

## What I wish to talk today

### オープンサイエンスと研究データ共有の 国際政策、動向について → Consideration on Possible Directions of International Open Science Scene and ICSU-World Data System 国際的なオープンサイエンス動向とICSU-WDSの方向性

村山泰啓

日本学術会議・連携会員  
 ICSU-WDS Scientific Committee ex officio  
 情報通信研究機構

WDS International Programme Office Hosted by **NICT**  
 Based in Tokyo, Japan

1

- Open Science, Research Data Sharing  
 オープンサイエンス、研究データ共有の流れ
  - 学術活動の価値の転換(または拡大): データの科学業績化
- What "data" is for OS & RDS  
 対象となる「データ」とそのアウトプット
  - 「データ」の「整理学」が不完全 Incomplete "data" taxonomy and beyond
    - Webテキストデータ、書誌(メタ)データ、数値測定データ、...
  - データ出版、データジャーナルの新しい考え方
- ICSU-WDS Activities and Discussion
  - WDS-International Technology Office
  - Metadata Catalogue、Knowledge Network
  - Scholix Initiative (framework for SCHolarly [data-literature] Link eXchange)
  - CoreTrustSeal: Intl. Certification of Trustworthy Data Repository
  - Brokering Approach → EarthCube?, next interoperabilityの考察

## Locating myself in Open Science scene

### Geophysics, Atmospheric/Space Science

- Board member of Japan Geoscience Union (2014-2015)
- Visiting Professor, Kyoto University, (2013-2014)
- AGU Earth & Space Science Informatics Section Exec. Committee ('18-)
- IUGG Union Commission of Data & Information (2017-)
- Member, Science Council of Japan
- Research Executive Director, Strategic Program Office, NICT
- ICSU-World Data System Scientific Committee, ex officio member
- Member, Expert Panel of Open Science, Cabinet Office
- EC's High Level Expert Group of European Open Science Cloud (2015-2016)

### Data Management Advocacy, Science Policy

3

## G8、G7におけるオープンサイエンス、科学データポリシー

### 国際動向、国際政策

- 2013年G8サミット(英): G8国オープンデータ合意
- 「データ」を重要な研究成果として位置付け。  
 急速に国内外の政策動向が変化しはじめた。
- 2016年G7科技大臣会合: RDA推進を明記  
 オープンサイエンスセッション(村山・林講演)   
 → G7オープンサイエンス部会の設置
- 2017年G7イタリア

- データ業務のインセンティブ、評価
- 国際共通データ基盤



WG:左から村山、原山議員(CSTI)、鳥尻前科技担当大臣

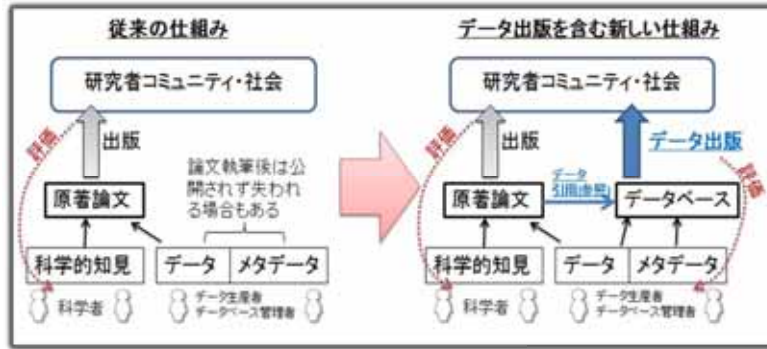
2016年G7科技大臣会合写真:内閣府・茨城県・つくば市提供

4

## 論文だけでなく、データも研究成果

### ■ データを公表(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。
- 信頼できるデータの生成・提供は現代では科学者の仕事ではないか。← 評価



[地球電磁気・地球惑星圏学会, 2013]

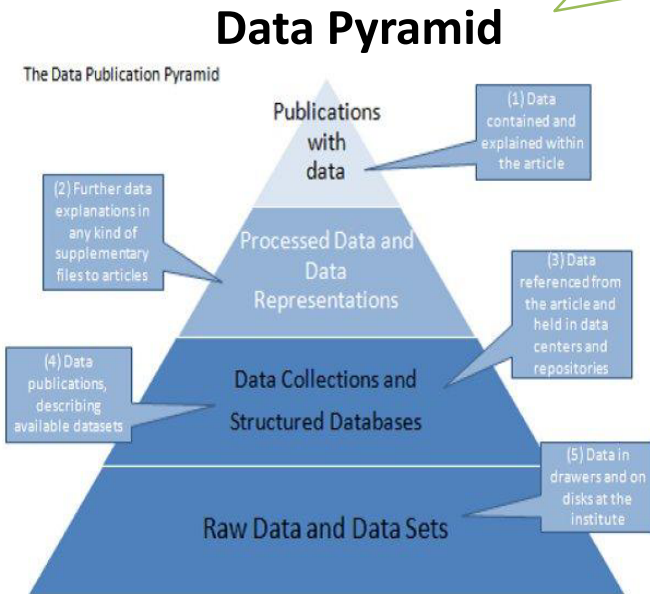
5

## 学術論文出版時のデータポリシー (地球惑星科学分野の大手ジャーナル誌の例)

- **Wiley/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..."

[H. Frederick Dylla, 2012]

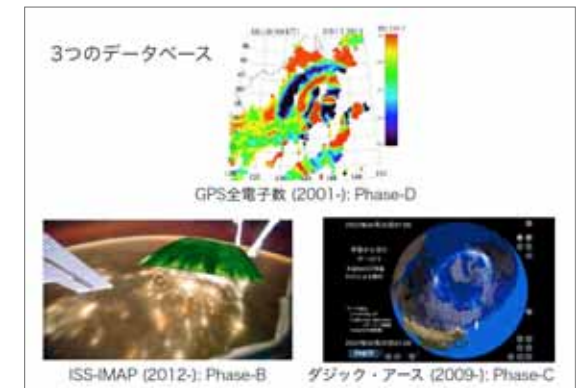
どんなデータが管理対象か



7

## 地球科学研究者の見るデータ整備のフェーズ例

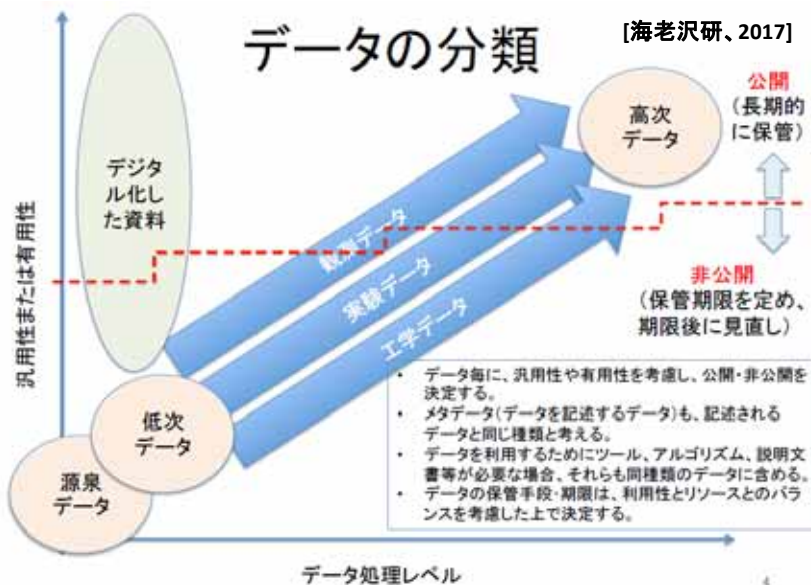
[齊藤昭則、内閣府オープンサイエンス検討会、2017/3/30]



「一般研究者が作成するデータが、4つのデータベースのPhaseを経て発展していくことにより、有益なデータが保持されるようになるのではないか。」(齊藤、2017)

- Pre Phase A : Source Data (検索性もない)
- Phase A : Raw Database (検索性はある)
- Phase B : Close Database (限定されたユーザーの利用)
- Phase C : Open Database (誰でも利用可能)
- Phase D : Permanent Open Database (機関による恒常的サービス)
- (Phase E : Open Open Database (誰でもデータ・サービスが可能))

# JAXA宇宙科学研究所におけるデータ整備論の例



# 「データジャーナル」という新しいメディア

nature publishing group **npg**

www.nature.com/scientificdata

**SCIENTIFIC DATA**

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

**「査読付きデータ・ペーパー」**

Author: Michael P. McKay & David S. Kaufman

Abstract: Modern reconstructions of the near-to-modern history of the Arctic region... (text truncated)

Subject Categories: Paleoclimatology, Climate Change, Meteorology, Climatology

Design Topics: Information Design, Longitudinal Data Collection Method, Data Integration

Measurement Types: Climate Proxy

Technology Types: Data Collection Method

Partial Topics: Arctic, Journal

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... (text truncated)

Author Contributions: N.P.M. and D.S.K. designed the study and prepared the manuscript... (text truncated)

Simplified Japanese version

翻訳・翻案: 村山・南山 (2017)  
南山他 (情報管理, vol60, no3, 2017) で使用

## “Data publication”の現代化 (Data Journal “overlay model”)

(c) 『新しいデータ提供(データ出版)』モデル中のデータ出版の内部ワークフロー

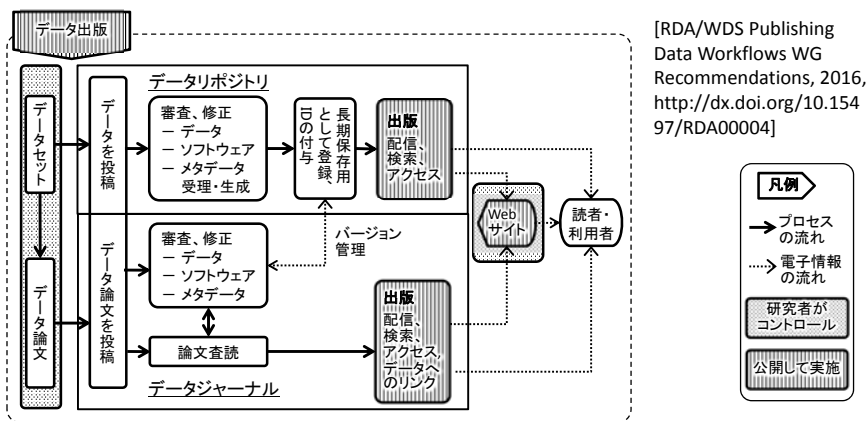
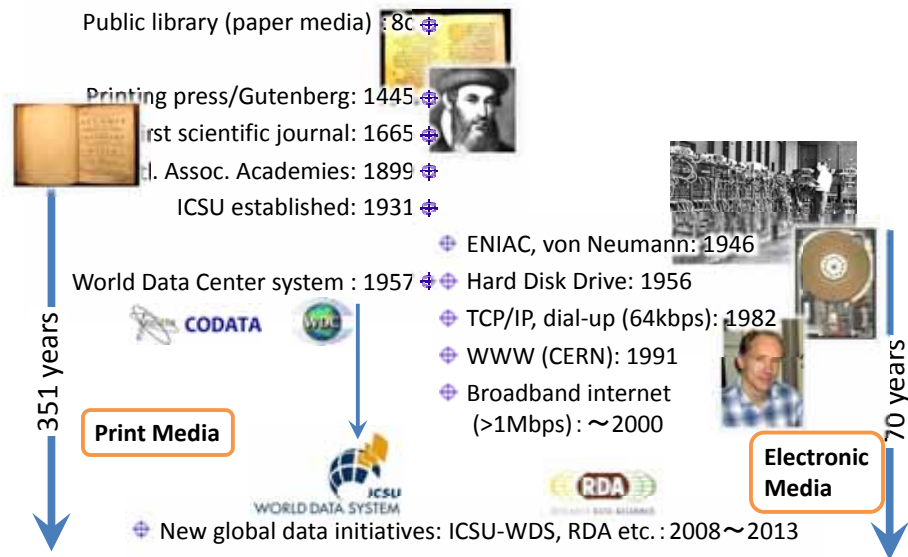


図2 RDA/WDSで検討されたデータ出版に関わるワークフローモデル  
RDA/WDS Publishing Data Workflows Working Group提言中のFigure 2を簡略化したもので、上から(a)従来の論文出版モデル、(b)データ再利用性向上を考慮した論文出版・データ出版モデル、(c) (b)中に示した「データ出版」ボックスの内部のワークフローモデル。(図2の作成に当たっては、原文献の著者・部会員のAmy Nurnberger氏、Mustapha Mokrane氏からご意見、ご助言を頂いた。)

## Print & Electronic Technologies as Social Info. Infrastructures

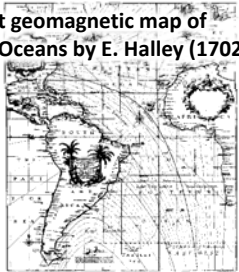
--- 百年の印刷文化の基礎支えと、成長途中のデジタル・サイエンス



科学データ共有の国際的枠組み：  
国際学術事業としてのICSU-World Data System

18th Century

E.g., First geomagnetic map of Atlantic Oceans by E. Halley (1702)



e.g., <https://www.ualberta.ca/~unsworth/UA-clases/210/notes210/O/21002-2008.pdf>

20th Century

国際地球観測年→国際組織設置

International Committees were established for data preservation/services.

International Geophysical Year (1957-1958)



<http://www.kakioka-jma.go.jp/intro/en/index.html>

Federation of Astronomical & Geophysical Data Analysis Services

21st Century

新たな委員会へ発展的改組

ICSU -WDS established (2008) toward a new data management.

情報・通信技術の発展  
分野横断・多分野連携の必要  
グローバル研究体制へ

Increasingly Required:  
- Use advanced IT/ICT  
- More multi & interdisciplinary  
- Globally interoperable system

13

Disciplinary coverage



WDS Membership (November 2017)

70 Regular	Data stewards and/or data analysis certified <b>Trustworthy Data Repositories</b> NSIDC, NASA-DAACS, IRIS-DS, PANGAEA, DANS, DANS, GBIF... )
11 Networks	Umbrella bodies of data stewards: accredited <b>Trustworthy Data Networks</b> ESDIS, NERC-EDC, IODE, IVOA...)
9 Partners	Contribute support to WDS Membership (DataCite, ORCID, IEDRO...)
18 Associates	Interested in the WDS endeavour



Working Groups

- Publishing Data (with RDA):
  - Scholarly Link Exchange (SCHOLIX; [www.scholix.org](http://www.scholix.org))
  - Assessment of Data Fitness for Use
- Audit and Certification of Data Repositories (with RDA)
- Citizen Science (with CODATA)
- Knowledge Network

データ出版

文献-データ  
相互リンク事業

データリポジトリ  
(保存機構)  
信頼性認証

市民科学

メタデータ・  
ナレッジネットワーク



# データリポジトリの国際信頼性認証



## WDS-International Technology Office (WDS-ITO)の設置について

1. 新たにITO(国際テクノロジーオフィス)の設置を計画中。
2. ITOに期待されるファンクション
  - A) GRDIの構成要素となる運用システムへWDSから貢献する。ナレッジネットワーク、GRDI構成要素(機関?システム?、アグリゲートされたメタデータ資源)のレジストリサービス等?  
(注 GRDI: Global Research Data Infrastructure)
  - B) GRDIの構成要素の開発・統合を他のプレイヤー(\*)とともに国際調整する。( \* DataCite、Re3data、ORCID、学術出版社、GEOやFuture Earth等国際事業)
  - C) Brokering サービスを扱うデータは主にメタデータや識別子を想定。
  - D) WDSと他の技術的なグループ(RDAのWG等)と協力する際の調整など
3. IPOの立場・職務:これまでどおり(国際的な「顔」として機能)。
4. ITOのリソース想定:
  - A) 一期3年を単位として覚書をかかず。フルタイム専門職を3~5名?

[Wim Hugo, 2017]

## Working Group Charter: Background

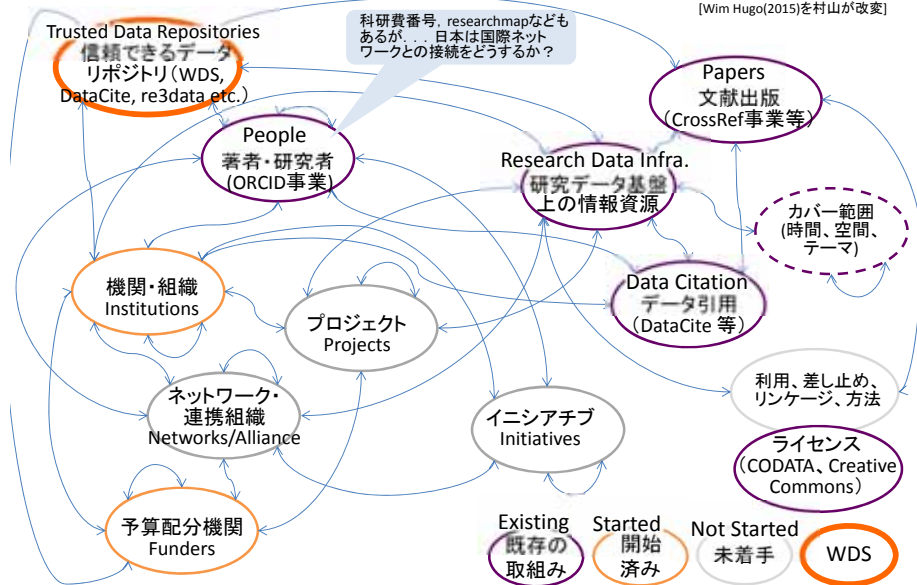
- The ICSU-WDS Scientific Committee identified at its 7th meeting (1-2 November 2012), the need to update the concept of the WDS Data Portal and build an online **open WDS Metadata Catalogue** that can be used to discover and access the quality-assessed scientific data and information holdings of WDS Regular and Network members.
- The WDS-SC also subsequently identified the need to test the feasibility of developing a scalable and agile **WDS Knowledge Network**, which draws information from the *WDS Metadata Catalogue* but which is heavily supplemented with additional data and information from a variety of other sources.

# WDSメタデータ・ナレッジネットワーク [Wim Hugo, 2015]

研究データ・メタデータのグローバルネットワーク構想

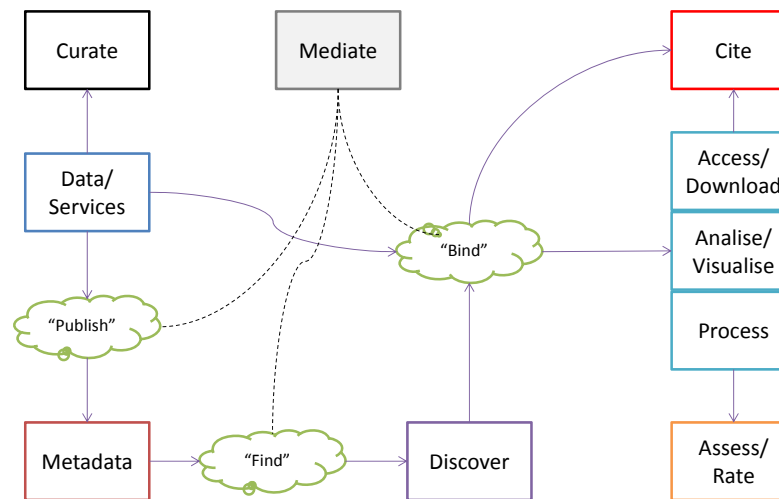
Concept of ICSU-WDS Metadata Catalogue Knowledge Network: The Fabric of Science

[Wim Hugo(2015)を村山が改変]

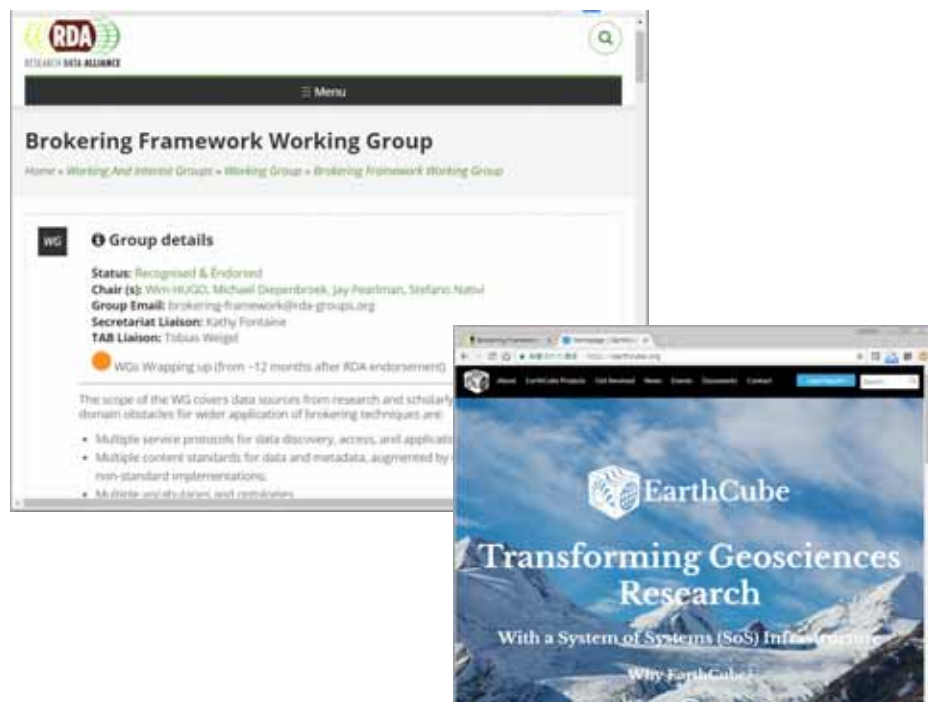


# Generalised Scientific Data Infrastructure Use Case

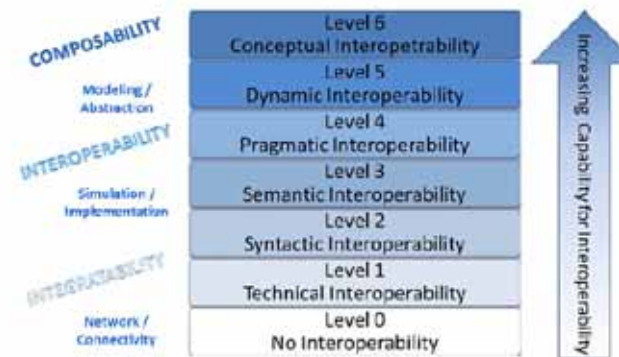
[Wim Hugo, 2017]



"Predictable Assembly from Reliable Components"



# Conceptual Data Interoperability Model



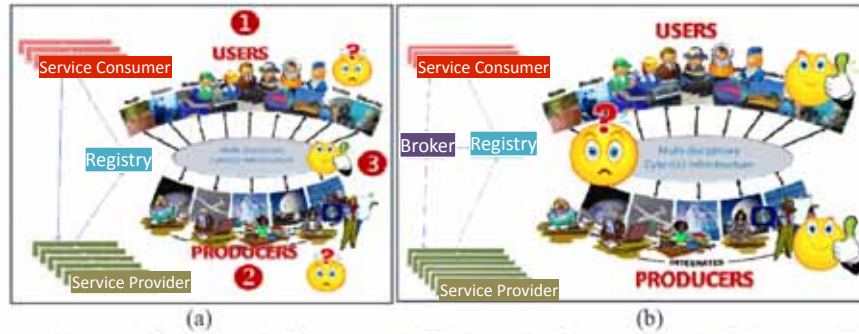
データ利用のための相互通用性 (Interoperability) モデル例 (The Levels of Conceptual Interoperability Model; LCIM)。

- Level 2 : Webのファイル交換・読み出し等。Level 3(パラメータ読みだし・同定等)は含まない
- データの高次利用において相互通用するためには上位階層の相互通用性が必要となる。
- 上位3層の相互通用性の実現は、国際的にまだ困難な状況と思われる。
- ブローキングモデルによってより上位の相互通用性によってデータの取得、読み取り、内容の同定・解釈、解析が可能になると期待される。

[S. Nativi, M. Craglia, J. Pearlman, 2013, Earth Science Infrastructures Interoperability: The Brokering Approach, IEEE JSTARS, Vol. 6 N. 3, pp. 1118-1129.]

# Brokering Approach for Optimized Cyber-infra.

[Nativi et al., NSF Earth Cube white paper, 2011]



多分野横断型データアクセスを可能にするブローキングシステムの考え方。

- (a) シンプルなプロトコル (HTTP+Web等) の場合: ブローカーがなくてもデータやサービスの所在情報から (Registry) アクセス可能 → しかしユーザ (①)、提供者 (②) の両側で interoperability 担保が不可欠。
- (b) プロトコル (利用ルール) が多数存在する場合: ブローカーを用いるほうが効率がよい。ブローカーサービスはユーザと提供者の間のデータ通用性を吸収する → ユーザー・提供者は最低限の相互通用性でよい。

[Nativi et al., NSF Earth Cube white paper, U.S. NSF [Online]. [http://semanticcommunity.info/@api/deki/files/13798/=010\\_Domenico.pdf](http://semanticcommunity.info/@api/deki/files/13798/=010_Domenico.pdf)]

## visual example: Traditional GIS view



## A visual example: Atmospheric Science view



### Over-simplified Worldviews

- To the Geographic Information community, the world is:
  - A collection of **features** (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).
  - The **features** are **discrete objects** described by a set of characteristics such as a **shape/geometry**
- To the Earth Science community, the world is:
  - A set of event **observations** described by **parameters** (e.g., pressure, temperature, wind speed) which vary as **continuous functions** in 3-dimensional space and time.
  - The behavior of the **parameters** in space and time is governed by a set of **equations**.

[from Ben Domenico]

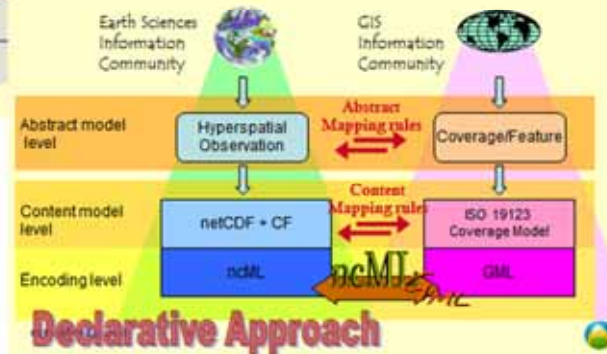
[Stefano Nativi, Atmospheric Data Access for the Geospatial User Community (ADAGUC), 2006]

## ES and GI Info realms

- Historical and technological differences:

	ES Realm	GIS Realm
Data aggregation levels	Hierarchical tree (multiparameter complex datasets) Simple trees (time series) Grid cell aggregations (regions, topological set) Fiber bundles (multichannel imagery)	Dataset Series Dataset Features
Data types	Multi-dimensional array D + time	

### Data model harmonization: Implementation style



[Stefano Nativi, Atmospheric Data Access for the Geospatial User Community (ADAGUC), 2006]

## WDSの今後?

- 「データ資源 (図書館、データセンター等) 管理技術・体制」と、「ICT/IT技術システム開発・体制」との結合が必要
- 抽象的にいうと: Global Research Data Infrastructure にとっての「OS (Operating System)」になっていくのではないかと。
  - E.g., CP/M: Integrating personal-computer components (CPU, CRT, keyboard, FDD...) to make a value of PC.
  - However, the value of the system is often measured via quality of "application software".

FIN.

Thank you for your attention.



Likely to be shifted to early November 2018!



Thank you!

ipo@icsu-wds.org @ICSU\_WDS



## 「デジタルデータの信頼性」のための考察例

### Context: What is a Trust of Digital Data

- 業務設計、予算計画... Works, workflow, budget plan in repository
  - データマネジメント設計 Design of metadata, IDs, data curation, and their management
  - 情報工学、ICT技術、サイバーセキュリティ技術など Cyber security technologies, electronic watermark, block chain...
- Creation 生成
  - Storing, Curation 保存、整備(キュレーション)
  - Long-term preservation/ archive, PID 長期保存、メタデータ整備、永続的識別子(PID)整備
  - Cyber security of data inf. system データ公開技術基盤のセキュリティ
  - Unauthorized access and alteration データや公開内容の詐称、不正な変更、不正なアクセスなど

[https://datascience.nih.gov/Trusted\\_Digital\\_Repository](https://datascience.nih.gov/Trusted_Digital_Repository)

	コア認証	「拡張」認証	本格的な認証
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	€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

2018年1月より有料

# AGU Data Policy

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。



- 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

# COPDESS Statement of Commitment

分野内のメタデータ標準化

- Promulgate metadata information and **domain standards**.
- Promote referencing of data sets using the **Joint Declaration of Data Citation Principles**.
- Include in research papers concise statements about **data location and availability**.
- Promote **education of researchers** in data management; organize and develop training and educational tools and resources
- Develop **workflows** within the repositories that support the peer review process (for example, embargo periods with secure access) and within the editorial management systems that will ease transfer of data to repositories.
- Promote use of **Persistent Identifiers** (ORCID, IGSN, FundRef).

データサイテーション

研究者への周知、啓蒙

データパブリケーション・ワークフロー

永続的識別子

[K. Lehnert, 2017]