

2018.9.14 平成30年度「IUGONET 研究集会」「科学と
データ研究集会」
合同研究集会@情報通信研究機構 小金井本部
本館 4 階 国際会議室



データ保存の信頼性向上 について

情報・システム研究機構 国立極地研究所
南山 泰之

minamiyama@nipr.ac.jp

ORCID ID:0000-0002-7280-3342

オープンサイエンスの潮流

オープンサイエンス = オープンアクセス + オープンデータ



知の創出、新たな価値



研究プロセスの透明化

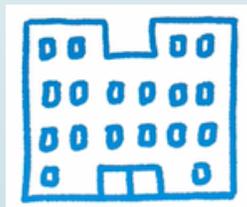


研究成果の幅広い活用

オープンイノベーション



国



資金配分機関



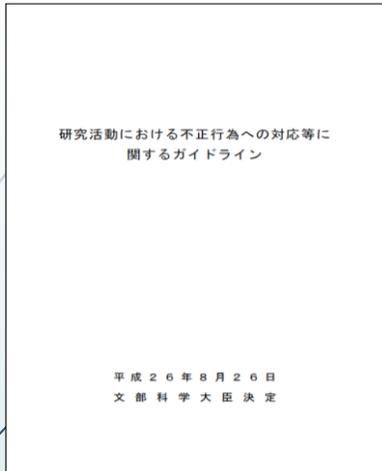
大学・研究機関



研究者

オープンサイエンスの推進体制の構築

不正対策



▶ 文部科学省

『研究活動における不正行為への対応等に関するガイドライン』
(2014年8月)

- **研究データの保存**
- **必要な場合のデータ開示義務**

http://www.mext.go.jp/b_menu/houdou/26/08/_icsFiles/afidfile/2014/08/26/1351568_02_1.pdf



▶ 日本学術会議

『科学研究における健全性の向上について』 (2015年3月)

- **原則、論文発表後10年間の保存期間**

<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>

セキュリティポリシーと研究データの管理

情報類型	格付け
個人情報を含むものや、漏洩によるインパクトの大きい研究資料やデータ	機密性 3 ・ 完全性 2 ・ 可用性 2
上記以外の非公開研究資料やデータ	機密性 2 ・ 完全性 2 ・ 可用性 2
リポジトリ等で公開中の研究資料やデータ	機密性 1 ・ 完全性 2 ・ 可用性 2
他の組織でも公開中の研究資料やデータ	機密性 1 ・ 完全性 2 ・ 可用性 1

機密性・完全性・可用性がともに 1 ⇒ 取扱制限を受けない非保護情報



機密性はないが公開を前提としない参考資料で、消失しても再作成又は再入手が容易で、研究遂行上支障のないような情報

- ➡ **ほぼ全ての研究データは要保護情報 = 取扱制限の対象**
- ➡ **定めた取扱制限を満足する研究データの管理が必要**

データの保存主体

- 情報・システム研究機構研究活動不正への対応に関する規程

第6条

第3条第2項の研究者は、一定期間研究データを保存し、必要な場合は開示しなければならない。

2 前項の保存又は開示すべき研究データの具体的内容やその期間、方法、開示する相手先については、データの性質や研究分野の特性等を踏まえて、最高管理責任者が別に定める。

- 九州大学・研究データの保存等に関するガイドライン

2. 研究活動の記録・保存

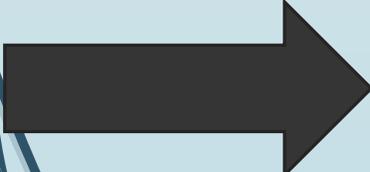
(5) 研究データ等は、それらを生み出した研究者自身が責任をもって保存・管理しなければならない。なお、転出や退職した後も本ガイドラインで定める期間は適切に管理しなければならない。

研究者が主体

(だが具体的な手続きや支援体制は不明瞭)

基準の策定

- ▶ データの信頼性を担保し長期保存していくために
 - ▶ 推奨される保存先の認証
 - ▶ メタデータ記述の充実
 - ▶ 推奨される保存フォーマットの策定



+ これらの基準を満たしているかどうか、
第三者による検証の体制構築

推奨されるデータリポジトリの認証

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 (C) ISO 16363 (C) ISO 16919 (C) ISO 17021 (C)
Audit Process	Self-assessment + independent peer review (2)	Self-assessment + independent peer review (2)	ISO certified accredited au
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

<http://www.re3data.org/search?query=&certificates%5B%5D=WDS>

re3data.org

Filter [Reset all](#)

- Subjects
- Content Types
- Countries
- AID systems
- API
- Certificates**
 - DIN 31664 (1)
 - DSA (9)
 - WDS (70)**
 - other (2)
- Data access
- Data access restrictions
- Database access
- Database licenses
- Data licenses
- Data upload
- Data upload restrictions
- Enhanced publication
- Institution responsibility type
- Institution type
- Keywords
- Metadata standards

Search...

← Previous **1** 2 3 Next →

Found 70 result(s)

The Fish Database of Taiwan

Subject(s)

Content type(s)

Country

The Fish Database of Taiwan is a complex of resea

NOAA Paleoclimatology Data

Paleoclimatology Data

Subject(s)

Content type(s)

Country

Paleoclimatology data are derived from natural sou

https://datascience.nih.gov/Trusted_Digital_Repository

CoreTrustSealとは

- ▶ 「信頼できるデータリポジトリを認定するための中核的な統一要件」を満たすリポジトリに対して付与される認証
 - ▶ 2016年11月に、Data Seal of Approval (DSA) とICSU-WDSにより策定された共通のフレームワーク
- ▶ 各リポジトリに対して16項目（3要素）の自己評価を実施。
 - ▶ 組織的な基盤（Organizational Infrastructure）
 - ▶ デジタルオブジェクト管理（Digital Object Management）
 - ▶ 技術（Technology）
- ▶ 認証により期待できること
 - ▶ 利害関係者からの信頼性向上
 - ▶ 透明性の確保や手続きの改善

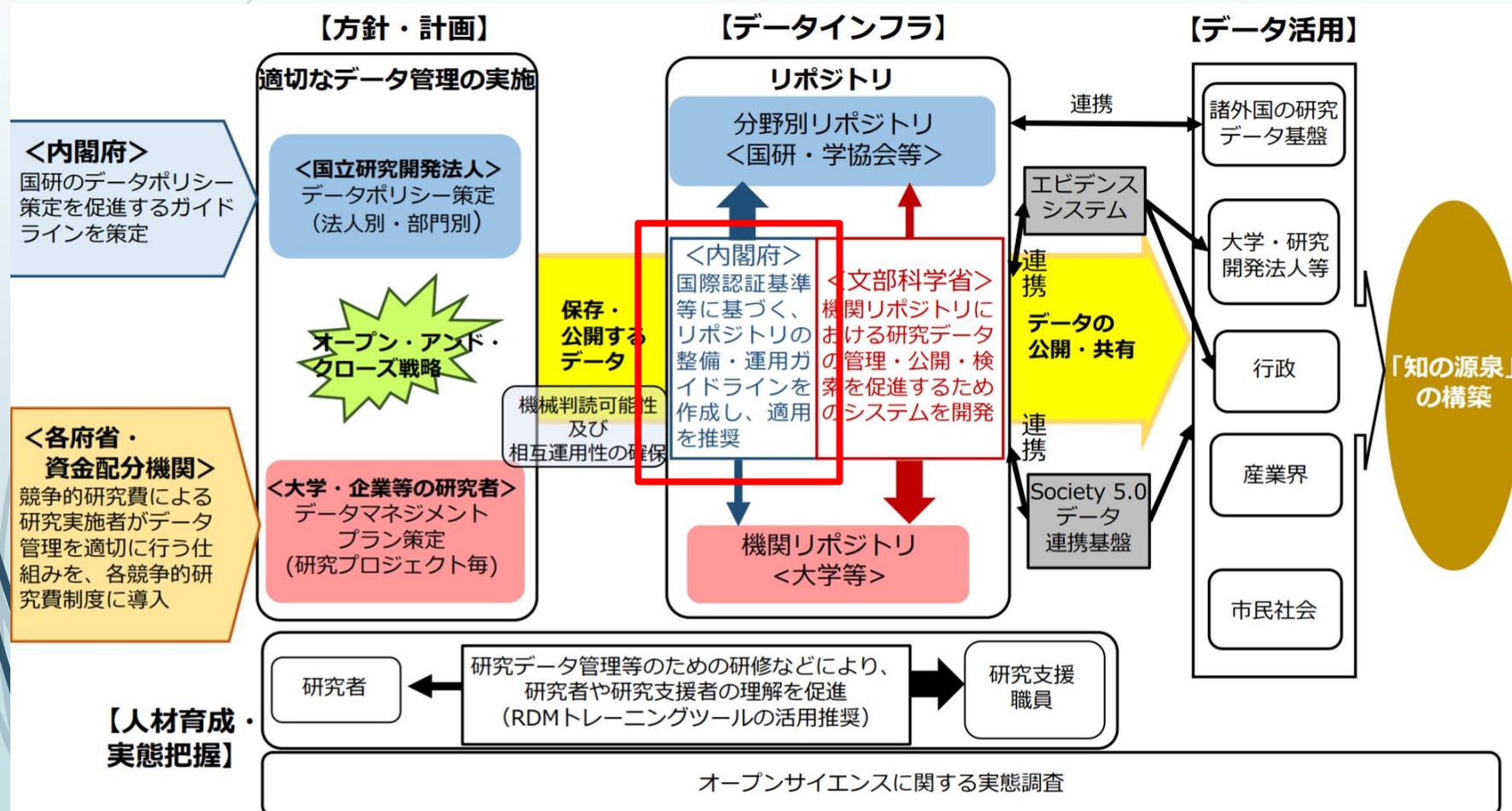


データの長期的な保全と提供

Ref: <http://current.ndl.go.jp/e1888>

政策面での動き

内閣府.国際的動向を踏まえたオープンサイエンスの推進に関する検討会 (第6回)
<http://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/6kai/siryo1.pdf>



国内における検討

- 研究データ利活用協議会・小委員会（リポジトリ）での議論
 - CTS要件の日本語訳
 - CTS認証済のリポジトリから、3つの自己評価例を抽出
 - DRUM（ミネソタ州立大学の機関リポジトリ）
 - Finnish repository（フィンランド社会科学分野の研究データリポジトリ）
 - NGDC（英国 地球科学データリポジトリ）
 - 各リポジトリについて、16項目の自己評価を分析
 - 分析結果からリポジトリに必要なアイテムを抽出
- **リポジトリの品質向上に資するためのガイドライン作り**

八塚茂.リポジトリの信頼性を高める為に必要なアイテム (Japan Open Science Summit 2018)

https://japanlinkcenter.org/rduf/doc/joss2018_c5_03.pdf

メタデータの充実

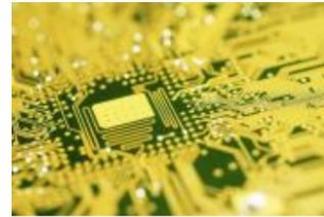
Search by Discipline



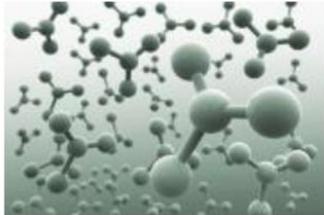
Biology



Earth Science



General Research Data



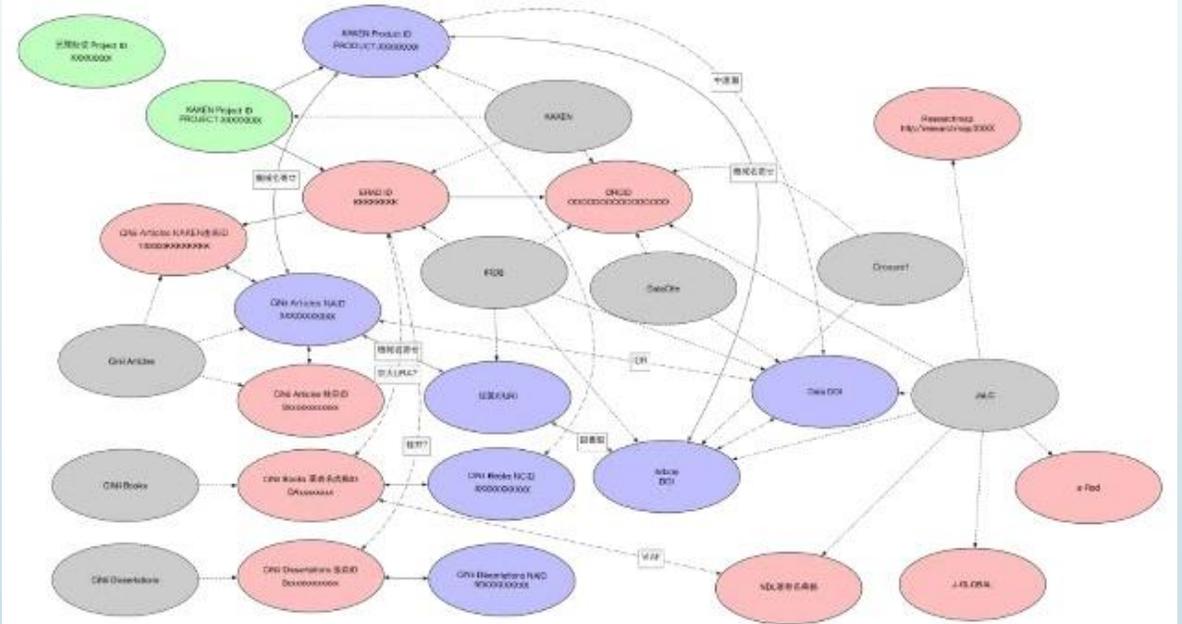
Physical Science



Social Science & Humanities

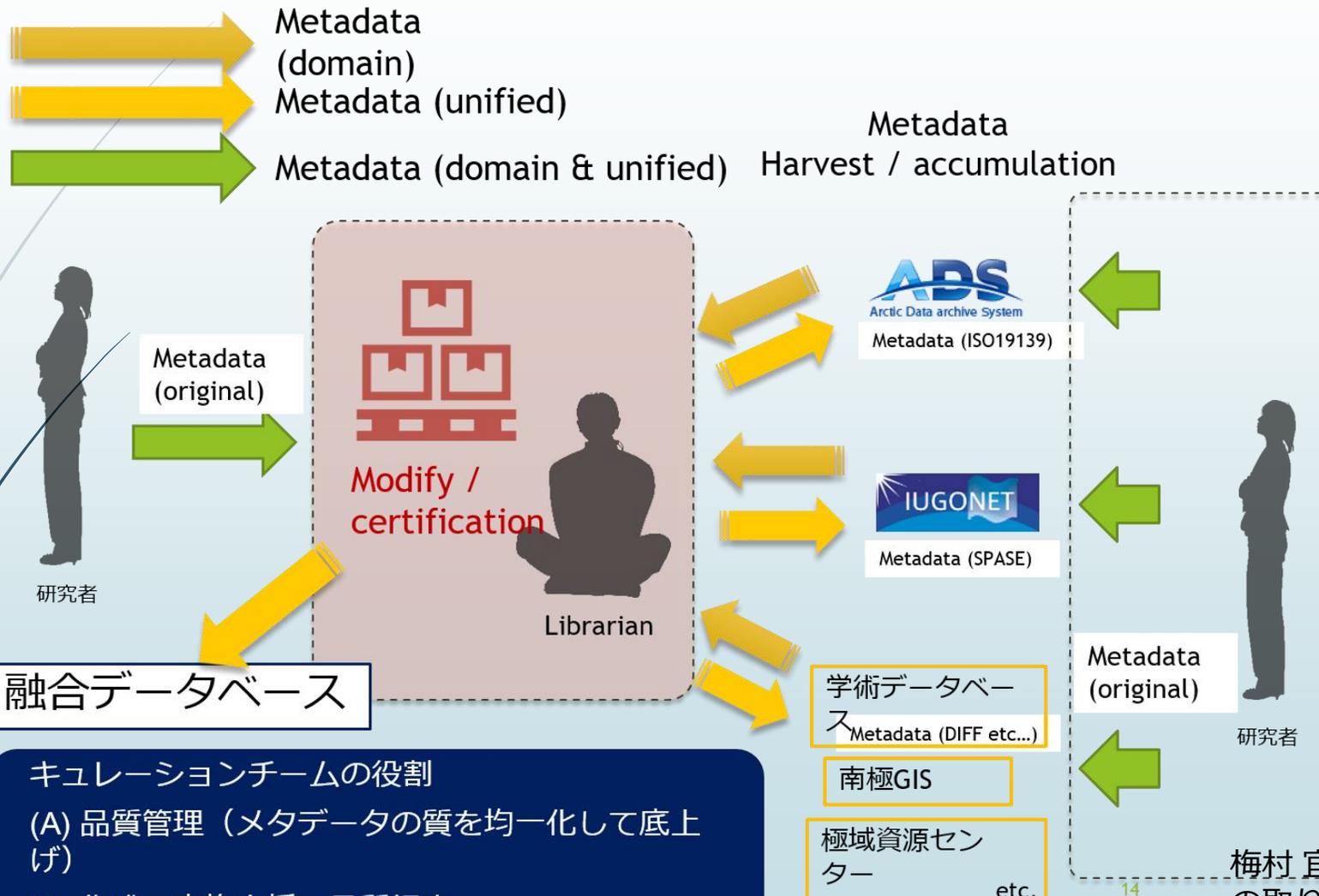
DCC. "Disciplinary Metadata".
<http://www.dcc.ac.uk/resources/metadata-standards>

ミッシングリンク



大向一輝. "CiNii Researchと大学図書館".
<https://www.slideshare.net/ikki.ohmukai/cinii-research2018618-103285531>

ドメインのメタデータ



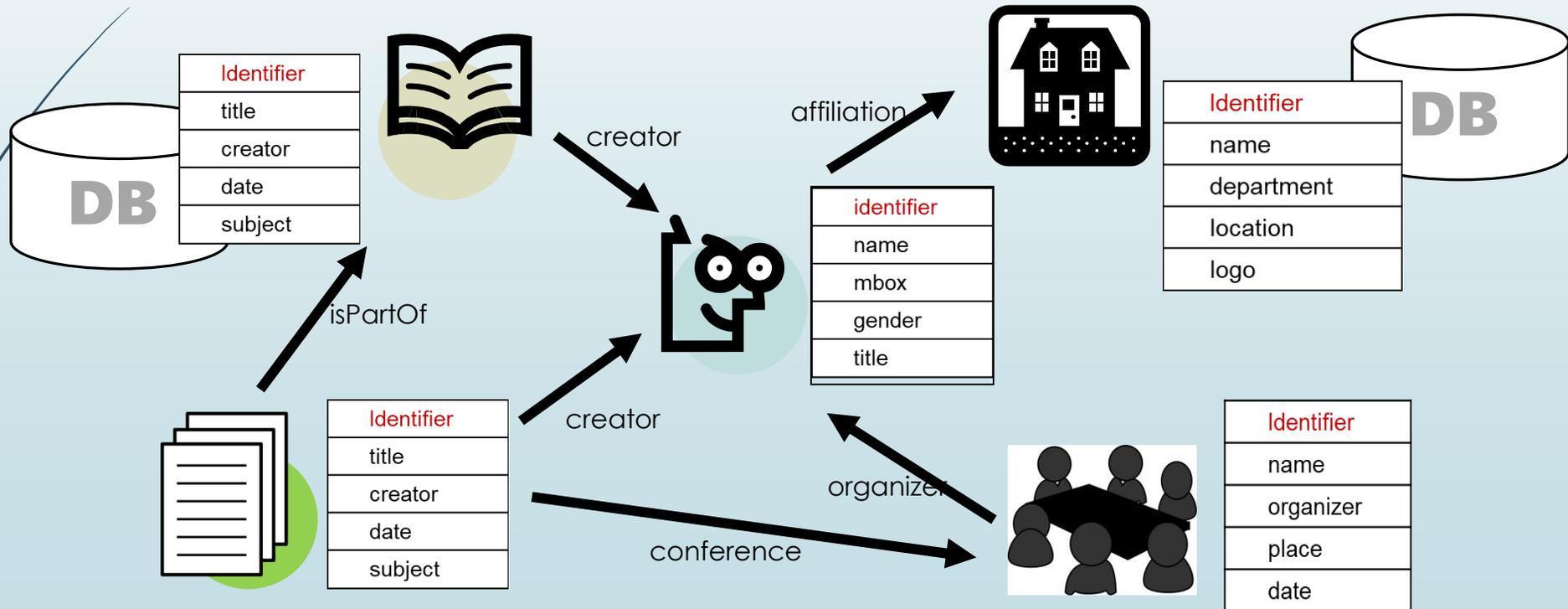
キュレーションチームの役割
 (A) 品質管理 (メタデータの質を均一化して底上げ)
 (B) 作成・変換支援、品質提案

梅村 宜生ほか. ROISデータサイエンス事業の取り組みにおける図書館員の役割
<http://id.nii.ac.jp/1291/00014980/>

リンク情報の充実

- 識別子 (ID) のつながりで表現する
 - 1つの事物・事柄に1つの識別子 (ID) を付与する
 - 識別子 (ID) と識別子 (ID) の関係を明確にする
 - 識別子 (ID) をどの単位に付与していくか、がポイント

高橋菜奈子. JPCOARスキーマ説明会資料より



推奨フォーマット

LIBRARY OF CONGRESS ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS Search GO

The Library of Congress > Preservation > Resources > Recommended Formats Statement

PRESERVATION

Print Subscribe Share/Save Give Feedback

« Back to Recommended Formats Statement

Main | Table of Contents | Introduction | Textual Works and Musical Compositions | Still Image Works | Audio Works | Moving Image Works | Software and Electronic Gaming and Learning | Datasets/Databases | Websites

Library of Congress Recommended Formats Statement 2018-2019

Introduction

Categories of creative content:

- I. [Textual Works and Musical Compositions](#)
 - i. [Textual Works - Print](#)
 - ii. [Textual Works - Digital](#)
 - iii. [Textual Works - Electronic Serials](#)
 - iv. [Digital Musical Compositions \(score-based representations\)](#)
- II. [Still Image Works](#)
 - i. [Photographs - Print](#)
 - ii. [Photographs - Digital](#)
 - iii. [Other Graphic Images - Print](#)
 - iv. [Other Graphic Images - Digital](#)
 - v. [Microforms](#)
- III. [Audio Works](#)

Related Links

- > [Donate](#)
- > [Audio-Visual Preservation](#)
- > [National Film Preservation Board](#)
- > [National Recording Preservation Board](#)

Library of Congress. "Recommended Formats Statement"
<http://www.loc.gov/preservation/resources/rfs/TOC.html>

UK Data Service

Site search Accessibility FAQ Help Contact New users Sitemap

About us Get data Use data **Manage data** Deposit data News and events

Home > Manage data > Format your data > Recommended formats

Recommended formats

SHARE

This table contains guidance on file formats recommended and accepted by the UK Data Service for data sharing, reuse and preservation.

You may need to convert your data files to a preservation file format.

We welcome [queries](#) from researchers about appropriate file formats for working and preservation, particularly early in the research process.

If you are unsure of the suitability of your file formats for the data you want to deposit with the UK Data Service, please [get in touch](#).

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings,	comma-separated values (.csv) tab-delimited file (.tab)	delimited text (.txt) with characters not present in data used as delimiters

[Data lifecycle](#)
[Plan to share](#)
[Legal and ethical](#)
[Rights](#)
[Document your data](#)
[Format your data](#)
[File formats](#)
[Recommended formats](#)
[Organising](#)
[Quality](#)
[Versioning](#)
[Transcription](#)
[Digitisation](#)
[Store your data](#)
[Collaborative research](#)
[Training](#)
[Tools and templates](#)

[LOGIN](#) / [REGISTER](#)
DISCOVER UK DATA SERVICE
 [GO](#)
 Data Website

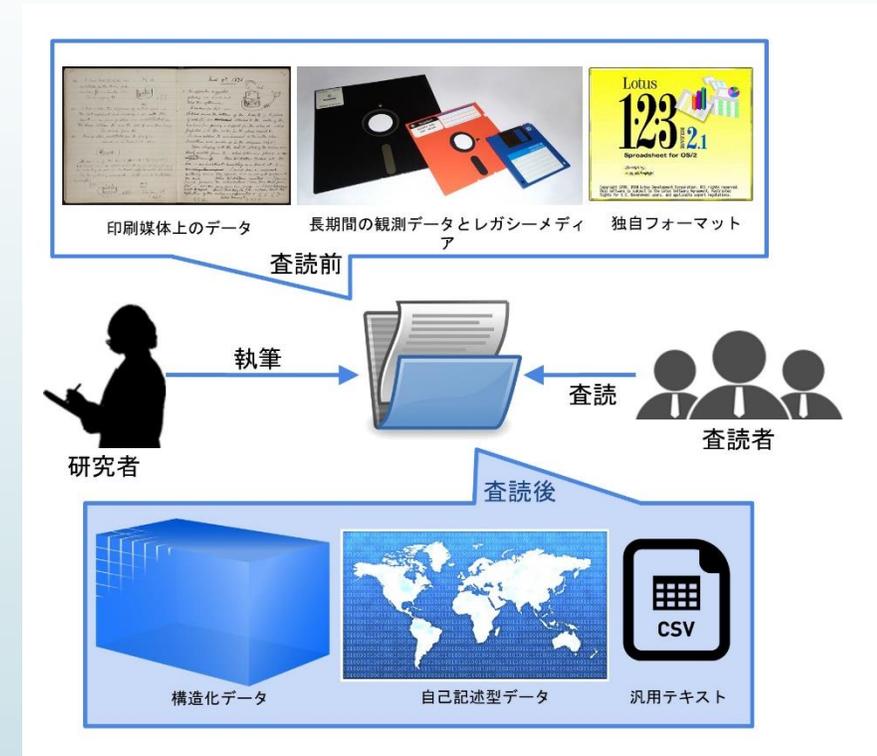
MANAGING AND SHARING RESEARCH DATA
 A Guide to Good Practice

UK Data Service. "Recommended formats"
<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>

データの「査読」

- ▶ Data Quality
 - ▶ **フォーマット**、データ値の一貫性、精度、粒度、論理値、独立性など
- ▶ Metadata Quality
 - ▶ **データ自体に関する記述**
 - ▶ データのフォーマット、物理量、来歴、プロセスの文書化、データ間の関連性など
 - ▶ **メタデータの再利用性**
 - ▶ 発見可能性の確保、スキーマ・統制語彙の活用、有効なリンク、標準化など
- ▶ General
 - ▶ データを活用するコミュニティの存在
 - ▶ **データソースの信頼性**
 - ▶ 知的財産権
 - ▶ 有効なURI

Ref: "Citation and Peer Review of Data:
Moving Towards Formal Data Publication"
<http://dx.doi.org/10.2218/ijdc.v6i2.205>

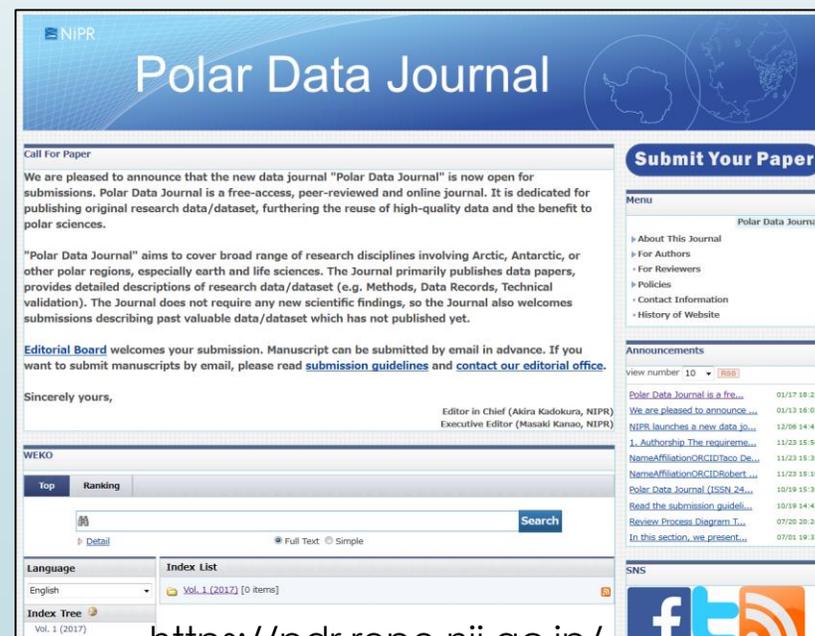


検証の試み：データジャーナル

- ▶ データジャーナルとは
 - ▶ データそのものを記述する論文（データ論文）を掲載するジャーナル
 - ▶ 原著論文と同じ出版プロセス（査読・審査）
 - ▶ データから得られる知見の記述は求めない（原著論文との違い）
 - ▶ 「査読」によるデータの品質保証、再利用性の向上
 - ▶ 査読レポートを公開する事例も



<http://www.nature.com/sdata/>



<https://pdr.repo.nii.ac.jp/>

まとめ

- ▶ データ保存は研究者主体
- ▶ 保存や公開を支援し、品質向上に資するための様々な取り組みが始まっている
- ▶ 取り組みを把握し、上手に使ってもらえると嬉しい
- ▶ Next Step
 - ▶ こういった情報を集約し、相談するためのコミュニティ構築
 - ▶ 各学会との接続