、科学データに適用。
 - *)(Internet Engineering Task Force
 - 実質的な国際標準・国際相互結合理体制の形成を目指す。
 - 研究者・技術者によるボランティアベースでの合意形成
 - ICSU、WDS、CODATA、社会科学分野などとも協力。
➔ 国際的な人材基盤・ノウハウ基盤を他組織と共有して推進。

RDAの組織構成、ガバナンス

[Mark Parsons, 2013]



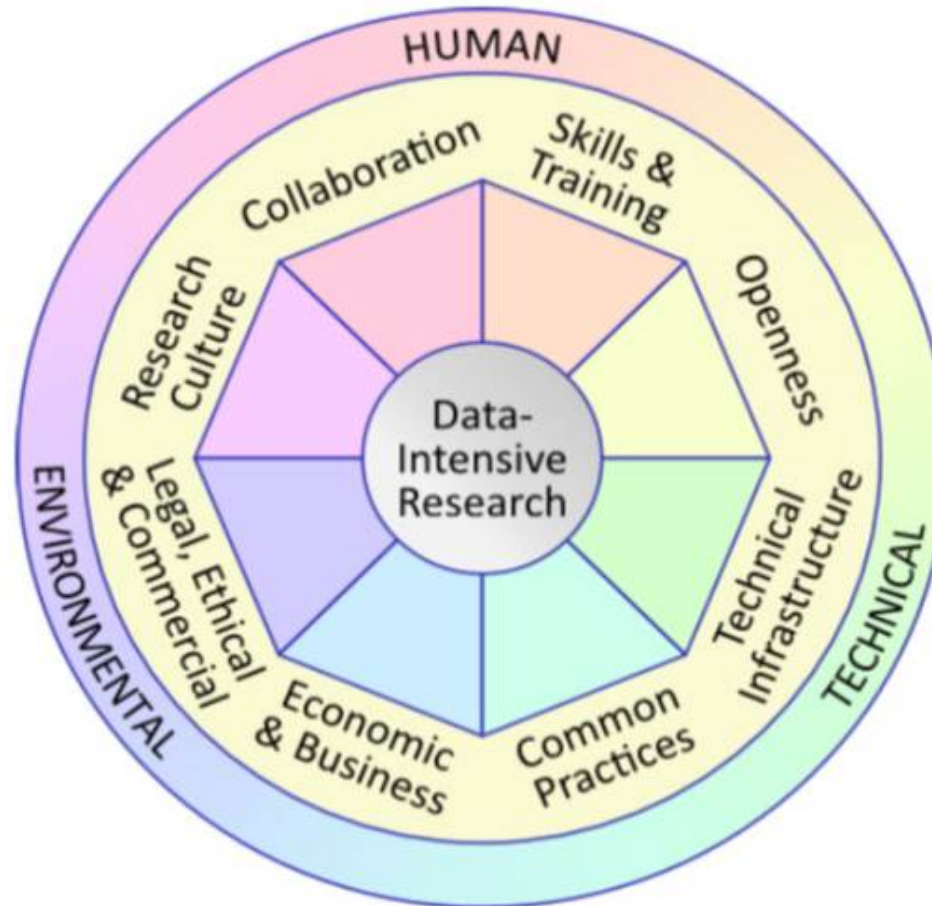
RDAと他組織のジョイントグループ

RDA/WDS: 6、RDA/CODATA: 2 (+ 審査中1)、DSA/WDS: 1 (2014年8月現在)

<u>Group Type</u>	Status		Chairs
RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World	Working Group	Under TAB review	Hugh Shanahan, Andrew Harrison, Simon Hodson
RDA/WDS Publishing Data Bibliometrics WG	Working Group	Recognised & Endorsed	Kerstin Lehnert, Sarah Callaghan
RDA/WDS Publishing Data Services WG	Working Group	Recognised & Endorsed	Hylke Koers, Adrian Burton
RDA/WDS Publishing Data Workflows WG	Working Group	Recognised & Endorsed	Sunje Dallmeier-Tiessen, Elizabeth Newbold, Theodora Bloom
Repository Audit and Certification DSA–WDS Partnership WG	Working Group	Recognised & Endorsed	Lesley Rickards, Mary Vardigan
RDA/CODATA Legal Interoperability IG	Interest Group	Recognised & Endorsed	Paul F. Uhlig, Enrique Alonso Garcia, Bob Chen
RDA/CODATA Materials Data, Infrastructure & Interoperability IG	Interest Group	Recognised & Endorsed	James A. Warren, Laura M. Bartolo
RDA/WDS Certification of Digital Repositories IG	Interest Group	Recognised & Endorsed	Micheal Diepenbroek, Ingrid Dillo, Mustapha Mokrane
RDA/WDS Publishing Data Cost Recovery for Data Centres	Interest Group	Recognised & Endorsed	Ingrid Dillo, Simon Hodson
RDA/WDS Publishing Data IG	Interest Group	Recognised & Endorsed	Micheal Diepenbroek, Eefke Smit, Jonathan Tedds, Mustapha Mokrane

RDAの取り組みから:

データ中心科学 (Data Intensive Science) と新たなイノベーションへの道のり



Microsoft
Research
Connections



https://www.rd-alliance.org/filedepot_download/383/230

- RDA Community Capability Model Interest Group
 - Secretary: Univ. of Bath & Microsoft Research Connections
- 技術的、人的、環境の課題を解決してはじめて、データ中心科学が実現する

參考資料

“Data Publication”、“Data Citation”

■ 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

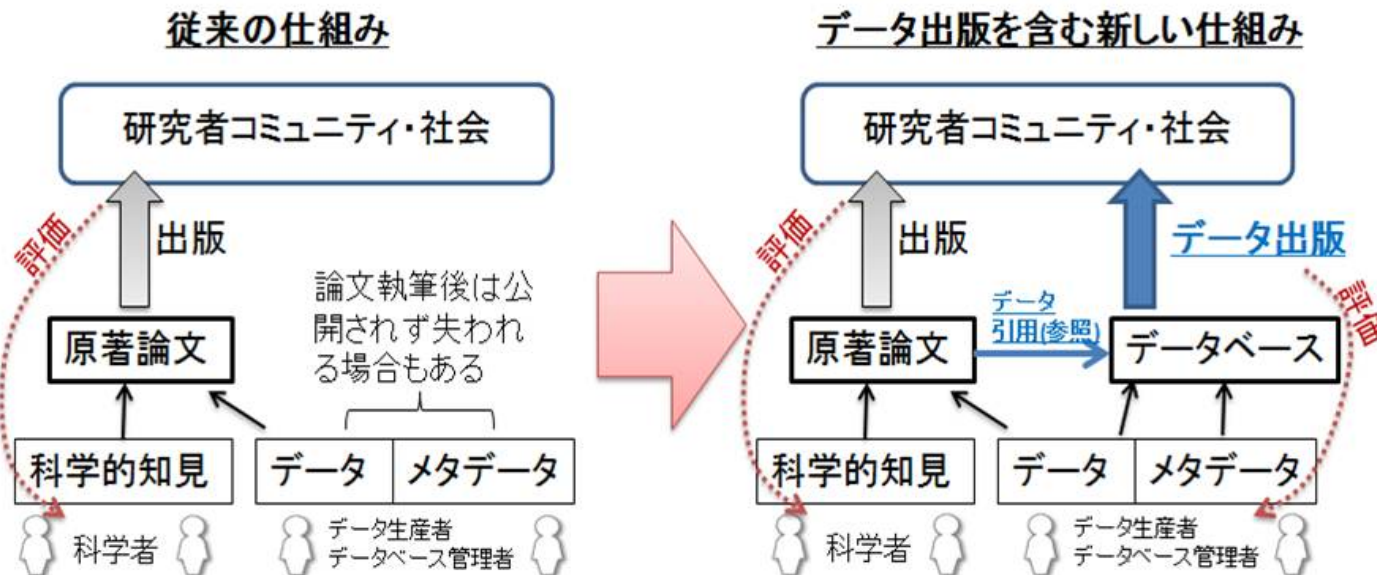


■ データサイテーション

- 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



[地球電磁気・
地球惑星圏学会,
2013]

(参考: 豪政府機関 Australian National Data Service による取組み)

○DOI (Digital Object Identifier) を論文だけでなく、データにも付与し、論文・文献で引用する取り組み
→ データ公開者・機関の活動評価 (論文と同等に) とクレジット

データ・サイテーション (データ参照・引用) 文化の形成へむけて

研究データ 取得・生成

研究データを取得・生成、
データをもとに論文出版

データを公開

データに対して
DOI付与

DOIをデータ引用に利用

DOI・メタデータを使って検索、
データ発見 → 研究実施

データ利用

研究発表時、データ参照に
DOIを使用

文献・データ被引用数の指標化
(サイテーション・インデックス等)

データ参照指標をつかって予算獲得など

論文とデータの
被引用度等で評価
(予算申請・研究業績
評価など)

業績/予算獲得

研究の評価

ands.org.au

ands
AUSTRALIAN NATIONAL DATA SERVICE

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

Changing standards and culture takes a long time.

- Metric system
 - Introduced 1795
 - Convention du Mètre, 1875
 - International System of Units (SI) 1960
 - One big holdout
- Time zones
 - 1st use of standard ("railway") time 1847
 - International Meridian Conference 1884 established GMT but did not alter local times
 - Final adoption of "standard offset" from GMT/UTC 1986
 - Current number of time zones in China and India: 1