# オープンサイエンスの推進と
# Data Citation Principles

村山泰啓
(ICSU-World Data System ex officio／京都大学生存圏研究所、
情報通信研究機構　統合データシステム研究開発室長)

# 目次

- オープンデータ、オープンサイエンス
  - 国際的な動向、G8の動き
  - 国内の状況
- 「科学」とデータ
  - なぜデータの引用（サイテーション）が重要か。
- Data Citation Principles

# G8（2013）における研究データオープン化の合意
## (↔ Open Government Data)

**G8 Science Ministers Statement** London UK, 12

**Introduction**

We, the G8 Science Ministers met in London on Wednesd
of our respective national science academies, as part of th
this unique meeting we discussed how our nations could le
transparency, coherence and coordination of the global sc
in order to address global challenges and maximise the so
of research.

**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**

G8 Open Data Charter will 'increase transparency' and 'fuel innovation'

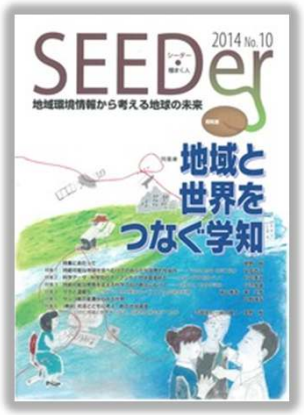
Five key principles outlines how governments should release datasets for economic and social benefits

# 科学技術オープンデータの背景

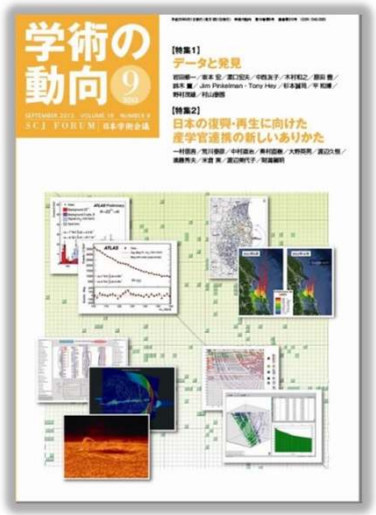
- 科学（技術）と社会
  - 社会と科学者の関わりが問われている
- データ
  - 科学技術活動の重要な成果物
  - 公的資金研究によるデータの公開原則
  - さらなる研究の加速（e-science, data intensive science…）

- オープンサイエンスデータの検討
  - World Data System事業（International Council of Science)等
  - G8+O6 data infrastructure WG
    ⇒ 国際コンソーシアム「RDA (Research Data Alliance)」
  - 内閣府・CSTI「国際動向を踏まえたオープンサイエンスに関する検討会」
  - 国立国会図書館「科学技術情報整備審議会」／第四期科学技術情報整備基本計画策定に向けた基本方針検討部会

# オープンサイエンス系のコンセプトの国内周知




SEEDer（シーダー） 10号(2014年4月）
特集：地域と世界をつなぐ学知
（総合地球学研究所・昭和堂）




学術の動向、2013年9月
（日本学術会議）

国立国会図書館月報
639号 2014.6


文科省「科学技術動向」誌」
連載（2014年9・10月号、
11・12月号．．．）

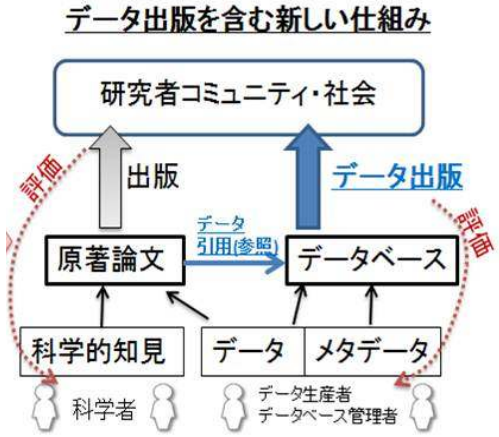
# 内閣府/CSTI：研究データ共有のガイドライン策定

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の開催について

平成 26 年 11 月 13 日
内閣府　政策統括官（科学技術・イノベーション担当）決定

1. 趣旨

オープンサイエンスに係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ため、「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」（以下「検討会」という。）を開催する。




Ⅰ．国際的動向からみたオープン化に関する現状認識
1．認識すべき重要性
2．基本概念
（1）オープンサイエンス
① オープンアクセス
② オープンデータ
③ オープン化すべき科学研究データ
3．国際的動向にみるオープン化の必要性

Ⅱ．国際的動向にみるオープン化に関する課題と検討すべき方向性
1．国際的動向にみる我が国の現状
2．日本におけるステークホルダーに求められる役割と課題

［内閣府、2014、
http://www8.cao.go.jp/cstp/sonota/openscience/1kai/1kai.html、2015年2月20日取得］

# 科学的方法論と情報共有

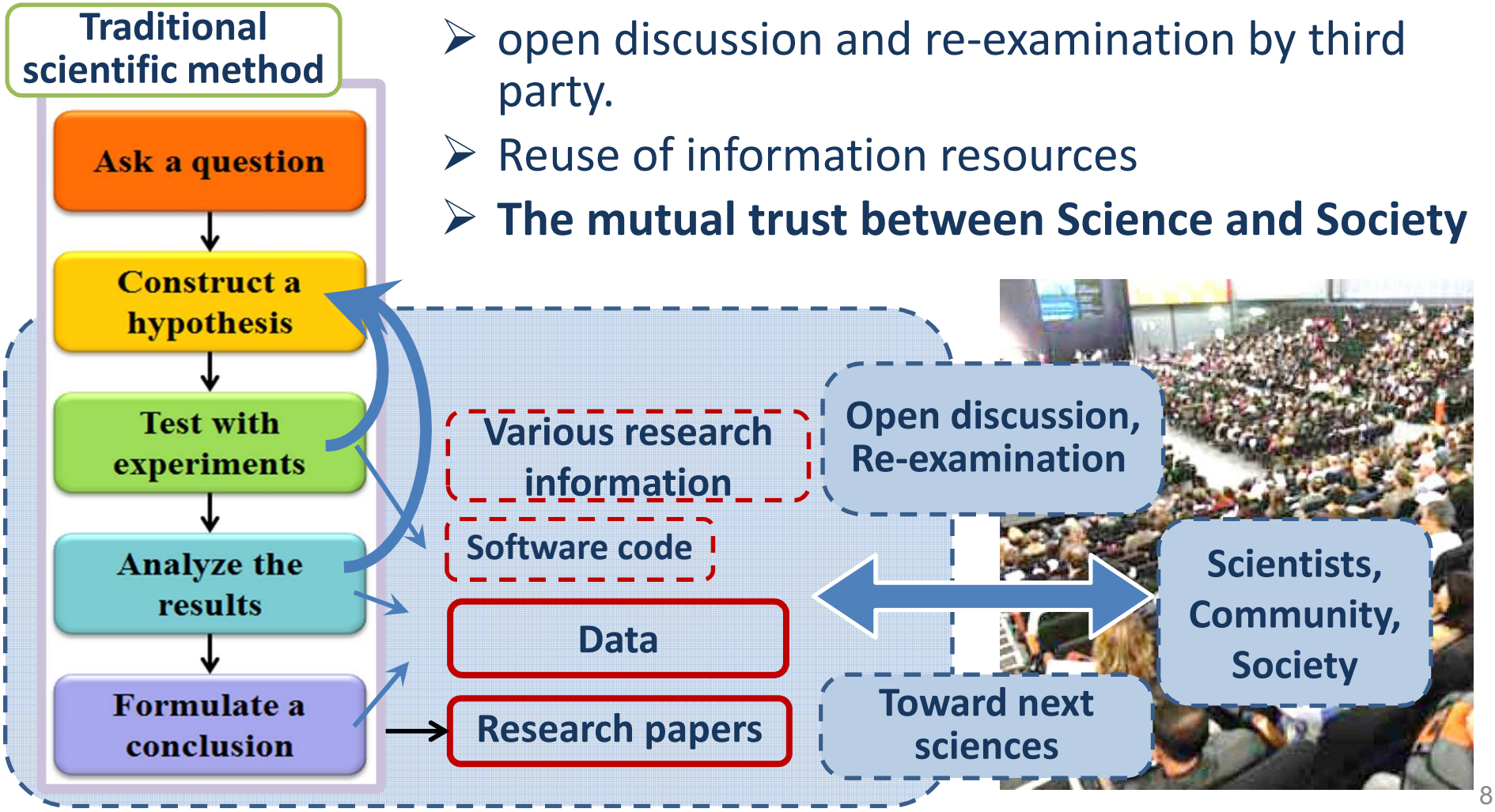
- 欧米の意識：
  科学的方法論（「科学」という「制度」）
  - 研究の方法・過程・論理・結論等の記録、相互批評
  - 情報がオープンに共有されることが必須
    - 第3者による再検証の担保
    - 研究者コミュニティでのコンセンサス形成
      ⇒社会との知識共有
  - 従来は、文献、口頭発表（ジャーナル、学会）での共有
    ⇒インターネット上での電子情報の共有
    (科学技術研究開発・イノベーションの新時代を目指して）
- 科学的発見（原著論文）と知の共有
  - 論文の固定、評価、公表、保存、引用、再利用（再検証）

# Why Open Data, Open Access?

Important are:

- science today made of the conventional method+ communication (sharing info.).
- open discussion and re-examination by third party.
- Reuse of information resources
- **The mutual trust between Science and Society**

http://www.getchemistryhelp.com/chemistry-lesson-scientific-method/

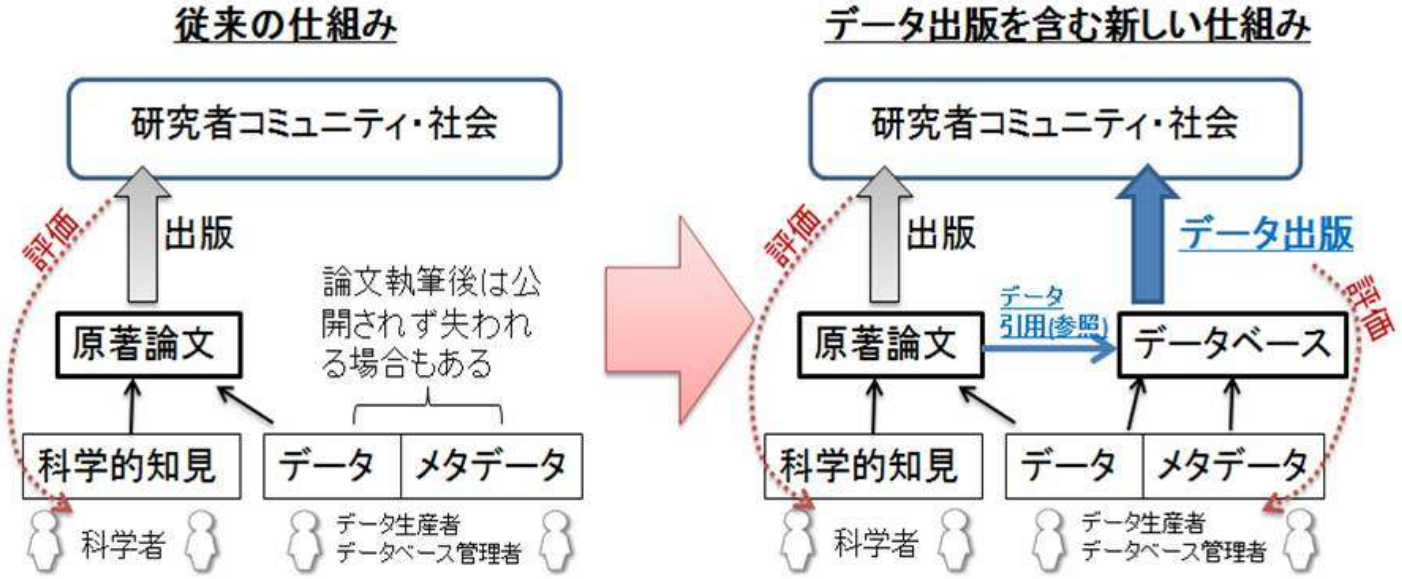


# 科学的方法論とデータの問題

- 根拠となるデータと知の共有
  - データセットの固定、評価、公表、保存、引用、再利用
    - 科学知の基礎として共有する必要性。
    - ⇔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
  - 注意点：公開データの範囲、猶予期間、利用条件、サービス設計. . .
    - cf. 図書館サービス

# “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

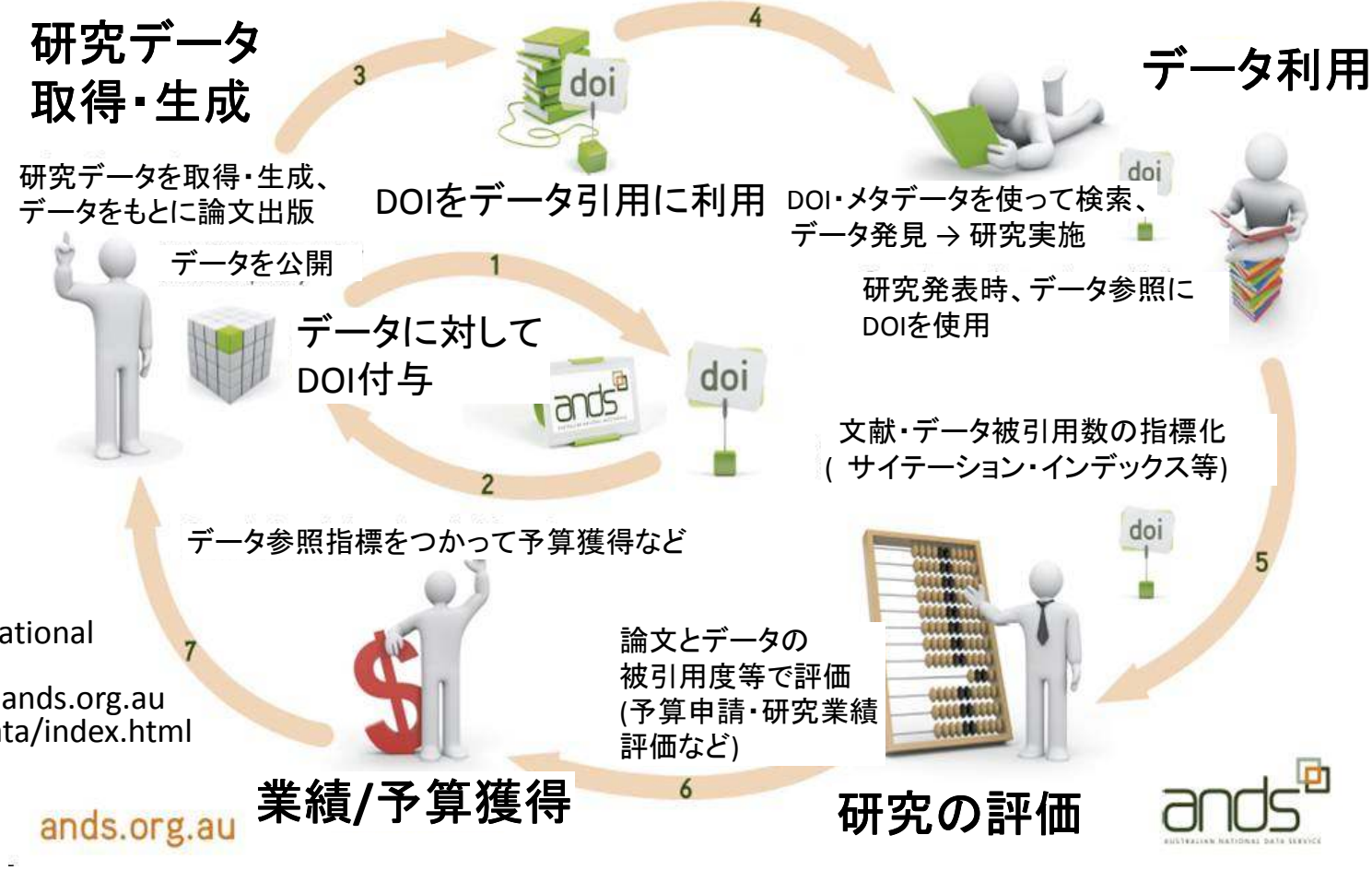



[地球電磁気・地球惑星圏学会, 2013]

# (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




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）

# Steps by Major scientific publishers
## encouraging data deposition

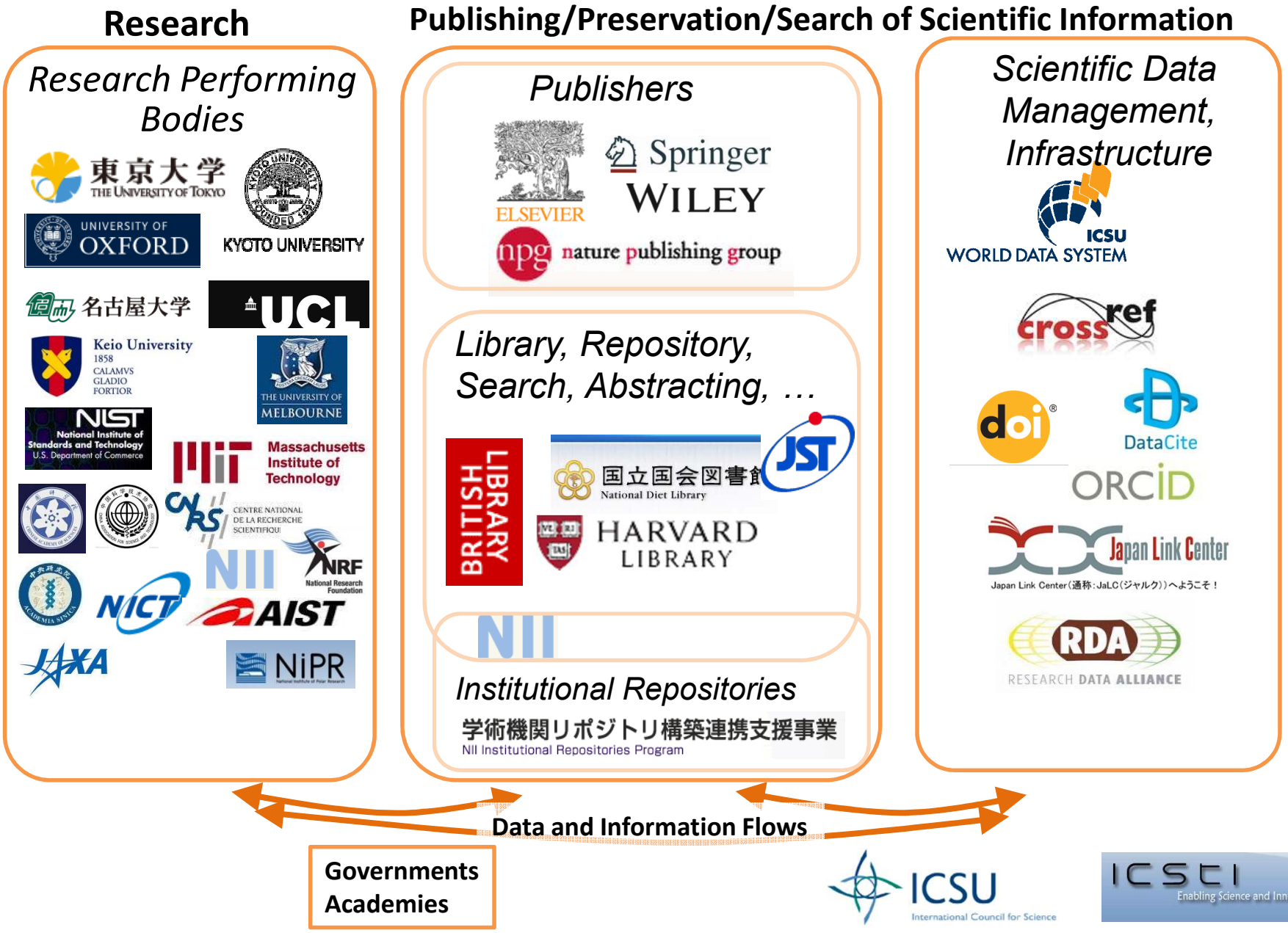
- **Willey/AGU** publication policy:
  ”…in AGU’s journals, **all data** necessary to understand, evaluate, replicate, and build upon the reported research **must be made available and accessible whenever possible**…”
- **SpringerOpen**/”Earth, Planets and Space”, “Geoscience Letters”…
  “…Electronic archiving of data enables readers to replicate, verify and build upon the conclusions published in papers in the journal. **It is recommended that all data** which are not directly attached to a publication as electronic supplementary files **be deposited**…”
- **Elsevier**/JASTP:
  “…Elsevier encourages **authors to deposit raw experimental data sets** underpinning their research publication in data repositories, and to enable interlinking of articles and data…”

# Joint Declaration of Data Citation Principles

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

1. Importance
2. Credit and Attribution
3. Evidence
4. Unique Identification
5. Access
6. Persistence
7. Specificity and Verifiability
8. Interoperability and Flexibility

# Science as a Social System (with "Print" Publication)



# データ・サイテーション：
## 今後どうやっていくべきか？

- まずは、論文中でデータを引用する
- 引用するときに識別子（例えばDOI）を使う
- DOIは1つの具体的な手法
  - どういうデータにDOIをつけるのか？
    ➔整理学、管理のための理念が必要
  - 理念、整理学は、研究分野ごとに違うのでは？
    ➔学会・研究コミュニティごとに、引用しやすい、データ作成者が報われやすい、方法をすこしづつみつけて実施していくべき。
- 国際的にどういう風に実践されていくのか？
- DOIだけでよいのか？
  - 詳細なContributorへのクレジットなどは次の課題。
  - まずは論文と同じ程度のクレジットをデータに。

# Placement of Citations

## Intra-work:

- *Should provide sufficient information to identify cited data reference within included reference list.*
- *Citation to data should be in close proximity to claims relying on data.* ***[Principle 3]***
- *May include additional information identifying specific portion of data related supporting that claim.* **[Principle 7]**

***Example:*** The plots shown in Figure X show the distribution of selected measures from the main data [Author(s), Year, portion or subset used].

## Full Citation:

*Citation may vary in style, but should be included in the full reference list along with citations to other types works.*

*Example:*

**References Section**

Author(s), Year, Article Title, Journal, Publisher, DOI.

Author(s), Year, Dataset Title, Data Repository or Archive, Version, Global Persistent Identifier.

Author(s), Year, Book Title, Publisher, ISBN.

[FORCE11, https://www.force11.org/node/4771, retrieved 27 Feb. 2015]

# Example

**The dataset:**

**Storz, D et al. (2009):**

**Planktic foraminiferal flux and faunal composition of sediment trap L1_K276 in the northeastern Atlantic.**

**http://dx.doi.org/10.1594/PANGAEA.724325**

**Is supplement to the article:**

**Storz, David; Schulz, Hartmut; Waniek, Joanna J; Schulz-Bull, Detlef; Kucera, Michal (2009): Seasonal and interannual variability of the planktic foraminiferal flux in the vicinity of the Azores Current.**

**Deep-Sea Research Part I-Oceanographic Research Papers, 56(1), 107-124,**

**http://dx.doi.org/10.1016/j.dsr.2008.08.009**

**[Jan Brase, 2013]**

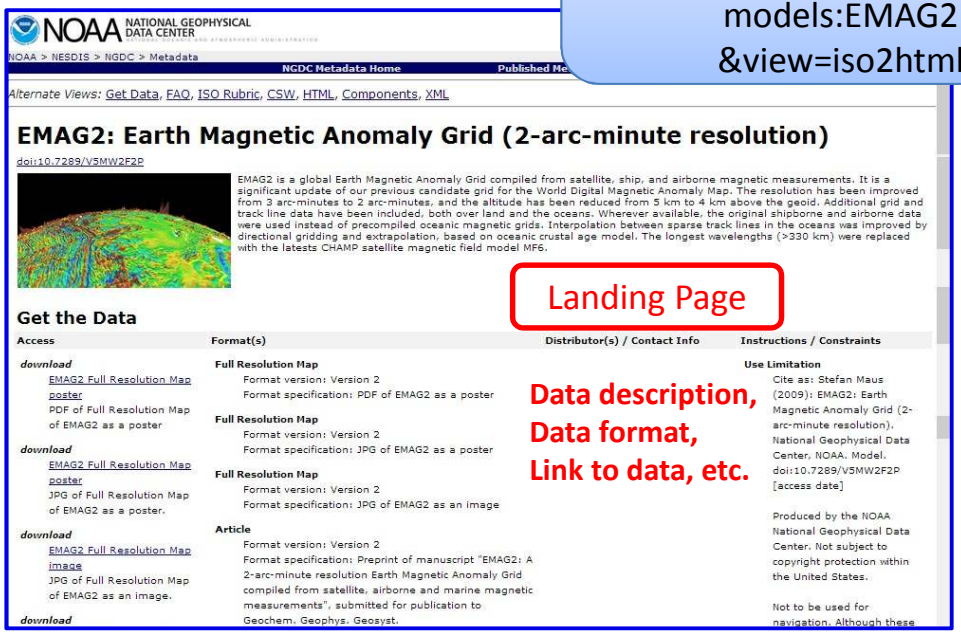
DataCite

# Example of DOI-minting to Earth Science database in NOAA/NGDC

- EMAG2: Earth Magnetic Anomaly Grid (2-arc-minute resolution)

doi:10.7289/V5MW2F2P

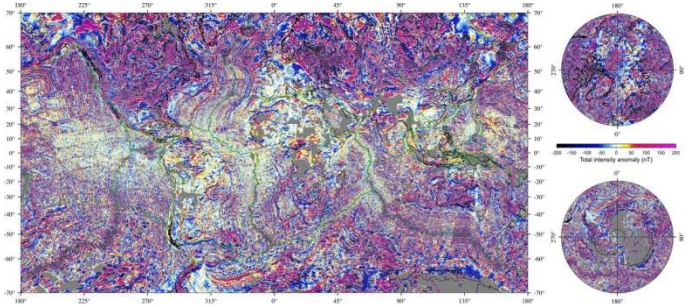
http://www.ngdc.noaa.gov/nmmrview/metadata.jsp?id=gov.noaa.ngdc.mgg.geophysical_models:EMAG2&view=iso2html




**Digital data**

```
-55.033333 -89.900000 -56.134989
-55.000000 -89.900000 -56.127400
-54.966667 -89.900000 -56.119808
-54.933333 -89.900000 -56.112213
-54.900000 -89.900000 -56.104616
-54.866667 -89.900000 -56.097016
-54.833333 -89.900000 -56.089413
-54.800000 -89.900000 -56.081806
-54.766667 -89.900000 -56.074197
-54.733333 -89.900000 -56.066584
-54.700000 -89.900000 -56.058968
-54.666667 -89.900000 -56.051348
-54.633333 -89.900000 -56.043725
-54.600000 -89.900000 -56.036097
-54.566667 -89.900000 -56.028466
-54.533333 -89.900000 -56.020832
-54.500000 -89.900000 -56.013193
-54.466667 -89.900000 -56.004623
```

**Data plot**



**Instruction of data citation**

Maus (2009): EMAG2: Earth Magnetic Anomaly Grid (2-arc-minute resolution). National Geophysical Data Center, NOAA. Model, doi:10.7289/V5MW2F2P [access date]

[ Nose et al., 2013]

# Example of data citation

Evaluation of the Solutrean hypothesis

2008 *Journal of the North Atlantic* 1:85–98

## The Solutrean Atlantic Hypothesis: A View from the Ocean

Kieran Westley[1,2,*] and Justin Dix[3]

**Abstract** - One current hypothesis for the Pleistocene peopling of the Americas invokes a dispersal by European hunter-gatherers along a biologically productive "corridor" situated on the edge of the sea-ice that filled the Atlantic Ocean during the Last Glacial Maximum (LGM). In this paper, we assert that critical paleoceanographic data underpinning this hypothesis has not yet been examined in sufficient detail. To this end, we present data which show that the corridor may not have existed, and that, if it did, its suitability as a migration route is highly questionable. In addition to demonstrating that the hypothesized migration was unlikely, this highlights the importance of integrating paleoceanographic and archaeological data in studies of paleo-coastal societies.

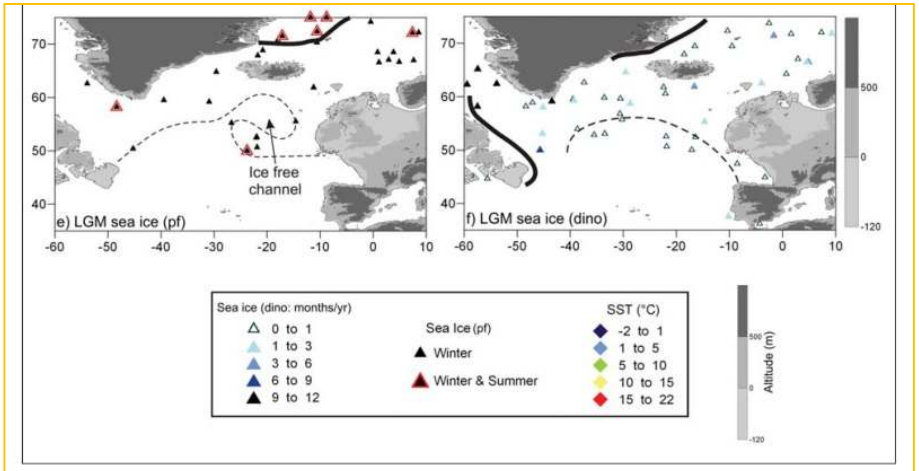



Figure 2. Quantitative reconstructions of LGM North Atlantic paleoceanography based on two different proxies: dinocysts and planktonic foraminifera. Data from De Vernal et al. (2004) and Weinelt (2004) (See also De Vernal et al. 2006, Kucera et al. 2005). a) Summer SSTs from planktonic foraminifera. b) Winter SSTs [illegible] from dinocysts. d) Winter SSTs from dinocysts. e) Sea-ice extents from planktonic foraminifera: triangles [illegible] with evidence of summer and winter ice. Heavy black line represents the extent of perennial ice, and dashed line is the maximum extent of winter ice (based on Sarnthein et al. 2003). f) Sea-ice extents from dinocysts: triangles show core sites with evidence of the duration of ice in months per year. Heavy black line represents the extent of perennial ice, and dashed line is the maximum extent of winter ice (based on De Vernal et al. 2006).

### References

variability on the Western Iberian margin during the last two glacial periods. Marine Geology 196:1–20.

De Vernal, A., and C. Hillaire-Marcel. 2000. Sea-ice cover, sea-surface salinity, and halo-/thermocline structure of the northwest North Atlantic: Modern versus full glacial conditions. Quaternary Science Reviews 19:65–85.

De Vernal, A., and T. Pedersen. 1997. Micropaleontology and palynology of core PAR87A-10: a 23,000 year record of paleoenvironmental changes in the Gulf of Alaska, northeast North Pacific. Paleoceanography 12(6):821–830.

De Vernal, A., F. Eynaud, M. Henry, C. Hillaire-Marcel, L. Londeix, S. Mangin, J. Mattheissen, F. Marret, T. Radi, A. Rochon, S. Solignac, and J.-L. Turon. 2004. MARGO (SST) unpublished data: Compilation of dinoflagellate cyst LGM SST data. doi: 10.1594/PANGAEA.127383. World Data Center for Mar[illegible] Environmental Sciences (WDC-MARE), P[illegible]ing Network for Geoscientific and Environm[illegible] Data (PANGAEA). Available online at http://www.pangaea.de/. Accessed June 2006.

De [illegible]ernal, A., F. Eynaud, M. Henry, C. Hillaire-Marcel, [illegible]. Londeix, S. Mangin, J. Mattheissen, F. Marret, T. Radi, A. Rochon, S. Solignac, and J.-L. Turon. 2005. Reconstruction of sea-surface conditions at middle to high latitudes of the Northern Hemisphere during the Last Glacial Maximum (LGM) based on dinoflagellate cyst assemblages. Quaternary Science Reviews [illegible]:897–924.

2000. Luminescence dating of sand deposits related to late Pleistocene human occupation at the Cactus Hill Site, Virginia, USA. Quaternary Geochronology 1:167–187.

Fiedel, S.J. 1992. Prehistory of the Americas. Cambridge University Press, Cambridge, UK..

Fiedel, S.J. 1999. Older than we thought: Implications of corrected dates for dates for Paleoindians. American Antiquity 64(1):95–115.

Fiedel, S.J. 2000. The peopling of the New World: Present evidence, new theories, and future directions. Journal of Archaeological Research 8(1):39–103.

Fladmark, K. 1979. Routes: Alternate migration corridors for early man in North America. American Antiquity 44(1):55–69.

Gamble, C., W. Davies, P. Pettitt, L. Hazelwood, and M. Richards. 2005. The archaeological and genetic foundations of the European population during the Late Glacial: Implications for "agricultural thinking." Cambridge Archaeological Journal 15(2):193–223.

Heaton, T.A., S.L. Talbot, and G.F. Shields. 1996. An Ice Age refugium for large mammals in the Alexander Archipelago, Southeastern Alaska. Quaternary Research 46:186–192.

Hemming, S.R. 2004. Heinrich events: Massive Late Pleistocene detritus layers of the North Atlantic and their global climate imprint. Review of Geophysics 42:RG1005.

Henshaw, A. 2003. Polynyas and ice-edge habitats in [illegible]

structions of LGM North Atlantic paleoceanc
. Data from De Vernal et al. (2004) and Wein
s from planktonic foraminifera. b) Winter SS



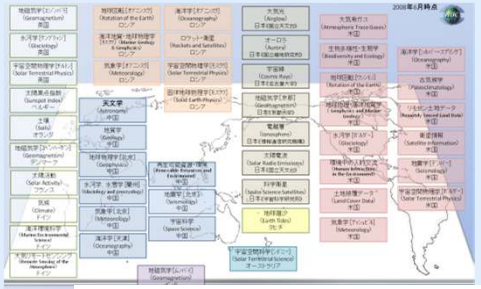
Westley and Dix [2008]

# 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。現在の形は1931年。国連等で科学技術分野代表などを務める。国際NGO（本部：パリ）。

ICSU-WDS members (加盟機関)：合計89メンバー（2015年1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。

# 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。

[恒松・村山、2014]

# Research Data Alliance Created to Accelerate Development of Research Data Sharing Infrastructure Worldwide

- RDA community efforts focus on building **social, organizational and technical infrastructure** to
  - *reduce barriers to data sharing and exchange*
  - *accelerate the development of coordinated global data infrastructure*

*RDA and RDA/US are supported in part by the National Science Foundation.*

Fran Berman

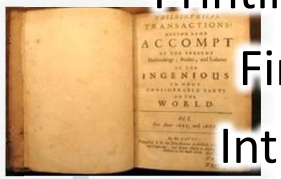
[Fabrizio Gagliardi , 2014]

# 以下、参考資料

# 科学技術イノベーションの基盤としての情報共有
## ～数百年の印刷文化の基礎支えと、成長途中のディジタル・サイエンス

Public library (paper media) : 8c

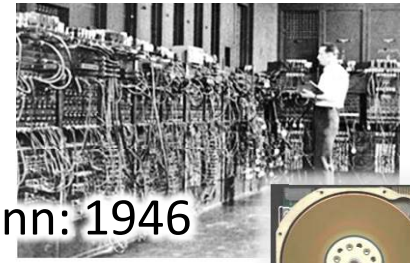
Printing press/Gutenberg: 1445



First scientific journal: 1665

Intl. Assoc. Academies: 1899

ICSU established: 1931



ENIAC, von Neumann: 1946

Hard Disk Drive: 1956

World Data Center system : 1957

CODATA

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

ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

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

# 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．．．

# 日本から見た関係国際機関の概観

地球科学
宇宙科学
物理学、地理学
気象、農業、情報科学

地震学

地球科学

地球
科学
気象

電離層・
宇宙天気

宇宙科学
計算機科学
物理学

言語学
歴史学
心理学

言語学

社会科学全般

. . . 等、計86機関加盟

ICSU
International Council for Science

協力要請

futurearth
research for global sustainability

Future Earth
(ICSU, UNESCO, UNEP,
UNU, Belmont Forum,…)

協力

協力

日本学術会議

協力

協力

DIAS
Data Integration & Analysis System

GEOSS/DIAS
（東大、JAXA、
JAMSTEC etc）

総合科学技術会議
COUNCIL FOR SCIENCE AND TECHNOLOGY POLICY

協力
要請

協力要請

G8科学大臣会合で
言及(2013.6)

MOU
RDAのWDS加盟
合同部会活動

RDA
(Research Data
Alliance)

RESEARCH DATA ALLIANCE

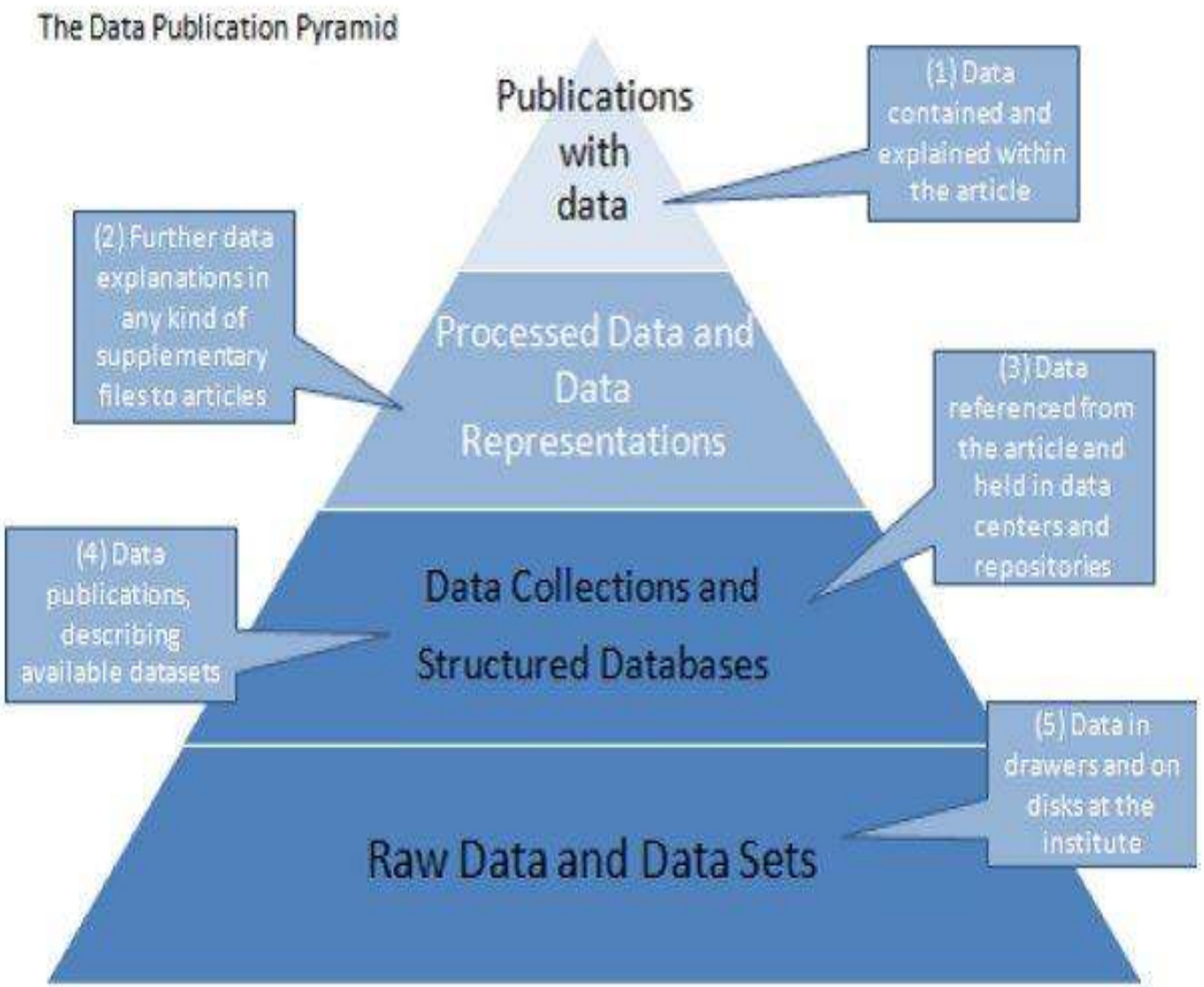
G8

G8 Open Data Charter will 'increase transparency' and 'fuel innovation'

Five key principles outlines how governments should release datasets for economic and social benefits

**[H. Frederick Dylla, 2012]**

# Data Pyramid

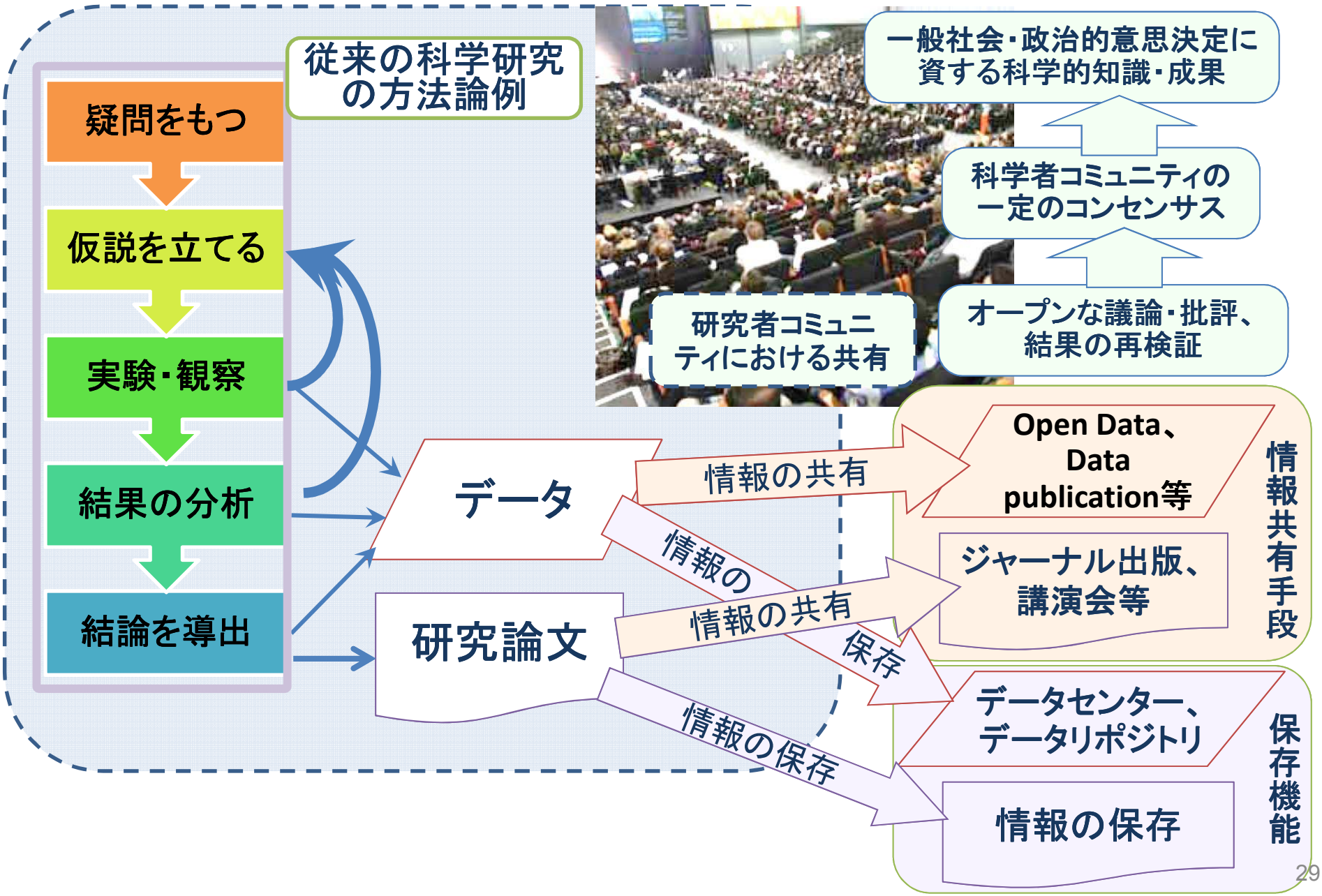


# Changing standards and culture takes a long time.

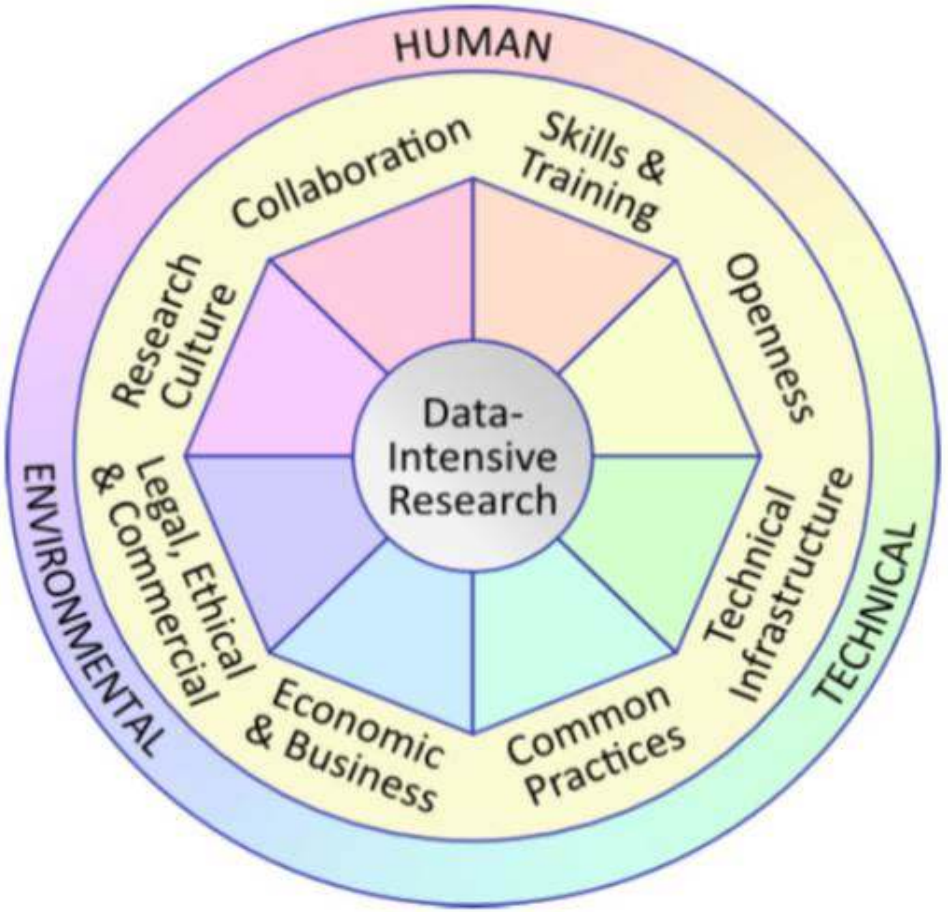
- Metric system
  - Introduced 1795
  - Convention du Mètre, 1875
  - International System of Units (SI) 1960
  - One big holdout
- Time zones
  - 1st use of standard ("railway") time 1847
  - International Meridian Conference 1884 established GMT but did not alter local times
  - Final adoption of "standard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

[Mark Parsons, 2013]

# 科学研究とデータ･情報


従来の科学研究の方法論例
疑問をもつ
仮説を立てる
実験･観察
結果の分析
結論を導出
データ
研究論文
研究者コミュニティにおける共有
一般社会･政治的意思決定に資する科学的知識･成果
科学者コミュニティの一定のコンセンサス
オープンな議論･批評、結果の再検証
情報の共有
情報の
情報の共有
保存
情報の保存
Open Data、Data publication等
ジャーナル出版、講演会等
情報共有手段
データセンター、データリポジトリ
情報の保存
保存機能

# 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.