


World Data Systemと 日本のオープンサイエンス (試論)

World Data System and Open Science in Japan (*an attempt of interpretation*)

村山泰啓 (NICT/ICSU-WDS)、渡邊堯 (WDS-IPO/NICT)

Today's topics:

- Status of WDS-SC/ WDSの近況
- Open Science Discussion in Japan/オープンサイエンスの議論の進展
- Action for “Openness”/ われわれは今後どうするべきか？

International Programme Office Hosted by 

Based in Tokyo, Japan

WDS-SCの近況

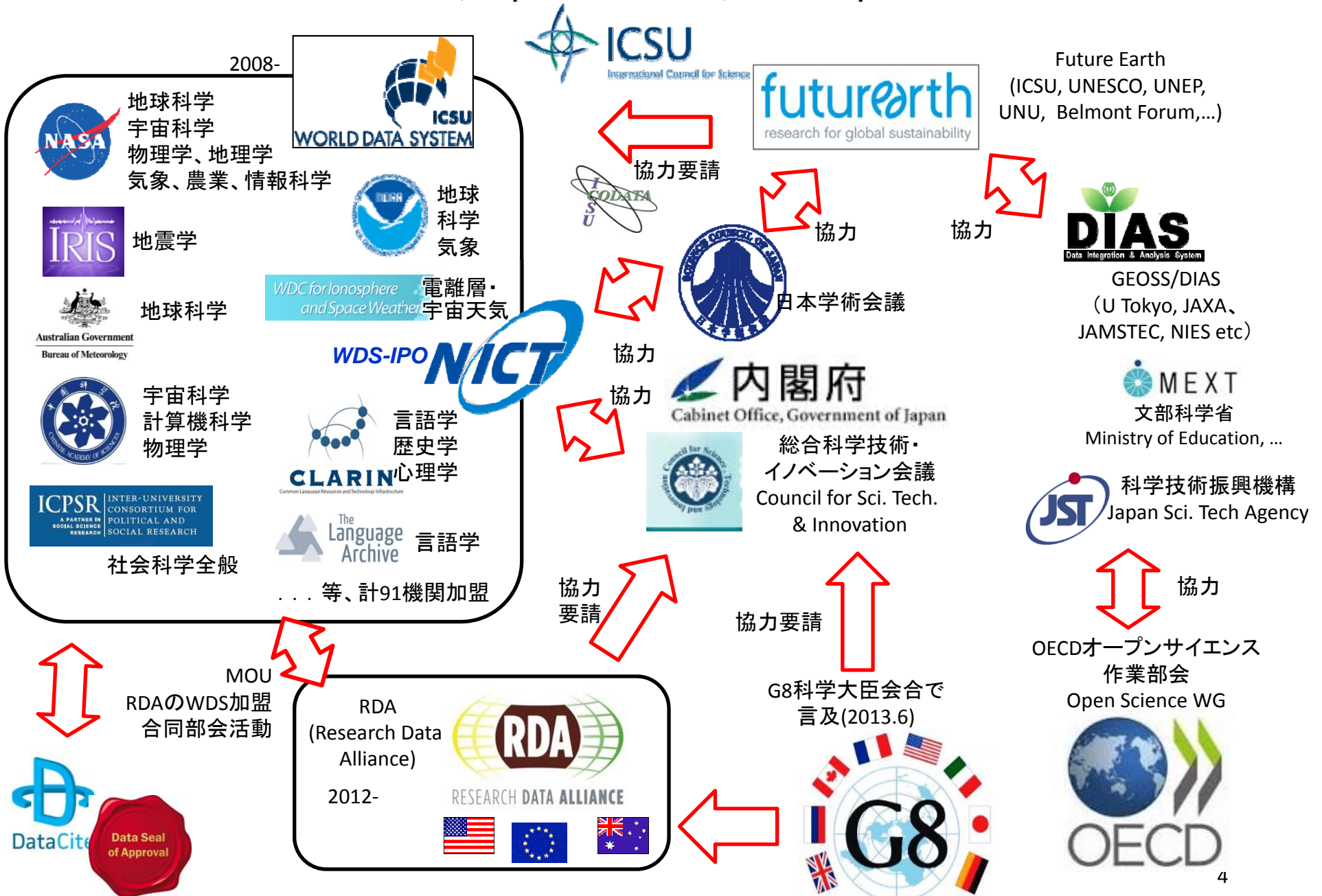
Status Report of
WDS Scientific Committee

Creation of ICSU-WDS (世界科学データシステム)



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、ワイリー社、エルセビア社、等が加盟している。

オープンサイエンスを巡る組織と日本 WDS, Open Science, and Japan



WDS-SC Members (2015-2018)

Committee Members 2015–2018



Chair: Sandy Harrison

Centre for Past Climate Change at the University of Reading, United Kingdom and Macquarie University, Australia

Sandy Harrison is a Professor of Palaeoclimates and Biogeochemical Cycles and Director of the Centre for Past Climate Change at the University of Reading in the UK, and also a Professor in Ecology and Evolution Centre for Research Excellence at Macquarie University, Australia. She studied geography at the University of Cambridge, did her Masters in Geomorphology at Macquarie University and a PhD at the University of Lund in Sweden.

WDS-Scientific Committee (2015-2018) (blue: 再任 / re-appointment from 2012-2015)

Sandy Harrison * (Chair)	U.K.	Centre for Past Climate Change at the University of Reading, U.K., and Macquarie University, Australia
Aude Chambodut	France	International Service of Geomagnetic Indices: ISGI
Arona Diedhiou	France, Côte d'Ivoire	Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE), University of Grenoble-Alpes,
Ingrid Dillo	Netherlands	Data Archiving and Networked Services (DANS)
Claudia Emerson*	Canada	Department of Philosophy, McMaster University
Elaine M. Faustman	U.S.A.	Institute for Risk Analyses and Risk Communication, School of Public Health, University of Washington
Wim Hugo*	South Africa	South African Earth Observation System (SAEOS)
Toshihiko Iyemori	Japan	WDC - Geomagnetism, Kyoto, Kyoto University
Rob Kitchin	Ireland	National Institute for Regional and Spatial Analysis, National University of Ireland Maynooth
Guoqing Li*	China	Institute of Remote Sensing and Digital Earth (RADI)
Alex de Sherbinin	U.S.A.	Socioeconomic Data and Applications Center: SEDAC
Sanna Sorvari	Finland	Finnish Meteorological Institute
Ex officio		
Heide Hackmann		ICSU
Yasuhiro Murayama		NICT

日本国内のオープンサイエンスの議論

“Open Science Discussion in Japan”

科学技術オープンデータの背景

- 科学(技術)と社会

Science/Scientist and Society

– 社会と科学者の関わりが問われている

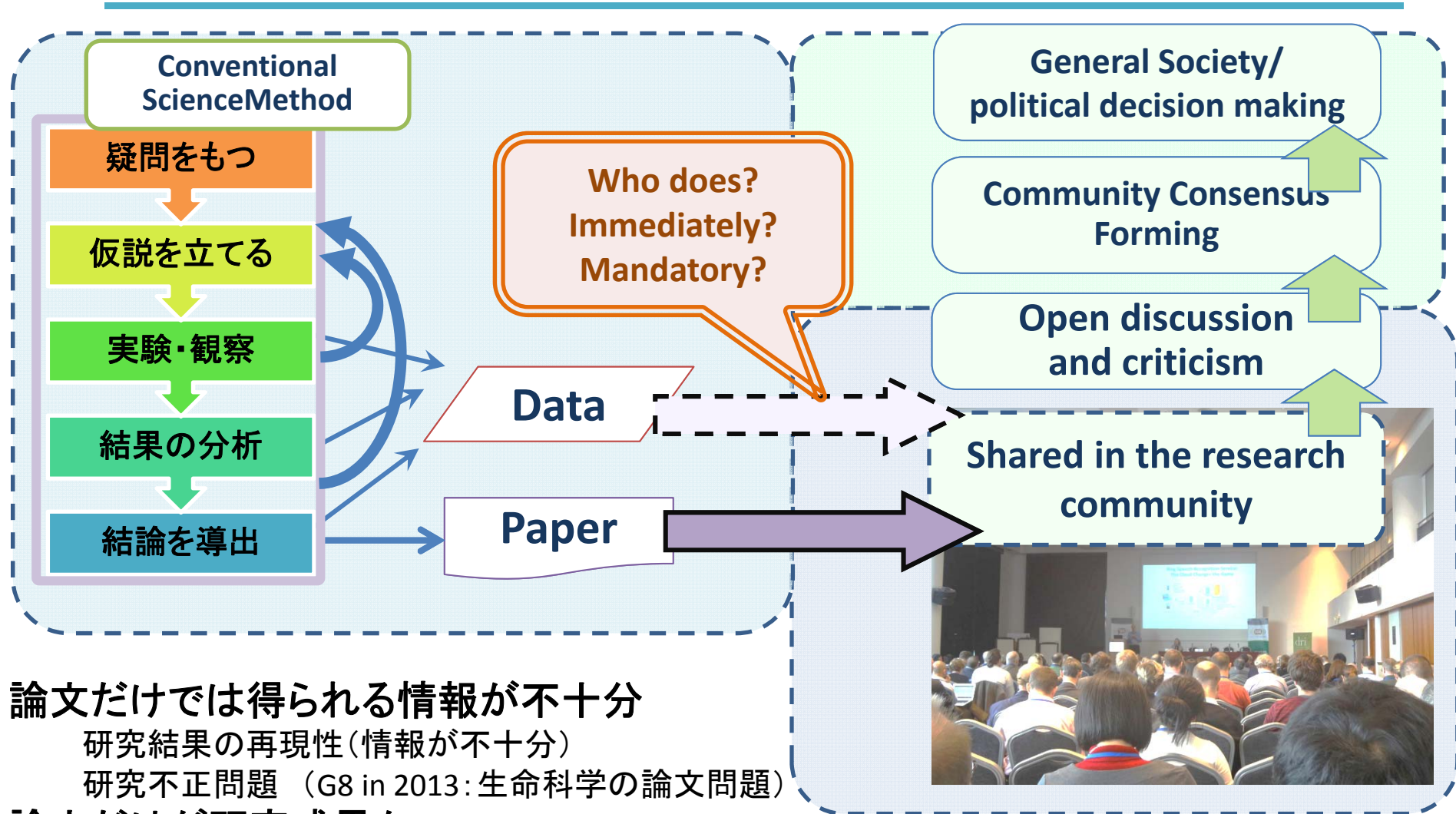
- データ Data

– 科学技術活動の重要な成果物 Output of science

– 公的資金研究によるデータの公開原則 Open as default

– さらなる研究の加速 (e-science, data intensive science...)

学術研究と論文・データを通じた社会との関わり



論文だけでは得られる情報が不十分

研究結果の再現性(情報が不十分)

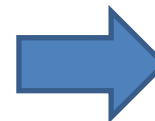
研究不正問題 (G8 in 2013: 生命科学の論文問題)

論文だけが研究成果か？

データは貴重な情報資源。社会へ発信できる成果。

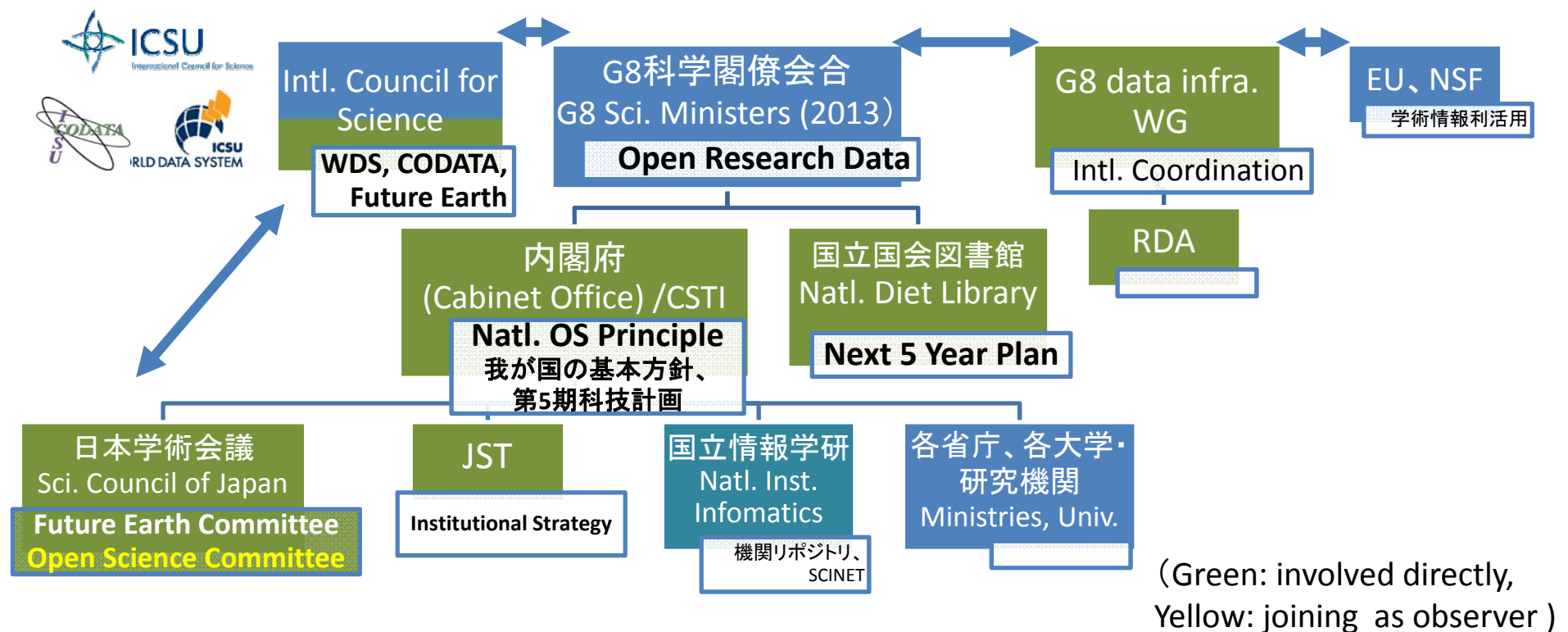
再現できない事象の検証をどうするか

例: 環境、地球・宇宙、生命・生体...



科学と社会の相互信頼、相互協力の問題でもある

“Open Science Data” Landscape Viewed from/in Japan in 2014



–関係する国内機関 Domestic Stakeholders

- 例: JST、AIST産総研、NII、ROIS情報・システム研究機構、MEXT-NISTEP / 科学技術政策研究所、NDL/国会図書館、Univ. Lib./大学図書館など。

–Promoting OA Journal: 国内的には、ようやく形が見えてきたところ。

–Res. Data Sharing: 今後の議論、方向付けが重要。Strategy, Practice? 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⇒変化には時間がかかる。

「オープン」ということについて

Action for Open Data Sharing?
(Immediate and Full “Openness”?)

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
 - 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
 - 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)

国内への国際議論の紹介



Political/
Gov.

文科省「科学技術動向」誌
シリーズ記事



Library
Scientists

国立国会図書館月報



Scholars

学術の動向
(日本学術会議)



Future
Earth?

SEEDer(シーダー)
(総合地球学研究所・
昭和堂)

□ 紹介:「国際的にはこうですよ」

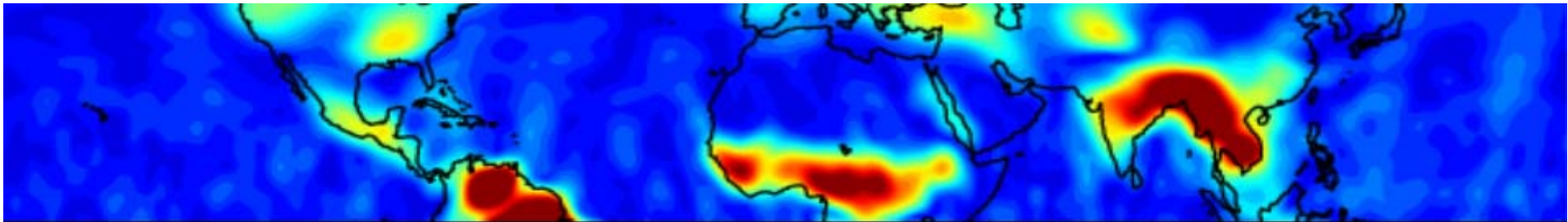
“This is the international discussion.”

➔国内では理解している人に会えなかった(2013年時点)

□ 問い:「日本はこれを受け入れますか?拒否しますか?独自に考えますか?」

(“Japan will accept it?/reject it?/create an original rule?”)

➔研究コミュニティが議論に参加して、よりよい仕組みに。



Data Sharing Principles

- ❧ Group on Earth Observations (GEO, 130+ nations) / Global Earth Observation System of Systems (GEOSS). 2010-present.
- ❧ **Equitable, unimpeded access to data for research and education**
- ❧ **Long-term data preservation**
- ❧ **Many exceptions (National security, privacy laws, commercial protection, ecological protection)**

WDS Data Policy

August 2012

The International Council for Science – World Data System (ICSU-WDS), recognizing the benefits and importance of contributing to the growing international efforts of data sharing, has adopted the following Data Policy based on the data-sharing principles of the Global Earth Observation System of Systems (GEOSS):

- **There will be full and open exchange of data, metadata and products shared within WDS, recognizing relevant international instruments and national policies and legislation;**
- **All shared data, metadata and products will be made available with minimum time delay and at minimum cost;**
- **All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.**

今後の課題の1つ: 公開に制約のあるデータについても考慮が必要
Consideration is required for restriction regarding Sensitive Data

科学が仕事としてなりたつための社会的構成要素例

研究事業

研究主体・科学研究者



研究情報の出版・保存・利用事業

出版社・出版組織



「ライブラリ」、 学術研究情報組織



NII 機関リポジトリ

学術機関リポジトリ構築連携支援事業
 NII Institutional Repositories Program

研究情報マネジメント、 インフラ構築



研究結果、データ、メタデータ、のやりとり



まとめにかえて：研究現場では今後？

- 「オープンサイエンス」の(来たるべき)功罪(?)
 - 「オープン」は絶対的善なのか
 - 制約付きのデータ共有も「あり」？ (cf.論文は過去「制約付き」)
 - 新しいスタイルのデータの扱いが重要⇒「新しいデータ共有」
- 「新しいデータ共有」に必要なものは何か
 - 面識がなくても、情報が流通すること。⇔学会・コミュニティつながり
 - データの付帯情報・説明情報(メタデータ)：分野ごと⇔分野横断
 - 50年後の(違う学会の)後輩が同じ解析ができるように。
- Important Dates
 - 29 Oct. 2015: CODATA Data Citation Workshop at NII (sponsored by CODATA-ICSTI Data Citation TG, +JpGU, +NII)
 - 29 Feb.—3 March 2015: RDA (Research Data Alliance) 6th Plenary in Tokyo, and side events:
the main host is JST! (co-sponsors will be NICT, NII, NIMS,...)



Research Data Alliance
Plenary Meeting 7
1- 3 March 2016, Tokyo, Japan

research data sharing without barriers
rd-alliance.org

シンポジウム:
科学の発展への起爆剤
～データ駆動型科学の推進に向けて～

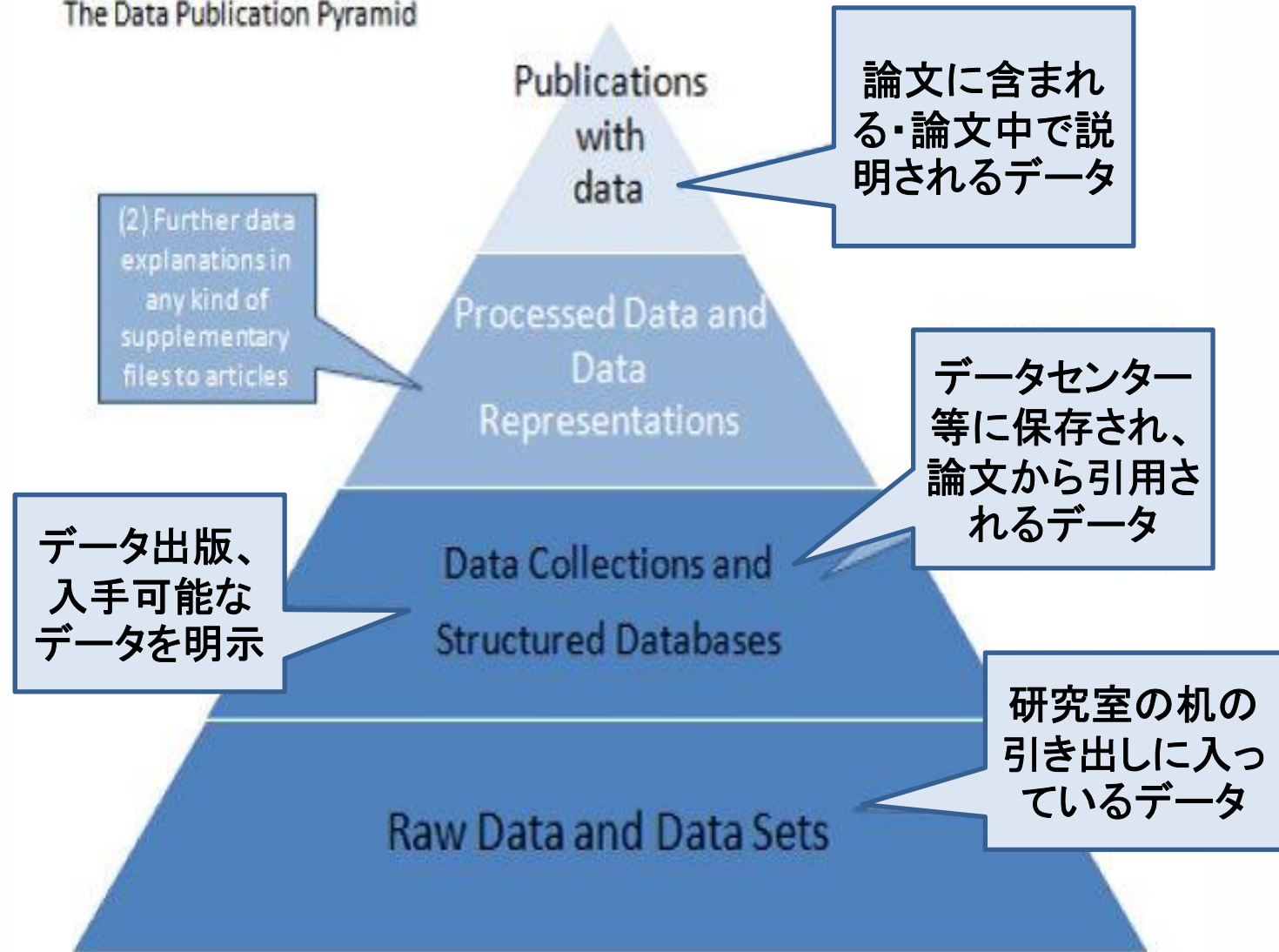
内閣府
文部科学省
国立研究開発法人 科学技術振興機構
国立研究開発法人 産業技術総合研究所
国立研究開発法人 情報通信研究機構
国立研究開発法人 物質・材料研究機構
大学共同利用機関法人 国立情報学研究所

2016年2月29日
一橋講堂

[H. Frederick Dylla, 2012]

The Data Publication Pyramid

Data Pyramid



SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

- » Palaeoclimate
- » Climate change
- » Palaeoceanography
- » Limnology

An extended Arctic proxy temperature database for the past 2,000 years

Nicholas P. McKay & Darrell S. Kaufman

Robust climate reconstructions of the most recent centuries and millennia are invaluable for placing modern warming in the context of natural variability. Here we present an extended and revised database (version 1.1) of proxy temperature records recently used to reconstruct Arctic temperatures for the past 2,000 years. The datasets are presented in a machine-readable format, and have been extended with the geochronologic data and consistently generated time-uncertain ensembles, which will be useful in future analyses of the influence of geochronologic uncertainty. A standardized description of the seasonality of the temperature response for each record, as reported by the original authors, is also included to motivate a more nuanced approach to integrating records with variable seasonal sensitivities. Despite the predominance of seasonal, rather than annual, temperature responders in the database, comparisons with the instrumental record of temperature suggest that, as a whole, the datasets best record annual temperature variability across the Arctic, especially in northeast Canada and Greenland, where the density of records is highest.

Received: 01 April 2014

Accepted: 22 July 2014

Published: 19 August 2014

Design Type(s)	observation design • longitudinal data collection method • data integration
Measurement Type(s)	Climate proxy
Technology Type(s)	data collection method
Factor Type(s)	resolution • period
Sample Characteristic(s)	Central Russia • Alaska • Canada • Scandinavia • Eastern Russia • Greenland • North Atlantic • Arctic Canada • Forest • Ice • Marine • Lake • Cave

School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA.

Correspondence and requests for materials should be addressed to N.P.M. (email: Nicholas.McKay@nau.edu)

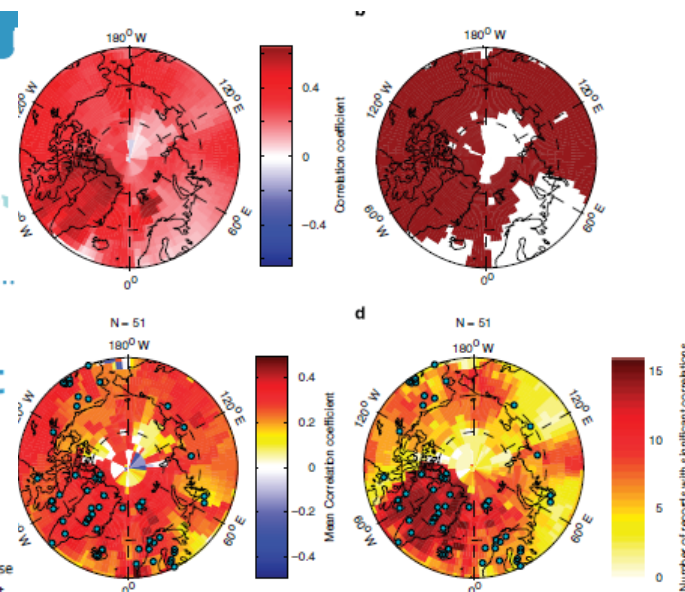


Figure 3. Spatiotemporal relation between annual, instrumental temperature and the PAGES Arctic 2k database. (a) Correlation coefficient between observed temperature at each grid cell and the revised Arctic temperature reconstruction between AD 1880 and 2000. (b) Grid cells with significant ($P < 0.05$; corrected for serial autocorrelation) correlations in (a) are shown in dark red. (c) Mean significant correlation coefficient at each grid cell for all records in the database AD 1880 and 2000. (d) Number of records with a significant correlation.

Data Citation

1. McKay, N. & Kaufman, D. *Figshare* <http://dx.doi.org/10.6084/m9.figshare.1054736> (2014).

Acknowledgements

We thank those who discovered and shared errors and updates to the original P. Code and support for BACON and BAM was kindly provided by Maarten Blaauw, respectively. Kevin Anchukaitis and participants of the PAGES 2k Climate Workshop contributed to this study. WDC-NOAA Paleoclimatology helped design the data files. Support for PAGES activities is provided by the US and European Foundations, US National Oceanographic and Atmospheric Administration Geosphere-Biosphere Programme. N.P.M. was supported by NSF award ARC-08-08281. Many colleagues who kindly made digital versions of their data available for this study and suggestions of two anonymous reviewers improved this data descriptor and

Author Contributions

N.P.M. and D.S.K. designed the study and prepared the manuscript. N.P.M. generated age ensembles, and formatted the data.

Scientific Data in Earth & Planetary Science.

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

Focuses to handle your research data

- 保存 Preservation
 - Technical infra., Function/Costs, Sustainability, Institutional management
- 公開 Publication
 - Review? Quality Assurance?
- 内容記述、属性 Metadata, data description
 - Data journal? “Readme”? For whom?
- 引用、参照 Citation, reference
 - データ提供者の評価、権利を確保する(≠法的保護)
Reputation of data producers (≠ legal right issue)
 - Persistent Identifiers (DOI?)
- 再利用 Reuse
 - Preservation/publication/metadata needed.

まとめにかえて：研究現場では今後？

- 研究データを配信する「お作法」？
 - 学術情報として広く研究データ情報を出すノウハウがない？
 - 整備すべきデータの整理学、理念がない？（粒度、処理レベル...）
- データの置き場、選別は？
 - データリポジトリ？ 各研究室でやりきれるのか？
 - 「図書館」的なしくみ？
 - 長期／短期保存データの選別プロセスは？ . . .
- メタデータ（データを使うための必須情報）の作成、管理は？
 - 研究者は苦手？書誌情報管理に似ている（司書的な仕事？）
 - 識別子（DOI?）を付与して被引用度で業績評価する？
- 「データ・キュレータ」（管理者？）
 - 各分野の専門家、かつ情報管理・情報処理スキル（研究とデータ管理が半々に必要？）
 - 人材育成、資金
- データ「出版」を研究成果に？
 - 査読プロセス？
 - 「データジャーナル」（Nature、Wiley、Elsevier...）
- 研究予算と、その成果となるデータ公開を一体に？
 - 米NIH、NSFなどはその方向へ推進中とのこと。JST? JSPS?...

⇔データ・サイエン
ティスト？

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)