研究室におけるデータ管理

―あなたと共同研究者を守るために―

東京大学 定量生命科学研究所
須谷尚史

自己紹介

生命科学分野の研究者（専門：染色体高次構造）

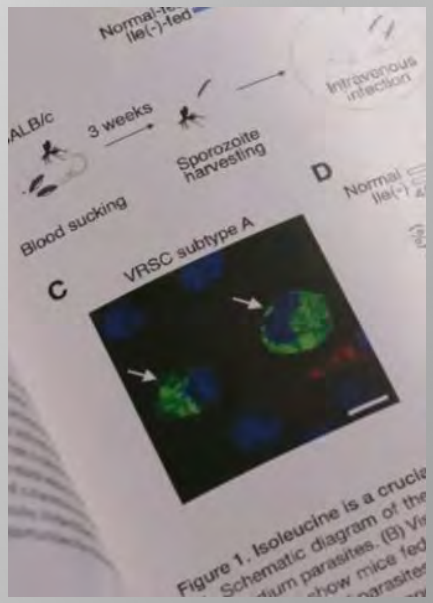
部局において研究倫理に関する業務を 10年近く担当

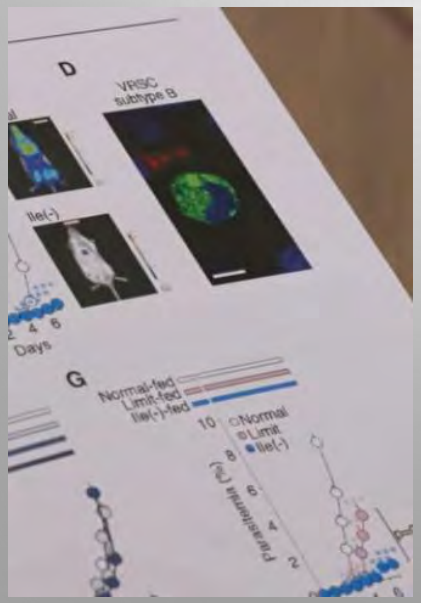
- ✓研究倫理推進室・倫理部門長
- ✓倫理セミナーの担当
- ✓画像処理, データ解析に関する啓蒙•相談
- ✓論文化された研究データの収集•保管

本講演の内容

- ✓ 研究室における研究データ管理の実態
- ✓ 失敗しない研究データ管理のための方策
- ✓ 東大•定量研における研究データ収集•保管の取り組みについて

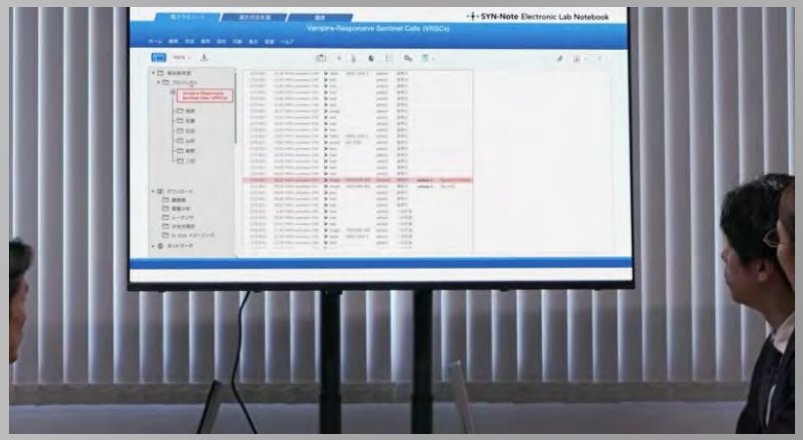
「倫理の空白V」責任ある研究活動
データ管理編





- 論文内に同じ画像があると指摘
- 研究データを使い回している、すなわち実験をでっち上げた(ねつ造) 可能性が疑われる

「倫理の空白V」責任ある研究活動
データ管理編



- 研究室主宰者は調査を決意
- 複数の人の「うっかり」が重なって起きた誤りとわかる
  →雑誌に修正依頼
- 原因究明にはデータ履歴を記録するシステムが奏功

研究データの実際

- 画像, 数値, 各種測定装置の出力ファイルなど多種類のデータを組み合わせて論文が作られる。
- データ解析, 加工の段階でさらに様々な種類のデータが生成される。

> 例：1 つの論文に 16 種類のデータ ~350点
> (データサイズ計 300 Gb)

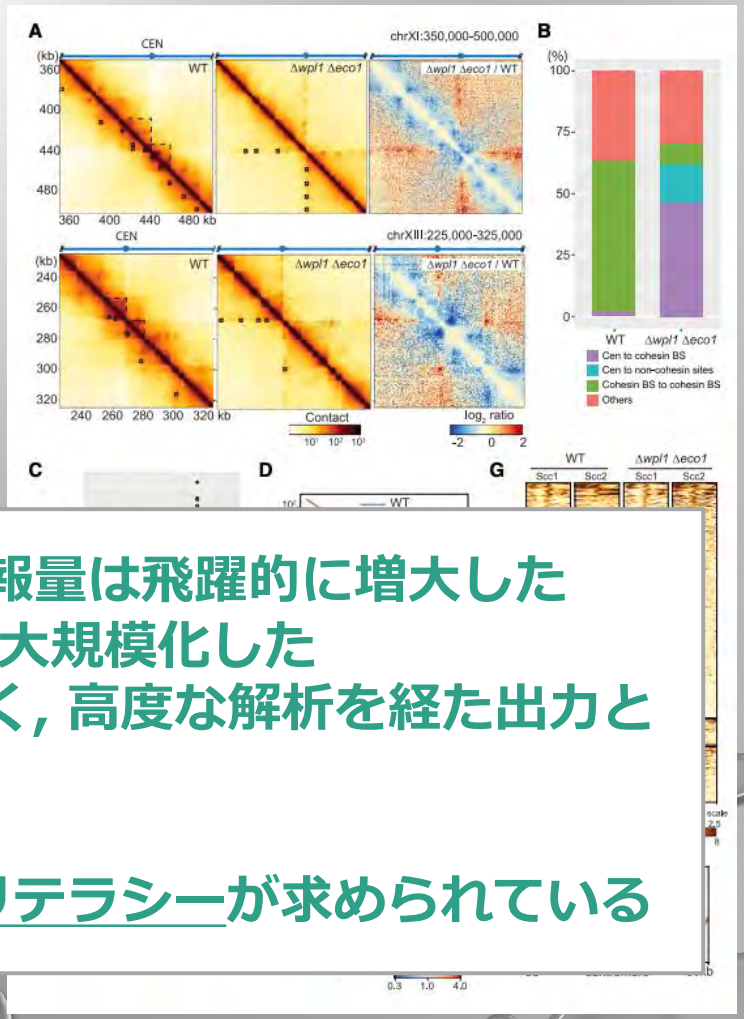
- 一つの論文ができるまでに数年かかる。
- データ取得, 解析に複数の人が関わるのが普通。途中で関係者の異動が起きることも。
- データの質, 量とも21世紀に入って大きく変化。特に年配の教員はその変化に十分対応できていない可能性

研究データはこの30年でどう変わったか

1994年

Y Saka, T Sutani,et al.(1994)
Fission yeast cut3 and cut14,members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis
〈Fig.5〉
https://doi.org/10.1002/j.1460-2075.1994.tb06821.x 参照

2025年



- ✓ この30年で，論文図に含まれる情報量は飛躍的に増大した
- ✓ 元となる研究データはデジタル化•大規模化した
- ✓ 論文図は実験結果そのものではなく，高度な解析を経た出力となった

研究者には新しい時代のデータ管理リテラシーが求められている

iScience (2025)
https://doi.org/10.1016/j.isci.2025.112322

「倫理の空白V」データ管理編

「ルールをわかっている
つもりでも ミスをして
しまうことはあります」

不正疑いの指摘を受けた研究者は、”ねつ造” ”改ざん”
ではないことを実証しなくてはいけない

→ **正しく実験が行われたことを示す証拠の提示**
(実験ノート, 実験データ, 研究材料)

研究データの適切な管理は「研究者を守る仕組み」でもある

- ✓ 研究データ (＋実験ノート、研究材料) を適切に管理•保存することは、研究者が自身の研究の公正性を証明するために必要不可欠

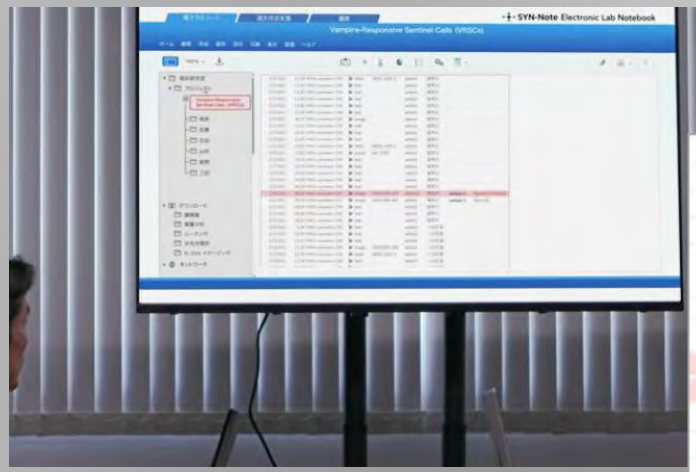
不正が疑われた際に適切な証拠の提出ができないと‥

→ **研究不正と認定されうる**

→ **失職や学位剥奪の可能性**

研究データ管理は難しい

- ✓ データ管理のための万能システムは存在しない
- ✓ ドラマ中のデータ管理システムも架空のお話



15:43 VRSCs activation 1218	text		added	紺野天		
16:02 VRSCs activation 1218	table	VRSC el34-3	added	紺野天		
17:03 VRSCs activation 1218	graph	act-2345	added	紺野天		
18:10 VRSCs activation 1218	text		added	紺野天		
19:20 VRSCs activation 1218	text		added	紺野天		
19:29 VRSCs activation 1218	text		added	紺野天		
20:20 VRSCs activation 1218	text		added	紺野天		
20:30 VRSCs activation 1218	text		added	紺野天		
20:32 VRSCs activation 1217	image	VRSC044-011	deleted	柳航介	subtype A	figure1C Deleted
20:33 VRSCs activation 1217	image	VRSC046-003	added	柳航介	subtype A	figure1C
20:34 VRSCs activation 1218	text		added	紺野天		
20:40 VRSCs activation 1218	text		added	紺野天		
20:57 VRSCs activation 1218	text		added	紺野天		
9:20 VRSCs activation 1220	text		added	三田若葉		
10:20 VRSCs activation 1220	text		added	三田若葉		
11:12 VRSCs activation 1220	text		added	三田若葉		
11:30 VRSCs activation 1220	image	VRSC045-005	added	三田若葉		
11:33 VRSCs activation 1220	table	VRSC el34-3	added	三田若葉		

「倫理の空白V」データ管理編

システムに頼った研究データ管理は難しい

- データ形式が多様
  顕微鏡画像, ゲル画像, シークエンサーデータなど
- 実験•解析手順も多様
  研究ごとにワークフローが異なる
- 研究は本質的に非定型業務
  会計システムのような標準的ワークフローが存在しない
- 保存すべきものはデータだけではない
  実験条件, 解析過程, 考察などの記録も重要
- 論文図と元データの対応付けは自動化が難しい
  人による継続的な記録•更新が必要

ルールに基づいたデータ管理 (1)

- **生データ (一次データ) は必ずオリジナルを保存**
  - オリジナルは変更しない
  - 加工は複製を作ってから行う。
- **生データは研究室の共有領域に保管する**
  - 定期的にバックアップを取る
- **ファイル名の命名規則を設ける**
  - 日付、作成者、実験ID などの情報を含める
  - フォルダ名でなくファイル名にも情報を含める
- **実験ノートから対応するデータを辿れるようにする**
  - データファイル名や保存場所を記録
- **ラボ単位で最低限のデータ管理ルールを決め、共有する**

ルールに基づいたデータ管理 (2)

(東大定量研で行っている取り組みです)

- **論文受理直後に, 論文作成に用いたデーター式を取りまとめ, 一括して保管する**
  - ✓ 生データ, 加工•解析を施した中間ファイルと最終の論文図を収集する
  - ✓ 論文図とデータの対応付けを記録する
  - ✓ 学生・ポスドクが在籍しているうちに実施する

論文受理直後は、データ整理の最後の好機である

定量研における研究データ収集•管理の取組み

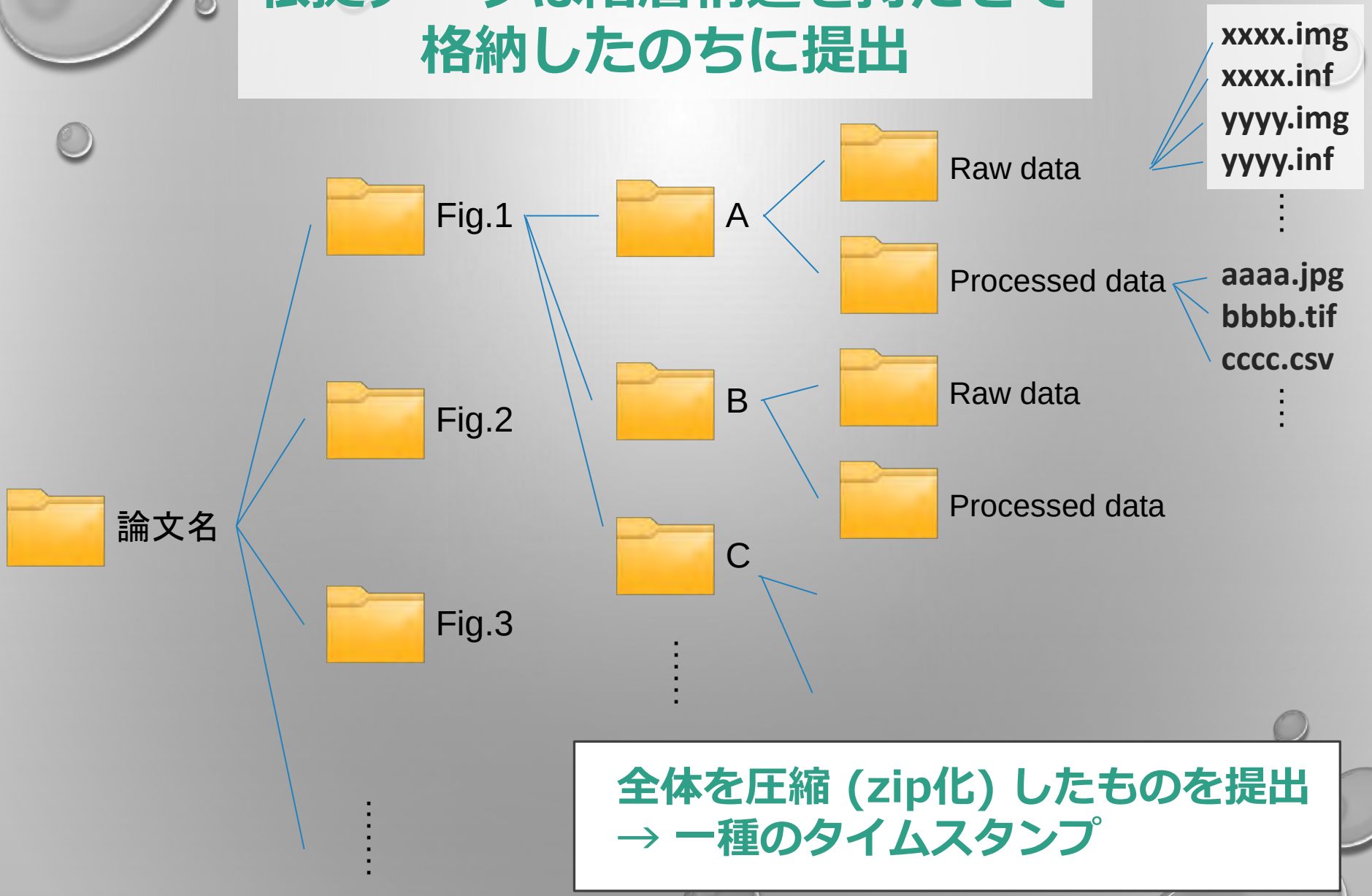
定量研メンバーが著者に入った論文が受理されたら、
１ヶ月以内に

- **論文の最終原稿と図**
- **論文に使った全ての根拠データ**
- **公正な研究論文であることのチェックシート**

を取りまとめて提出する。

■ 根拠データ
- **生データ**：実験で取得されたオリジナルのデータ。無加工のもの。デフォルトのファイル形式。
- **中間加工ファイル**：論文図を作成する過程で、生データを加工して得られた全ての途中ファイル。

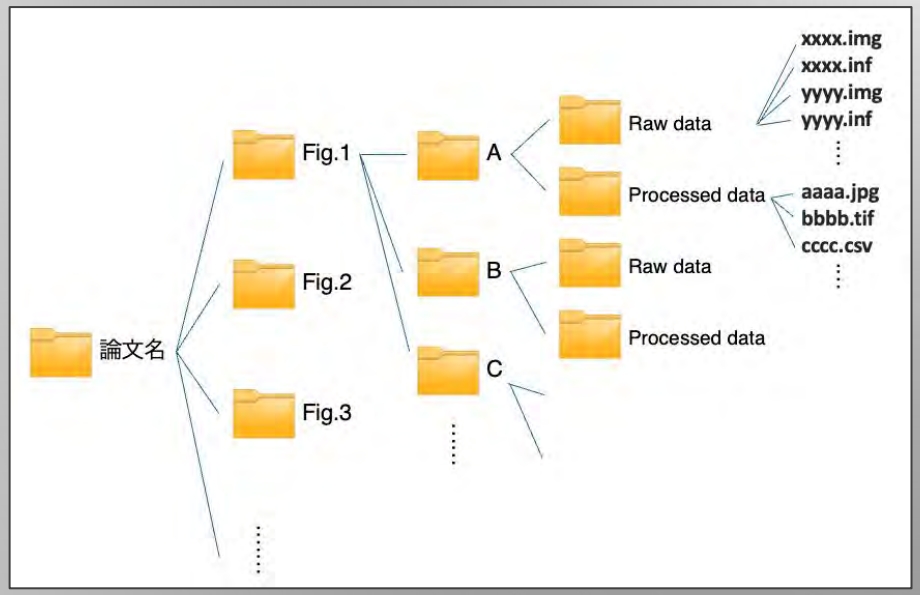
根拠データは階層構造を持たせて格納したのちに提出



全体を圧縮（zip化）したものを提出
→ 一種のタイムスタンプ

各データの作成者や作成に用いた機器•アプリの情報も収集

ファイル名と階層構造を専用マクロで抽出



各データの作成者, 作成に用いた機器•アプリを記入し提出

定量研における研究データ収集•管理の取組み

- 論文受理直後のまだ記憶の新しいうちに行う
- ある程度の労力は必要。一度経験すると、データ作成時/論文作成時から「データ管理」を意識した行動が取れるようになる
- データの保存義務は各研究者にある。データの複製を部局に提出する (バックアップとして機能)
- 収集プロセスを管理するのは「研究倫理推進室」。省力化のため、 GakuNin RDM上で動くシステムを構築•運用している

IQB-RIMS (定量研 研究公正管理システム)

- 国立情報研が提供するデータ管理基盤 GakuNin RDM に専用アドオンを導入することで実装
- 進捗管理や画像スキャンの機能を内包
- メッセージ機能
- 履歴管理
- ストレージとして外部サービス (Google ドライブ等) を選択可能
- 将来的には電子実験ノートとのシームレスな運用も視野に
(論文化の前からデータをクラウド上で管理)
- 他機関での利用を想定した「運用コストが小さく汎用性の高いバージョン」を今後開発予定とのこと

本講演のまとめ

- ✓ 「責任ある研究活動」を実現するためには正確で誠実なデータ管理が必要
- ✓ データ管理の完璧なシステムは存在しない
- ✓ 研究現場に合った、無理なく続けられるデータ管理ルールを定め、それを組織として運用することが重要である

研究データの適切な管理は
研究を推進するためにも重要である

研究データの適切な管理•保管は：

- ✓ **研究の公正性を保証するために必要**
- ✓ **論文を書くために必要**
- ✓ **将来の研究のための基盤として必要**
  様々な試行錯誤の知見が、後の研究の進展に必要となることがある
- ✓ **データそのものの再利用や活用 (AI 解析など) に必要**

「世界中に人々と共有することで私という人間が去った後も研究はつながっていく」

「データの保存は研究者としての責任だけじゃなく希望でもあるんですね」

「倫理の空白V」データ管理編