

シンポジウム・ワークショップ「大学における研究 データサービス」

Imker, Heidi
University of Illinois at Urbana-Champaign

Mischo, William H.
University of Illinois at Urbana-Champaign

Schlembach, Mary C.
University of Illinois at Urbana-Champaign

Miyazaki, Yuta
Kyushu University Library

他

<https://hdl.handle.net/2324/2547228>

出版情報 : 2019-12-05. Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

バージョン :

権利関係 :

December 6, 2019

Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University & Kyushu University Library (Organizer: Ishita Emi)
Workshop "Research Data Services in Universities"

Session 3: Curating Research Data

Heidi Imker, PhD
Director, Research Data Service

Bill Mischo
Head, Grainger Engineering Library and Information Center

Mary Schlembach
Chemistry & Physical Sciences Librarian



2019年12月6日

九州大学統合新領域学府ライブラリーサイエンス専攻・附属図書館共同開催 (代表者 石田栄美)
ワークショップ「大学における研究データサービス」

※ このスライドの日本語訳は、ワークショップの配布資料として作成したものです。
専門家による訳ではないため、誤りがある場合があります。

日本語訳：九州大学附属図書館eリソース課 芦北卓也

Session 3: 研究データのキュレーション

Heidi Imker, PhD
Director, Research Data Service

Bill Mischo
Head, Grainger Engineering Library and Information Center

Mary Schlembach
Chemistry & Physical Sciences Librarian

Session 3 Outline

1. What is data curation?
2. Data Curation Network
3. Activity: data curation checklist
4. Discussion questions

Session 3 アウトライン

1. 研究データのキュレーションとは?
2. データキュレーションネットワーク
3. 実習: データキュレーションチェックリスト
4. 議論の課題

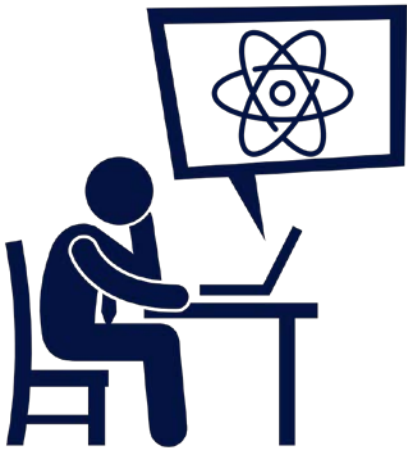
Session 3 Outline

1. What is data curation?
2. Data Curation Network
3. Activity: data curation checklist
4. Discussion questions

Session 3 アウトライン

1. 研究データのキュレーションとは?
2. データキュレーションネットワーク
3. 実習: データキュレーションチェックリスト
4. 議論の課題

Quality Data Sharing is a Challenge



Researcher works for many years and knows their data too well

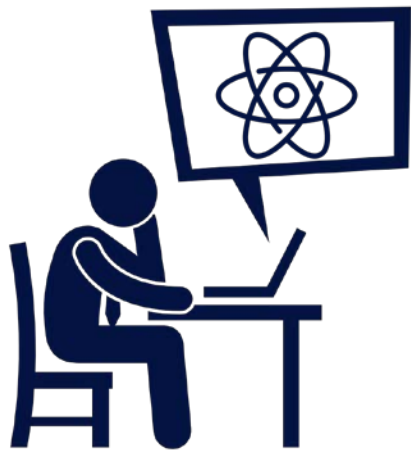


Researcher share data that is difficult to understand



People who try to use the data are confused and frustrated

質の高いデータ共有という課題



研究者は長年の研究で
自分のデータのことなら
とてもよく知っている



研究者が共有するデータ
には理解困難なものもある



そうしたデータを使おう
とする人たちは混乱し、
イライラする

Example Problems

Sometimes when researchers share their data:

- They provide only the data files, with no documentation to explain details
- Data variables, data structures, and even the file names are not explained
- Files may be missing or incomplete
- Code may not run outside researchers' environment

問題の例

研究者がデータを共有する際にありがちなこと

- データファイルだけで、詳細を説明するドキュメントがない
- データの変数や構造、ファイルネームですら説明されていない
- ファイルが欠落していたり不完全だったりする
- プログラミングコードが外部の研究者の環境では実行できないこともある

Data Curation

Data curation prepares and maintains research data in ways that make it **findable, accessible, interoperable and reusable** and include a wide range of possible activities including **developing metadata, associating documentation, providing access, or supporting preservation**.

Note: similar to the term “research data,” “data curation” is often used to describe a wide range of activities and the term itself may have different meanings depending on the context and environment in which it is used.

Johnston LR, Carlson JR, Hswe P, Hudson-Vitale C, Imker H, Kozlowski W, Olendorf RK, Stewart C. Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions’ Data Repository and Curation Services. *Journal of eScience Librarianship* 6(1): e1102. <https://doi.org/10.7191/jeslib.2017.1102>

データキュレーション

データキュレーションは、**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で、**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**といった幅広い活動からなる。

Note: 「研究データ」という用語と同様に、「データキュレーション」も幅広い活動を説明するのに使われることが多く、その用語自体も使用される文脈や環境に応じて異なる意味をもつ場合がある。

Johnston LR, Carlson JR, Hswe P, Hudson-Vitale C, Imker H, Kozlowski W, Olendorf RK, Stewart C. Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services. Journal of eScience Librarianship 6(1): e1102. <https://doi.org/10.7191/jeslib.2017.1102>

Data Curation Workflow at Illinois



University researchers submit their datasets to the Illinois Data Bank

→
curators receive an email notification



Curators log into the Illinois Data Bank

→
access deposited dataset



Download all dataset files for review, including README

イリノイにおけるデータキュレーション ワークフロー



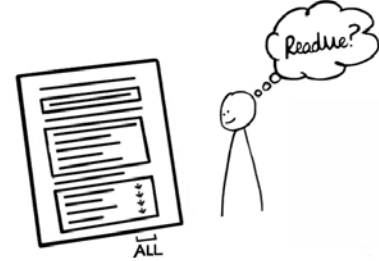
大学研究者が彼らのデータセットを
Illinois Data Bankに提出

→
キュレーターが
メール通知を受信



キュレーターがIllinois Data
Bankにログイン

→
デポジットされた
データセットに
アクセス



レビューのため、READMEを含む
すべてのデータセットを
ダウンロード

Data Curation Workflow at Illinois



University researchers submit their datasets to the Illinois Data Bank

curators receive an email notification



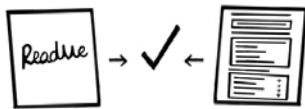
Curators log into the Illinois Data Bank

access deposited dataset



Download all dataset files for review, including README

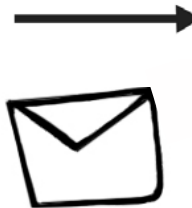
- # files
- open files (sample)
- functions/headers
- organization
- description



Final check: does the README match up to the dataset?



email summary notes to researcher with recommendations



changes to dataset, if necessary

- record process
- check off ticket
- constant monitoring

イリノイにおけるデータキュレーション ワークフロー



大学研究者が彼らのデータセットを
Illinois Data Bankに提出

キュレーターが
メール通知を受信



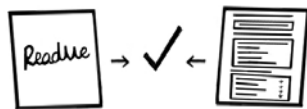
キュレーターがIllinois Data
Bankにログイン

デポジットされた
データセットに
アクセス



レビューのため、READMEを含む
すべてのデータセットを
ダウンロード

- # files
- open files (sample)
- functions/headers
- organization
- description



最終チェック: READMEがその
データセットに一致しているか?



レコメンド付きのサマリーノート
を研究者宛にメール



必要があればデータ
セットを変更

- 記録処理
- チケットのチェックオフ
- 恒常的なモニタリング

Session 3 Outline

1. What is data curation?
2. Data Curation Network
3. Activity: data curation checklist
4. Discussion questions

Session 3 アウトライン

1. データキュレーションとは何か?
2. データキュレーションネットワーク
3. 実習: データキュレーションチェックリスト
4. 議論の課題

Data Curation Network

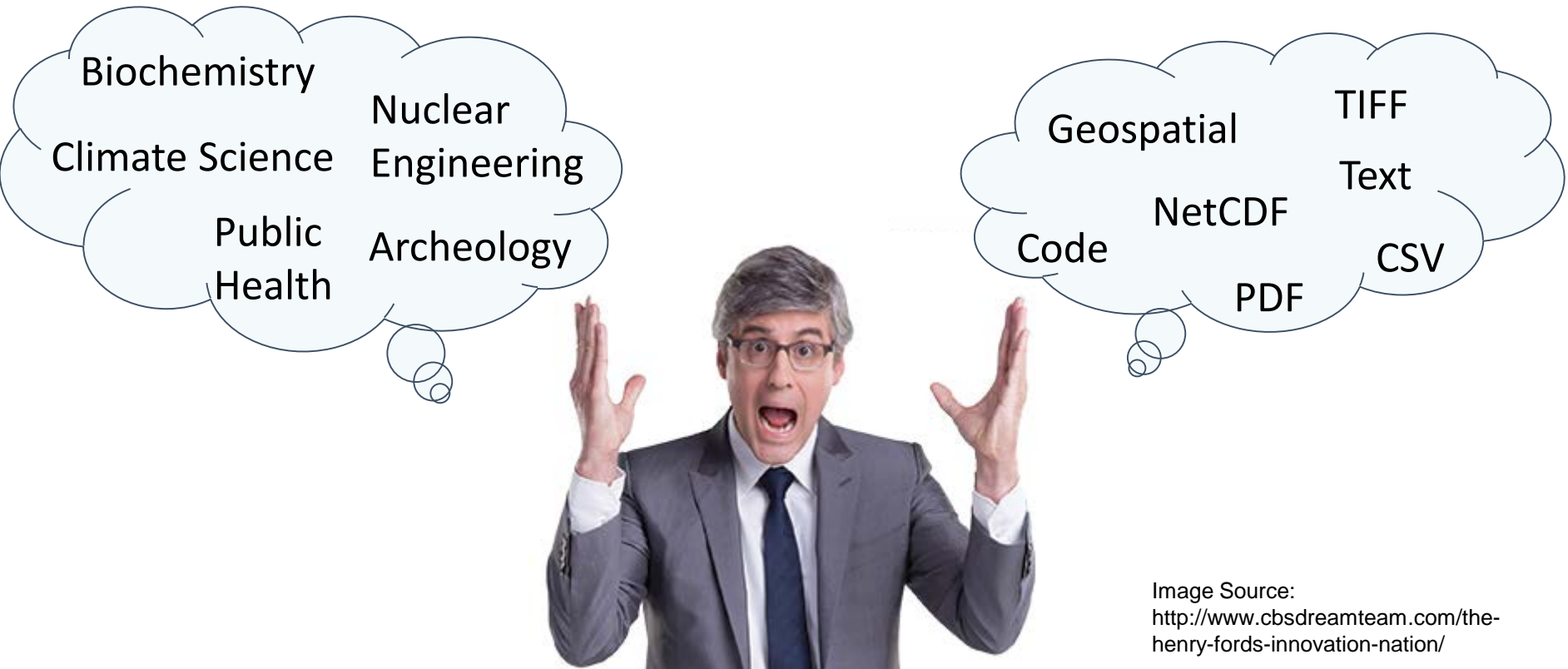


Image Source:
<http://www.cbsdreamteam.com/the-henry-fords-innovation-nation/>

Solution: Shared staffing across institutions to expand our capacity to curate data from different disciplines and of different types.

データキュレーションネットワーク

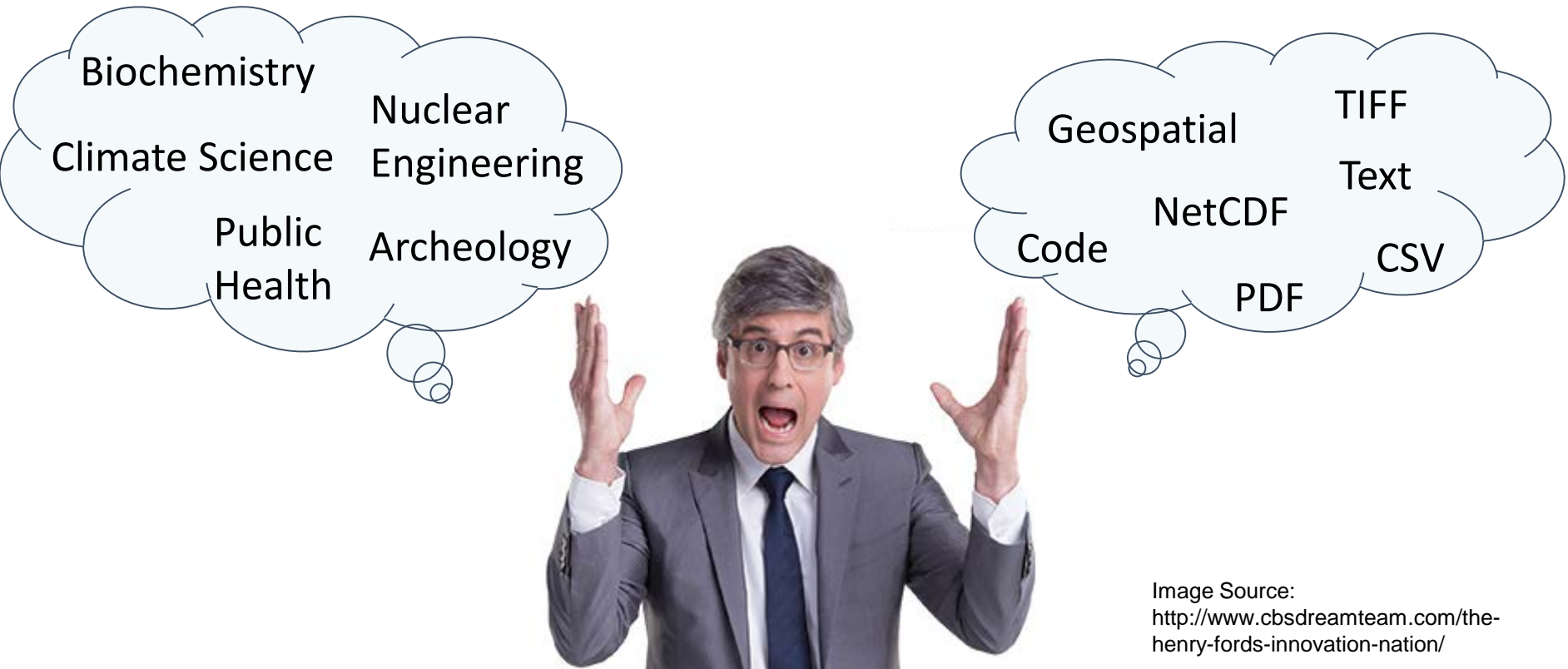


Image Source:
<http://www.cbsdreamteam.com/the-henry-fords-innovation-nation/>

Solution: 様々な分野、様々な形式からデータのキュレーションのキャパシティを拡げるため、機関間でスタッフを共有する。

Data Curation Network

Shared staffing across institutions to expand our capacity to curate data from different disciplines and of different types.

DATA CURATION NETWORK

Aims to develop the “human layer” in local data repository stacks through

- expert services
- incentives for collaboration
- normalized curation practices
- professional development



データキュレーションネットワーク

様々な分野、様々な形式からのデータのキュレーションのキャパシティを拡げるため、機関間でスタッフを共有する。

DATA CURATION NETWORK

以下を通じてローカルのデータリポジトリ
スタックにおける“human layer”の開発を
目指す

- エキスパートサービス
- コラボレーションの促進
- 標準化されたキュレーション手法
- 専門能力開発



Data Curation Network's CURATED Checklist

- C Check** data files and read documentation
- U Understand** the data (try to), if not...
- R Request** missing information or change
- A Augment** the submission with metadata for findability
- T Transform** file formats for reuse and long-term preservation
- E Evaluate** and rate the overall submission for FAIRness
- D Document** all curation activities throughout the process

データキュレーションネットワークの CURATEDチェックリスト

- C Check** データファイルを**確認**してドキュメントを読む
- U Understand** データを**理解**(しようと)する、できなければ…
- R Request** 欠落した情報を**要求**または変更する
- A Augment** 検索できるようにメタデータでその提出物を**補足**する
- T Transform** 再利用および長期保存のためファイルの形式を**変換**する
- E Evaluate** FAIR原則のため提出物全体を**評価**および査定する
- D Document** プロセス全体を通じてすべてのキュレーション行為を
文書化(記録・管理)する

Data Curation Network's CURATED Checklist

CHECK Step

CURATE Action	Curator Checklist
<p>Check data files and read documentation</p> <ul style="list-style-type: none">• Review the content of the data files (e.g., open and run the files or code).• Verify all metadata provided by the author and review the available documentation.	<ul style="list-style-type: none"><input type="checkbox"/> Files open as expected<ul style="list-style-type: none"><input type="checkbox"/> Issues _____<input type="checkbox"/> Code runs as expected<ul style="list-style-type: none"><input type="checkbox"/> Produces minor errors<input type="checkbox"/> Does not run and/or produces many errors<input type="checkbox"/> Did not try to run code<input type="checkbox"/> Metadata quality is rich, accurate, and complete<ul style="list-style-type: none"><input type="checkbox"/> Metadata has issues _____<input type="checkbox"/> Documentation Type (circle) Readme / Codebook / Data Dictionary / Other: _____<ul style="list-style-type: none"><input type="checkbox"/> Missing/None<input type="checkbox"/> Needs work<input type="checkbox"/> Human subjects data present<ul style="list-style-type: none"><input type="checkbox"/> Request consent form / participation agreement

**DATA
CURATION
NETWORK**

データキュレーションネットワークの CURATEDチェックリスト

CHECK Step

CURATE Action	Curator Checklist
<p>データファイルをチェックしてドキュメントを読む</p> <ul style="list-style-type: none">データファイルの内容をレビューする (ファイルを開き実行する、またはコード化する等)著者から提供されたメタデータを検証し、利用可能なドキュメントをレビューする	<ul style="list-style-type: none"><input type="checkbox"/> ファイルが期待通りに開くか<ul style="list-style-type: none"><input type="checkbox"/> 問題点 ()<input type="checkbox"/> コードが期待通りに実行できるか<ul style="list-style-type: none"><input type="checkbox"/> 軽微なエラーが生成される<input type="checkbox"/> 実行不能または多数のエラーが生成される<input type="checkbox"/> 実行を試すこともできない<input type="checkbox"/> メタデータの品質について、豊富で、正確で、完全か<ul style="list-style-type: none"><input type="checkbox"/> メタデータの問題点 ()<input type="checkbox"/> ドキュメントタイプ (該当するものに○) Readme / Codebook / Data辞書 / その他: ()<ul style="list-style-type: none"><input type="checkbox"/> 欠落/なし<input type="checkbox"/> 要作業<input type="checkbox"/> 人が対象のデータが存在する<ul style="list-style-type: none"><input type="checkbox"/> 合意フォームや参加合意書をリクエストする

**DATA
CURATION
NETWORK**

Data Curation Network's CURATED Checklist

CHECK UNDERSTAND Step

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

CURATE Action

Understand the data (or try to)

- Check for quality assurance and usability issues such as missing data, ambiguous headings, code execution failures, and data presentation concerns.
- Try to detect and extract any "hidden documentation" inherent to the data files that may facilitate reuse.
- Determine if the documentation of the data is sufficient for a user with similar qualifications to the author's to understand and reuse the data. If not, recommend or create additional documentation (e.g., a readme.txt template).

Curator Checklist

Varies based on file formats and subject domain. For example...

Tabular Data Questions (Microsoft Excel)

- Organization of data well-structured
 - Not rectangular
 - Split tables into separate tabs
- Headers/codes clearly defined
 - Define headers
 - Clarify codes used _____
 - Clarify use of "blanks"
 - Clarify units of measurement
- Quality control clearly defined
 - Unclear quality control
 - Update/add Methodology

**DATA
CURA
NETWORK**

データキュレーションネットワークの CURATEDチェックリスト

CHECK UNDERSTAND Step

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the authenticity of available

CURATE Action

データを理解(しようと)する

- データの欠落、あいまいな見出し、コードの実行エラー、データ表現に関する懸念等、品質保証とユーザビリティに関する問題をチェックする
- かんたんに再利用できそうなデータファイルにありがちな「隠れたドキュメント」を見つけ抜き出す
- 作成者と同様の能力を持つユーザーがデータを理解して再利用するのに、データのドキュメントが充分かどうかを判断する。そうでない場合は、追加のドキュメント(readme.txtテンプレートなど)を推奨または作成する

Curator Checklist

データ形式および主題分野によって異なる。
例えば…

テーブル形式の問題 (Microsoft Excel)

- 適切に構造化されたデータ構成になっているか
 - 矩形でない
 - タブ区切りでない
- ヘッダー/コードは明確に定義されているか
 - ヘッダーを定義しているか
 - 使用コードは明確か ()
 - 「空白」の使用は明確か
 - 計測単位は明確か
- 品質コントロールは明確に定義されているか
 - 品質コントロールが不明確
 - 手法の更新/追加

DATA
CURA
NETWORK

Data Curation Network's CURATED Checklist

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data author's intent) to ensure the data is usable and the documentation is clear. This includes:

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST Step

CURATE Action

Request missing information or changes

- Generate a list of questions for the data author to fix any errors or issues.

Curator Checklist

Narrative describing the concerns, issues, and needed improvements to the data submission.

Sample email to researcher:

Dear [name of the data set author or contact],

Thank you for depositing your data set, [title of the data set] to the [name of repository].

After we receive a data set, we review it to ensure that the data sets we host are as complete, accessible and understandable as possible. We have reviewed your data set and have the following recommendations for you:

- Recommendation #1
- Recommendation #2
- Recommendation #3
- Recommendation #4

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about our recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Curator]

**DATA
CURA
NETWORK**

データキュレーションネットワークの CURATEDチェックリスト

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation conceals important information
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST Step

CURATE Action

- 欠落した情報を**要求**または変更する
- データ作成者がエラーや問題点を修正するための質問リストを生成する

Curator Checklist

データ提出に関する懸念、問題点および必要な改善点のざっくりぼらんな記述

研究者宛メールのサンプル

[データ作成者名または窓口担当者名]様

[リポジトリ名]への[データセットのタイトル]のご提出、ありがとうございます。

私たちは、データセットを受領後レビューして、ホストするデータセットができるだけ完全にアクセス可能で理解可能であることを確認します。お送りいただいたデータセットをレビューしたところ、次の推奨事項があります。

- 推奨事項1
- 推奨事項2
- 推奨事項3
- 推奨事項4

質問や追加情報の要求に対するご回答をお待ちしております。ご質問やご要望がありましたらお知らせください。ご希望に応じて、データセットに対する私たちのレビューについて電話か直接お会いして話し合うことができます。

よろしくお願いいたします。

[キュレーター名]

DATA
CURA
NETWORK

Data Curation Network's CURATED Checklist

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the author's intent) for reuse. Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation conceals important information. Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse. Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. README.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a list of the data author's issues.

AUGMENT Step

CURATE Action

Augment the submission

- Enhance metadata to best facilitate discoverability.
- Create and apply metadata for the data record, including descriptive keywords.
- When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems.

Curator Checklist

- Discoverability sufficient
 - Recommend (circle one) full-text index / file rename / file reorder / file descriptions / zip files into one archive
 - Other _____
- Keywords Sufficient
 - Suggestions _____
- Linkages Sufficient
 - Link to report/paper
 - Link to related data sets
 - Link to source data
 - Link to other _____

**DATA
CURA
NETWORK**

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about our recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Curator]

データキュレーションネットワークの CURATEDチェックリスト

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST AUGMENT Step

CURATE Action

Request missing information changes

- Generate a request for the data author or issues.

CURATE Action

提出物を補足する

- 発見可能性が最も高くなるようにメタデータを強化する
- キーワードの記述等、データレコード向けのメタデータを作成し適用する
- 必要に応じて、他のシステムとの相互運用性を促進するために、領域特有のスキーマでメタデータを構成および表示する

Curator Checklist

- 発見可能性は充分か
 - 推奨 (該当するものに○)
 - 全文インデックス / ファイル名変更 / ファイルの並べ替え / ファイルの説明文 / zipファイルにまとめてアーカイブ / その他 ()
- キーワードは充分か
 - 提案 ()
- リンクによる関連付けは充分か
 - 報告書/論文へのリンク
 - 関連するデータセットへのリンク
 - ソースデータへのリンク
 - その他へのリンク ()

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about our recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Curator]

DATA
CURA
NETWORK

Data Curation Network's CURATED Checklist

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation conceals important information
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. README.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a list of the data author's questions or issues.

AUGMENT Step

TRANSFORM Step

CURATE Action

Transform file formats

- Identify specialized file formats and their restrictions (e.g., Is the software freely available? Link to it or archive it alongside the data).
- Transform files into open, non-proprietary file formats¹ that broaden the potential audience for reuse and ensure that preservation actions might be taken by the repository in later steps. Retain original files

Curator Checklist

- Preferred file formats in use
 - Recommend conversion from _____ to _____
 - Retain original formats
- Software needed is readily available
 - Unclear version of software
 - Unclear software used
- Visualization of data easily accessible
 - Recommend graphical representation _____
 - Recommend web-accessible surrogate _____

**DATA
CURATION
NETWORK**

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about our recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Curator]

データキュレーションネットワークの CURATEDチェックリスト

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a request for the data author or issues.

AUGMENT Step

TRANSFORM Step

CURATE Action

- ファイル形式を変換する
- 特殊なファイル形式や制限を特定する (ソフトウェアはフリーで使えるか、データにリンクまたはデータと併せてアーカイブされているか等)
 - ファイルをオープンでプロプライエタリでないファイル形式に変換することで、再利用の潜在的な対象者を広げ、後のステップでリポジトリが保存アクションを取れるようにする。オリジナルのファイルは残す。

Curator Checklist

- 優先的なファイル形式はあるか
 - 推奨される変換 from _____ to _____
 - オリジナルの形式を保持する
- 必要なソフトウェアはすぐに使えるか
 - ソフトウェアのバージョンは明確か
 - 使われているソフトウェアは明確か
- 視覚化によりアクセスを容易にできるか
 - 推奨されるグラフィカルな表現 ()
 - 推奨されるWebアクセス可能な代替 ()

**DATA
CURA
NETWORK**

We look forward to hearing your response to our questions and requests for additional information.

Please do let us know if you have any questions about our recommendations. We would be happy to talk with you over the phone or meet with you in person to discuss our review of your data should you wish to do so.

Sincerely,

[Name of Curator]

Data Curation Network's CURATED Checklist

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a request for the data author or issues.

AUGMENT Step

TRANSFORM Step

EVALUATE Step

CURATE Action

Evaluate and rate the overall data record for FAIRness.*

- Score the dataset and recommend ways to increase the FAIRness of the data and become "DCN approved."

Curator Checklist

Findable -

- Metadata exceeds author/ title/ date,
- Unique PID (DOI, Handle, PURL, etc.).
- Discoverable via web search engines.

Accessible -

- Retrievable via a standard protocol (e.g., HTTP).
- Free, open (e.g., download link).

Interoperable -

- Metadata formatted in a standard schema (e.g., Dublin Core).
- Metadata provided in machine-readable format (OAI feed).

Reusable -

- Data include sufficient metadata about the data characteristics to reuse
- Contact info displayed if the direct assistance of the author needed.
- Clear indicators of who created, owns, and stewards the data.
- Data are released with clear data usage terms (e.g., a CC License).

DATA
CURA
NETWORK



データキュレーションネットワークの CURATEDチェックリスト

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data changes)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data that may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a request for the data author or issues.

AUGMENT Step

TRANSFORM Step

EVALUATE Step

CURATE Action

FAIR原則のためデータレコード全体を評価および査定する

- データセットにスコアを付け、データをFAIR原則に準拠させ「DCN承認」を得るための方法を推奨する

Curator Checklist

Findable (検索可能性)

- メタデータが優良か 作成者/タイトル/日付
- 一意のPIDの有無 (DOI, Handle, PURL等)
- Web上の検索エンジンで発見可能か

Accessible (アクセス可能性)

- 標準的なプロトコルで取得可能か (HTTP等)
- フリー、オープンか (ダウンロードリンク等)

Interoperable (相互運用性)

- メタデータが標準的なスキーマでフォーマットされているか (ダブリンコア等)
- メタデータが機械可読の形式で提供されているか (OAIフィード)

Reusable (再利用可能性)

- 再利用のためのデータ特性に関するメタデータがデータに充分含まれているか
- 作成者の直接の支援が必要な場合に連絡先は表示されているか
- データの作成者、所有者、管理者のインジケータは明確か
- データは明確なデータ使用条件でリリースされているか

DATA
CURA
NETWORK



Data Curation Network's CURATED Checklist

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a list of the data author's or issues.

AUGMENT Step

TRANSFORM Step

EVALUATE Step

DOCUMENT Step

CURATE Action

Document throughout curation activities.

- Record all necessary information capturing who did what to the dataset and when

Curator Checklist

- Accessioning & deposit records (Names, dates, contact information, submission agreements, etc)
- Repository collection metadata
- Provenance logs
- Service workflow
- Preservation packaging
- Any additional requirements at your institution

- Metadata provided in machine-readable format (OAI feed).
- Reusable -
 - Data include sufficient metadata about the data characteristics to reuse
 - Contact info displayed if the direct assistance of the author needed.
 - Clear indicators of who created, owns, and stewards the data.
 - Data are released with clear data usage terms (e.g., a CC License).

**DATA
CURA
NETWORK**



データキュレーションネットワークの CURATEDチェックリスト

CHECK

CURATE Action

Check data file documentation

- Review files (e.g. or code)
- Verify the author's availability

UNDERSTAND

CURATE Action

Understand the data (or the data's context)

- Check for quality as usability issues such as data, ambiguous headers, execution failures, a presentation concern
- Try to detect and explain "hidden documentation" inherent to the data may facilitate reuse
- Determine if the documentation of the data is sufficient with similar qualified author's to understand and reuse the data. If not, recommend or create documentation (e.g. readme.txt template)

REQUEST

CURATE Action

Request missing information or changes

- Generate a request for the data author or issues.

AUGMENT Step

TRANSFORM Step

EVALUATE Step

DOCUMENT Step

CURATE Action

キュレーション行為全体を通して
文書化(記録・管理)する

- 必要な全情報を記録する。記録にはデータセットにいつ誰が何をしたのかキャプチャーする。

Curator Checklist

- 受入、デポジットの記録 (名前、日付、連絡先、提出同意書等)
- リポジトリが収集するメタデータ
- 由来 (出所) の記録
- サービスのワークフロー
- 保存のパッケージング
- 機関における何らかの追加要件

- Metadata provided in machine-readable format (OAI feed).
- Reusable -
 - Data include sufficient metadata about the data characteristics to reuse
 - Contact info displayed if the direct assistance of the author needed.
 - Clear indicators of who created, owns, and stewards the data.
 - Data are released with clear data usage terms (e.g., a CC License).

**DATA
CURA
NETWORK**

Session 3 Outline

1. What is data curation?
2. Data Curation Network
3. Activity: data curation checklist
4. Discussion questions

Session 3 アウトライン

1. データキュレーションとは何か?
2. データキュレーションネットワーク
3. 実習: データキュレーションチェックリスト
4. 議論の課題

Session 3 Activity – Data Curation Checklist

We have created a mock dataset for you to review.

There are three printed handouts for this mock dataset:

1. data (a spreadsheet)
2. metadata (as text)
3. CHECK and UNDERSTAND checklists

Take 15 minutes to review the dataset you've been given. Work through as much of the CHECK and UNDERSTAND checklists as you can. Feel free to ask questions!

Session 3 実習 - データキュレーションチェックリスト

レビュー用の模擬データセットを用意しました。

模擬データセットについての3種類の配布資料があります。

1. データ (スプレッドシート)
2. メタデータ (テキストとして)
3. CHECKおよびUNDERSTANDチェックリスト

与えられたデータセットを15分でレビューしてください。
できるだけ多くのCHECKおよびUNDERSTANDチェックリストをこなしてください。質問はお気軽に!

Session 3 Outline

1. What is data curation?
2. Data Curation Network
3. Activity: data curation checklist
4. Discussion questions

Session 3 アウトライン

1. データキュレーションとは何か?
2. データキュレーションネットワーク
3. 実習: データキュレーションチェックリスト
4. 議論の課題

Discussion Questions - CHECK

Curator Checklist

- Files open as expected
 - Issues _____
- Code runs as expected
 - Produces minor errors
 - Does not run and/or produces many errors
 - Did not try to run code
- Metadata quality is rich, accurate, and complete
 - Metadata has issues _____
- Documentation Type (circle)
Readme / Codebook / Data Dictionary
/ Other: _____
 - Missing/None
 - Needs work
- Human subjects data present
 - Request consent form / participation agreement

1. Which items on the checklists were not relevant to this dataset?
2. Which items were relevant to this dataset?

議論の課題-CHECK

Curator Checklist

- ファイルが期待通りに開くか
 - 問題点 ()
- コードが期待通りに実行できるか
 - 軽微なエラーが生成される
 - 実行不能または多数のエラーが生成される
 - 実行を試すこともできない
- メタデータの品質について、豊富で、正確で、完全か
 - メタデータの問題点 ()
- ドキュメントタイプ (該当するものに○)
Readme / Codebook / Data辞書
/その他: ()
 - 欠落/なし
 - 要作業
- 人が対象のデータが存在する
 - 合意フォームや参加合意書をリクエストする

1. このデータセットに無関係だったチェックリストの項目はどれか?
2. どの項目がこのデータセットに関連しているか?

Discussion Questions - UNDERSTAND

Curator Checklist

Varies based on file formats and subject domain. For example...

Tabular Data Questions (Microsoft Excel)

- Organization of data well-structured
 - Not rectangular
 - Split tables into separate tabs
- Headers/codes clearly defined
 - Define headers
 - Clarify codes used _____
 - Clarify use of "blanks"
 - Clarify units of measurement
- Quality control clearly defined
 - Unclear quality control
 - Update/add Methodology

1. Which items on the checklists were not relevant to this dataset?
2. Which items were relevant to this dataset?

議論の課題 – UNDERSTAND

Curator Checklist

データ形式および主題分野によって異なる。
例えば…

テーブル形式の問題 (Microsoft Excel)

- 適切に構造化されたデータ構成になっているか
 - 矩形でない
 - タブ区切りでない
- ヘッダー/コードは明確に定義されているか
 - ヘッダーを定義しているか
 - 使用コードは明確か ()
 - 「空白」の使用は明確か
 - 計測単位は明確か
- 品質コントロールは明確に定義されているか
 - 品質コントロールが不明確
 - 方法の更新/追加

1. このデータセットに無関係だったチェックリストの項目はどれか?
2. どの項目がこのデータセットに関連しているか?

Discussion Questions

1. If this had been a real dataset, would someone who wanted to reuse the data struggle to understand it?
2. Which items seem most critical to you? Do we all think the same things are the most critical?
3. What do you see as challenges to curation?
4. What do you see as benefits?

議論の課題

1. これが実際のデータセットだった場合、データを再利用したい人はそれを理解するのに苦労するでしょうか？
2. 最も重要と感じる項目はどれですか？ 私たち全員が同じように最も重要と考えるでしょうか？
3. キュレーションの課題は何だと思えますか？
4. 利点は何だと思えますか？

Bonus – Compare to Published Data

Your mock dataset was based on a dataset from our colleague at Illinois, Christie Wiley, and her study “Assessing Research Data Deposits and Usage Statistics within IDEALS.”

Her full dataset is published here:

Wiley, Christie (2017): Data from: Assessing research data deposits and usage statistics within IDEALS. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-1235375_V1

Compare her real data and documentation with the example you were given.

おまけ – 公開データとの比較

模擬データセットは、イリノイの同僚、Christie Wileyと彼女の研究「IDEALSにおける研究データの蓄積と利用統計の評価」からのデータセットに基づいたものでした。

彼女の完全なデータセットは以下に公開されています。

Wiley, Christie (2017): Data from: Assessing research data deposits and usage statistics within IDEALS. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-1235375_V1

彼女の実際のデータとドキュメントを与えられた例題と比較してみてください。