研究実施中

研究データの保存

Data Preservation
- Execution Phase -

このブックでのトピック

- 研究データ管理における所属機関の役割
- 研究データのリスク評価
- センシティブデータの取扱い
- 研究データの保存先
- 研究データの情報セキュリティ対策
- 研究データのバックアップ
- 研究データの保存フォーマット
- 研究データ管理に有用なツール

研究データ管理が求められる背景を理解する

オープンサイエンスの流れ

- 内閣府『我が国におけるオープンサイエンス推進のあり方について』（2015年3月）
- 日本学術会議『オープンイノベーションに資するオープンサイエンスのあり方に関する提言』(2016年7月)

研究活動の不正行為への対策

- 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)
- 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**

  →**教職員の異動・退職後も、機関として保存又はアクセスを確保**

所属機関の研究データ保存に関するガイドライン、研究助成機関のデータポリシーを把握する

- 例：京都大学

『京都大学における公正な研究活動の推進等に関する規程』(2015年7月)　第７条

- 「教職員等は、適正な保存方法により、一定期間研究データを保存し、必要に応じて当該研究データを開示しなければならない。」

http://www.kyoto-u.ac.jp/uni_int/kitei/reiki_honbun/w002RG00001165.html

- 例：科学技術振興機構

国立研究開発法人 科学技術振興機構 Japan Science and Technology Agency

『オープンサイエンス促進に向けた研究　成果の取扱いに関するJSTの基本方針』（2017年4月）

- JSTが研究資金を配分するプロジェクトについては、データマネジメントプランを作成、提出
- データの適切な保管・管理
- エビデンスデータの公開を推奨

→関連：「研究後　研究データの公開」-「データ共有・公開にかかる義務・制限の確認」

所属機関の研究データ保存に関するガイドライン、研究助成機関のデータポリシーを把握する

- 関連する多くのポリシー、規程、ガイドラインが既に制定されている。
  - 公正な研究活動のためのガイドライン
  - 個人情報保護に関する規程
  - 情報セキュリティポリシー 等

- 研究データポリシー
  - 「研究データ」の取り扱い全般について、機関の基本方針を定める

内閣府「国立研究開発法人におけるデータポリシー策定について」
https://www8.cao.go.jp/cstp/stsonota/datapolicy/datapolicy.html

内閣府「第6期科学技術・統合イノベーション基本計画(2021年)」では、「機関リポジトリを持つ大学等」にデータポリシーの策定を求めている(本文 59ページ)
https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html

研究データに関する倫理的な問題・法的な問題を理解し、適切な利用・管理が必要なことを認識する

- 倫理的な問題
  - データの保存、共有について利害関係者から同意を得ているか
  - 個人識別性は保護されているか（データの匿名化、不要な項目の削除等）
  - データの安全な保管・移動のための取り組みがされているか

- 法的な問題
  - データの所有者は誰かはっきりしているか
  - 再利用のためにどのようなライセンスを付与するか
  - 第三者による再利用について何らかの制限を定めるか

研究データを機密性、完全性、可用性の観点から格付けする

機密性 Confidentiality	• アクセスを認可された者だけが、情報にアクセスできるようにすること • 情報漏えい防止、不正アクセス対策
完全性 Integrity	• 情報や情報の処理方法が正確で完全であるようにすること • 改ざん防止、データの差異の検出などの対策
可用性 Availability	• 許可された者が、必要な時に、情報や情報資産にアクセスできるようにすること • バックアップ、システムの二重化などの対策

Ref:上原孝之 情報処理安全確保支援士2021年版 翔泳社

機密性についての格付けの定義例

機密性格付け区分	分類の基準
機密性３情報	秘密文書※に相当する機密性を要する情報 ※「行政文書の管理に関するガイドライン」では二つに分類： • 極秘文書：秘密保全の必要が高く、その漏えいが国の安全、　　利益に損害を与えるおそれのある情報を含む行政文書 • 秘文書：極秘文書に次ぐ程度の秘密であって、関係者以外には知らせてはならない情報を含む極秘文書以外の行政文書
機密性２情報	秘密文書に相当する機密性は要しないが，その漏えいにより利用者の権利が侵害され又は大学や研究機関としての活動の遂行に支障を及ぼすおそれがある情報
機密性１情報	機密性２情報及び機密性３情報以外の情報

Ref:内閣サイバーセキュリティセンター サイバーセキュリティ戦略本部　政府機関等のサイバーセキュリティ対策のための統一基準群 https://www.nisc.go.jp/policy/group/general/kijun.html
内閣総理大臣決定「行政文書の管理に関するガイドライン」(2011年4月)
http://www8.cao.go.jp/chosei/koubun/hourei/kanri-gl.pdf

完全性・可用性についての格付けの定義例

完全性格付け区分	分類の基準
完全性２情報	改ざん，誤びゅう又は破損により，利用者の権利が侵害され又は大学や研究機関としての活動の遂行に支障（軽微なものを除く）を及ぼすおそれがある情報
完全性１情報	完全性２情報以外の情報（書面を除く）

可用性格付け区分	分類の基準
可用性２情報	滅失，紛失又は当該情報が利用不可能であることにより，利用者の権利が侵害され又は大学や研究機関としての活動の安定的な遂行に支障（軽微なものを除く）を及ぼすおそれがある情報
可用性１情報	可用性２情報以外の情報（書面を除く）

Ref:内閣サイバーセキュリティセンター サイバーセキュリティ戦略本部　政府機関等のサイバーセキュリティ対策のための統一基準群
https://www.nisc.go.jp/policy/group/general/kijun.html

情報の種類に応じて適切に研究データを管理する

- 非保護情報（機密性1・完全性1・可用性1）**以外**の情報は「**要保護情報**」にあたる
- 「**要保護情報**」は所属機関の情報セキュリティーポリシーやガイドラインに定義された**取扱制限**に基づき、適切に管理する必要がある

区分	取り扱い制限の例
機密性に基づく制限	複製、配布、暗号化、印刷、転送、転記、再利用、送信の禁止・要許可・必須などの制限を規定
完全性に基づく制限	保存期間、保存場所、書き換え、削除、保存期間満了後の措置などについて規定
可用性に基づく制限	復旧までに許容できる時間、保存場所などについて規定

センシティブデータとは

- 個人情報
- 国家安全保障に関するもの
- 知的財産権
- その他特に配慮を必要とするもの（例：絶滅危惧種の生息地情報）

センシティブデータを取り扱う際に注意すること

- 取得したデータの管理方法や利用方法を説明し、本人及び関係者から同意を得る
- 所属機関の情報セキュリティポリシーを確認し、データの取り扱い制限（保存先・保存期間・セキュリティ対策・アクセス可能な者の範囲等）に従う
  → 機関の知財担当部門・共同研究者等への確認が必要な場合もある
- データの解析・保存等で外部サービスを利用する場合、データが公開・漏洩される恐れがないかサービスの利用規定や利用条件等を確認する
- 研究データを公開・共有する際には適切な匿名化処理がされていることを確認する

Ref:オープンサイエンスのためのデータ管理基盤ハンドブック～学術研究者のための“個人情報”の取扱い方について～，オープンサイエンスを推進するデータ基盤とその利活用に関する検討委員会
https://repository.nii.ac.jp/records/2000318
https://doi.org/10.20736/0002000318

→関連：「研究後　研究データの公開」-「データ共有・公開にかかる義務・制限の確認」

センシティブデータを取り扱う際に注意すること

- センシティブデータを預かる情報システムについては、特に以下の項目に対する配慮が必要となる。
  （JIS Q 27002の項目）
  - 8　資産の管理　（例　データの保存期間、データの廃棄方法、等）
  - 9　アクセス制御　（例　アクセス権の管理方法、等）
  - 10　暗号　（例　記録メディアに対する暗号化、等）
  - 11　物理的及び環境的セキュリティ　（例　データセンターへのアクセス(出入り)、等）
  - 12　運用のセキュリティ　（例　アクセスログの記録、監査、等）
  - 13　通信のセキュリティ　（例　暗号化した経路での通信、等）

- 必要であれば、データセンター運営や、データの取り扱いに関する第三者機関の認証取得も検討する。
  - ISMS 認証を取得した事例
    - NII RCOS (GakuNin RDMやJairo Cloud を運用)
    - 東北大メディカル・メガバンク機構 https://www.megabank.tohoku.ac.jp/news/14648

保存場所を選択する

- 事前に所属機関の情報セキュリティポリシーや研究助成機関のデータポリシーなどで、データの保存・公開に関する規定を確認する
- ストレージに求める条件（利用料金・容量・保存期間・バージョン管理の有無・対応フォーマット・共有機能等）を考慮してストレージを選ぶ
- 所属機関のネットワークドライブや、組織として契約したオンラインストレージが条件を満たしていれば優先的に使用する
- 外部の商用クラウドサービス等を利用する場合は、情報セキュリティに関して条件設定の余地があるものを選び、機関内の外部委託に関する規程やクラウドサービスの利用に関する規程に反していないかあらかじめ確認する

研究データの保存場所の種類

保存場所	データの保存先としての適正	
PC本体のHDD 外部記憶デバイス （USBメモリ・外付けHDD）	△	破損・障害によるデータ消失のリスク、紛失・盗難などのセキュリティリスクに留意する
所属機関が提供する オンラインストレージ	○	所属機関の情報基盤部門が提供し、安定的に運用されているものであれば保存先として最適 機関が外部クラウドサービスと契約している場合、その利用条件にしたがって利用可否を判断
外部事業者が提供する オンラインストレージ	？	機関の規程への適合確認、リスク評価が必須 個人で登録・利用することは慎重に (シャドウIT化の防止)

ハードウェアの購入や、オンラインストレージの契約上の注意

- 容量や費用、利用のしやすさ以外に、障害発生時のサポート内容、データが破損・消失した場合の補償、情報セキュリティ対策なども考慮する
- オンラインストレージの場合、サービスのガイドライン・利用規約・約款等に目を通し、所属機関の情報セキュリティポリシーに反しないか、預けたデータが公開・漏洩される危険性がないか、保管や共有に対して制限がないか等を確認する
- 納品後またはサービス開始後に、提供された内容と契約時に提示された内容に差異がないか確認する

研究の段階に応じて保存先を選択する

- 研究中はデータの更新頻度も高く、複数のメンバーで作業することも想定されるため、バージョン管理やチーム内でのファイル共有機能等、必要な機能を備えたストレージを利用する
- 研究後は長期保存を目的とした、データを頻繁に更新しない長期的・安定的にデータを維持可能なストレージを利用する
- 公開が可能なデータであれば、データリポジトリ上での公開を検討する

→関連：「研究実施中　データの収集・分析・整理」-「データ分析環境の構築」　「研究後　研究データの公開」-「リポジトリへのデータ登録」

研究の段階に応じて保存先を選択する

- 研究の段階に応じ、データの利用方法は異なる。
  - 多量のデータを収集し、一時保管する場合
  - データを共同研究者間で共有する場合
  - データに様々な処理を施し、分析する場合
  - データを長期保存する場合　等

  それぞれのシーンに応じた、適切な情報サービスを提供、または取り扱い方法を指導する必要がある。

- 研究分野、研究者それぞれの研究の進め方にもよるが、
  - 最もニーズの高い個所
  - 費用対効果の高い個所

  を理解し、整備を進めることが重要。

研究活動に用いられるデバイスの情報セキュリティ対策

物理的な対策

- 研究活動に使用する部屋は施錠し、関係者以外の入室を制限すること
- 端末を操作可能な状態で放置しない、モバイルPCやUSBメモリ等を放置しない
- 外部記憶デバイス（外付けHDD・USBメモリ等）によるデータの持ち出しは極力避け、使用する場合はデバイスとファイルの両方にパスワードを設定する

システム的な対策

- OS・ウイルス対策ソフトの定期的なアップデートを心掛け、パスワードは容易に推測できない文字列を使用する
- ネットワークには必要なデバイスのみ登録し、定期的にアクセスログをチェックする

研究活動に用いられるデバイスの情報セキュリティ対策

- 機関で研究データ管理システムを提供している場合、そのセキュリティ対策が問われる。
  - セキュリティ要件をどのように満たしているかをSLA（Service Level Agreement）またはSLO(Service Level Object)や外部認証等を示しつつ説明できるようにしておくと良い。また、BCPに関連して、SLAやSLOにある目標復旧時間(RTO)、目標復旧レベル、及び目標復旧時点(RPO)について説明できるようにしておくと良い。

- ISMS適合性評価制度、及び、情報セキュリティ監査制度
  - ISMS適合性評価制度は「保証型監査」であり、情報セキュリティマネージメントが確立され適切に維持・管理できていることを審査登録機関の審査員が認証する制度。
  - 情報セキュリティ監査制度は「助言型監査」であり、情報セキュリティマネジメントにおける欠陥箇所を発見し、それを改善する活動を繰り返すことで理想的な情報セキュリティマネジメントを目指す制度。
  - 例えば、NII RCOS (GakuNin RDMやJairo Cloud を運用)はISMS認証取得済となっている。ただし、すぐにISMS認証を目指すのは困難な場合は、情報セキュリティ監査によって問題点を一つづつクリアしながらISMS認証を目指すという方法も現実的。

Ref:　上原孝之　情報処理安全確保支援士2021年版　翔泳社

データのバックアップ

- バックアップの重要性
  - 不慮のトラブルで研究データを失った際にも、トラブルの影響を最小限に抑えることができる
  - 不用意なファイルの削除や上書き、システム障害にも対応できる

- バックアップの設定
  - 既に利用しているストレージに付随するバックアップ機能がないか確認する
  - バックアップは自動か手動か、OSの標準機能か専用のソフトウェアが必要か等を決める。ストレージの容量が不足していれば新たに用意する
  - 指定した範囲・頻度・保存先でバックアップがとれるよう設定する
  - バックアップを手動で実行し、バックアップにかかる時間、容量、保存先などが想定どおりか確認し、作成したバックアップから復元できるかどうかも確認する
  - バックアップの対象となるデータの容量、バックアップにかかる時間等を考慮して頻度を調節する（毎日差分バックアップで毎月1回フルバックアップをする、など)

バックアップをする上で注意すること

- 重要なデータの場合には災害対策として、物理的に離れた複数の場所にバックアップを取っておく
- バックアップ先のストレージを定期的に確認し、障害が発生していないか、データ容量を超過していないか確認する
- バックアップの実行ログを取得できるようにし、定期的にバックアップが実行されていることを確認する

バックアップサービスに求められること

機関によるバックアップサービスは、研究者から見たサービスの位置づけに従い、適切に設計される必要がある。

- バックアップ可能な容量
- バックアップ保持期間
- バックアップデータを預かる方法
- バックアップデータを預かる単位(ex. 一つの巨大なzipファイル単位)
- 利用にかかる費用

- バックアップの消失に備えた対策の明示
- バックアップ内容の真正性保証
  - チェックサムの確認
  - タイムスタンプの維持

各種フォーマットの長所・短所等を理解し、研究のニーズに合ったフォーマットを決定する

- 自分の研究分野ではどのようなフォーマットが用いられているか、分野のデータリポジトリなどで調べて把握する

- 長期保存するデータのフォーマットには特定の商用ソフトウェアでしか読み込めない形式は使わず、仕様が国際標準となっているフォーマットや広く普及しているフォーマットを選択する
  - Recommended Formats Statement (The Library of Congress)
    https://www.loc.gov/preservation/resources/rfs/data.html

- 直接開いて参照されるデータには可読性の高いテキスト形式を使用し、解析・分析に使用されるデータには情報量の多いバイナリ形式を使用する

→関連：「研究実施中　データの収集・分析・整理」-「データリポジトリからデータを検索・発見する」

研究データアーカイブ

- 研究データが第三者から利用されるためには、アーカイブとして整理される必要がある

- アーカイブを構成するには、
  - 研究データにメタデータを付与し、整理、分類し、更に検索されやすくする
  - データの閲覧を容易にする

  等の整備が必要

- 広く普及しているデータ、メディア形式に対し、メタデータ付与、閲覧に対する機能拡大を検討する
  - 画像では、IIIFフォーマットの普及(メタデータの埋め込み、利用シーンの拡大)
  - 音声、動画のストリーミング再生(大容量データから特定個所の取り出し)

電子ラボノートの導入・活用を検討する

- 導入にあたって、機関内での利用見込みを予想して導入単位（個人単位・研究室単位等）を決定する
- 性能や使いやすさ、得意な分野、導入費用と使用料などを比較し、導入する製品を決定する
- オンプレミス型の場合はサーバの設置に関して機関内の各担当部署と調整する
- 複数名での利用が想定される場合、導入時に利用方法についてルールを定め、マニュアル等の存在を周知する

電子ラボノートの導入・活用を検討する

- 電子ラボノート(Electronic Laboratory Note, ELN)とは、研究データに加え、実験手順や議論のメモを電子的に記録するソフトウェアである
- 専用のアプリケーション、クラウドサービスではなく、Microsoft Onenote や Github のコミット記録で代用するという事例もある
- より統合的な研究環境、実験プロセス管理ツールとして、Laboratory Information Management System (LIMS) を導入している研究室もある
- 機関単位での提供について問われる場合がある
  - 機関で導入を検討、または、予定している場合は、スケジュールを確認し、今後の計画を答えられるようにしておく

この文書について

「情報基盤スタッフ向けの研究データマネジメント －研究実施中:研究データの保存－」

教材作成TFメンバー：元木正和、古川雅子、青木学聡(2022)
JPCOAR-研究データ作業部会、AXIES-RDM部会の共同制作

"RDM Tutorials for ICT Staff - Data Preservation on Execution Phase"

Member of Teaching Material Making TF: MOTOKI Masakazu, FURUKAWA Masako, AOKI Takaaki (2022)
Co-Produced by JPCOAR-Research Data Working Group, AXIES-SIG-RDM

文書の一部は「研究者のための研究データマネジメント」の内容を編集、再構成しています。
(http://id.nii.ac.jp/1458/00000247/)