



HOKKAIDO UNIVERSITY

Title	「大学における研究データサービス」ワークショッププログラム（2019/12/6(金)）参加報告
Author(s)	堀越, 邦恵
Description	令和元年度研修出張報告会（特別編）：オープンサイエンス・研究データ関係. 2020年2月4日. 北海道大学附属図書館本館 4階大会議室, 札幌市
Issue Date	2020-02-04
Doc URL	https://hdl.handle.net/2115/76792
Rights(URL)	https://creativecommons.org/licenses/by-sa/4.0/
Type	lecture
File Information	03_Horikoshi.pdf



「大学における研究データサービス」 ワークショッププログラム (2019/12/6(金)) 参加報告

2020/2/4 (火)

研究支援課研究支援企画担当 堀越 邦恵

2019/12/6 ワークショップ

- 6~7名程度・6班
- 図書館員・研究者・学生
- セッション
 - Session 1 研究データとは何か
 - Session 2 ジャーナルのデータポリシー
 - Session 3 研究データのキュレーション

Session 1 研究データとは何か

- 「研究データ」の定義
 - いろいろ
 - 「なんでもあり」…？
- raw data / underlying data / figures, tables
 - それぞれの関係
 - 保存しておくべき「データ」はどれ？

Session 2 ジャーナルのデータポリシー

- Transparency and Openness Promotion (TOP) Guidelines
 - 「透明性とオープン性促進ガイドライン」
 - <https://cos.io/top/>
 - Center for Open Science（非営利団体）策定
 - 厳格さのレベル3段階で規定（カテゴリごとに組み合わせ可能）
- これに署名する出版社が増えている

TOP Guidelines: カテゴリー

- 引用
- **データの透明性**
- **分析手法（コード）の透明性**
- **研究材料の透明性**
- デザインと分析の透明性
- 研究の事前登録
- 分析計画の事前登録
- 複製（追試）

	Not Implemented	Level I	Level II	Level III
Citation Standards	Journal encourages citation of data, code, and materials, or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	Journal encourages data sharing, or says nothing.	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic Methods (Code) Transparency	Journal encourages code sharing, or says nothing.	Article states whether code is available, and, if so, where to access it.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Research Materials Transparency	Journal encourages materials sharing, or says nothing.	Article states whether materials are available, and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Design and Analysis Transparency	Journal encourages design and analysis transparency, or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Study Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Analysis Plan Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies, or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts results blind review.	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

TOP Guidelines: データの透明性のレベル

- (採用なし)
 - ポリシーがない / 言及してない / データ共有の「推奨」はするが、具体的な指示がない
- Level 1: 最低限
 - データが利用可能か、利用可能ならそのアクセス先を提示
 - 「あれば必須」レベル。
 - 「著者に連絡すれば取得可能」でも可

TOP Guidelines: データの透明性のレベル

- Level 2

- データは信頼できるリポジトリに掲載されなければならない
 - 義務(must)「著者に連絡すれば～」はNG
 - 権利関係等で公開できない場合、NGな理由を論文投稿時に説明する必要あり

- Level 3: もっとも厳格

- Level 2 + 出版に先立ち、そのデータを用いて当該研究の検証（結果の再現の確認）を行う
 - 「再現性」に関しては努力目標

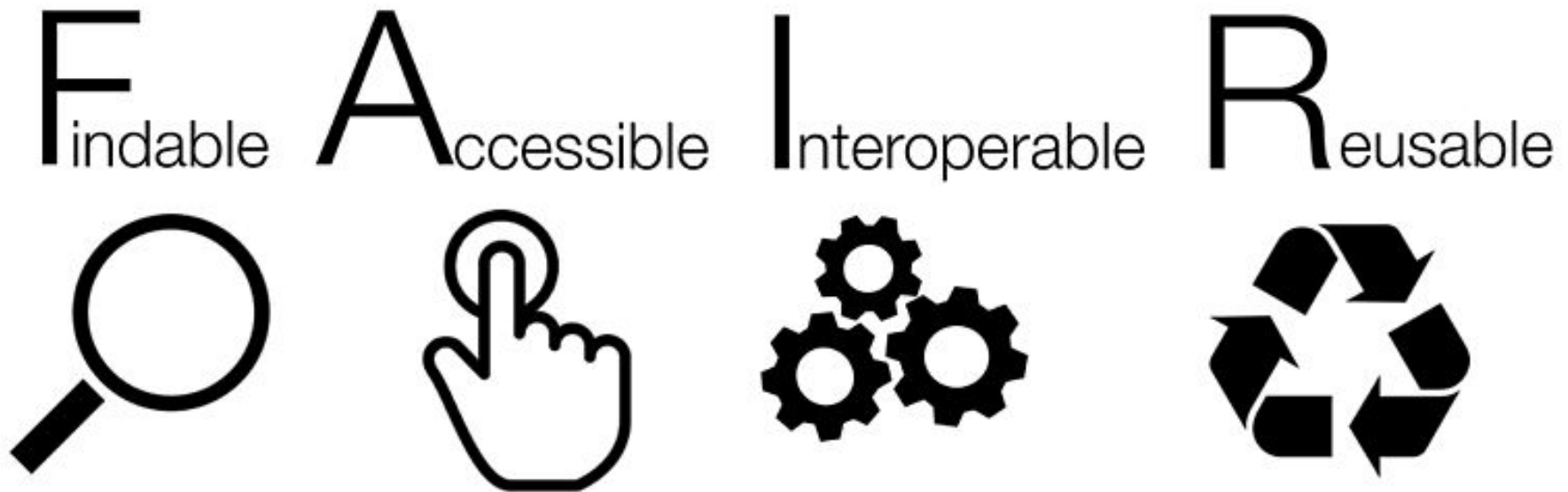
Session 3 研究データのキュレーション

- よくある「問題」
 - ドキュメント (README.txt) がない
 - データファイルの命名規則がない
 - どこにでもありがちなファイル名: 例: final.csv
 - データの不備 (ファイル欠落, 不完全なデータ等)
 - 特殊な環境でのみ利用・再生可能なデータ

データキュレーションとは

- 検索、アクセス、相互運用、再利用できるように研究データを準備し、維持すること
 - メタデータの開発
 - ドキュメントの関連付け
 - アクセスの提供
 - 保存の支援

参考: FAIR原則



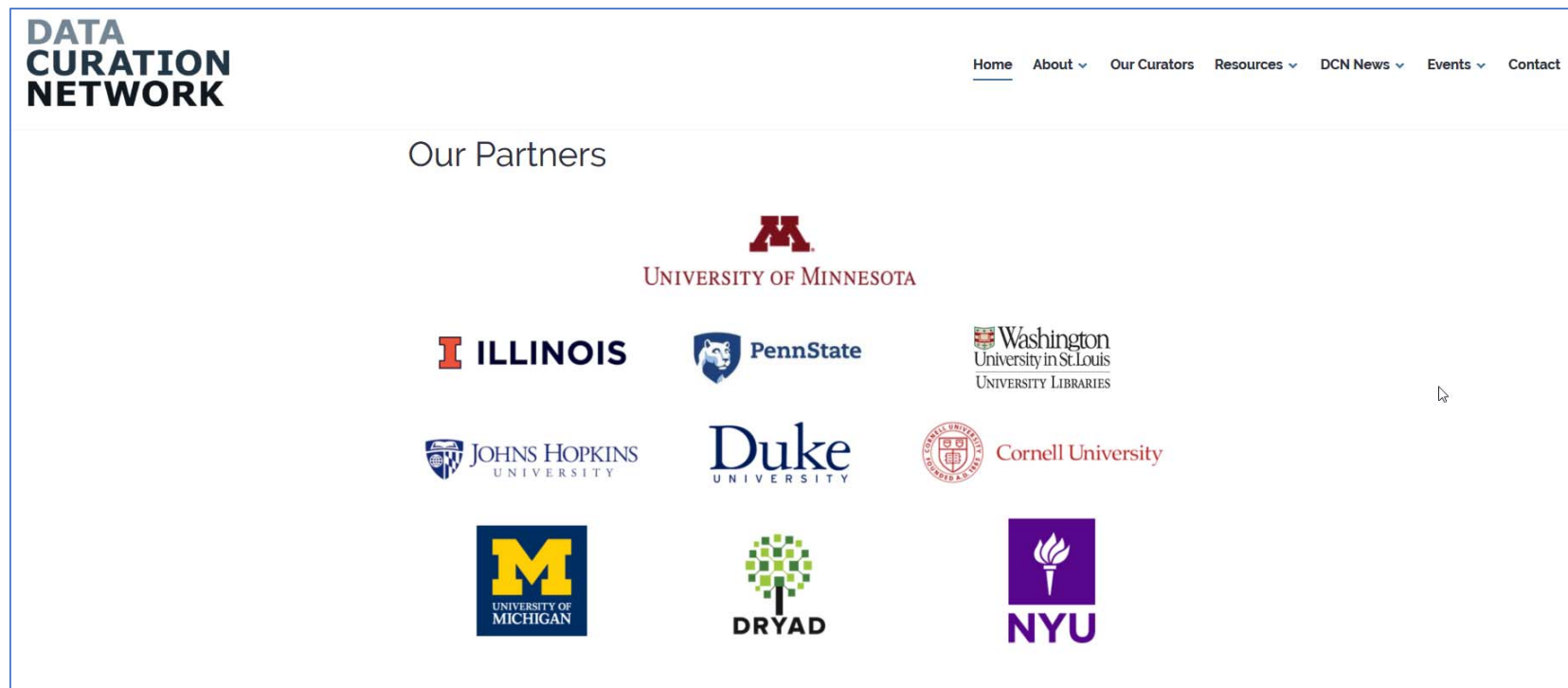
https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

SangyaPundir [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)]

Findable（見つけれられる）、Accessible（アクセスできる）
Interoperable（相互運用できる）、Reusable（再利用できる）

データキュレーションネットワーク

- <https://datacurationnetwork.org/>
 - “human layer”



CURATED チェックリスト

- **C**heck データファイル確認 & ドキュメントを読む
- **U**nderstand データを理解(しようと)する
- **R**equest 欠落した情報を要求・修正する
- **A**ugment 検索できるようにメタデータで補足
- **T**ransform 再利用／保存のためにファイル形式変換
- **E**valuate FAIR原則に則しているかを評価・査定
- **D**ocument キュレーション作業全体を記録・管理（文書化）
- <https://datacurationnetwork.org/resources/resources-2/>

ワークショップ討議中に出た話題 1

- サンプル例「A氏が1994年に作成・論文発表したデータを使用し、B氏が新たに2000年に論文を発表した」
 - A氏のデータそのものの保存に関しては、B氏はコントロールできる立場にない。この場合、B氏はA氏のデータをコピーして手元に保存しておかなければならないか？（A氏の「10年保存」期限が過ぎたら消されるかも→2004年には消える？）
 - 保存しておくデータは raw data / underlying data のどちら or 両方？
- メタデータはどこまで詳しく記述すればいいのか
 - データファイル本体からとれる情報は非常に少ない（論文と比べて）

ワークショップ討議中に出た話題 2

- サンプル例「ある特定の日時に見たWebページ掲載の情報を集約して解析した」
 - → Webページ自体は公開されているが、時間経過で内容が変わる。
「研究時に見たページ内容」は「raw data」？
- CURATEDのR（データの修正）
 - データに手を入れる・修正することに関して、教員へ事前に十分説明しておく必要があるのでは（センシティブな問題）
- 教員からの「論文根拠データの登録依頼」の実際の例

- シンポジウム・ワークショップの資料：
<http://hdl.handle.net/2324/2547228> (九州大学学術情報リポジトリ)