

産学連携・研究者支援の現場から見た 研究データ管理の新たな課題： ミクロからマクロへの課題の逆展開と 望ましいデータ管理基盤のあり方の模索



オンライン名刺(Eight)

大阪大学産業科学研究所
新産業創成研究部門 知的財産研究分野／戦略室
特任准教授(常勤)／リサーチ・マネージャー
加藤 久明

「ミクロからマクロへの課題の逆展開」

ボトムアップのデータマネジメント環境の変化

- 1990年代～2016年頃からのAI・ビッグデータブーム前まで、**SINETに繋がる各研究機関が個々に準備をした学術情報基盤のメニューを研究者が活用する「マクロからミクロへ」というスキームが主体だった**



- しかし、**2018年頃から爆発的なデータ増加の起点が「個々の研究者(研究室)」レベルで起きる傾向が増加し続けていることを数多の案件を処理しながら現場の肌感覚で感じている。そして、「ミクロからマクロへの課題の逆展開」が起きていると考えている**
(個々の研究者がAI・ビッグデータを続べながら、大規模なイメージングデータ、PHR/PLR等を生成し、そこに生成AIの活用による桁違いのデータ追加ができる時代だからこそ発生した課題の逆展開)

自己紹介(加藤久明)

大阪大学 産業科学研究所
新産業創成研究部門 知的財産研究分野 / 戦略室
特任准教授(常勤) / リサーチ・マネージャー
安全保障輸出管理アドバイザー

→「田舎の駅の駅長さん」のような存在です(笑)

〔知の放浪歴〕

- 2004.3 駿河台大学大学院修了 修士(文化情報学) ※学士も文化情報学(レコード・アーカイブズコース卒)
- 2008.3 千葉商科大学大学院修了 博士(政策研究)
- 2007.11~2009.5 立命館サステナビリティ学研究センター 客員研究員
- 2009.6~2011.7 立命館グローバル・イノベーション研究機構ポスト・ドクトラル・フェロー
- 2011.8~2016.3 総合地球環境学研究所(地球研) 研究部 プロジェクト研究推進支援員
- 2016.4~2020.3 大阪大学産業科学研究所 特任助教
- 2018.4~現在 大阪大学D3センター(旧 データビリティフロンティア機構) 企画室(兼任)
- 2020.4~現在 大阪大学産業科学研究所 特任准教授(常勤) / リサーチ・マネージャー

→図書館情報学(+ 計算機科学)・文書館学・社会情報学、非営利組織論、環境政策(エネルギー政策)、水文・水質・水理学、同位体を活用した環境分析(主として酸素水素安定同位体比など)、農地観測、シミュレーション(+ スパコン運用管理)、産学連携マネジメント(国内外)、競争的外部資金獲得支援、研究支援におけるデータマネジメントプラン作成支援、研究所評価情報の管理(各種IR)、ネットワークやDB構築支援等、多様な領域とご縁をいただいで現在に至っています



オンライン名刺(Eight)

研究者としての詳細は、
<https://researchmap.jp/hisaaki.kato>
をご覧ください

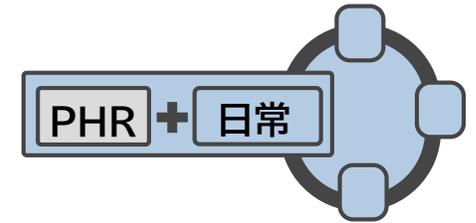


過去の仕事例(1): データプラットフォーム

■ PLRプラットフォーム



文部科学省 Society 5.0 実現化研究拠点支援事業
大阪大学 ライフデザイン・イノベーション研究拠点
Initiative for Life Design Innovation (iLDi)



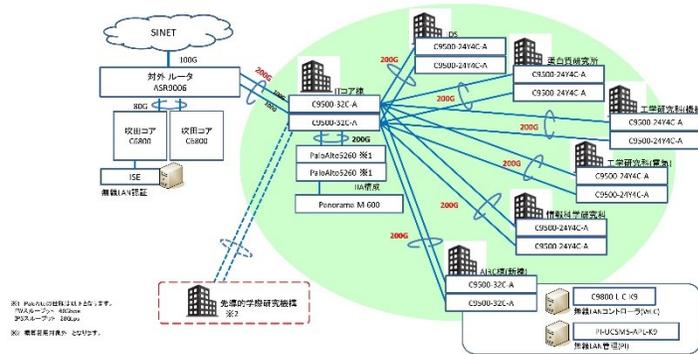
個人 社会
病院ネットワークとセンサネット
ワークによるPLRデータベース

<https://www.ildi.ids.osaka-u.ac.jp/>

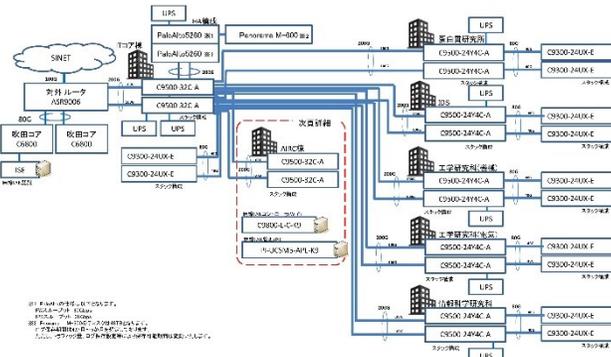
- 2018年度 文部科学省の「Society 5.0実現化研究拠点支援事業」の申請支援
→採択後は「企画支援室」+ 外郭団体の事業化(法人設立+スタートアップ事業
立ち上げ)を担当(FY2019まで)
- 特に、外郭団体として、①AI人材育成、②Society5.0事業から産み出される
データ・ハンドリング事業を担う(一社)データビリティコンソーシアム
(<https://cde.or.jp/>)の設立・登記を担当
- 日常的な生活データであるパーソナル・ライフ・レコード (Personal Life
Records: PLR) と医療系のPHR(Personal Health Records)のハンドリ
ングに関する複雑な問題に振り回されながら事業化を担当

過去の仕事例(2): 高速ネットワーク整備計画

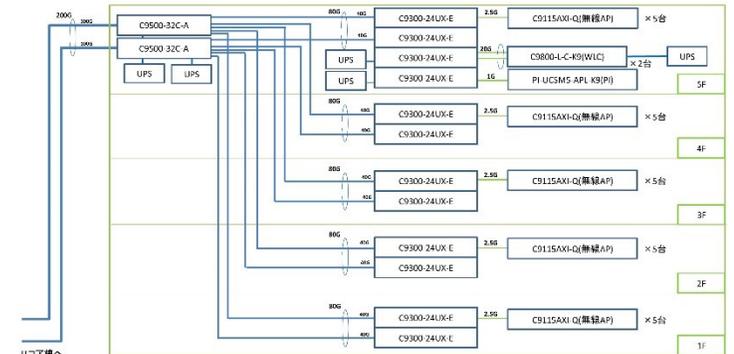
- AIデータ解析用高速ネットワーク設備 [R3年度 共通政策課題(基盤的設備等整備分)概算要求]



100Gbps部局間接続

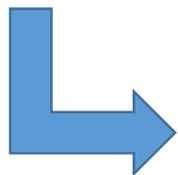


各部局との接続構造

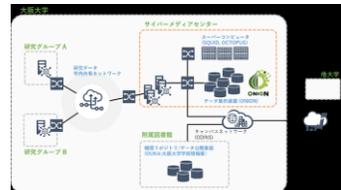


産業科学AIセンター棟ネットワーク構造

- ①サイバーメディアセンターを中心に産業科学研究所「産業科学AIセンター棟(同時に増改築予算申請)」、蛋白質研究所、工学研究科、データビリティフロンティア機構、情報科学研究科との間を100Gbpsの光ファイバーで繋ぐ部局間接続設備を設計・提案(NTT西日本・伊達先生@D3センター・相原様@産研技術室との共同作業)
- ②全学ネットワークODIN第9期に求められる環境の先行提案として、膨大な全学AIデータ処理の拠点となる産研「産業科学AIセンター棟」(R2概算要求)内部のデータトラフィックを支えるための棟内高速ネットワーク設備



要求は不採択だったが、伊達先生@D3センターが構想を引き継いでくださり、
100Gbps接続の大阪大学データ基盤「RED-ONION」へ
https://jpn.nec.com/press/202311/20231106_01.html



産学連携版「田舎の駅の駅長さん」の 主な仕事から見るデータマネジメントの課題

1. 年間100件以上の様々な問い合わせ対応(電話・メール等色々…)
2. 年間約70件の競争的外部資金申請書作成支援
3. 年間数十件の産学連携契約締結(協議・法務対応・契約書起案含む)
→外部資金は研究所として年間30億円以上稼ぐ必要があり、件数も増加傾向にある
4. 年間二桁件の安全保障輸出管理書類のチェック(技術の輸出としてのデータ輸出なども)
5. 各種IR(Institutional Research):研究所内の様々な評価データの収集・分析・管理
6. 所内インキュベーション施設の管理、入居者からの産学連携やデータ管理等に関する相談対応
7. 国プロや産学連携プロジェクトにおけるデータ管理基盤や高速ネットワークの検討(時に設計補助もします)

→ 個別の産学連携案件では、個々の研究室が機微なデータを保有しているが、それらのデータを大学全体の情報基盤に預けてマネジメントするような体制ができていない状況が一定以上の割合で存在する(DMP作成の負担も大きい)

大学全体の情報基盤とデータの相性

1. 大学全体に跨る情報基盤は、全学に跨る解析プラットフォームなどとは相性が良い(というか全学で統合したほうが解析機器のネットワーク化が進む今日では効率的)
(例:大阪大学コアファシリティ機構、マテリアル先端リサーチインフラ設備供用拠点(阪大ARIM)、産業科学研究所総合解析センター、超高压電子顕微鏡センター等…)
2. 反面、個別の研究室単位で動く案件、例えば大型の国プロ(NEDO・JST・JST K Program・文科省系国プロ等)、大型産学連携案件(様々なデータ取得を伴う共同研究案件)、規模を問わない医工連携案件やバイオ系案件等では機微なデータが個々の(研究室サーバー等の)プラットフォームに止まることが多く、全学の情報基盤に預けることができない(or できていない)
→オンプレ、AWS、その他置き場所はバラバラ…(正直、「大丈夫かな？」と思う案件もあります) ※事務部は事務ローカルも加わるのでさらに複雑…

課題は山積する、人手も資源も足りない…

- 全学と繋がる共用施設等の解析プラットフォームは、全学の情報基盤との接続が進展している(ネットワーク化だけでなく、スタンドアロン機器のペーパーレス化や多様なメディアへの書き出しの手間の削減等のOutcomeも多い)
- だが、個々の研究者が持つ機微なデータをどこに預けるのか?というミクロから出てきた課題に全学レベルの基盤は上手く繋がっているとは言えない(上手に繋がれば個別の情報基盤の維持負担を軽減できるはずなのだが…)
- 部局レベルで見るとGakuNin RDMの活用だけでなく、そもそも研究者レベルでの認知も停滞している(部局レベルの議論も停滞気味というか、メリットを研究者が理解しきれていない状況も見受けられる…)
- そもそも、研究者にデータガバナンスと所有者としての役割を全て負わせてしまう大学システムの実態(制度としてそれを意図していなくてもそうなっていることが多い)と定年退職/異動によるデータの管理すら追いつかない現状…



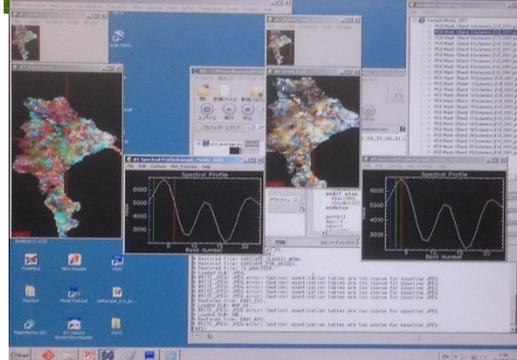
- さらに、個々の研究者に外部競争資金などで求められる「データマネジメントプラン」の作成や検討が追い打ちをかける状況…

「データマネジメントプラン」をめぐる教員との伴走

- 地味な仕事なのだが、問い合わせが多い仕事のひとつ
- 「正直、どう書いたらよいか分からないのですが、何か近い分野の他の先生方のお手本のようなものはありますか？」という問い合わせをサポートすることがよくある(年間二桁件)
- 申請前も採択後も急ぎの案件が多く、検討にあまり時間がかけられない
- そのため、その教員の研究内容や方法論を理解し、申請書の内容等から生成されるであろうデータ群を推定・定義し、「データマネジメントプラン」として具体的にドラフトを起案し、提案を行うのが現状
- これらの作業を全て研究者が考えること自体、研究時間が足りない現状では問題となるため、支援人材が伴走するしかない(正直、手探り状態…)
- データ内容は分野によってバラバラ(例: 一細胞単離の領域の場合、従来の細胞関係のデータだけでなく、有機化学系の合成データ等も入ってくることもあり、時に医療系の画像データが混ざり合う)
- いずれ、フィールドとデータ処理基盤を往復してきたような支援人材がその経験を上手に活用して支援が増えることが望ましい(と考えています)

フィールドワークとスパコン管理に明け暮れた ポストドク時代(FY2011-2016)

- フィールドで観測データや試料を得て、ラボにおいて試料やデータ解析を行い、ラボを支えるスパコン資源やシミュレーション環境に奔走したポストドク時代
→この時の経験が阪大産研におけるデータマネジメントを絡めた研究者支援に繋がっている



農地や湖沼における観測データの取得・分析

スパコン管理とPJにおけるシミュレーション環境構築

まずは、研究者のデータマネジメントプラン作成の負担だけでも軽減する仕組みが欲しいです…

- RDMを促進する人材は足りない。特に「プランの意思決定」を支援する人材が足りないのだが、すぐに増員をすることは難しい

- ① 現場では、実際にRDMプランを組んだ研究者から「これで問題無いですか？」という問いに回答をする責任を支援人材は負う→誰かがOKを出さないと進まないし、各種の競争的外部資金事業では採択後の各種書類の提出期限が常に迫っている(申請前にもRDMを突貫工事で作成し、実際の契約段階でも走りながら考えている状況)
- ② 研究シーズを知り、研究現場と発生するであろうデータを想像できないといけないのだが、そのような人材育成がどこまでできるのか…

- GakuNin RDMの機能として「データマネジメントプランの自動生成機能」があると嬉しいです(あくまでも個人的な願望です)

- ① NEDOやJSTのデータプランフォーマットをベースとした、RDMプランを研究者と共に整理してくれる生成AIを活用した支援ツールがGakuNin RDMで実現すると、外部資金を獲得する研究者が救われます
- ② RDMプラン作成機能を通じて、GakuNin RDMと個々の機微なデータを持つ研究者たちが新たな関係づくりができるのではないかと

大阪大学産業科学研究所
新産業創成研究部門 知的財産研究分野／戦略室
〒567-0047 大阪府茨木市美穂ヶ丘8-1
E-mail: hisaaki@sanken.osaka-u.ac.jp



オンライン名刺(Eight)

特任准教授(常勤)
リサーチ・マネージャー
加藤 久明

ご清聴ありがとうございました

謝辞: 脇 様@NetApp、伊達先生@阪大D3センター、iLDiの皆様、
(一社)データビリティコンソーシアムの皆様、
その他産学連携案件でお世話になっている先生方と企業の皆様