

オープンサイエンス、オープンデータの国際潮流と日本における現状について

村山泰啓
（日本学術会議連携会員、
G7オープンサイエンスWG共同議長、
情報通信研究機構）



Acknowledgement: This work was supported by Japan Science and Technology Agency as part of the Belmont Forum PARSEC project.



私自身の位置づけ

大気科学、超高層物理学、大気レーダー等



京都大学生存研・客員教授（H25）、国立極地研究所南極観測審議委員



日本地球惑星科学連合 理事、情報システム委（'14-）他



AGU Earth & Space Science Informatics Section Exec. Committee（'18-）



IUGG Union Commission for Data & Info., member (2017-)



ISC-World Data System Scientific Committee member

情報通信研究機構（NICT）



日本学術会議 連携会員（国際サイエンスデータ分科会・委員長、等）



国立国会図書館 科学技術情報整備審議会・委員

内閣府 オープンサイエンス検討会・委員（2014-）



欧州委員会 European Open Science Cloud・欧州外委員（'15-'16）



G7科学技術大臣会合 オープンサイエンス部会（2016-、共同議長2019-）

研究データ／オープンサイエンス、科学政策論、科学と社会

村山版「オープンサイエンス」（と呼ばれるもの）の議論：
その全体像へむけて（暫定、改訂中）

1. 国内外の科学政策動向～「デジタル」科学政策としてのオープンサイエンス
2. 近代科学の「健全性」～科学研究論文と研究データ
3. 印刷文化から電子技術へ～時代の曲り角と新しい社会基盤上での科学
4. データを科学業績へ近づける取組み例～データDOI、データ出版
5. 国際学会・国際社会における規範の変化
6. 「オープン化」できる研究データとは～研究現場での「データ」の多様性
(6.5 データの多様性を読み解く～データ分類学のすすめ)
7. 新たな研究業績評価のあり方を模索する
～あるべき科学に戻すために、健全な科学のあり方のために
8. オンライン科学データの信頼性を担保する
9. 社会システムとしての「科学」再考～研究する、出版する、保存・再利用する
10. 科学データ基盤への社会投資～資源枯渇社会のデジタルデータ基盤

お話ししたいことの予定

1. 論文出版とデータ
2. データの格納場所（データリポジトリ）
3. 海外学協会での議論例
4. 学術ジャーナルのあるべき姿
5. 科学政策の動向

（補足）

- 「オープンサイエンス」
 - 「近代科学の原理はよいが、実行方法はいまのままでよいの？」
 - 科学と社会の関係の歪み、科学者の成果と評価
 - 「こうあるべき」という「べき」論を中心にやってきた

“オープンサイエンス”（～新しいデータの取り扱い）の諸外国動向の例 (JST科学技術情報委員会調査より) H27.4

(1) OECD 科学技術政策委員会閣僚級会合宣言¹

2006年12月、OECD加盟国において“OECD Principles and Guidelines for Access to Research Data from Public Funding(公的資金による研究データへのアクセスに関する原則及びガイドライン)”が承認された。研究データへのオープンアクセスを実現し、その広範な利用を促すことが、研究の質と生産性の向上につながるという認識のもとに、「公開性」、「透明性」、「遵法性」、

(4) 米国国立衛生研究所(NIH) データ共有ポリシーと実施要綱²

2003年10月、NIHはData Sharing Policyを制定し、研究資金公募時におけるデータ管理計画(Data Sharing Plan)の提出を義務化した。本要綱では、NIHの助成研究によるデータの共有理由として、「科学研究のオープン化の強化」、「分析や見解の多様化の奨励」、「新たな研究の推進」、「新規又は代替仮説と分析方法の検証」、「データ収集方法や手段に関する研究の支援」、「研究者育成」、「最初の研究者達が想定しえなかった課題の模索」、「複数のソースからのデータを組み合わせる新しいデータセットの作成」を挙げている。

(6) 英国研究会議(RCUK) Common Principles on Data Policy³

2011年、RCUKは「公的助成を受けた研究データは公益財であり、できる限り制限なく、適時にまた知的財産を害することのないよう責任ある方法でオープンに利用できるようにする」という基本原則のもと、“Common Principles on Data Policy”を公表した。本ポリシーは、政府助成研究成果に対する透明性の確保と統一的研究基盤の整備を目的として、研究データのオープン化に関する英国内各研究会議に共通の原則を定める包括的枠組みと位置づけられている。この枠組みに沿って、下記参考のように、英国における多くの研究助成機関ではデータポリシーが発行されており、データマネジメントおよびシェアリングの計画書提出を要求しているものが

5

G8、G7における科学データ政策潮流とオープンサイエンス（＝データ問題）



2013年

- ・ 2013年 G8サミット（英）：G8国研究データオープン化合意
- ・ 「データ」を重要な科学技術の研究成果として位置付け。
→国内外の科学政策動向へ影響。



2016年

- ・ 2016年 G7科技大臣会合（日本）：→G7オープンサイエンスWGの設置

村山：2018年より
OS-WG共同議長



科学大臣会合(2016)にて講演。村山と林上席研(NISTEP)。
写真：2016年G7科技大臣会合写真：内閣府・茨城県・つくば市提供



G7部会にて。左から村山、原山CSTI 議員、島尻科技担当大臣
(肩書は2016年当時のもの)



2017年

- ・ 2017年G7科学大臣会合@イタリア
- ・ 2018年G7科学シエルパ会合@カナダ
- ・ 2019年G7科学シエルパ会合@フランス
- ・ 2020年G7科学大臣会合@米国
- ・ 2021年英国、2022年独、2023年日本...



2019年

オープンサイエンス、
デジタルデータ問題：
「Incentive, rewards
＝やる気・業績化」と
「インフラ」

(→100年単位での情報保存) 6

論文投稿・出版と研究データ ～ジャーナル出版ポリシー上のデータ提供、 論文におけるデータ引用など

AGU出版誌の投稿規程の変化

[能勢正仁(名大ISEE)、科学データ研究会・WDS国内シンポジウム、2010]

- 2019年8月から、FAIR Data Policyに沿うことが求められた。
- 論文に用いたデータをリポジトリに登録する必要がある。
- リポジトリのリストとして、Repository Finder(<https://repositoryfinder.datacite.org/>)が挙げられている。
- 実際に、投稿の際に受け付けてもらえなかったり、受理を保留されるケースが起きている。

AGU PUBLICATIONS

論文投稿ページ

JOURNAL OF GEOPHYSICAL RESEARCH
Space Physics

Thank you for choosing AGU Publications.

FAIR Data Policy

As of August 1, 2019 AGU is adhering to the common enabling FAIR data policies from community data repositories (not in supplements). All authors must indicate where readers can access the data or have you intend to not yet available in a repository, include the data as supplements for review. Further guidelines are <https://www.agu.org/pubs/authors/faq-fair-data>. See section 3 for information on data availability.

世界的な動向や取り組み

[三好由純(名大ISEE)、科学データ研究会・WDS国内シンポジウム、2010]

1. FAIR Data Standardsの徹底化が、研究成果の公表時に求められている

- ・米地球物理学連合：2019/08よりポリシーの改訂
- ・認証や、doiの取得（ERGサイエンスセンターでも対応を行っている）

背景:

- 科学データ全般に追跡・追試可能性が求められるようになった。
- 米地球物理学連合の雑誌では、以下が**投稿の際の必要条件**となっている。
 - ・データを開示
 - ・出版時にデータファイルにフリーにアクセスできる
 - ・データ引用をきちんと行うこと

今後の研究論文出版では、データ取り扱いが不可欠になる趨勢

“Copernicus publication”社からEditor, refereeへのメッセージ(2018/11/8、村山仮訳)

- 学術成果は、科学情報同士の「相互ネットワーク」により成り立つ（「引用の連鎖」）
 - **FAIRデータ原則**：データは発見可能(findable), アクセス可能(accessible), 相互利用可能(interoperable), 再利用可能(reusable)であるべき
 - データが備えるべきもの：
 - **ユニークな永続的識別子 (PID) ← DOI, ORCIDなど**
 - 適切なメタデータ（データ検索を可能に）
 - 引用可能であること
 - 論文執筆・投稿時の著者に求められること：
 - **データ（+もとになる資料類）はFAIR原則に沿ったデータ置き場(data repository)に置かれること → 出版社の論文PDFと同じクオリティの保存場所は確保可能か？**
 - 論文中で利用・言及したデータを引用すること
 - そうしたデータについての論文・解説等は参考文献リストに適切に掲載する
 - データが入手可能かどうか、また入手方法などについての説明もふくめること
- こうした要請は論文原稿の査読時にチェックされる**

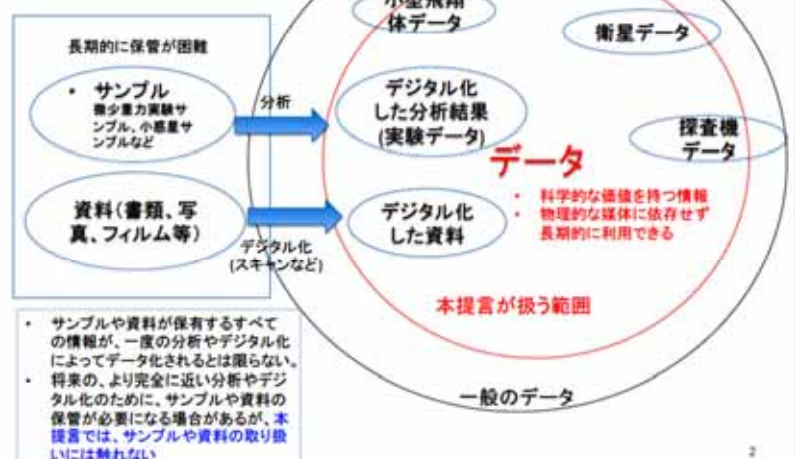
9

科学データキュレーションのプロセス (宇宙科学研究所の事例)

データの分類

- データを、種類とデータ処理レベルによって分類する
- **デジタル化した資料**
 - 汎用性・有用性には幅がある。通常、データ処理は必要ない。
- **源泉データ**
 - 衛星テレメトリなど、データ処理の源泉になるもの。
- **低次データ、高次データ**
 - データ処理のレベルに応じた分類
 - 一般に、データ処理レベルが上がると、データの汎用性・有用性が上がる
- **観測データ**
 - 制御できない対象の性質を記述する、天体や宇宙現象の観測データ。
- **実験データ**
 - 観測者が、対象に何らかの意図的な操作を加えることによって得られるデータ
 - ISS与圧部実験データ、ロケット・大気球を用いた実験データ、サンプルの分析
- **工学データ**
 - 衛星・探査機・装置の軌道・姿勢・性能等、データを取得する側の状態を記述
- **探査データについて**
 - ここでは、探査データという用語は用いない
 - 上記のどれにあてはまるか、その都度検討して分類する

本提言が対象とするデータの範囲



論文投稿時のデータ保存要請に困っている人のために： 当面、とりあえずの対応（数か月～1、2年）

- 既存の外部データを引き受けてくれるデータリポジトリへ寄託
 - Zenodo（OpenAIRとCERN共同事業、Dryadとの統合？）
 - DOI付与。原則50GB未満データ
 - データ内容のチェックなどキュレーションはなし
 - EU公的資金で運営（OpenAIR）⇒予算終了後の対応が課題。
 - Figshare（Natureと同じMcMillan系商業資本による）
 - 原則、コミュニティ研究者のデータを無料でひきうけるとのこと
 - DOI付与
 - 村山の個人的心配：商業ベースだとコミュニティの意識とは別に将来何か違うことにならないか
 - PANGAEA
 - （独<https://www.pangaea.de/>。アルフレッドウェーゲナー研究所、ブレーメン大学の共同運営。
 - コミュニティ研究者のデータを引き受ける。原則無料（らしい。寄付は歓迎？）
 -

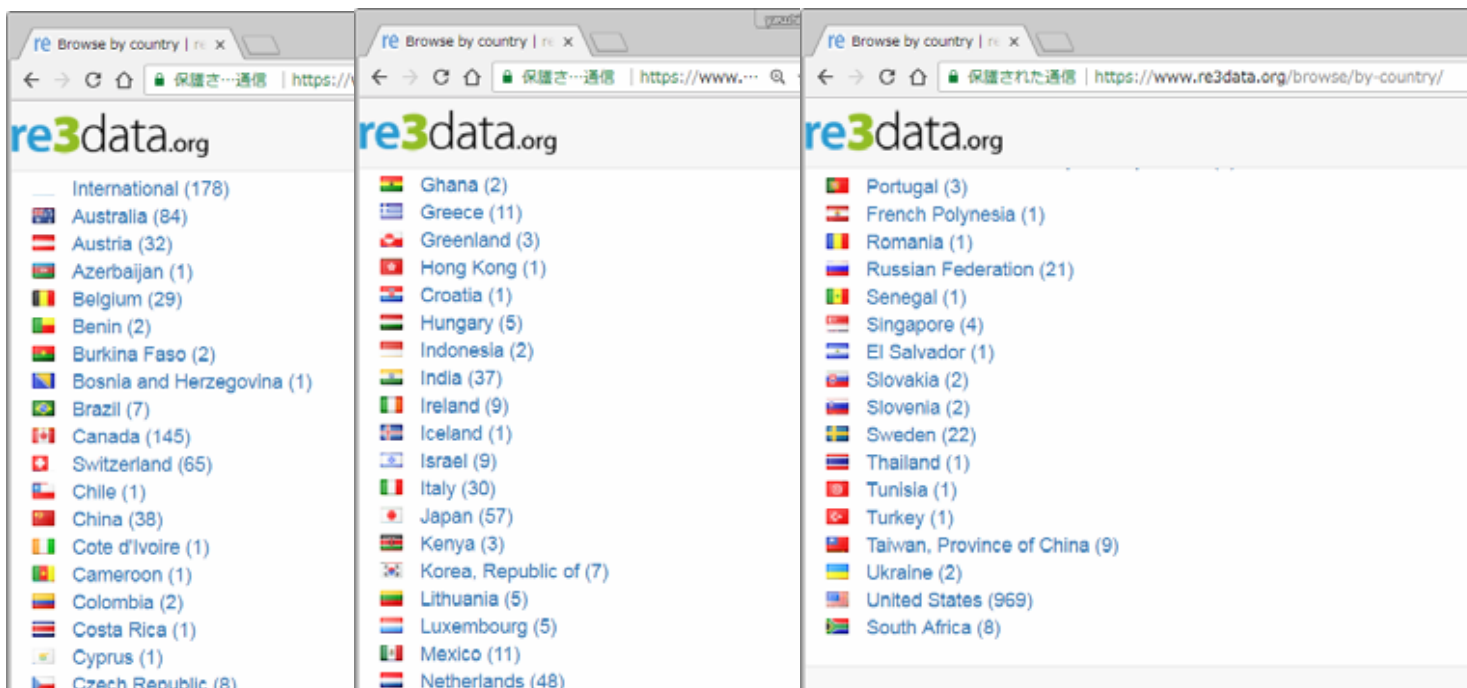
もうすこし先を見た対応

- 日本で生成したデータを国外リポジトリに集積してよいのか？
 - 政府内では問題意識がある。（国のデータ資産は国内に）
 - 国内学術データ保存管理基盤への投資、戦略が必要
 - 各分野ごとに異なる管理方針、管理要請の詳細へ対応が必要
- 科学データ保存・管理拠点が必要
 - 分野ごと拠点にするか、ナショナルライブラリ的にするのがよいか
 - データ管理・整備にはその専門分野の専門家が必要ではないか
 - 分野の知識がなければ、必要なデータ整備作業はできない
 - 村山の試案：研究半分／データ管理半分エフォートの研究者ポストを作るなど
 - 分野の拠点的な機関から、基盤整備費と人員の配備を要求するなど？
（共同利用研などは研究分野の拠点である？もしそうならその役割を担えないか？）

研究データの格納場所、管理機構の議論 (データ・リポジトリ)

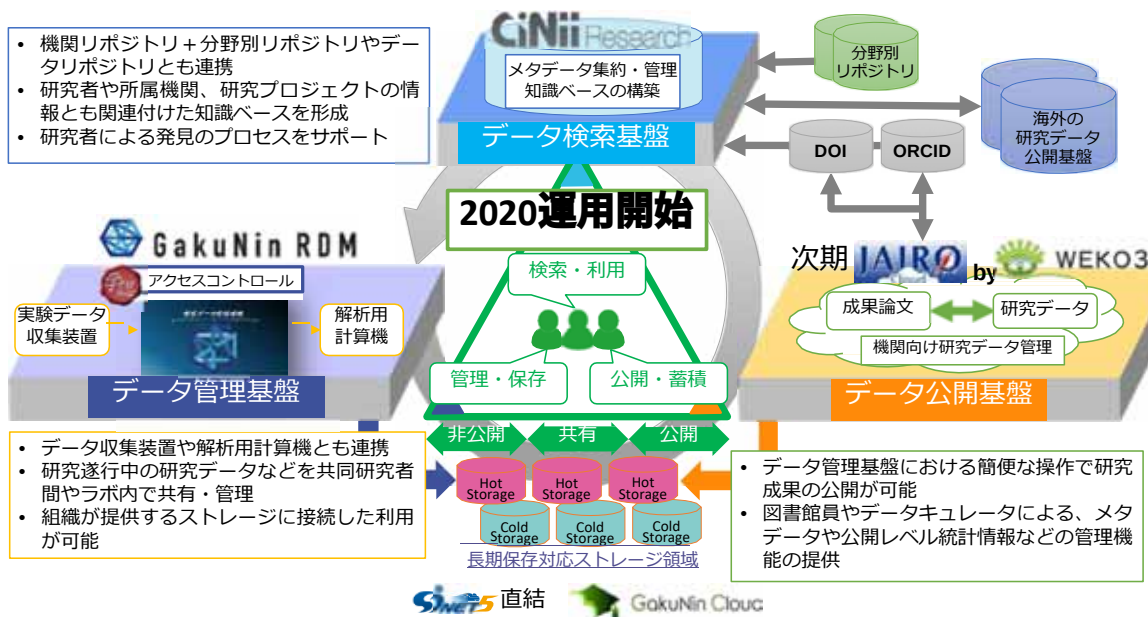
データリポジトリのレジストリ（カタログ） 例

- re3data.org (directory search of DataCite)



NII Research Data Cloud

NII: 情報・システム研究機構
国立情報学研究所

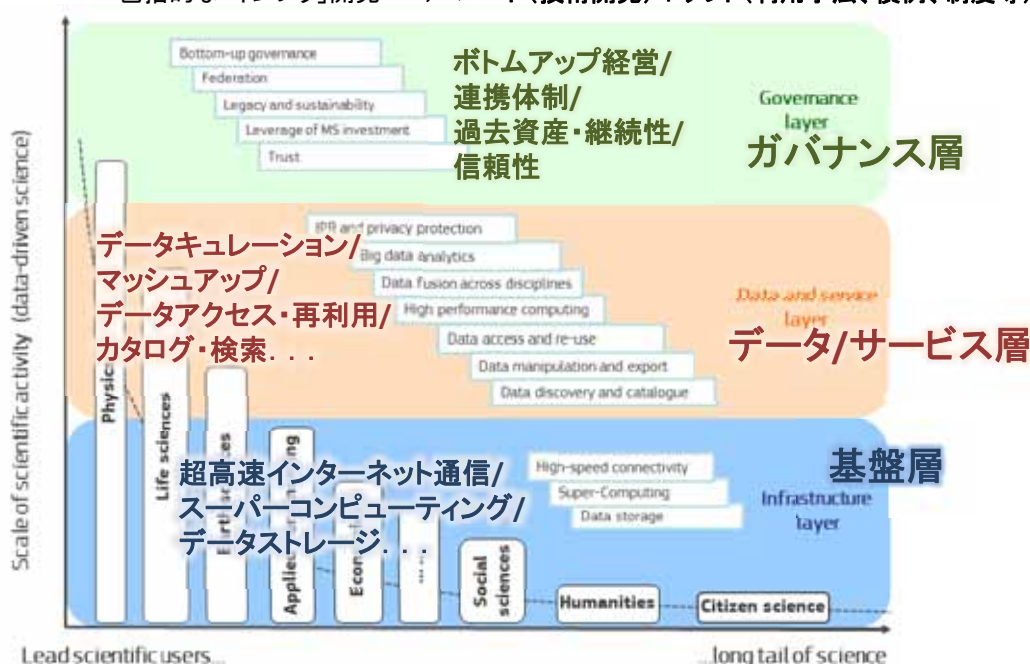


[Kazutsuna Yamaji (NII), 2019]

1

EU European Open Science Cloud 施策の概念 (J.C.Burgelman, 2016)

- ① 従来のデータ・インフラは基盤層+α程度だった
- ② 国際的なデータ利用の流れは、ソフトなサービス層・ガバナンス層までふくめた包括的な「インフラ」開発 → ハード(技術開発)+ソフト(利用手法、慣例、制度等)が肝となる



欧州オープンサイエンスクラウド計画 (EUは67億€投資。内20億€はHorizon2020)。

re3data.org

Search Browse Suggest Resources Contact DataCite

South Africa (1)
Spain (4)
Sweden (3)
Switzerland (8)
Taiwan (1)
United Kingdom (8)
United States (23)
Viet Nam (1)

AID systems
API
Certificates
CoreTrustSeal (2)
DSA (1)
WDS (2)
other (5)

Data access
Data access restrictions
Database access
Database access restrictions
Database licenses
Data licenses
Data upload
Data upload restrictions
Enhanced publication
Institution responsibility type
Institution type
Keywords
Metadata standards
PID systems
Provider types
Quality management

Evaluated Nuclear Data File
ENDF

Subject(s) Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas Physics Natural Sciences

Content type(s) Archived data Plain text Scientific and statistical data formats Raw data Structured graphics
Standard office documents

Country France Austria United States

Core nuclear reaction database contain recommended, evaluated cross sections, spectra, angular distributions, fission product yields, photo-atomic and thermal scattering law data, with emphasis on neutron induced reactions. The data were analyzed by experienced nuclear physicists to produce recommended libraries for one of the national nuclear data projects (USA, Europe, Japan, Russia and China). All data are stored in the internationally-adopted ENDF-6 format maintained by CSEWG.

International Service of Geomagnetic Indices
ISGI

Subject(s) Physics Astrophysics and Astronomy Geosciences (including Geography) Geophysics
Natural Sciences Geophysics and Geodesy

Content type(s) Raw data Structured graphics Structured text Scientific and statistical data formats
Software applications

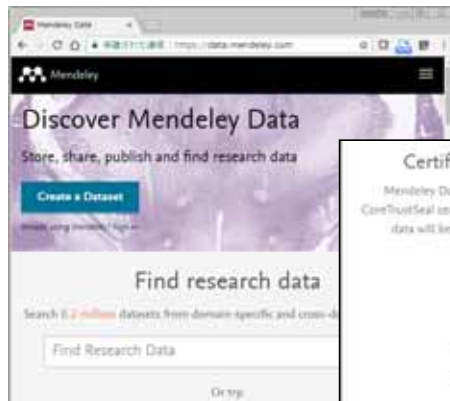
Country Russian Federation Japan Spain International France Denmark Germany

The International Service of Geomagnetic Indices (ISGI) is in charge of the elaboration and dissemination of geomagnetic indices, and of tables of remarkable magnetic events, based on the report of magnetic observatories distributed all over the planet, with the help of ISGI Collaborating Institutes. The interaction between the solar wind, including plasma and interplanetary magnetic field, and the Earth's magnetosphere results in a transfer of energy and particles inside the magnetosphere. Solar wind characteristics are highly variable, and they have actually a direct influence on the shape and size of the magnetosphere, on the amount of transferred energy, and on the way this energy is dissipated. It is clear that the great diversity of sources of magnetic variations give rise to a great complexity in ground magnetic signatures. Geomagnetic

“Certified”
(信頼性認証済み)マーク

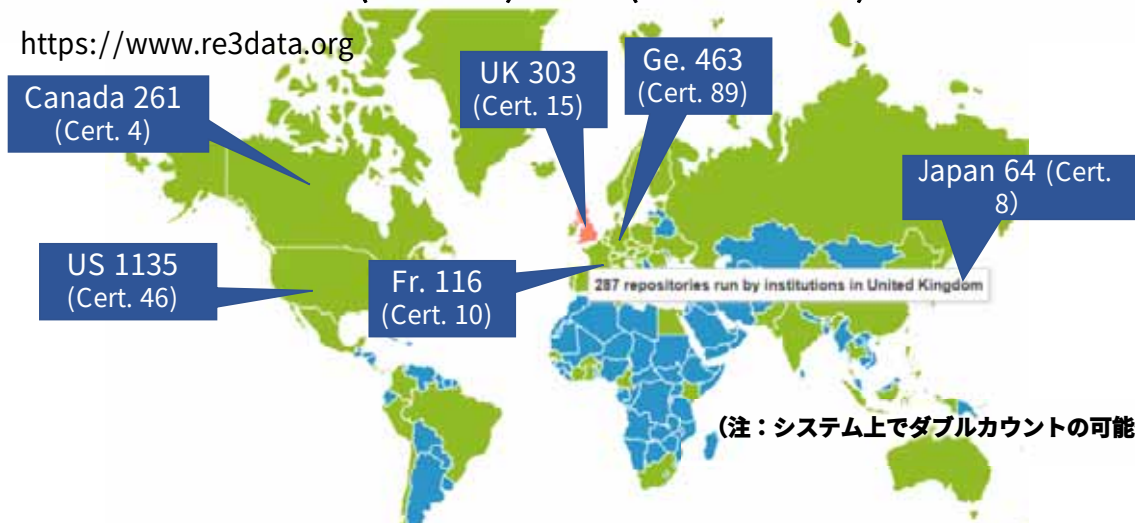
CoreTrustSeal～研究コミュニティによるデータ置き場（リポジトリ）の国際認証

- WDSとDSA（Data Seal of Approval）が統合
- 国際的な認知が進んでいる（欧州連合の標準採用など）
- 非営利機関による認証（⇔ 商業学術事業）



世界の「データリポジトリ」（データ保存機構）カタログ（e.g., re3data） 上の登録数と国際認証済み(Certified) 機関 (12 Nov. 2021)

<https://www.re3data.org>



(注：システム上でダブルカウントの可能性あり)

データリポジトリ信頼性の国際認証

- WDSとDSA（Data Seal of Approval）が統合、よりユニバーサルな認証機構に。
- 国際的に認知が進んでいる（EU、米など）
- その他の認証機構も進展中



Core level Certification

コア・レベル認証



- **16 Requirements** 16個の要求要件
- Context 背景
- Organizational infrastructure (6) 組織
- Digital object management (8) デジタルデータ整備、管理
- Technology (2) 技術基盤
- Applicant feedback 申請者からのフィードバック



- www.icsu-wds.org/services/certification

コア認証

「拡張」? 認証

本格的な認証

米National Institute of Health (国際的にオープンサイエンスの牽引機関の1つ) のブログより (ちょっと古い。2017年4月版。)



LEVEL	CORE	EXTENDED	FORMAL
Organization(s)	WDS: ICSU World Data System DSA: Data Seal of Approval	DIN : German Institute for Standardization	ISO : International Organization for Standardization
No. of Requirements	16	34	100+
Standards	Mandatory Requirements	DIN 31664	ISO 14721 (OAIS) ISO 16363 ISO 16919 ISO 17021
Audit Process	Self-assessment + independent peer review (2)	Self-assessment + independent peer review (2)	ISO certified audit with accredited auditors
Certification Cost	Free 2018年1月より有料	€500	\$10,000
Designation	World Data System logos or Data Seal of Approval	nestor Seal for Trustworthy Digital Archives	TBD
Certification lifespan	3 years	Indefinite	3 years
No. of Certified Repositories	130+ (WDS DSA)	2	Coming Soon



研究データ「オープン化」「共有」

- 過去にもデータ共有は特定分野では行われている Data sharing has been successful practices in several scientific domains.
- データ共有・公開 Open/Sharing Data
 - 従来のデータ公開類型例 Past “data sharing” examples
 - 方法論を共有するコミュニティ内 Sharing data inside a community
 - 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)
Metadata or attribution was not important. There are communities where “meta” information of data is **implicitly** shared. Dataset and its metadata are not “organized” for wider use.
 - 近年のオープンリサーチデータ例 In the context of “Open Science”
 - データを原著論文と同様の研究成果物と位置付け
Data is research output as important as original paper.
 - 論文のように、後々まで、誰がリファアー・利用してもよい
Data needs to be referred/cited/reused (and preserved) like papers.
 - 付帯情報、データ生成に関わる情報の記録が重要 (metadata、「data journal」 ”Data Descriptor” article、等)
Data’s description, metadata, attribution are as important as data itself.

「データジャーナル」という新しいメディア

- ・査読付きの「データ論文」という形でデータの成り立ち、性質、詳細を論述。
- ・データセットの公開もあわせて要請される

- Geoscience Data Journal (Wiley、Royal Meteorological Society)
 - ・ 投稿するデータにDOI付与を推奨（義務付けるジャーナルもあり）
 - ・ データは指定のリポジトリ（データセンター）に置いて公開すること
- Data in Brief (Elsevier)
 - ・ Elsevier系データリポジトリへデータアップロードを推奨
- Scientific Data (Springer-Nature)
 - ・ データ論文を「Data Descriptor」と称する。
- Earth System Science Data (Copernicus Publications)
 - ・ インパクトファクター：11.333！！
- Polar Data Journal (国立極地研究所)
 - ・ 小さいデータならJAIRO Cloudに収容してもらえる？

Open Science



Good for science: limit research duplication, ensure verifiability, ensure replicability,



Good for the economy: uptake of results by businesses, esp. SMEs → innovation potential



Good for society: higher level of citizen and civil society trust in science, open and collaborative research practices lead to high degree of responsiveness and adaptability to societal challenges

[Celina Ramjoué, Deputy Head of
Data Policy and Innovation Unit,
DG CONNECT, 2017]

海外学協会におけるオープンサイエンスへの対応例

AGU Position Statement

(村山仮訳)

AGU Position Statement (1997~) : 地球科学・宇宙科学のデータは複数の形式でアクセス可能であるべき。その長期間保存は科学者とその所属機関がかならず負うべき責任である。



米国地球物理連合

- 地球・宇宙科学データは、人類の資産である。
- 科学者・一般社会が入手可能であるべき

Earth and Space Science Data Should Be Widely Accessible in Multiple Formats and Long-term Preservation of Data is an Integral Responsibility of Scientists and Sponsoring Institutions

Earth and space sciences data bases are a world heritage that should be made available to the scientific community and public as soon as possible (in some cases in real-time), should be organized and preserved in useable format, and should be conserved long-term for future use. The responsibility for achieving this falls upon individual scientists and their sponsoring institutions, and should be considered an integral part of conducting scientific research.

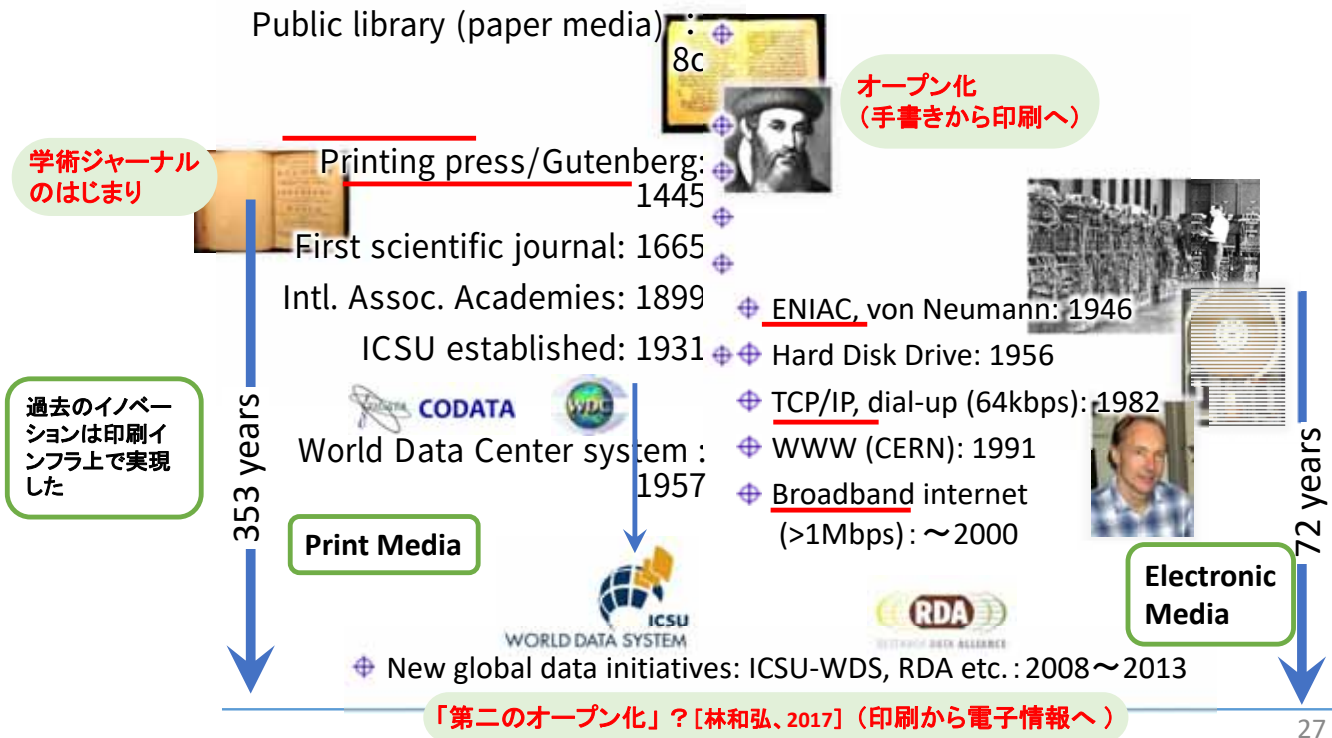


- Earth and space science data collection, analysis, and archiving are essential to our understanding of the natural environment and how it changes with time. AGU policy is

Print & Electronic Technologies as Social Info. Infrastructures

--- 百年の印刷文化の基礎支えと、成長途中の電子基盤上のサイエンス

Public library (paper media) :



FAIR Guiding Principles

[Shelley Stall (AGU), 2018]

FAIR is...

Findable

Accessible

Interoperable

Reusable



Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).



[Sarah Jones, 2018]

永続的識別子 (PID)
(DOIとは書いてないが)

Findable

FL (meta) data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

Interoperable

ii (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

12. (meta)data use vocabularies that follow FAIR principles.

13. (meta)data include qualified references to other (meta)data.

doi: 10.1038/sdata.2016.18

Accessible

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

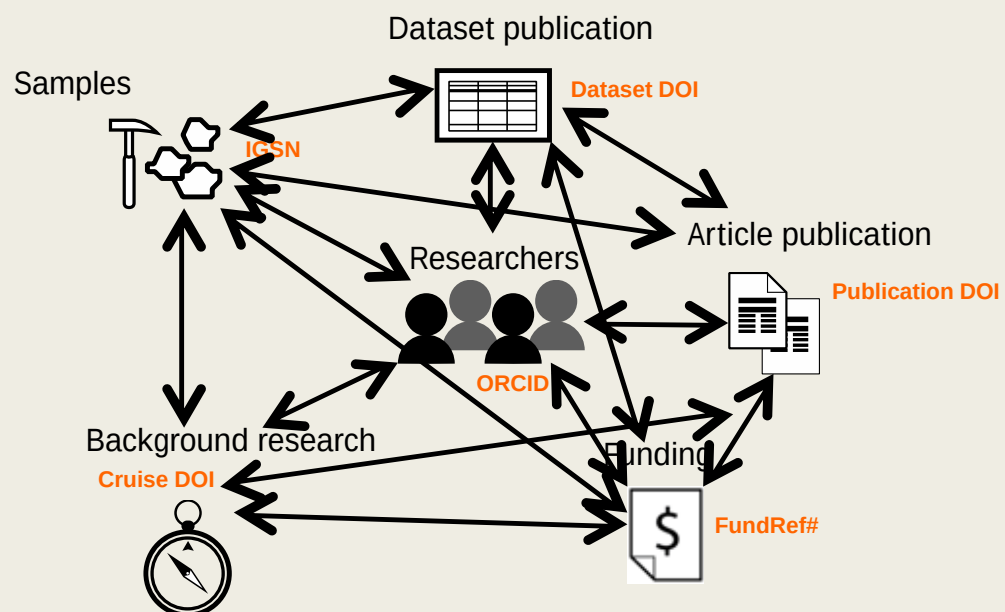
Reusable

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3 (meta)data meet domain-relevant community standards.



DOI (digital object identifier)と恒久性

重要なことは

- DOIの不変性は、「国際DOI財団(IDF)」と加盟組織が担保する
- URLは変化しても、DOIは不変。URL変更時は登録変更。

文献のDOI (恒久的識別子)

10.1002/2014JD022647

「解決(resolve)」されたURL

<http://onlinelibrary.wiley.com/doi/10.1002/2014JD022647/abstract;jsessionid=453F8F55438032C647492980DD0AB02.f02t01>

ランディングページ (論文書誌情報など)

実際の 論文PDF



データのDOI (恒久的識別子)

doi:10.17591/55838dbd6c0ad

「解決(resolve)」されたURL

http://www2.nict.go.jp/isd/doi-landingpage/wds/10.17591_55838dbd6c0ad.html

国内登録機関JALC 経由の最初のデータDOI ランディングページ (メタデータ情報)

実際の データ



データへのDOI付与時の課題

議論と課題:

- DOI-URLの対応関係は永続維持されるが、
- ランディングページやデータコンテンツは、出版論文と異なり、永続保存の状況は国、機関、分野ごとに異なる。
- 「恒久的」な保存は、容易でないのはよくわかる
 - WDCなどは50年以上してきた
(例外的? → だからこそその国際組織)
- **大規模なデータは今後、政策的な課題としてとらえていく必要がある?**
- **小サイズのデータは、もっと容易な扱いはできないか?**

DOI (恒久的識別子)

doi:10.17591/55838dbd6c0ad



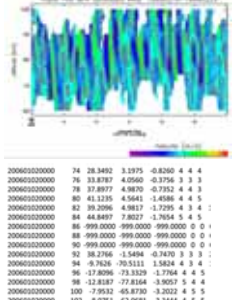
「解決(resolve)」されたURL

http://www2.nict.go.jp/isd/doi-landingpage/wds/10.17591_55838dbd6c0ad.html

ランディングページ



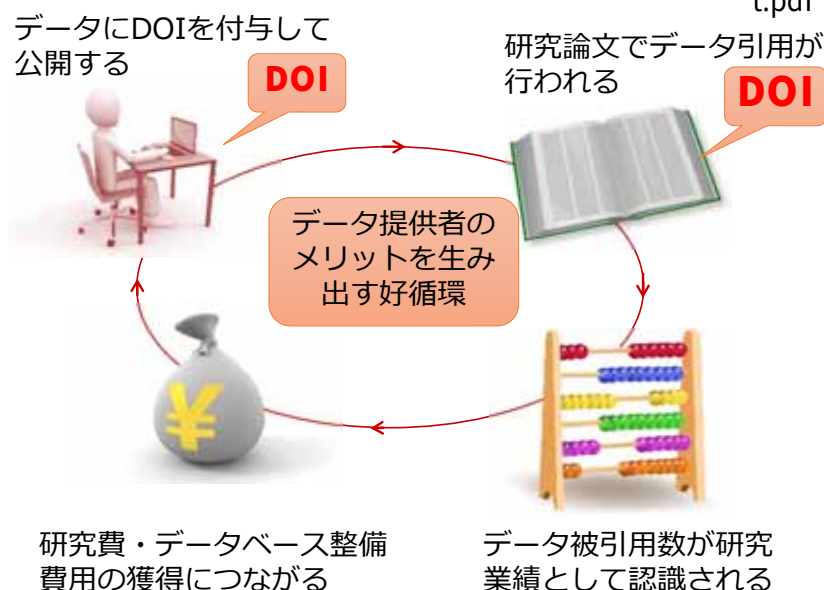
データ



データにDOIを付与、参照するメリット

研究データ利活用協議会（RDUF）・リサーチデータサイテーション小委員会（能勢正仁委員長）
作成リーフレット「研究データにDOIを付与するには？－5分で分かる研究データDOI付与－」
（2019年12月）より https://japanlinkcenter.org/rdof/doc/rdof_rdc_doilead

https://japanlinkcenter.org/rduf/doc/rduf_rdc_doileaflet.pdf

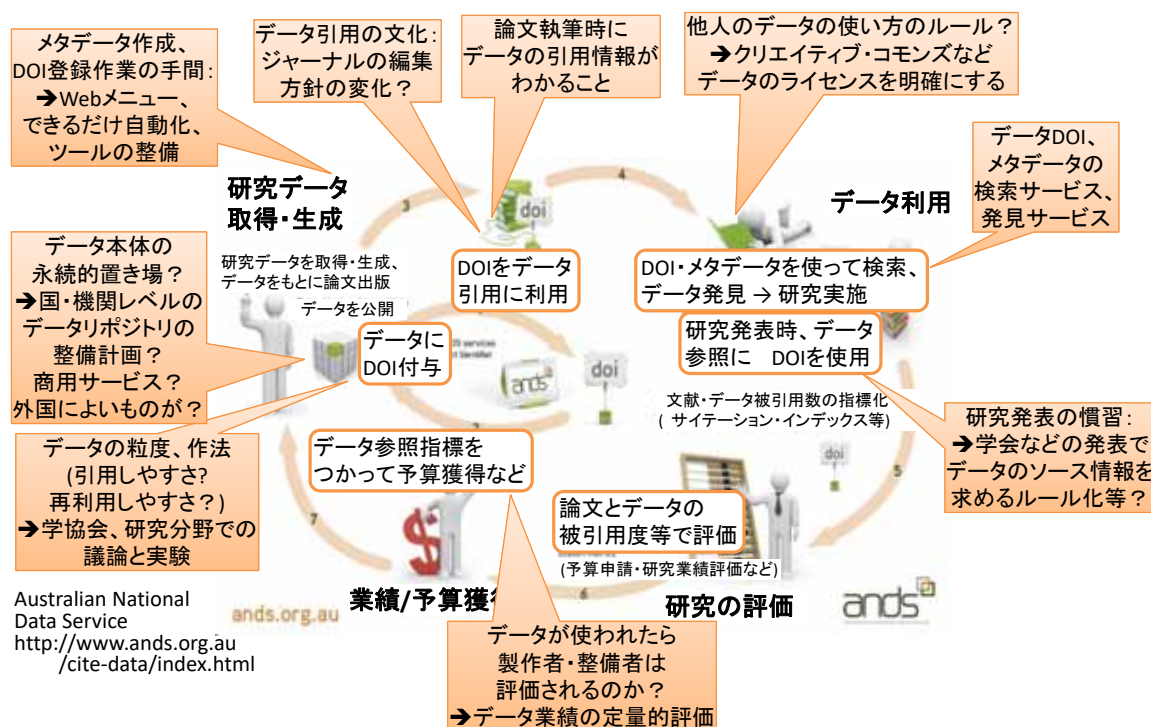


- データ提供者のメリット
 - DOIを付与することにより、データ「出版」として認識されるようになります。
 - DOIを用いて研究論文にデータが「引用」されることで、研究成果として評価されるようになります。
- 研究機関のメリット
 - 自機関が生成した研究資源・学術資産の管理・利用を促進出来ます。
 - 研究成果の散逸を防止することが出来ます。
- ユーザーのメリット
 - 研究資源へのアクセスしやすさ、利便性が向上します。
 - 論文と同じようにデータを引用することが出来ます。



<https://japanlinkcenter.org/rduf/>

データサイテーションのエコサイクルができるために必要なこと(の例)



(和訳は能勢(京大、2014)を参考にした)

AGU、EGUのアクティビティ事例

- AGU Earth & Space Science Informatics (ESSI) Section



Earth & Space
Science Informatics

- AGU Fall Meeting 情報学セクション: セッション提案61 (2018)、実績35程度? (2016、2017)

- EGU Earth & Space Science Informatics (ESSI) Division



Earth & Space
Science Informatics

- EGU General Assembly 2017
情報学カテゴリ: 22

- JpGU-AGU meeting 2018にて: 「JpGUでも情報・データに関するコミュニティをつくるべき」(AGU Board Dr. Kerstin Lehnert 他)

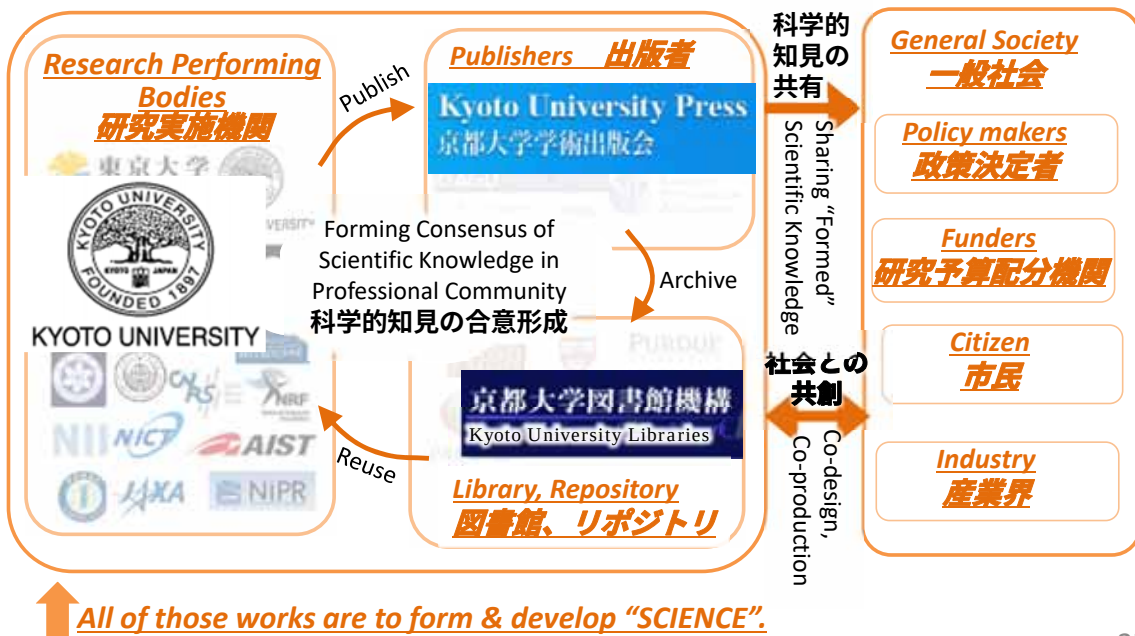
35

**電子メディア時代の学術ジャーナル
～学術情報配信メディアおよび
学術評価尺度としてのジャーナル**

Science as a Social System (at least on the “Print” technology so far) 社会的仕組みとしてのサイエンス

注：ここでいうサイエンスは、近代科学原理に基づくすべての学問領域を指す。

NB: “Science” in this talk is all the scholarly/scientific research domains (including humanity, social sci., eng.,...)



37

Peer-reviewed scientific journals

French Mathematician, CNRS Sci. Director, Advisor to the EU ministers

[Marie Farge, 2017]



The publication of research results in peer-reviewed journals is the backbone of the present research system.

It was founded on January 5th 1665 for sharing ideas and results.

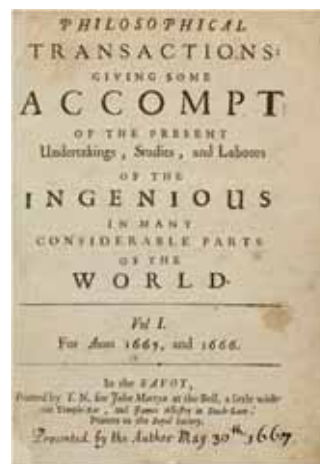
It is also used today for evaluating researchers and projects.

<http://wavelets.ens.fr/>

- 学術出版システムは印刷ビジネスモデルに基づく
- ジャーナルは出版社に所有されている
- 査読者＝研究者への報酬は無いが、
- 出版者は論文を印刷して図書館へ売っている



Paris, 5th January 1665



London, 6th March 1665

The scientific publishing system is based on the business model of printing: journals belong to publishers, researchers peer-review articles without being paid by publishers for this, then articles are printed and sold by publishers to academic libraries.

Researchers want to recover control

[Marie Farge, 2017]

‘Neither author nor reader should have to pay to publish and a journal should not belong to its publisher but to its editorial board.
The dissemination of the peer-reviewed articles should be done using public infrastructures, from where articles should be accessible for free.’

Marie Farge, Note for the French Minister of Research, June 29th 2012
http://openscience.ens.fr/MARIE_FARGE/

- ・著者も読者も出版費を支払わないほうがよいモデル？
- ・ジャーナルは出版者の所有物でなく、Editorial Boardに帰属すべき。
- ・論文配信は公的な基盤によっておこなわれるべきであり、
- ・個々の論文は無料でアクセス可能であるべき’

(Marie Farge, 研究担当相への覚書, 2012年6月29日
http://openscience.ens.fr/MARIE_FARGE/)

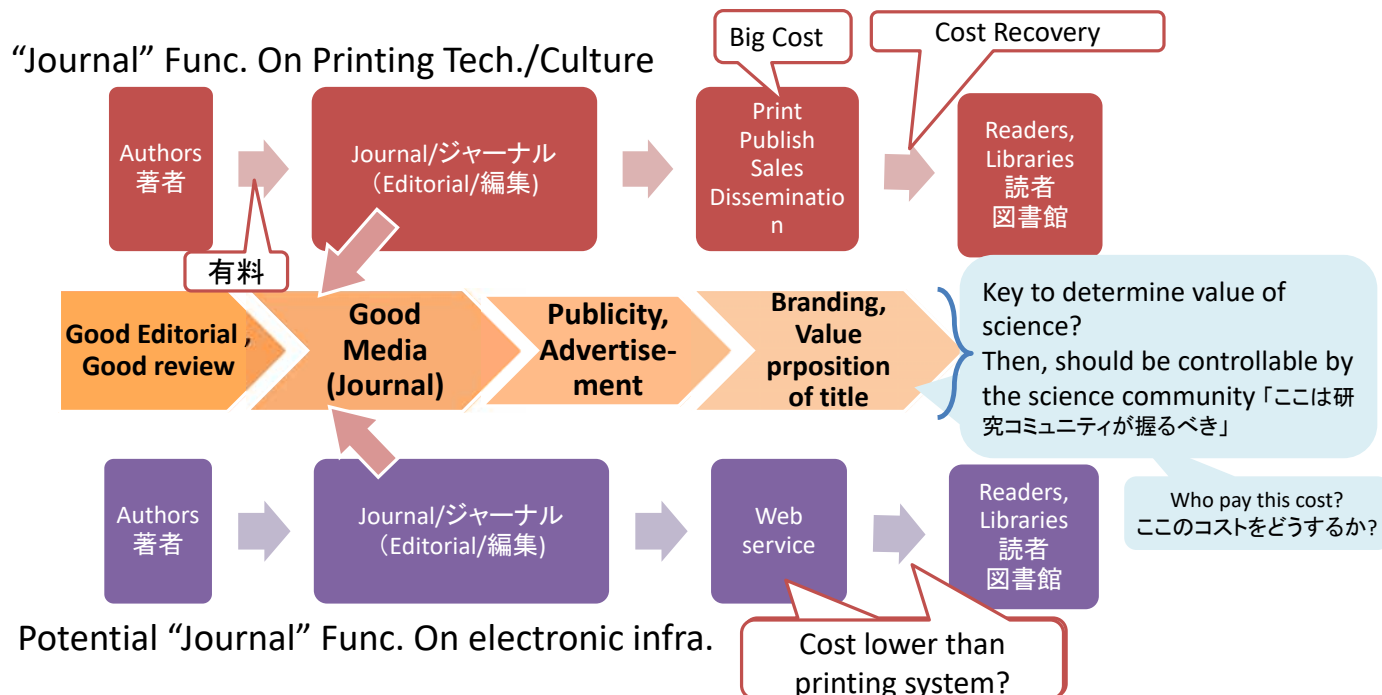


Researchers proposed an alternative model : *Diamond Open Access*

Diamond Sutra, the earliest complete survival of a dated printed book, China, 11th May 868 British Library, London

Publication Process (Academic Output Sharing Mechanism) and “Journal”

(The author’s personal opinion based on articles of Marie Farge etc.)

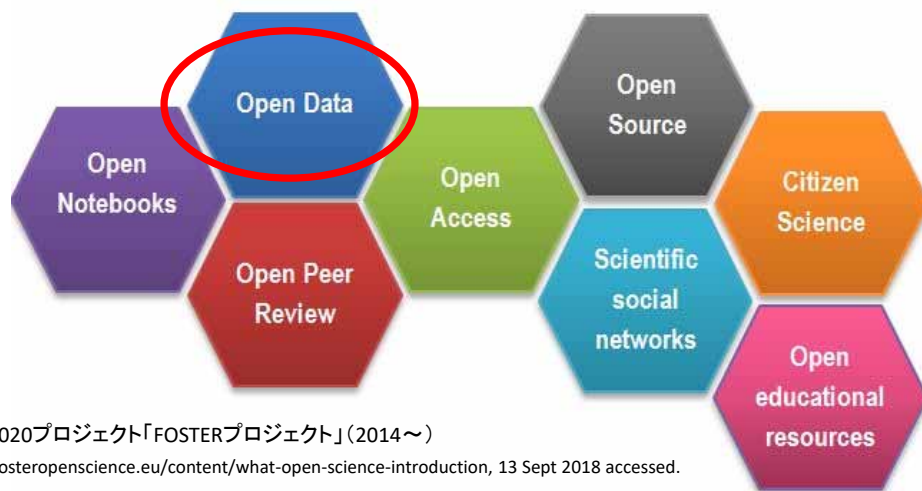


別視点のジャーナルの価値

- 「査読付き論文が学術評価の基盤」(Marie Fargeなど)
- Editorial Policy: どういう論文を採択するか、判断基準をもつ
＝科学のあり方を規定する
 - － (投稿者は採択されるように書く傾向)
- ジャーナルの編集、査読のあり方は対象分野の科学研究方向性、科学のあり方の基準となりうる
 - － (だからこそ、学術会議の規定する協力学術研究団体の規定に「学術研究(論文等)を...発行している」がある?)
- 誰が「科学のあり方」を決めるべきか?
 - － 科学者本人が、「科学のあり方」の決め方に関与すべきではないのか。
→ 科学者がジャーナル方針に関与すべきではないのか。
 - － 何が科学の成果なのか、考えるのも決めるのも科学者自身であるべきではないのか。
 - － 同時: 「社会の知的資産としての科学者」の役割: 科学がどうあるのが社会と科学にとってベストか、を追及する責務があるのではないか。

**科学政策、国際機関等における
オープンサイエンス、データ共有**

オープンサイエンスのさまざまな切り口



EC Horizon2020プロジェクト「FOSTERプロジェクト」(2014～)

<https://www.fosteropenscience.eu/content/what-open-science-introduction>, 13 Sept 2018 accessed.

科学者、専門家による価値判断が必要なデータ

本議論の主対象

(気象庁、国土地理院等のデータは?)

注意: Open Research Data
研究/科学データ共有

Open Government Data
公共データのオープン化

43

科学政策におけるオープンサイエンス

• 欧州委員会：

オープンサイエンスは、デジタルテクノロジーを用いた相互協力および知識の伝搬に基づく、科学研究への新しいアプローチといえる。

(European Commission (2016), Open innovation, Open Science, open to the world. A vision for Europe, 2016b:DOI: 10.2777/061652.).

• OECD:

オープンサイエンスは公的資金による主な研究成果（論文出版と研究データ）を一般にたいして、最低限の制約をもってデジタルフォーマットでアクセス可能とすること

(OECD (2015), Making Open Science a Reality, <http://dx.doi.org/>



(image by OpenAIRE, CC-BY)

“オープンサイエンス”（～新しいデータの取り扱い）の諸外国動向の例 (JST科学技術情報委員会調査より) H27.4

(1) OECD 科学技術政策委員会閣僚級会合宣言¹

2006年12月、OECD加盟国において“OECD Principles and Guidelines for Access to Research Data from Public Funding(公的資金による研究データへのアクセスに関する原則及びガイドライン)”が承認された。研究データへのオープンアクセスを実現し、その広範な利用を促すことが、研究の質と生産性の向上につながるという認識のもとに、「公開性」、「透明性」、「遵法性」、

(4) 米国国立衛生研究所(NIH) データ共有ポリシーと実施要綱²

2003年10月、NIHはData Sharing Policyを制定し、研究資金公募時におけるデータ管理計画(Data Sharing Plan)の提出を義務化した。本要綱では、NIHの助成研究によるデータの共有理由として、「科学研究のオープン化の強化」、「分析や見解の多様化の奨励」、「新たな研究の推進」、「新規又は代替仮説と分析方法の検証」、「データ収集方法や手段に関する研究の支援」、「研究者育成」、「最初の研究者達が想定しえなかった課題の模索」、「複数のソースからのデータを組み合わせる新しいデータセットの作成」を挙げている。

(6) 英国研究会議(RCUK) Common Principles on Data Policy³

2011年、RCUKは「公的助成を受けた研究データは公益財であり、できる限り制限なく、適時にまた知的財産を害することのないよう責任ある方法でオープンに利用できるようにする」という基本原則のもと、“Common Principles on Data Policy”を公表した。本ポリシーは、政府助成研究成果に対する透明性の確保と統一的研究基盤の整備を目的として、研究データのオープン化に関する英国内各研究会議に共通の原則を定める包括的枠組みと位置づけられている。この枠組みに沿って、下記参考のように、英国における多くの研究助成機関ではデータポリシーが発行されており、データマネジメントおよびシェアリングの計画書提出を要求しているものが

45

G8、G7における科学データ政策潮流とオープンサイエンス（＝データ問題）



2013年

- ・ 2013年 G8サミット（英）：G8国研究データオープン化合意
- ・ 「データ」を重要な科学技術の研究成果として位置付け。
→国内外の科学政策動向へ影響。



2016年

- ・ 2016年 G7科技大臣会合（日本）：→G7オープンサイエンスWGの設置

村山：2018年より
OS-WG共同議長



科学大臣会合(2016)にて講演。村山と林上席研(NISTEP)。
写真：2016年G7科技大臣会合写真：内閣府・茨城県・つくば市提供



G7部会にて。左から村山、原山CSTI 議員、島尻科技担当大臣
(肩書は2016年当時のもの)



2017年

- ・ 2017年G7科学大臣会合@イタリア
- ・ 2018年G7科学シエルパ会合@カナダ
- ・ 2019年G7科学シエルパ会合@フランス
- ・ 2020年G7科学大臣会合@米国
- ・ 2021年英国、2022年独、2023年日本...



2019年

オープンサイエンス、
デジタルデータ問題：
「Incentive, rewards
＝やる気・業績化」と
「インフラ」

(→100年単位での情報保存) 46

G7 Science Communiqué (Turin, Italy, 2017): Paragraph 19

An international approach can help the speed and coherence of this transition (to Open Science).

It should target two aspects:

- **Incentives for the openness** of the research ecosystem
 - **evaluation of research careers** should better recognize and reward Open Science activities.
 - **Infrastructures** for an optimal use of research data
 - all researchers should be able to deposit, access and analyse scientific data across disciplines and at the global scale
 - research data should adhere to **the FAIR principles** (findable, accessible, interoperable, and reusable).
- 「研究エコシステム」のオープン化へのインセンティブ
 - オープンサイエンス活動を業績としてみとめる+研究者のキャリア形成に貢献させる
 - 研究データの保管、アクセス、解析が行える、国際的に通用するデータ基盤（広義）の必要性
 - FAIRデータ原則

47

[林和弘、内閣府CSTI政策討議、2018/1/25]

7 政策討議でのオープンサイエンス

- Open ≡ 戦略的開放 ≠ 誰でも自由に
 - 我が国の研究、産業、文化振興と社会が発展する戦略・方針(シークレット→クローズ→オープン)を持った上で
 - ・ 研究者(関連分野、非関連分野)への開放
 - ・ ステークホルダー(研究助成団体等)への開放
 - ・ 広く産業・社会への開放
 - ・ (その中の1オプションとしてフルオープンも含まれる)
- 戦略を支える基礎的な方針の重要性和留意点
 - 研究データを資源とし最大限活用する
 - これまでのオープン・クローズ戦略だけでは通用しない
 - 新たな科学研究(と関連産業)が加わる
 - ・ オープンな情報基盤を活用したビッグデータ研究、市民科学等

総理官邸、CSTI議員、内閣府審議官等が参加しての議論。官邸側も大筋合意。

48

欧・日：科学技術による社会・経済活性化戦略

[村山・林(2016),
STIホライズン、Vol.2, No.3,
doi:10.15108/stih.00044]

・日本・第5期科学技術基本計画 超スマート社会、Society 5.0の実現

- サイバー空間と現実社会を融合した超スマート社会の未来
- サービスや事業の「システム化」、システム間の連携協調



翻訳するならば、「社会・経済・市民全体で活用できるデジタルデータプラットフォーム→「デジタルで社会の力を取り戻す」ではないか。

[内閣府,
<https://www8.cao.go.jp/cstp/kihonkeikaku/5gaiyo.pdf>]

欧州デジタル単一市場戦略 (2015～) と、 欧州オープンサイエンス・クラウド

デジタル単一市場戦略：
欧州クラウドを支援する



欧州オープンサイエンスクラウド計画
(EUは67億€投資。内20億€はHorizon2020)。

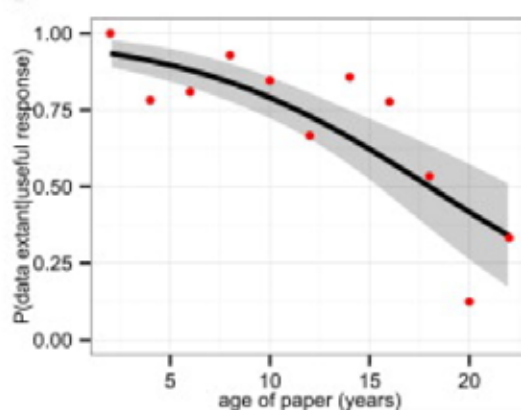
[from <http://www.eurogeography.eu/SoC/guidelines/stakeholders.html>,
translated and modified by Y. Murayama]

- ・「社会の基盤」⇒科学インフラも含めた投資。基盤の変化により新たな発展を目指す。
- ・欧米では新たな“ルールメイキング”を狙う？ (コミュニティ統合・包摂的な戦略決定⇒地域ガバナンスにも有用？)49

Credibility of Science

インターネット上のデータは、18年で50%が失われる？

- Access to original and complete data sets for reproducibility
- Re-usability declines with time
- Availability declines with age



<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308#pone-0000308-g002>

<http://www.sciencedirect.com/science/article/pii/S0960982213014000>

[Wim Hugo, UN World Data Forum, 2017]

テクノロジーだけで変革する世界ではなくなっている



Session 5: STI capacity building for achieving the SDGs

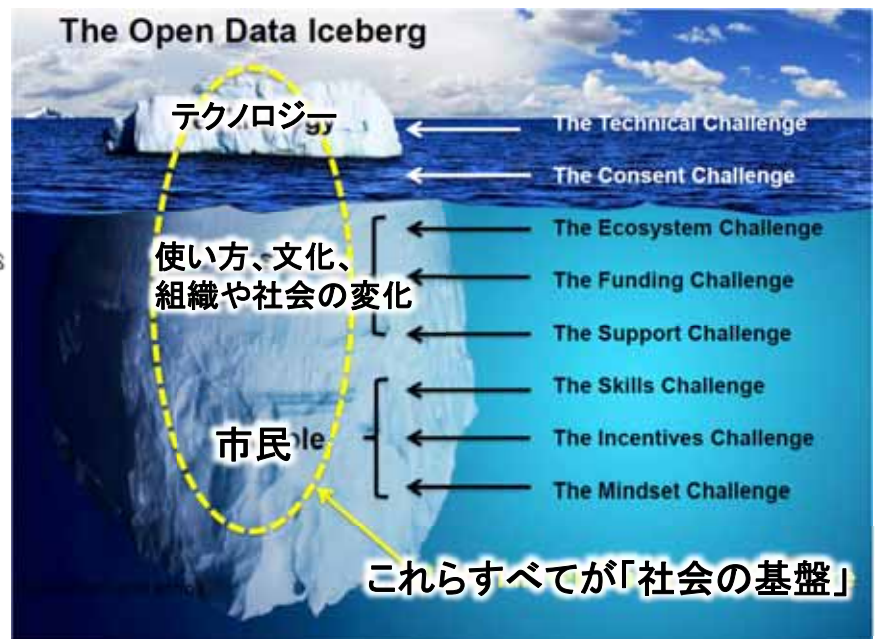
This session will build upon the previous one to explore existing STI policies for capacity building and science advisory systems, and discuss ways and means to bridge the gaps. The session will also review what it takes to

TUESDAY, 16 MAY 2017

12:00 PM - 1:00 PM

CONFERENCE ROOM 4, UNHQ

ジェフリー・ボールドン教授(英エジンバラ大、2017)の講演スライドより:



https://sustainabledevelopment.un.org/content/documents/26001Session_5_Geoffrey_S._BOULTON.pdf に村山が和訳、補足を追記

51

デジタルの束縛 Curse of “Digital”?

- 長期的なデジタルストレージ維持管理、利用のコスト
 - 紙媒体保存よりコストがかかるのではないかと
Costly more than paper-based preservation
 - データ読み書き、環境は維持可能かと
Long-term accessibility to data and data-contents
 - 新規アーキテクチャへの乗り換え(マイグレーション)を続けるのか
Eternal iteration of storage/data system migration?
 - これを凌駕できる社会経済的ベネフィットを生み出すには⇒そのためのFAIR原則
Ultimately can produce socio-economic benefit to cover them?

“オープンサイエンス”推進運動に伴う 研究評価の変化（＝学術界の文化・規範変容） （の兆し）

San Francisco Declaration of Research Assessment (DORA)

多様な科学研究活動のアウトプット（論文、データ、ソフトウェア、試料、etc.）が成果として認められるべき

- Promotes value of all scholarly outputs
 - Journal articles
 - Preprints
 - Datasets
 - Software
 - Protocols
 - Research materials
 - Well-trained researchers
 - Societal outcomes and policy changes
- Focuses on the merits of the work
 - Reduces JIF-chasing
 - Facilitates Open Science practices
 - Improves rigor and reproducibility
 - Enhances collaboration

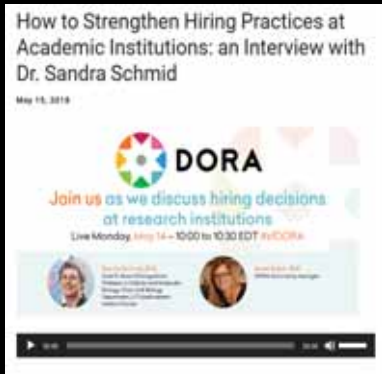
インパクト
ファクターによる
研究評価は
やめよう



Change is happening: Research Institutes

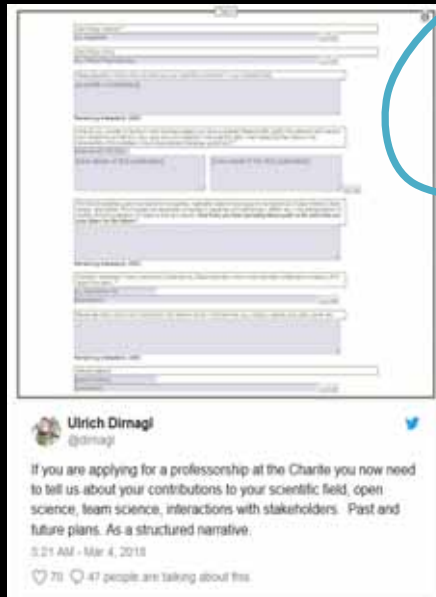
学術業績評価にオープンサイエンス活動を含める事例がではじめている

<https://sfdora.org/>
日本語コメントは
村山追記



UT Southwestern Medical Center
Dallas, Texas

Candidates receive questions before Skype interviews, so they have time to reflect. The goal is to identify thoughtful individuals, in addition to candidates who process information quickly.



Charité University Hospital
Berlin, Germany

Application includes

- Contribution to research field
- Open Science
- Team Science
- Interactions with stakeholders

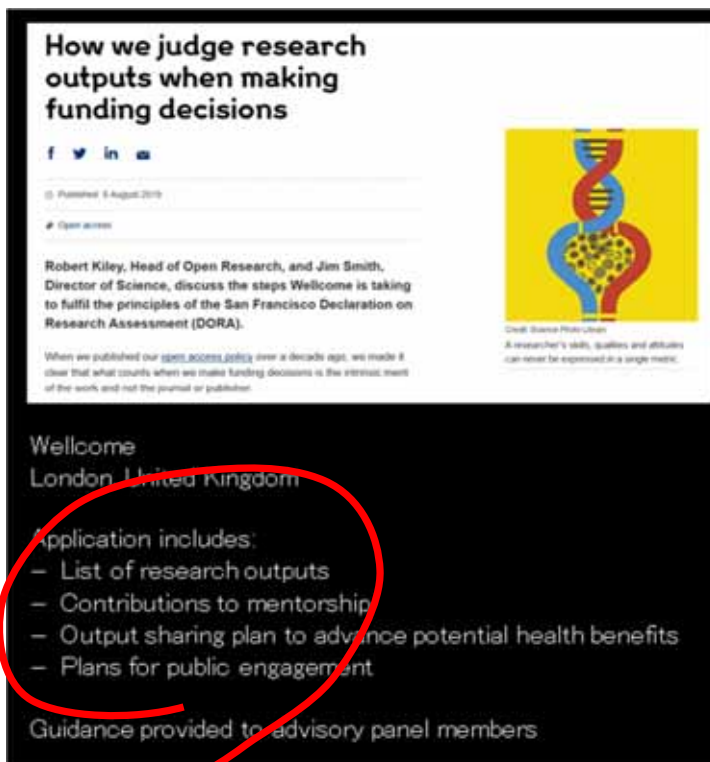
Staff member sits on hiring deliberation meetings as a neutral party to promote balanced discussions

独Charité Univ.
Hospital :
研究職応募書類の
業績に「オープン
サイエンスの活
動」記述欄を設け
ている



55

Change is happening (Funders)



Wellcome Trust Web site
(2018)

- **Improvements to our grant application and decision process**
- “We’ve modified our **grant application forms** so that now we ask researchers for their outputs, which may include (but are not limited to) **preprints, datasets, software, research materials and inventions, patents and other commercial activities**. Previously, we just asked researchers to cite their research publications.”

英Wellcome
Trust(研究資金配分
機関):
• 研究費申請書類
の業績欄に多様
な研究成果(論文、
データ、コード、特
許...)を記載できる。
など。

https://sfdora.org/wp-content/uploads/2018/10/18-1016-DORA-slides_short-presentation_FINAL.pptx



GDS (Greater Data Science)

[David Donoho (2017) 50 Years of Data Science, *J. Comput. Graph. Stat.*, 26:4, 745-766,
doi:10.1080/10618600.2017.1384734]

1. *Data Gathering, Preparation, and Exploration*
データ収集、データ整備、データチェック
2. *Data Representation and Transformation*
データ・DBの形式整備、再構成
3. *Computing with Data*
データ処理
4. *Data Modeling*
データ生成・予測のモデル (?)
5. *Data Visualization and Presentation*
データの可視化 (通常のデータ描画～高度化)
6. *Science about Data Science*
データサイエンスに関するサイエンス



<https://profiles.stanford.edu/david-donoho>

David Donoho
(Dept. Statistics of
Princeton University,
winner of von Neumann
Prize, Gauss Prize, and
fellow of American
Mathematical Soc.)

まとめにかえて (1)

- オープンサイエンスの近年の科学政策的動向
 - 国際的に推進合意、G7国内でも徐々に推進
 - 日本：内閣府、CSTIが主に推進（近年は文科省も）。デジタルデータの重視。
- オープンサイエンス、研究データの業績
 - 近代科学原理のもとで、新しいデジタル時代での最適化
⇒科学の原理を問い直す ⇒ あらためて基礎研究、積上げがいかに重要かを再認識すべき時
 - 公共データと科学データはポリシーが違う ←
- 新しい科学のシステムへむけて (事例)
 - 新しいデータ取扱いの作法、システム
 - データ引用、データ出版、データ業績 ⇒人事評価などへも反映させる方向
 - 欧米は学協会、学術会議、政府などが主導する方向
 - 予算配分からジャーナルポリシーまで、統一的な科学システムの変化
- 社会は「デジタル変革 (Digital Transformation)」：
 - →社会と科学の共創、協働、相互信頼：科学の再認識

(研究評価に関する
サンフランシスコ宣言、
2012)



<https://sfdora.org>

まとめにかえて（２）

- オープンサイエンス推進の流れ
- 科学研究における研究成果の位置づけ
 - 論文（ジャーナル出版）
 - データ、ソフトウェア、サンプル資料、etc.
- 科学の全体像のあり方
 - 三種の神器：研究＋出版・共有＋保存・再利用
- 科学成果の評価、「測り方」
 - 「科学成果」の拡張と評価
 - ジャーナルメディアあるいは科学成果「出版」の将来像
 - 科学成果のあり方を、科学者が社会とともに考える

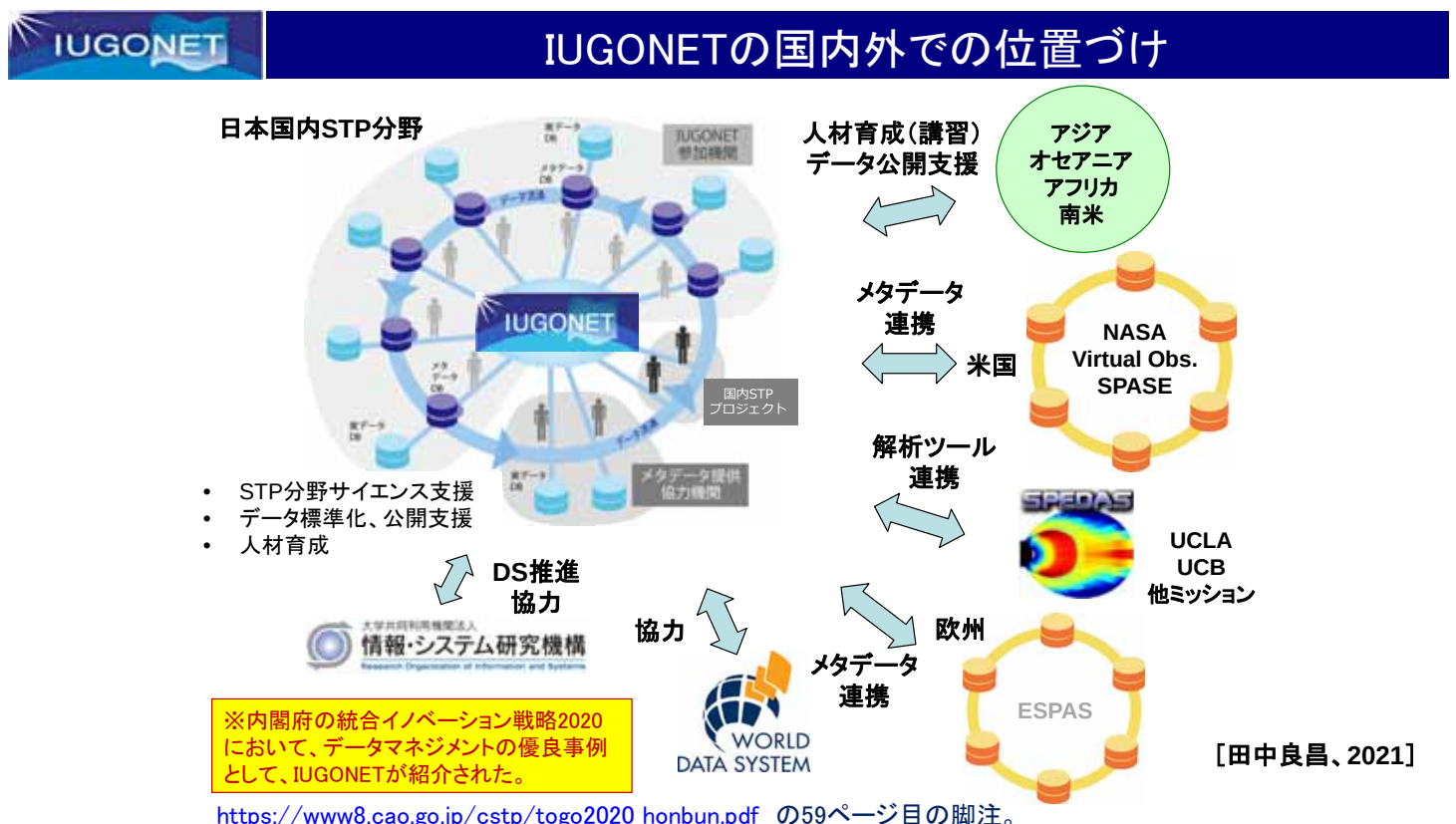
まとめにかえて（３）

- 「オープンサイエンス」という名前
 - 近代科学のやり方を、いまの社会で最適化したい
 - 論文だけで科学・社会はいいの？
 - データを科学者の成果、評価に
- データ（やその他の科学研究過程での生成物）を出すこと、出さないこと、評価されること
 - 研究分野に要請がある分野、ない分野
 - 海外の学会で推進
- 実践上の諸課題をどうするか
 - いよいよ具体的な作業段階へ？
 - 誰がデータを保存するのか（捨てるのか）
 - 保存するための社会的仕掛け＝信頼度評価CoreTrustSeal、WDS、DataSeal of Approval. . .

Fin.

Questions?

Contact: murayama@nict.go.jp
or
murayama.yasuhiro.r92@kyoto-u.jp



まとめにかえて：いくつかの注意点

- “オープンサイエンス”（仮称？）や「オープンデータ」は絶対正義か？
 - タイムスケール？：明日、1年後、3年後、30年後
 - オープン・クローズは後回しでも、データを整えるのが大事
（データの相互通用性=Interoperabilityは本質の1つ！）
- 優先すべきは、長期的な科学の健全な発展
 - 人口（学生）減、資源枯渇、気候変動による経済損失：
社会のためにも子孫の代で日本の科学者職がなくならないように！
 - 国際交流・連携：研究プロジェクトや成果（論文等）だけではなく、「科学のあり方」も国際学会・機関と意識をあわせる必要（← 輸入文化としての科学問題）
⇔ 国内の科学の未来、科学と社会との相互信頼（社会における科学という存在）
- データマネジメント人材は、研究者か研究支援者か？
 - **研究者**としてやるべきデータマネジメント部分は大きい！（ポスト問題）
- **DON'T regulate what we don't yet understand.**
「理解していないものを規制してはいけない」
（“The Data Harvest”, RDA Europe, December 2014）

SDGsと
科学！

Examples to approach Open Science (data sharing etc.)

- Academies **アカデミー・学術会議の例**:
 - US: Board on Research Data and Info. (Natl. Academies) 2007 or 2011(?) ~
 - JP: Committee for examining Open Science 2015~
日本学術会議 オープンサイエンスの取組に関する検討委員会2015年～、等
It was followed by committees for OS promotion/platform successively.
 - Flexibility of public funding is sometimes limited for community engagement and consensus forming.
公的科学研究資金でオープンサイエンスの議論・合意形成等に使えるお金が限定的
- Institutional/General Data Repository/ **機関・総合データリポジトリの例**
 - UK、Europe、US：EUDAT、OpenAIRE、EOSC、Univ./institutional data repositories, Zenodo, figshare...
 - JP: Often project-based data storage, not very “FAIR” nor very sustainable.
プロジェクト資金でデータ保存、“FAIR”原則・持続性に問題。
Recent trials of data preservation at dept. and/or library in universities.
大学や公的研究機関の部署や図書館で研究データ保存の試みが近年始まっている
 - Development of NII Research Data Cloud
 - Kyoto Univ.'s KURENAI, Nagoya Univ., Tokyo Univ. ...

いくつかの注意点

- 「オープンデータ」は絶対正義か？
 - 優先すべきは、長期的な科学の健全な発展
 - タイムスケール？：明日、1年後、3年後、30年後
 - 「誰が」「どう使ってよいか」などルール（and/or ライセンス）が必要
 - オープン・クローズは後回しでも、データを整えるのが大事
 - データの相互利用性=Interoperabilityが本質
- FAIR data principle :
Findable, Accessible, Interoperable, Reusable
- **DON'T regulate what we don't yet understand.**
「理解していないものを規制してはいけない」
(“The Data Harvest”, RDA Europe, December 2014)

多くの疑問

- 自分のHDDの中のごまんとあるデータがすべてオープン化対象なのか？
- 研究データが研究者・大学・機関の業績として評価されることなんてあるのか？
- 科学者の（雑用の）負担が増えるだけでメリットがないのではないか？
- データ保存の予算もデータ整理の人員費もかかるが予算は減る一方。一体、現実的な話なのか？