

オープンサイエンスの最新動向と 地球環境研究、Future Earth

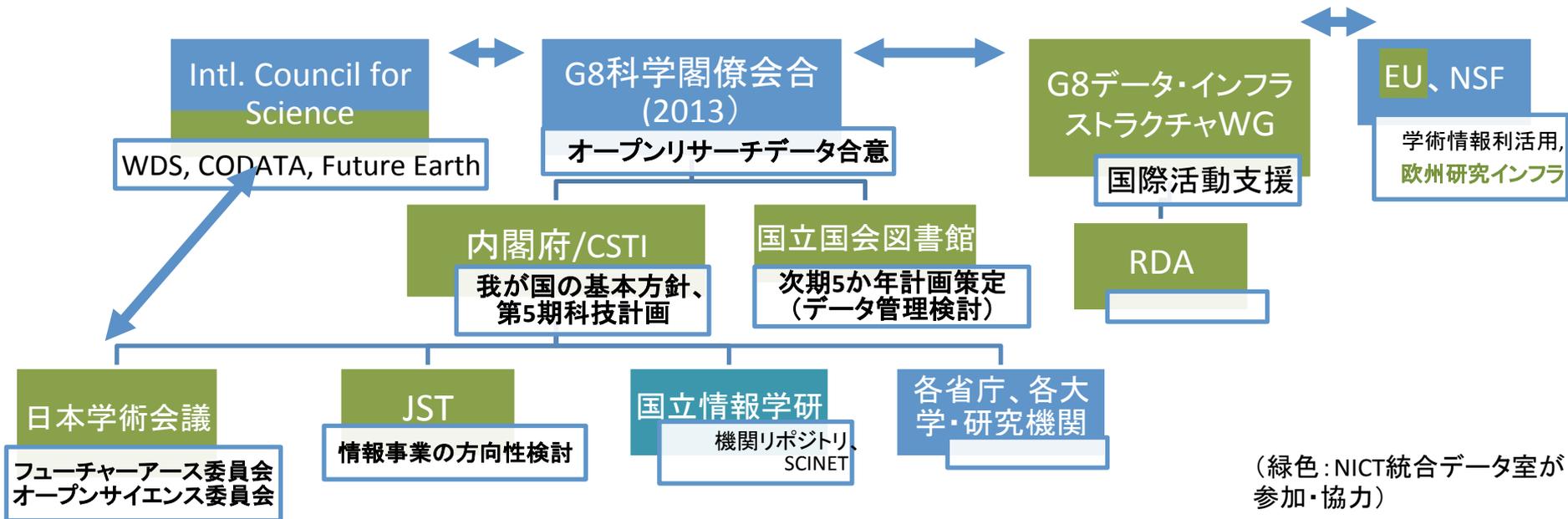
村山泰啓

(WDS科学委員会 ex officio委員／
日本学会協議特任連携会員／
情報通信研究機構)

オープンサイエンスの動向：国内の状況（H26年度～）

検討状況

－国際動向への対応、国内での研究データ基盤の検討



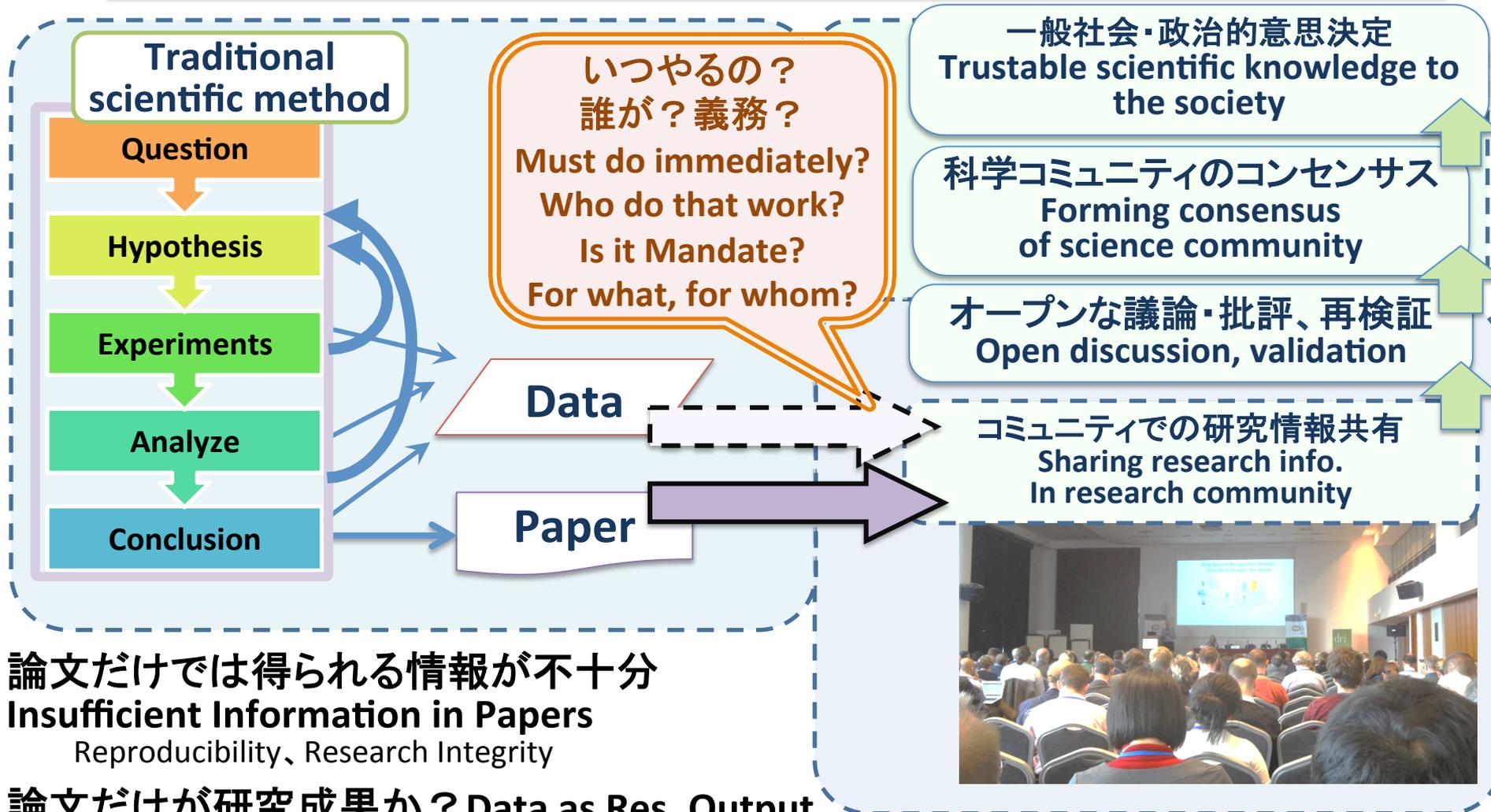
－関係する国内機関

- ・例：JST、産総研、NII、情報・システム研究機構、科学技術政策研究所、国会図書館、大学図書館など。

－国内の研究情報の公開は、OAジャーナル活動がようやく形になったところ。

－研究データの共有については今後の議論、方向付けが重要。
研究者にとってよい研究文化を醸成していく必要がある（ただ、変化には時間がかかる）。

ある類型：サイエンスのあり方と社会との関わり方



論文だけでは得られる情報が不十分
Insufficient Information in Papers
Reproducibility、Research Integrity

論文だけが研究成果か？ Data as Res. Output
Important Intellectual Asset, to be Shared with Society

再現できない事象の検証をどうするか
Research Target not Reproducible
e.g., Environment, Earth/Space, Living Organisms...

Data Management is also Question of Mutual Trust of Science and Society
データ問題⇔科学と社会の相互信頼

Scientific Data

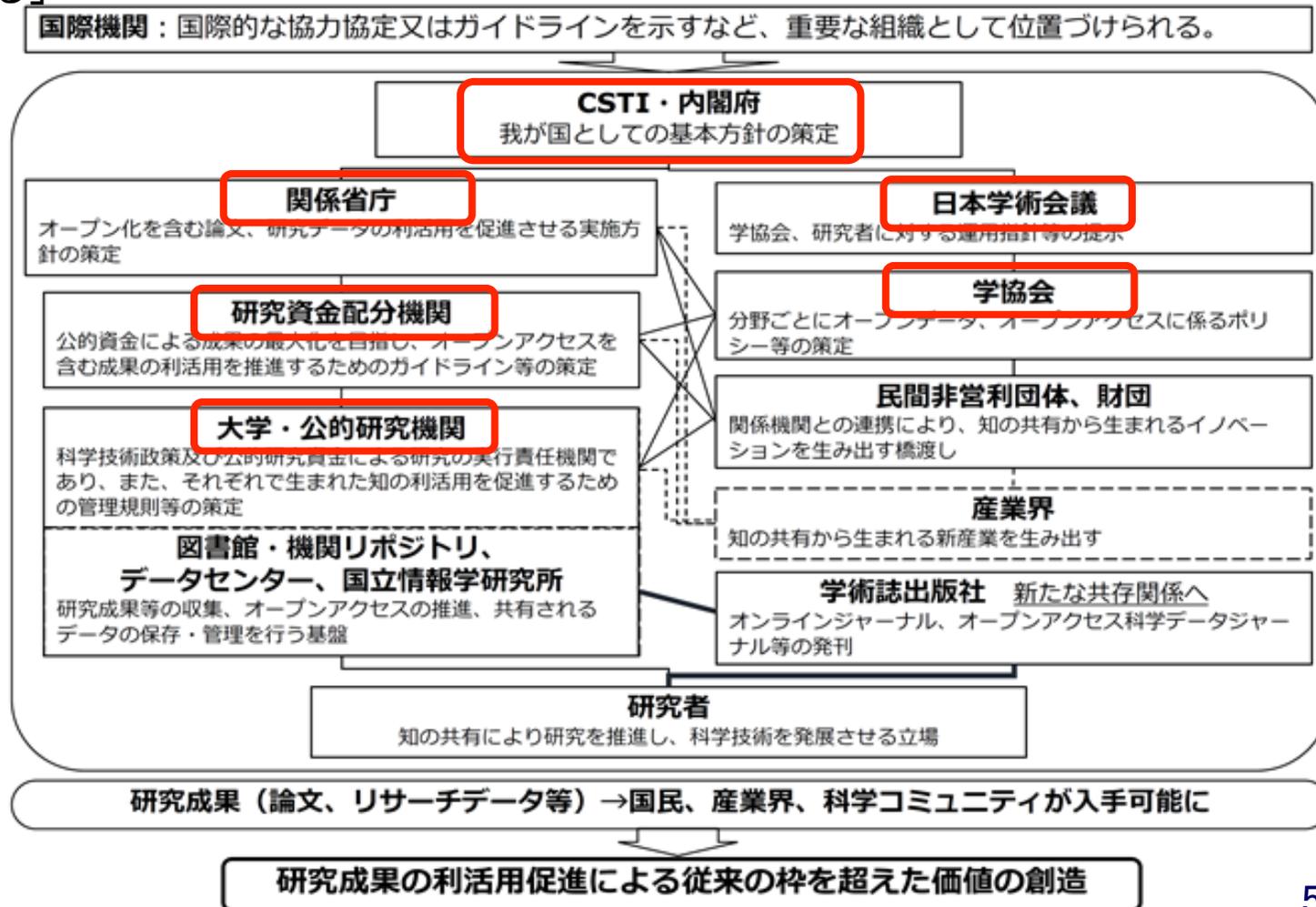
1. 研究活動の素材・イノベーション基盤としてのデータ
Data to be reused as important material or resource for scientific research
 2. 論文・研究の裏付けとしてのデータ
Data to secure reproducibility of results of scientific research
 3. 再現しないデータ(「歴史資料」と認識)
Data to study phenomena which never occur again/are difficult to be made again
(e.g. : climate change, earthquakes, ...)
- ➔ In Earth, Planetary, Environmental Science, data are often
- time series (“time-conscious”), or
 - “historical” records, or
 - record of the planet

内閣府/CSTI: 我が国の基本方針策定 (H27.3.30に報告書)

内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」('14/12~'15/3)

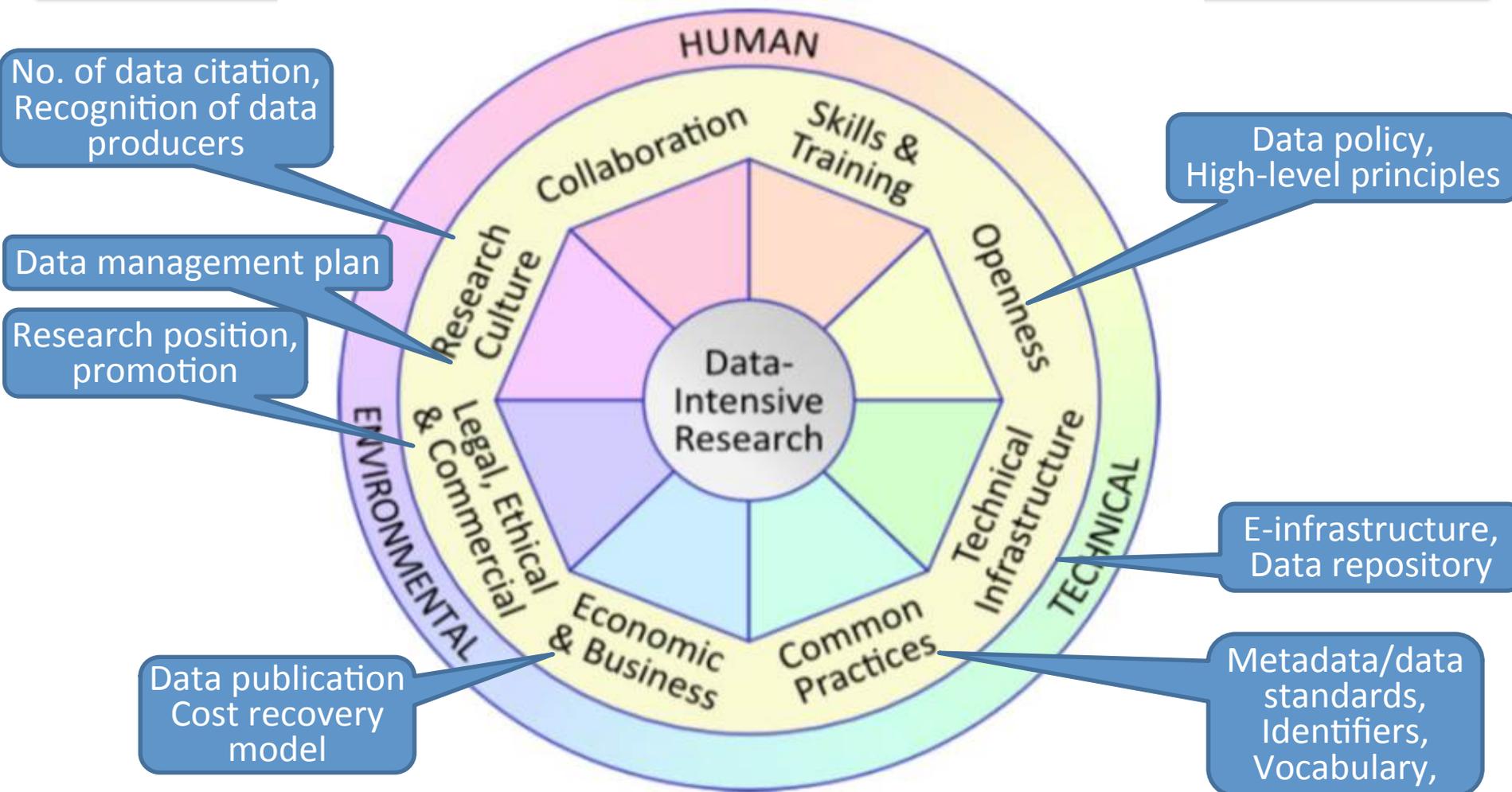
<http://www8.cao.go.jp/cstp/sonota/openscience/>

「オープンサイエンスにかかる世界的議論の動向を的確に把握した上で、我が国としての基本姿勢を明らかにするとともに、早急に講ずべき施策及び中長期的観点から講ずべき施策等を検討する」



第5期科学技術基本計画へ書き込み。

“Toward Data Intensive Science”



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

- RDA Community Capability Model Interest Group
 - Secretary: Univ. of Bath & Microsoft Research Connections
- Big data science/data intensive science become reality when the human, environmental, and technical difficulties are overcome.

Open Research Dataの現状分析

- 過去にもデータ共有は特定分野では行われている
 - 物理、化学、遺伝学、地球・宇宙科学...
 - ではなぜ、今大きな関心が寄せられるのか？ : WDS、G8、RDA、etc.
- データ共有・公開の試論
 - 従来のデータ公開類型の例 (研究分野のニーズ)
 - コミュニティ内で共有 : 分野内・方法論が共有されている場合等
 - データ利用のための付帯情報は最小限 :
コミュニティ内の規律、共有知としてデータ利用情報がある等
 - 近年のオープンリサーチデータ議論の例 (情報管理の視点?)
 - データを原著論文と同様の **研究成果物** と位置付け
 - **論文のように**、後世まで、誰がリファーマー・利用してもよい前提 (にしたい)
 - 付帯情報、データ生成に関わる情報の記録が重要
(メタデータ、ID (識別子)、
「データジャーナル」"Data Descriptor" 文献等)

[荒木徹、2015]

地球物理学の国際共同観測事業

1882.8-1883.8 第1回国際極年計画：11国 北極12点, 中緯度約30点
気象・地磁気・極光

工部省用地 (東京赤坂今井町42) で地磁気毎時観測開始

1932.8-1933.8 第2回国際極年計画：44国 110点 (半数が極地域)
気象・地磁気・極光・電離層

富士山頂気象観測所開設, 豊原(樺太)/阿蘇/仙台に地磁気観測所

1957.7-1958.12 国際地球観測年(IGY)：66国, 約4000点
気象・地磁気・極光・電離層・大気光・太陽活動・宇宙線・
ロケット・人工衛星・緯度・経度・氷河・海洋・地震・重
力・大気放射能

1957 ICSU-World Data Center発足：IGYデータの保存・利用

IUGG(国際測地学・地球物理学連合) IATME(国際地球電磁気学協会；現IAGA) 総会勧告(1939.9, Washington DC)

「海上・陸上におけるすべての磁気測量結果は2ヶ所の中央局 (Copenhagen と Washington DC) に送付、研究者への資料提供の便宜を図る」

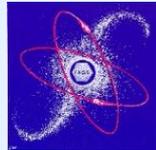
世界資料センター(WDC)の実質的スタート

ICSU-WDS (世界科学データシステム) の創設

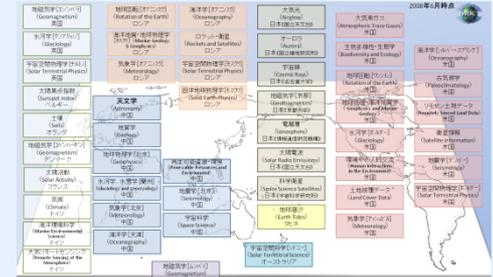
かつて
(1950年代～)



WDC (World Data Center)
・(紙やフィルム等をベースにして)
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、地球物理学データの解析サービスの提供態勢



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～)



WDS創設: 2008年10月28日
ICSU* 第29回総会にて決議



(*) ICSU: 国際科学会議。源流は1899年。1931年に現在の形となる。国連等で科学技術分野代表などを務める。国際NGOとして活動(本部:パリ)。

ICSU-WDS members (加盟機関): 合計92メンバー(2015年6月現在)。
NASA, 中国科学院、京大、バーミンガム大、国連、等の内部データ機関や、ワイリー社、エルセビア社等大手学術出版社、等が加盟している。

Examples of WDS's Implementing Actions

- ❑ WDS–RDA Data Publication Working Group
(データパブリケーションの運用、実例、計量のモデル検討)
- ❑ Global Registry for Trusted Digital Services
(信頼できるデジタルデータサービスの世界規模レジストリ)
- ❑ DSA–WDS Certification Working Group
(信頼できるデータ保存機関の認定プロセスの検討)
- ❑ WDS Metadata Catalogue, WDS Knowledge Network
(WDS加盟機関のメタ情報サービス、複合メタデータネットワーク)
- ❑ Support international initiatives (Future Earth, Urban Health, other programmes...)
(関連国際活動のデータマネジメント面で連携、支援)

など...

- ❑ ICSU requires Future Earth be supported by WDS and CODATA regarding research data management.



Data challenges facing Future Earth



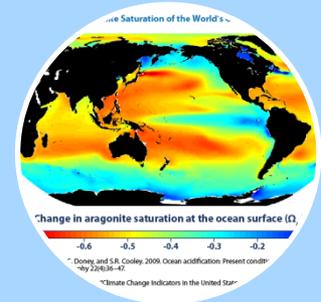
Existing
Data



Existing
Knowledge



New
Observations



Models

Integrated and interdisciplinary datasets, indicators,
visualizations, scenarios, information...

Social, economics, ecosystems, geophysics

(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



業績/予算獲得

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)



研究の評価

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

ands.org.au



(和訳は能勢(京大、2014)を参考にした)

Example of DOI-minting to Earth Science database

- Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska

First Data-DOI Registration by Japanese Platform

doi:10.17591/55838dbd6c0ad



http://www2.nict.go.jp/isd/doi-landingpage/wds/10.17591__55838dbd6c0ad.html

200601020000	74	28.3492	3.1975	-0.8260	4	4	4
200601020000		Digital data	0.3756	3	3	3	
200601020000	78	37.8977	4.9870	-0.7352	4	4	3
200601020000	80	41.1235	4.5641	-1.4586	4	4	5
200601020000	82	39.2096	4.9817	-1.7295	4	3	4
200601020000	84	44.8497	7.8027	-1.7654	5	4	5
200601020000	86	-999.0000	-999.0000	-999.0000	0	0	0
200601020000	88	-999.0000	-999.0000	-999.0000	0	0	0
200601020000	90	-999.0000	-999.0000	-999.0000	0	0	0
200601020000	92	38.2766	-1.5494	-0.7470	3	3	3
200601020000	94	-9.7626	-70.5111	1.5824	4	3	4
200601020000	96	-17.8096	-73.3329	-1.7764	4	4	5
200601020000	98	-12.8187	-77.8164	-3.9057	5	4	4
200601020000	100	-7.9532	-65.8730	-3.2022	4	5	5
200601020000	102	-8.0751	-62.0681	-3.3444	4	5	5
200601020000	104	-13.6694	-61.8343	-3.4641	5	5	5

Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska

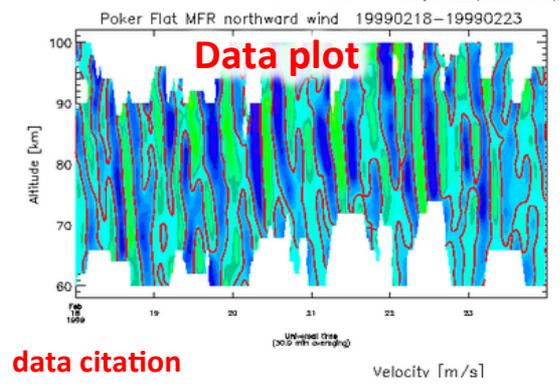
Horizontal wind velocity in the altitude range of approx. 60-90 km is observed with Poker Flat MF (medium frequency) radar, using the radar wave at 2.43 MHz. The radar receives weak radio echo signals returned from the weakly ionized atmosphere (ionospheric D-region) at the target altitudes, to deduce horizontal air motions (Murayama, Y., K. Igarashi, D. D. Rice, B. J. Watkins, R. L. Collins, K. Mizutani, Y. Saito, and S. Kainuma, Medium Frequency Radars in Japan and Alaska for Upper Atmosphere Observations, IEICE Trans., E83-B, pp.1996-2003, 2000). Poker Flat MF radar has been constructed as part of Japan-US joint research program of Arctic middle & upper atmosphere ("Alaska Project") in collaboration between National Institute of Information Technology, Japan (formerly Communications Research Laboratory), and Geophysical Institute, University of Alaska Fairbanks.

Landing Page

Data description, Data format, Link to data, etc.

Data Citation
Citation: Alaska Project of NICT (CRL)-GI/UAF, Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska, doi:10.17591/55838dbd6c0ad

General Characteristics
Parameters: Mesospheric horizontal wind velocity
Processing level:
Latitude: 65.1
Longitude: -147.5
Temporal resolution: 30 minutes



data citation
Cited by Kinoshita, T., Y. Murayama, and S. Kawamura (2015), Tidal modulations of mesospheric gravity wave kinetic energy observed with MF radar at Poker Flat Research Range, Alaska, J. Geophys. Res., 120, doi: 10.1002/2014JD022647

CODATA-ICSTI データサイテーションテストグループ 地域 (regional) ワークショップ計画

- **開催予定** : China, Taiwan, **Japan**, India, Australia, South Africa, Israel, Brazil, EU, USA (2015年後半～2016年前半)
- **趣旨** :
 - 各国・各地域において、データ引用(サイテーション)の普及・実践を進めたい。
 - “From Principles to Practice, A Focus on the Research Policy and Funding Community”
→ 科学技術政策・研究予算配分の当事者機関)との議論に焦点をあてたい
 - 国際共同議長やコアメンバーが最低1名参加すること。
- **国内取り組み** :
 - **ジャパン・データサイテーションワークショップ**
日時: 2015年10月29日午後(13時予定)
場所: 国立情報学研究所12階会議室(1208、1210室)
 - 世話人: 村山泰啓(情報通信研究機構/TGメンバー)、
安達淳(国立情報学研)、小関敏彦(東京大学、
CODATA Japan)、水野充(科学技術振興機構参事役)
 - **基調講演**:
Christine Borgmann (UCLA Distinguished Professor/
Librarian Science),



まとめ

- オープンサイエンスを進める大きな流れがある
- しかし具体化が問題である(国内外とも！)
- 研究データ、学術データを成果として扱う方向
- 「データを公表、引用、研究成果として評価」
- 理想と現実:いま何ができるのか、時代は本当に曲り角に来ているのか。

一方で:

オープンサイエンスは絶対正義か？

- 理想像と現実
- 成果があがらない取り組みはしてもしようがない？
- 「全体最適」と「局所最適」の両方が大事
- 遠い理想を心に抱きつつ、今できるわずかなことを行う
- 研究者が、自分の研究・学会・社会のためにベストだと思うやり方を探してみよう・考えておこう
- 自分たちの後輩にも「理想像」を伝えていければ。

FIN.

PLEASE KEEP IN MIND:

- **DON'T discourage/decelerate research.**
 - Any rules or regulations should not discourage scientists/researchers.
 - Which data should be open/close, embargo period, use condition, designing services for users...
- **DON'T regulate what we don't yet understand.** (“The Data Harvest”, RDA Europe, December 2014)

Joint Declaration of Data Citation Principles

(Data Citation Synthesis Group, hosted by FORCE11, March 2014)

Preamble:

(注意！村山私訳・意訳)

健全で再現可能な学術研究は、アクセス可能なしっかりしたデータが基礎になければならない。

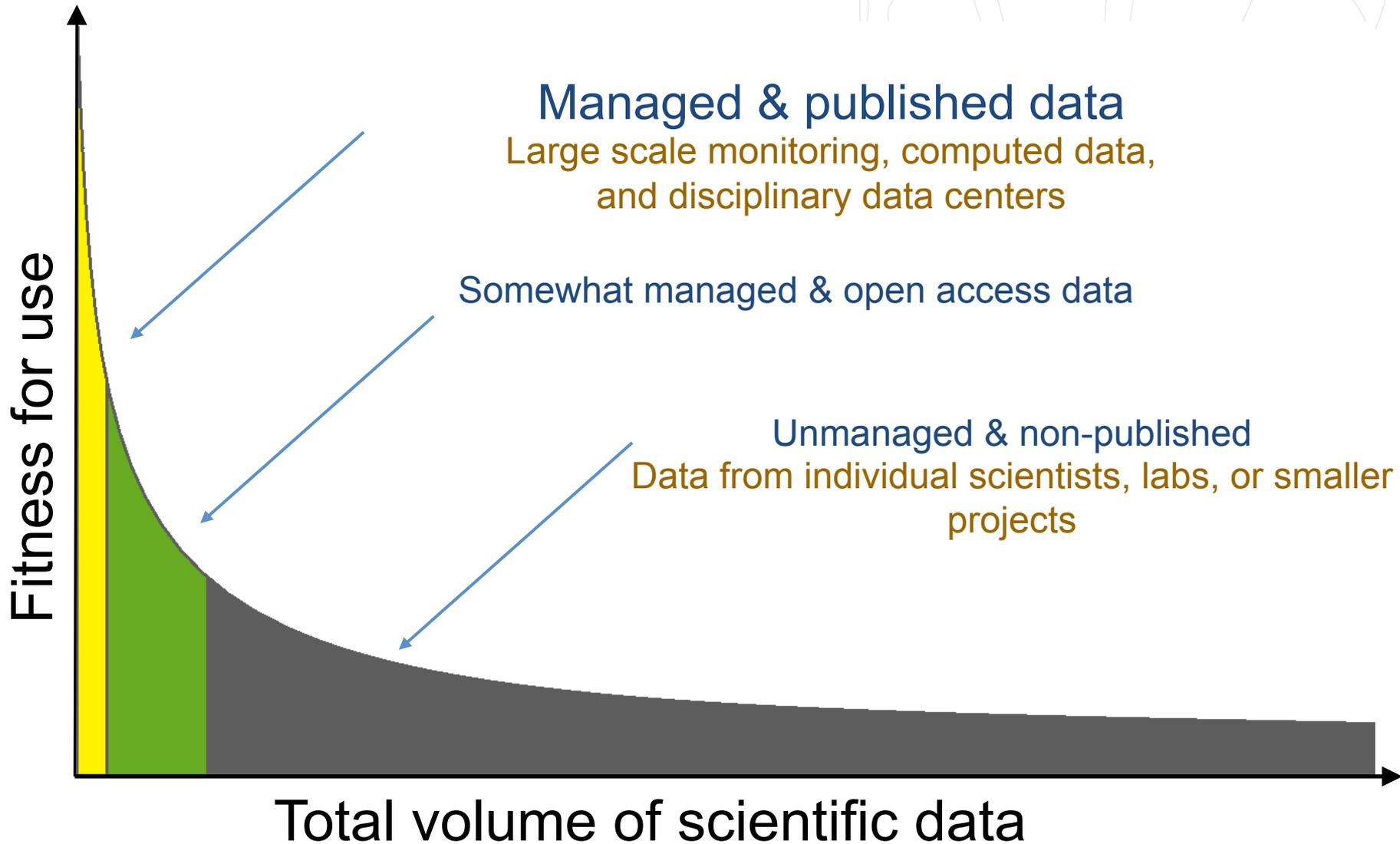
データは、引用され(評価され)るべき正当な研究プロダクトである。

データサイテーションは、研究上の他の引用と同じように行われるべきである。これはデータ再利用を支える学術的エコシステムの一部といってもよい。

私たちは、この価値ある活動を奨励し、学術文献や他のデータセットやさまざまな研究上のものにふくまれるデータについて、以下の指針を提案する。

この指針は多くの関連グループの活動を統合したものである。次のステップへ進むために、みなさんから、この指針を支持頂き、活動に参加していただくことを願っています。

The Long Tail



国内理解の推進

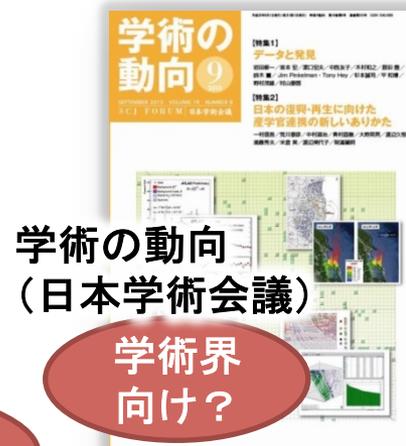


文科省「科学技術動向」誌
シリーズ記事

政治家、
行政トップ
向け



国立国会図書館月報
情報資産管理
の専門家向け



学術の動向
(日本学術会議)

学術界
向け?



SEEDer(シーダー)
(総合地球学研究所・
昭和堂) Future
Earth?

- 内閣府:各省庁検討のための「フォローアップ委員会」
- 日本学術会議:
 - オープンサイエンス検討委員会
 - フューチャーアース推進委員会:日本が国際事務局を誘致。データ活動は重点の1つ。
- 科学技術振興機構(JST):科学データの国際会議誘致
- 研究を後退させては意味がない。
→が、長い目で重要なことは進める必要がある?
- 拙速なルール化は望ましくない
→研究コミュニティが議論に参加して、よりよい仕組みに。
- 様々な“実験”:研究データDOI、データ出版、データジャーナル(Nature, Wiley, Copernicus, etc.)、国際相互運用・標準化の動き、図書館におけるデータ管理