



Title	オープンサイエンス時代における研究データマネジメント基礎
Author(s)	甲斐, 尚人; 西森, 哲也; 神崎, 隼人 他
Citation	
Version Type	VoR
URL	https://hdl.handle.net/11094/101413
rights	This article is licensed under a Creative Commons Attribution 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

研究データの取得・収集および保存・共有

本コンテンツは「AI等の活用を推進する研究データエコシステム構築事業」における大阪大学の成果である研究データ管理教材「オープンサイエンス時代における研究データマネジメントの基礎について学ぶ」を基に、大阪大学・九州大学・オープンアクセスリポジトリ推進協会が連携して改修したものです。

研究データの取得・収集、保存・共有

◆ DCC(Digital Curation Centre;英)のDMPテンプレート(一部)

大項目	詳細	備考
データ収集	どのようなデータを収集・生成するのか	データの種別、フォーマット、容量etc.
		フォーマットやソフトウェアのデータ共有・長期アクセスへの適性
		再利用できる既存のデータの有無
	どのようにデータを収集・生成するのか	用いる標準や方法論
		フォルダやファイルの構造化の方法
		バージョン管理の方法
保管とバックアップ	研究期間中のデータの保管・バックアップ方法	十分なストレージを有しているか
		付加サービスの利用に必要な料金を使途に含める必要があるか
		データのバックアップ方法
		障害発生時のデータの復元方法
	アクセス制限やセキュリティ確保の管理	
選定と保存	どのデータを保持・共有・保存すべきか	契約、法律、規制等の理由で保持あるいは破棄すべきデータ
		その他の保存するデータの決定方法
		データの保持・保存の期間
	データセットの長期保存計画の内容	リポジトリやアーカイブ等のデータの保存場所

- 内閣府研究データ基盤整備と国際展開ワーキング・グループ(2019, 10)「研究データ基盤整備と国際展開ワーキング・グループ報告書—研究データ基盤整備と国際展開に関する戦略—」, p.20-24. <https://www8.cao.go.jp/cstp/tyousakai/kokusaioopen/houkokusho.pdf>
- この内閣府の資料は、英国のDCC(Digital Curation Centre)のDMPプレートを参考に作られているため、合わせて参照のこと。

研究データの取得・収集

どのようなデータを収集・生成するのか

- データの種別、フォーマット、容量 etc.
- フォーマットやソフトウェアのデータ共有・長期アクセスへの適性
- 再利用できる既存のデータの有無

データの種別は概ね下記のいずれかに分類

集計データ、臨床試験データ、編集データ、符号化データ、実験データ、地理空間データ、測定・評価データ、観測データ、記録データ、シミュレーションデータ、調査データ

「JPCOARスキーマガイドライン」に基づく研究データのdataset区分に関しては、Cf., オープンアクセスリポジトリ推進協会(2019, 11, 27)「資源タイプ語彙別表」. [https://schema.irdb.nii.ac.jp/ja/2.0/resource type vocabulary](https://schema.irdb.nii.ac.jp/ja/2.0/resource%20type%20vocabulary)

どのようにデータを収集・生成するのか

- 用いる標準や方法論
- フォルダやファイルの構造化の方法
- バージョン管理の方法

- 関係する法令、学内規則、研究倫理その他の規範の遵守
- FAIR原則に則りデータ来歴を明らかにした上で保存
- 保存期間終了後には適切に分類(保存する研究データ、破棄する研究データなど)・保存し、適切な処理が必要

研究データの取得・収集

3.1 フォルダ構造

フォルダ構造の決定(組織化)

研究が進むに従って、数多くのデータ(ファイル)が生成されます。一貫性のある適切なルールを設けてフォルダ構造を組織化することで、効率的なアクセスが可能になり、効率的に研究を進めることにも繋がります。

- **フォルダ構造の組織化**とは
 - データファイルをフォルダに格納する際のルールを定めること
 - フォルダ名等でフォルダ間の関係を明確にしておくこと
 - 定めたルールに従ったフォルダ名にすること
- **メリット**
 - 一定のルールの下でファイルが保存され、場所の特定と閲覧が容易
 - 作成者だけでなく他の人もファイルを容易に区別・発見することが可能
 - 誤った上書き、削除の防止
 - フォルダ名等でバージョン違いを明確に区別することが可能
 - 別のプラットフォームに移行しても文脈を維持することが可能

研究データの取得・収集

3.1 フォルダ構造

フォルダ構造の決定(組織化)

- フォルダ構造に関するルールの例
 - 研究プロジェクトごとにフォルダを作成し、フォルダ名には日付をいれる
 - 実験や調査ごと、またはメンバーそれぞれでフォルダを作成する
 - フォルダの下にはそのフォルダの中身を説明する文書を作成しておく
 - 実験の場合は、実験のフォルダの下に、生データ、加工データ、論文等のフォルダをそれぞれ作成する
 - 様々な実験や調査をする場合には、実験や調査の違いが分かるような情報をフォルダ名に含める
 - 深すぎる構造にしない
 - ファイルの命名法も参考にする(ファイルの命名法に関しては次で説明)

適切なフォルダ構造は、研究内容によっても異なります。
研究プロジェクトごとに、相談して一定のルールを決めましょう。

研究データの取得・収集

3.2 ファイルの命名法

ファイルの命名法(1/3)

ファイルの命名は、研究データの出所を保証したり、その後の利用をする際に間違ったデータを使用するのを避けるためにも重要です。命名時は、主に以下を考慮しましょう。

1. ファイル名の構成

- 保存先システムのファイル名の制約を考慮する
- &*\$%\$£]{!@などOSで特別な意味を持つ場合がある特殊記号は用いない
- ピリオド、スペースの代わりにアンダーバーを用いる
- ファイル名は短く示唆的なものにする
- ファイル名の適切な長さは約25文字程度で内容がわかるものにする

研究データの取得・収集

3.2 ファイルの命名法

ファイルの命名法(2/3)

1. ファイル名の構成(つづき)

- 保存先を変更する際に混乱が生じるような一般的なファイル名は使用しない(例:sheet1.xlsx)
- 実験機器等が出力するファイル名をそのまま利用しない。
- ファイル名を付ける際に考慮すべき要素
 - プロジェクト番号
 - 作成者名、データに関与した研究チーム、部局の名前
 - データ内容の記述
 - データ作成日、公開日
 - バージョン番号
- ファイル名の例
[作成者]_[作成方法]_[キーワード]_[YYYYMMDD]_[バージョン].拡張子

研究データの取得・収集

3.2 ファイルの命名法

ファイルの命名法(2/3)

2. 文脈の保存

- 保存場所に関係なくデータを識別できるように内容情報をできるだけ含める
- ファイルの拡張子は、ファイルが作成された物理的環境を正確に示すものを用いる 例).por, .xls, .xlsx

3. 一貫性のある命名

- 命名ルールに従い、常に同じ情報(日付や時間)を同じ順番で含む
- 日付を用いる場合、フォーマットを統一する。例)20221201 (yyyymmdd)

研究データの取得・収集

4.5 バージョン管理

バージョン管理

- 多くの場合、研究者は独自の方法を利用し、データセットを始めとするファイルのさまざまなバージョンを識別。
- 適切に設計されたバージョン管理システムを利用すれば、ファイルに加えられた変更をバージョン番号、タイムスタンプ、変更箇所の説明とともに記録できる。変更は容易に比較することができ、必要に応じて復元することも可能。
- バージョン管理システムの一般的な例…Git

研究データの取得・収集

4.6 フォーマット

ファイルフォーマット

論文を執筆するまでの短期的な視点ではなく、長期的に研究データが利用できる状態を確保するためには、フォーマットによる特徴の違いを考慮し、適切なフォーマットを選択する必要があります。

- 推奨されるファイルフォーマット
 - 特定の商用ソフトウェアから独立したフォーマット
 - 仕様が国際標準化されているフォーマット
 - 広く普及しているフォーマット
- フォーマット変換のリスク
 - フォーマットを変換することにより、ファイル内部に記録されたメタデータや画像の品質、文章構造などの情報を失う可能性大。
- テキストとバイナリ
 - テキストファイルは環境に依存せず開くことができ、人間可読性高。
 - バイナリファイルはよりリッチな情報を包埋しているが、人間可読性低。

研究データの保存

研究期間中のデータの保管・バックアップ方法

- 十分なストレージを有しているか
- 付加サービスの利用に必要な料金を使途に含める必要があるか
- データのバックアップ方法
- 障害発生時のデータの復元方法

どのデータを保持・保存すべきか

- 契約、法律、規制等の理由で保持あるいは破棄すべきデータ
- その他の保存するデータの決定方法
- データの保持・保存の期間

「大阪大学における研究データの保存等に関するガイドライン」
https://www.osaka-u.ac.jp/ja/research/researchintegrity/files/data_guideline.pdf

アクセス制限やセキュリティ確保の管理

ストレージ選定における留意点

- 情報セキュリティポリシーに適應したシステム・ストレージを選定
- 求められるアクセス制限やアカウント管理のレベルに応じた選定
- 個人でインフラ整備の必要があれば、関連部門と相談のうえセキュリティ強化されたストレージを選定

「大阪大学情報セキュリティーポリシー」
(大阪大学情報推進本部)

<https://www.oict.osaka-u.ac.jp/securitypolicy/>

研究データの保存

4.3 保存場所

データの保存場所(1/2)

容量や費用、簡便さに加え、信頼性や安全性など各保存場所の特徴を考慮し、保存先を選択しましょう。

代表的な4つの保存場所

1. デスクトップ・モバイルPC

簡便な保存先ではあるが、ハードウェアが冗長化されておらず、障害によりファイルを失う可能性大。研究資料のマスターコピーを保存する場所としては不適切。

2. 外部記憶デバイス

CDやDVDはサイズが十分ではなく、書き込みエラーへの対処や確認が必須。USBハードディスクやUSBメモリ、SDカードなどは、安価で大容量化してきたが、PC・ラップトップと同様の問題あり。ポータビリティに優れる反面、紛失や盗難などのセキュリティリスク大。

研究データの保存

データの保存場所(2/2)

3. クラウドストレージ (例:OneDrive, Dropbox)

インターネット上のディスクスペースを利用できるサービス。一定容量までは無料で利用できるものもあるが、機関のセキュリティポリシーやガイドラインに適合するものを選択すべき。

*オンラインストレージサービスの比較:

https://en.wikipedia.org/wiki/Comparison_of_online_backup_services

4. ネットワークドライブ (例:研究室のNAS)

LANなどのネットワーク経由で接続できる他のコンピュータのディスクスペース。安定的に運用される所属機関のシステム管理部門が提供するものを利用するのが理想。

外部のクラウドストレージを利用する際は、機関のセキュリティポリシーに合致するかを事前に確認しましょう。所属する研究機関が提供するネットワークドライブ、クラウドストレージがあれば、安心して研究データを保存できます。クラウドストレージやネットワークドライブ等を利用して複数の環境で作業する際には、確実にファイルを同期させるようにしましょう。

研究データの保存

4.4 バックアップ

研究データのバックアップ(1/3)

不慮のトラブルで研究データを失った際にも、バックアップデータがあれば、トラブルの影響を最小限に抑えることができます。
定期的にバックアップを取りましょう。

- バックアップの場所
 - ネットワークドライブ、クラウドストレージなどの選択肢から、研究内容に応じて適切なものを選択
 - 情報セキュリティや災害リスクも考慮
 - 絶対に漏洩してはいけないデータは、オンライン上ではなく、ローカルのコンピュータやストレージに保管
- 複数のバックアップを取る
- 機密性のレベル等について各機関が提供する情報を確認する

研究データの保存

4.4 バックアップ

研究データのバックアップ(2/3)

- バックアップの方法
 - 手動ではなく、バックアップソフトウェアなどを利用し、定期的なバックアップを取得する
 - 定期的にバックアップされていることを確認。
 - バックアップからリストアできることを確認。
 - ディスク容量やバックアップに必要な時間を勘案し、**フルバックアップ**、**増分・差分バックアップ**から適切な方法を選択する
 - ファイルの更新頻度などから、適切なバックアップ頻度を設定。
例) 日次:差分バックアップ、月次:フルバックアップ

研究データの保存

4.4 バックアップ

研究データのバックアップ(3/3)

- 災害等を考慮し、複数の場所にバックアップを保管しましょう。
- 最低でも**3つのコピー**があることが推奨されています。
 - 「**オリジナル(original)**」:最初のコピー、作業用のデータセットと関連ファイル。通常、研究者が主に使用するコンピュータに保管。
 - 「**ニア(near)**」:オリジナルの近くになければならないが、物理的な場所が同じでないことが理想的。このコピーは自動バックアップソフトか手動のどちらかで更新。多くの場合、ニア・コピーは外部ハードドライブや研究者の属する施設内の共有ファイルサーバに保管。
 - 「**ファー(far)**」:オリジナル・コピーからもニア・コピーからも物理的な場所が遠い。災害の脅威が異なる場所に置くのが理想。

研究データの保存

4.7 セキュリティ対策

セキュリティ対策

セキュリティポリシーに沿った研究活動や環境を整備することで、トラブルを回避し、総合的には円滑な研究活動に繋がります。

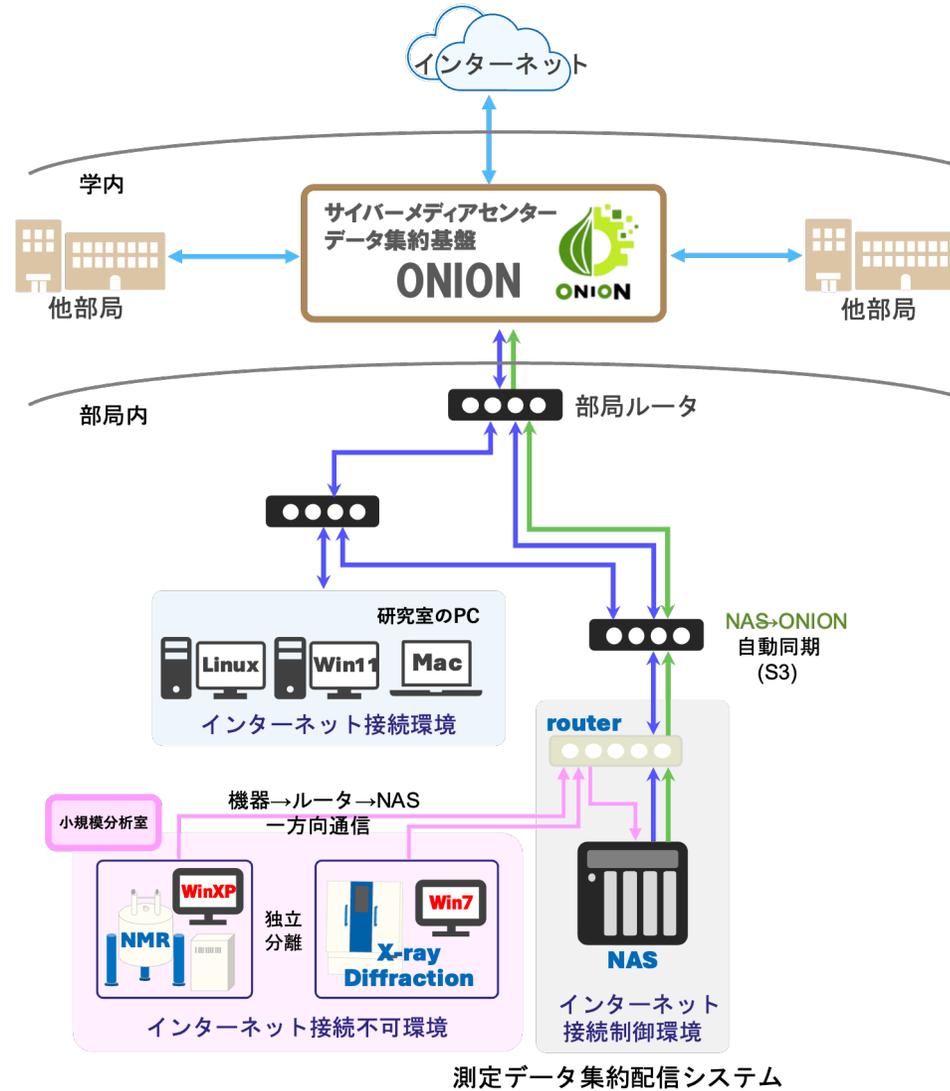
- 物理的なセキュリティ対策
 - 離席する際のコンピュータをロック
 - 部屋の施錠
 - モバイルPCを放置しない
 - USBメモリ等を紛失しないための対処
- PC上でのセキュリティ対策
 - アンチウイルスソフトウェア／ファイアウォールの導入
 - ソフトウェアやOSのアップデート
 - 適切な強度のパスワードや認証方法の利用
 - ファイルやハードディスク、USBメモリ等の暗号化

研究データの共有



研究データ共有 非公開共有 自然科学分野

- ◆ 大阪大学コアファシリティ機構が推進する測定データの集約・管理を例に



実践的な教材の整備・展開の具体例(実験データ)

実験系研究データ管理の実践

小規模分析室測定データ集約・配信システムを用いた
測定データの管理 ~データ取得から共有・公開~

大阪大学 コアファシリティ機構
オープンサイエンス推進室



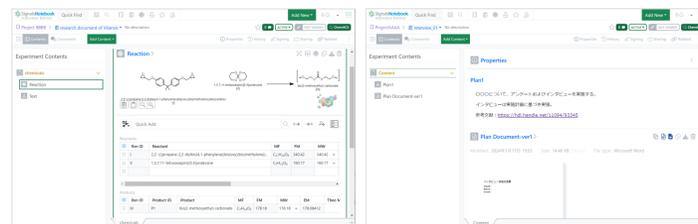
データを記録・管理する：電子実験ノートの利用

■ Signals Notebook Individual版の利用案内

- 本学教職員はアカウント登録のみで無料で利用可能 (令和4年から令和7年度の期間)

マイハンダイ [Software]アイコン

[Software]



実験系研究データ管理の実践

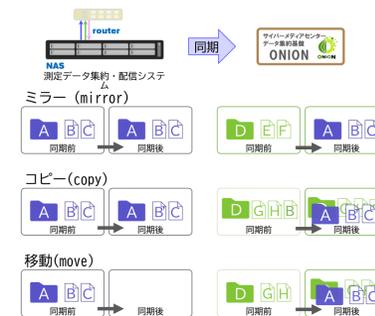
利用詳細は利用講習会動画で！ <https://el.hanna-nw.org/osaka-u/27/menu>

11

NASの設定 ONIONとの同期設定

■ 同期アクション

- ミラー
NASのデータをONIONにコピーする
→同期後の双方のフォルダの内容物は同一
ONIONに保存されていたデータは削除される
- コピー
NASの新規および更新データをONIONにコピーする
→ONIONに保存されていたデータに追加で新規および更新データがコピーされる
- 移動
NASのデータをONIONに移動する
→ NASに保存されていたデータは削除され、データはONIONでのみ保存される



実験系研究データ管理の実践

QNAP Hybrid Backup Sync, QNAP, n.d. https://docs.qnap.com/application/hybrid-backup-sync/9/21_xfm-us-creating-a-one-way-sync-job-624EE600.html. (2024, Jan 25)

24

小規模分析室測定データ集約・配信システムについて

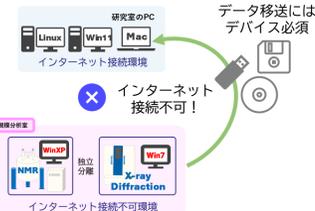
<背景>

- 分析機器制御PCの一部にはサポート対象期間外のものが稼働 (Win7・XP etc.)

インターネット接続は基本的に禁止

測定データの移送はUSBメモリなど
デバイスが必要

- 手間がかかる
- 紛失リスク
- ウイルス感染リスク …



小規模分析室向け
測定データ集約・配信システム

実験系研究データ管理の実践

4

データを記録・管理する：電子実験ノートの利用

■ 紙媒体と電子媒体のそれぞれの特徴 (一例)

	紙媒体	電子媒体
メリット	<ul style="list-style-type: none"> • いつでもどこでも使用できる (電源/ネットワーク不要) • 利用に他の媒体が不要 • データ付加が容易 (糊付けなどができる) • 経年劣化が緩やか (消失リスクが低い) 	<ul style="list-style-type: none"> • 検索機能がある • 証跡機能がある • リンク (電子的紐づけ) 機能がある • 複製利用が容易 • 保管場所が不要が省スペース
デメリット	<ul style="list-style-type: none"> • 検索・リンク機能がない • 証跡機能がない • 複製利用が難しい • 保管に場所 (スペース) が必要 	<ul style="list-style-type: none"> • 電源 (およびネットワーク) が必要 • 利用にデバイスが必要 • 記録媒体の劣化がある (消失リスクが高い) • (セキュリティ)コストが高め

実験系研究データ管理の実践

12

- 甲斐尚人 et al.(2024, 6, 20)「E1:研究データの未来を築く:研究データエコシステム構築のための人材育成」, p.6. <https://joss.rcos.nii.ac.jp/session/overview/>
- 吉賀夏子 et al.(2024, 8)「実験系研究データ管理の実践(研究者向け・管理者向け)」, <https://hdl.handle.net/11094/97150>

研究データ共有 公開共有 人文学・社会科学分野

◆ 日本学術振興会『人文学・社会科学におけるデータ共有のための手引き』



・「自然科学と同様、**人文学・社会科学においても、研究結果の根拠となるデータを研究者のあいだで共有できるようにすることは、研究者コミュニティにおける相互批判と研究の発展にとって不可欠のものです。**」(p.3)

・「管理対象となるデータには数量的データの他にもさまざまな種類のデータや資料、またいくつもの媒体手段(復元・再生手段の指定を含む)が存在します。」(p.9)

・「データ作成の有用性は人文学においても同様です。研究資料を整理・分析するにあたり研究している時点での効率や気づきの可能性を高めただけでなく、再利用しやすい形で作成しておけば、将来の自分の研究を発展させるにあたって有効に働いてくれます。」(p.4)

・「人文学においては、文化的な活動を記録したデータであれば、なんであれ研究対象となり得るため、データのフォーマットとしてはあらゆるものを想定する必要があります。」(p.18)

研究データ共有 公開共有 人文学・社会科学分野

◆ 人文学・社会科学におけるデータ共有の例

人文学・社会科学総合データカタログ



人文学オープンデータ共同利用センター



ROIS-DS人文学オープンデータ共同利用センター (Center for Open Data in the Humanities / CODH) は、情報学・統計学の最新技術を用いて人文学資料 (史料) を分析する「データ駆動型人文学」や、人文学研究の成果に基づき構築したデータセットを超学際的に活用する「人文学ビッグデータ」など、オープンサイエンス時代の新しい人文学研究を展開します。[もっと詳しく..][CODHパンフレット..]

- 左) 日本学術振興会(n.d.)「人文学・社会科学総合データカタログ(JDCat)」. <https://jdcatalog.jp/>
- 右) データサイエンス共同利用基盤施設(n.d.)「人文学オープンデータ共同利用センター」. <http://codh.rois.ac.jp>

実践的な教材の整備・展開の具体例(人文学データ)

人文学研究者の多くが一次資料の代替となる高精細な画像を用いた分析を求めている。

⇒画像および動画の相互運用を推進する国際的な枠組みであるIIIF(International Image Interoperability Framework)に着目。

※IIIF画像を用いたキュレーションや教育への応用、ならびに機械学習のためのアノテーションデータ付与機能などの開発が盛んに行われている。

AI等の活用を推進する研究データエコシステム構築事業
研究データ管理教材 実践編
OSAKA UNIVERSITY

グローバル日本学教育研究拠点「拠点形成プロジェクト」
一人文学分野向け研究データ管理を促進するデジタル・ヒューマニティーズ学習教材開発

人文学研究者必見！研究データ管理ことはじめ
--- OUKAで始めるIIIF画像の公開と利活用 ---

データ公開の可否を決める

1. 公開制約条件の解除を試みる
 - 個人情報を含むデータを匿名化
 - データについて権利者の許諾
 - 猶予期間を設けて公開を延期
2. 公開できる場合: 公開先を決める
 - ストレージあるいはリポジトリが必要
 - データの形式やサイズ、保存期間、アクセス制御(書誌項目)の記述などの要件に合致すること

アノテーションと共有

くずし字に楷書の文字をあてている例 (Kuzushiji.workより)

2 138 Go 3 4 5

MORE INFORMATION 4 5

About the Item

Title
懐徳堂考

Source
https://hdl.handle.net/11094/89231

Author
西村, 天因; ニシムラ, テンシュウ

Collection
懐徳堂文庫

Date of Issue
1910

書誌情報の表示

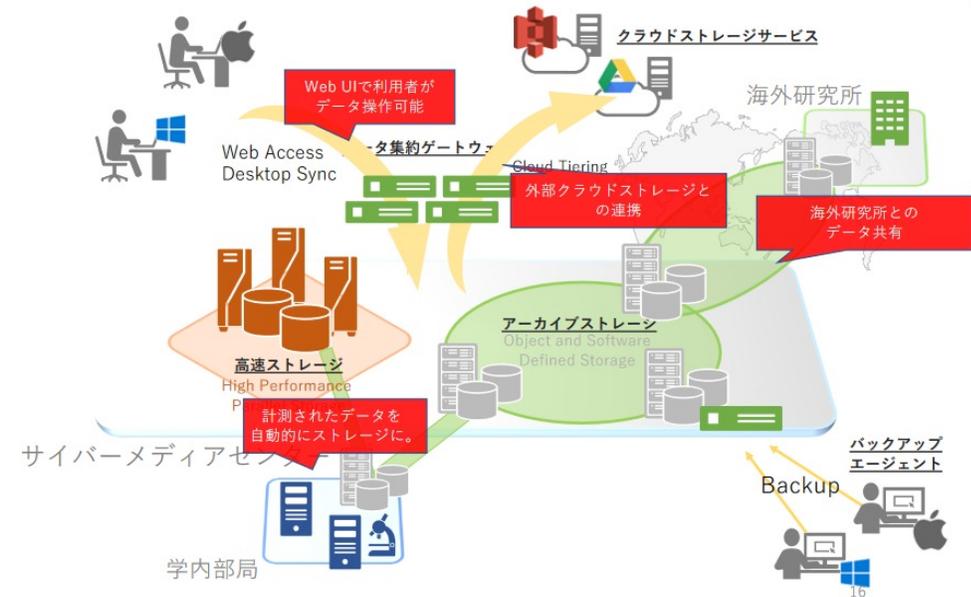
- 甲斐尚人 et al.(2024, 6, 20)「E1:研究データの未来を築く:研究データエコシステム構築のための人材育成」, p.6. <https://joss.rcos.nii.ac.jp/session/overview/>
- 吉賀夏子 et al.(2024, 8)「人文学研究者必見！研究データ管理ことはじめ：OUKAで始めるIIIF画像の公開と利活用」, <https://hdl.handle.net/11094/97757>

ストレージサービス(大阪大学の場合)

◆ ONION(研究データ集約基盤)



データ集約基盤 ONION 概要



DDN Japan User Forum 2021 - 2021.07.16 -

- ONIONを活用することで、利用者の環境とスーパーコンピュータ間のデータ移動が簡単になる。また、SQUIDやOCTOPUSのアカウントを持っていない海外や企業の共同研究者と即座に計算結果を共有する、スマートフォンからデータ操作をすることができるなど、様々な利用方法が考えられる
- 研究室の研究データの保存・共有などにも利用可能。

- 大阪大学サイバーメディアセンター(n.d.)「ONION」. <http://www.hpc.cmc.osaka-u.ac.jp/onion/>
- 伊達進・寺前勇希・勝浦裕貴・木越信一郎・木戸善之(2022)「大阪大学のデータ集約基盤ONION」『学術情報処理研究』26(1):87-96. http://www.hpc.cmc.osaka-u.ac.jp/public_data/axies2021_paper_onion/

参照資料

- 大阪大学サイバーメディアセンター(n.d.)「ONION」. <http://www.hpc.cmc.osaka-u.ac.jp/onion/>
- オープンアクセスリポジトリ推進協会(2019, 11, 27)「資源タイプ語彙別表」. https://schema.irdb.nii.ac.jp/ja/2.0/resource_type_vocabulary
- 伊達進・寺前勇希・勝浦裕貴・木越信一郎・木戸善之(2022)「大阪大学のデータ集約基盤ONION」『学術情報処理研究』26(1):87-96. http://www.hpc.cmc.osaka-u.ac.jp/public_data/axies2021_paper_onion/
- データサイエンス共同利用基盤施設(n.d.)「人文学オープンデータ共同利用センター」. <http://codh.rois.ac.jp>
- 内閣府研究データ基盤整備と国際展開ワーキング・グループ(2019, 10)「研究データ基盤整備と国際展開ワーキング・グループ報告書—研究データ基盤整備と国際展開に関する戦略—」, p.20-24. <https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/houkokusho.pdf>
- 西園由依 et al.(2020, 10, 29)「研究者のための研究データマネジメント_研究前_所属機関のインフラ活用」『教材「研究者のための研究データマネジメント」』, p.6. <https://jpcoar.repo.nii.ac.jp/records/294>
- 日本学術振興会(2021, 11, 8)『人文学・社会科学におけるデータ共有のための手引き』<https://www.jsps.go.jp/j-di/guide.html>
- 日本学術振興会(n.d.)「人文学・社会科学総合データカタログ(JDCat)」. <https://jdcat.jsps.go.jp/>
- 吉田幸苗 et al.(2021, 2, 10)「研究データ管理サービスの設計と実践_第3章_第2版」『教材「研究データ管理サービスの設計と実践」』, p.23. <https://jpcoar.repo.nii.ac.jp/records/607>
- 九州大学データ駆動イノベーション推進本部研究データ管理支援部門, 九州大学附属図書館図書館DX支援室(2023, 08)「はじめよう、研究データ管理」, 3_研究データの整理法_202407改訂, <https://doi.org/10.15017/6796147>
- DMPonline(n.d.)「Funder requirements」. https://dmponline.dcc.ac.uk/public_templates