

2023 年度 理研ハッカソン 公開シンポジウム

開催日時： 2023 年 10 月 24 日（火） 13:00~17:20 (12:30 開場)

場 所： 神戸キャンパス(融合連携イノベーション推進棟 (IIB) 8 階講堂)・オンラインのハイブリッド開催

主 催： 国立研究開発法人 理化学研究所 情報統合本部 基盤研究開発部門

参加費： 無料

参加登録： 下記サイトにて、事前参加申し込みをお願いします。

https://krs2.riken.jp/m/hackathon_r5

理化学研究所では、研究分野を超えた効果的な研究加速や研究成果の社会還元等を図るため、研究のデジタル変革（研究 DX）やオープンサイエンスを推進している。これらの活動の中で、信頼性の高い良質なデータの利活用が重要な課題の一つとなっている。

本シンポジウムでは、研究分野を問わず、研究データの収集、解析、利活用それぞれにおいて世界の最前線で活躍する研究者にご講演頂き、新たな知の獲得に向けた研究データ利活用の現状と将来について議論を行う予定である。

プログラム 令和 5 年 10 月 24 日（火曜日） 12:30 開場

13:00	開会挨拶 美濃 導彦（理化学研究所 情報統合本部）
セッション 1 基調講演 1	
13:10	演題: Knowledge Representation: the role of semantics in scientific research 講師: Ana-Claudia Sima（SIB Swiss Institute of Bioinformatics） 要旨: The Knowledge Representation Unit at the SIB Swiss Institute of Bioinformatics provides expertise and tools that improve scientific data retrieval, exploration and reuse by people and computer systems. Through semantic enrichment, ontology engineering and expressive search interfaces, we strive to bring the most added value to scientific datasets in Switzerland and worldwide. In this talk, we will present ongoing and future activities of our unit, from creating domain-specific, interoperable knowledge graphs, such as digital libraries of chemical compounds based on botanical garden collections, to creating rich metadata and tools that enable asking questions over data, as part of larger international collaborations. In particular, we will focus on synergies with the RIKEN Data Knowledge Organization Unit.
セッション 2 データコンテンツセッション	
13:40	演題: 生命科学データの統合作業を通して見えてくるもの What we have through the process of integrating RDF datasets 講師: 山本 泰智（ライフサイエンス統合データベースセンター） 要旨: 生命科学分野のデータが RDF として公開される例は多いが、概念を示す識別子である URI について、同一概念でもデータセット間で異なる場合が散見される。このため、RDF 化することで得られる大きな利点である、複数データの結合による統合が十分に行われないという課

	<p>題が生じる。そこで、ライフサイエンス統合データベースセンター (DBCLS) では、さまざまな識別子の間の関係を、公開されている RDF データセットから抽出し、同義であるなどの意味づけをした上で、誰でも検索可能な TogoID というサービスを立ち上げた。</p> <p>さらに、同一概念であるにもかかわらず、URI としては全く異なるものとして認識されてしまう、例えば、http と https の違いなど、プレフィックス間同義関係を調査し、それらの相違を吸収する作業を進めている。</p> <p>この結果として、生命科学分野において、これまで見落とされてきた可能性のある概念間の関係を顕在化できると期待している。</p>
14:00	<p>演題: 転写制御理解のための統合的データ基盤の開発 Development of an integrated data platform for understanding transcriptional regulation.</p> <p>講師: 粕川 雄也 (理化学研究所 生命医科学研究センター)</p> <p>要旨: 転写制御メカニズムの理解を目指し、現在、我々のグループでは統合的なデータ基盤 INTRARED (https://www.intrared.org/) を構築している。INTRARED はゲノム中の転写制御に関わる領域であるシスエレメント、転写制御等のゲノムに結合して転写制御を行うトランス因子、ゲノム DNA のメチル化やヌクレオソーム構造などのエピゲノム状態に関する様々なデータをカバーすることを目的とし、公共の ChIP-Seq データやエピゲノム解析データをまとめた既存のデータベースである ChIP-Atlas (https://chip-atlas.org/) と、新規作成するシスエレメントについてのデータベース fanta.bio (https://fanta.bio/) で構成される。本発表では INTRARED について、プロジェクトの概要や現在の状況について説明する。</p>
<p>セッション3 基調講演2</p>	
14:20	<p>演題: 我が国がもつべきデータ管理戦略について - 諸外国との連携を含めて - Academic Data Management Strategies required in Japan - From domestic perspective to international interoperability -</p> <p>講師: 山地 一禎 (国立情報学研究所 オープンサイエンス基盤研究センター)</p> <p>要旨: 欧米では、データ主権という言葉が、産業界だけではなく学术界にも浸透しつつある。組織や研究者は、データを適切に管理し、戦略的に共有・公開していくことに、これまで以上に意識を高めなければならない。本講演では、海外の状況を踏まえながら、我が国がもつべきデータ管理戦略について議論する。</p> <p>Data sovereignty is becoming a more popular term used in industry and then installed in academia. Organizations and researchers must be more aware than before of managing research data properly and strategically share and open it in public. This presentation will focus on research data strategies we must have in Japan while considering the international perspective.</p>
<p>14:50 休憩 (10分)</p>	
<p>セッション4 データ管理セッション</p>	
15:00	<p>演題: 研究者に役立つ研究データ流通・管理システムって何だろう? - 全学機器共用の立場から What is a useful research data distribution and management system for researchers? A standpoint of university-wide equipment sharing</p> <p>講師: 古谷 浩志 (大阪大学 コアファシリティ機構)</p> <p>要旨: 研究にとって「研究データが重要」であることは言うまでもありません。特に実験系の研究者にとって研究データは研究の命です。データなしには論文は書けません。今年の5月に仙台で開催された G7 仙台科学技術大臣会合においても「オープンサイエンスの推進」が宣言されました。「大切! 研究データ」ですが、忘れてならないことは、「研究データは、使って・使われてなんぼのもの」という点です。</p> <p>さて、皆が共に使う研究機器・設備である「共用機器」ですが、実はその分、「皆さんの研究データを生み出す場」でもあります。私達コアファシリティは、全学の共用機器の利用促進や拡大、それによる研究力の向上を目指していますが、研究データを生み出す研究現場に非常に近</p>

いこともあり、研究データに関与し始めました。

共用機器で得られた測定（研究）データは、更に分析・解析・統合を行うため、研究者の PC/Mac に集められます。しかし、その手間が非常に大きいことが最初の問題でした（共用機器の多くには、レガシー OS が使われており、装置のパフォーマンス向上やセキュリティ確保のためネットワークに接続されていない場合が殆どです。そのため、USB メモリや DVD 焼きなどが測定データの移動・集積に多用されていました）。

この問題を解決するために、小さくシンプルなローカルシステムを作り、ネットワーク経由で測定データが簡単に取得できるようにしました。ネットワークに接続している—ということ、そのローカルシステムを大阪大学の全学データ集約基盤 ONION（サイバーメディアセンターが構築・運営）に接続させ、更には GakuNin-RDM との接続（附属図書館が担当）にさせるまでになりました。これにより、研究の現場から、測定（研究）データが“ネットワーク経由で一気通貫”に部局内—大学内—学外の連携が実現されました。

一気通貫のネットワークでのデータ流通は、研究の効率化という意味では非常に大きな“役立ち”でした。しかし、ある意味、大学に特化した Cloud ストレージができただけです。本丸は、これらの測定（研究）データを“使って・使われて”、「新しい価値」を生み出すオープンサイエンスを進めることです。

「他人の研究データを使う」や「自分の研究データを他人に使ってもらう」がオープンサイエンスの肝ですが、ピアレビューの研究の世界において、研究者は「信頼できるデータしか使わない」ことは言うまでもありません。すなわち、クオリティーや妥当性を判断できる研究データしか使われませんし、それを判断するためには、充実したメタデータが必要となります。

さて、充実したメタデータは誰が作成するのでしょうか？ 研究者が新たな大きな手間を負う必要があるのでしょうか？ “研究者に役立つ”研究データマネジメントの枠組みが必要であることを痛感しています。それについて（研究現場に近い）全学機器共用の観点から、最近の取組等について紹介していきます。

15:20

演題: **R2DMS の機能および発展的利用法**

Features of R2DMS and its usage including new features

講師: 實本 英之（理化学研究所 情報統本部）

要旨: R2DMS を利用するための体制の構築および機能について紹介する。また R2DMS を用いたデータ集積の用例や、現在テスト中のメタデータに関する機能について合わせて紹介する。

セッション 5 データ利活用・応用事例紹介

15:40

演題: **マルチオミクス解析から農業デジタルツイン開発へ**

Multi-omics study toward development of digital twins in agriculture

講師: 市橋 泰範（理化学研究所 バイオリソース研究センター）

要旨: 「食料生産と地球環境保全の両立による持続可能な農業の実現」は人類が早急に達成すべき課題の一つであり、その解決策として多様な農地環境への最適化を目指した栽培のオーダーメイド化が考えられます。そこで私たちは、農業現場でのマルチオミクス解析により農業生態系のデジタル化を進めており、これらのデジタルデータを使って、農業生態系をサイバー空間でシミュレーションできる農業デジタルツインを開発するプロジェクトを開始しています。本システムでは、収穫時期までの気象予測とその土地の土壌データに基づき、作物の収量や品質さらに環境負荷を予測して、その土地に合った最適な栽培管理を導き出せるように設計しております。今回はこれまでの大規模圃場試験等の研究進捗を共有します。

Technological innovation has been demanded to achieve a sustainable crop production fed the growing human population without generating environmental burdens. We aim to develop digital twins that simulate crop production in cyberspace utilizing integrated reduction models with large-scale digital data by field multiomics analysis. The system integrates weather prediction and soil data to provide the best solution for the farming management practice. Today I will share our progress in large scale field trials dissected by multiomics and development of

	integrated model framework.
16:00	<p>演題: 環境・材料データサイエンスと予測科学</p> <p>Prediction science for environmental and material data science</p> <p>講師: 菊地 淳 (理化学研究所 環境資源科学研究センター)</p> <p>要旨: 演者らは環境複雑系に対峙する1手法として、「恒常性評価」を提言し続けている。環境複雑系に摂動を与える因子群は相互作用しており、多因子を同時に評価するデータサイエンス手法が有効と思われる。例えば昨今、世間を騒がせている炭素循環で言えば、地球上の約半分の炭素は海中に吸収・固定される。特に沿岸に藻類を繁茂させる「ブルーカーボン」の投資熱が盛んであるが、水温・酸素濃度・栄養塩・動物食害等々、沿岸生態系の将来予測に関わる環境データサイエンス技術が価値を生じることになる。他方で汎用プラスチックの多くはサーマルリサイクルされてきたが、焼却時の炭素源放出の懸念もあり、サーキュラーエコノミー化へのリサイクル戦略も議論されている。講演では石油依存を脱却すべくバイオエコノミー化を意識した、材料データサイエンスの試みも紹介したい。</p>
16:20	<p>演題: Designing a Semantic-Enhanced Hybrid Search Engine for Biomedical Events</p> <p>講師: Rangel Reyes Julio Cesar (理化学研究所 情報統合本部)</p> <p>要旨: In the realm of biomedical research, there's a pressing need for efficient search systems. Traditional methods often oscillate between exact match searches, which can be too narrow, and semantic similarity searches, which can be too broad. This gap can hinder the extraction of interconnected knowledge from the vast expanse of scientific literature.</p> <p>Driven by the desire to enhance the search experience for researchers and healthcare professionals, our team sought to develop a system that combines the best of both search methodologies. The goal was to provide relevant search outcomes for a wide array of queries while ensuring precision and comprehensiveness.</p> <p>Enter BEHF. This system integrates the precision of exact match search with the broader scope of semantic similarity search. By leveraging network graphs and a specialized dataset, BEHF efficiently taps into the interconnected knowledge of scientific literature. Its user-centric design, complete with an adaptable scoring system, positions BEHF as a versatile tool in the biomedical domain.</p> <p>Future prospects for BEHF encompass the addition of more manuscript datasets, refining event embedding techniques, and introducing a feature for query graph-based searches. These enhancements aim to further the system's capabilities.</p> <p>The talk will also shed light on the applications of BEHF for Retrieval-augmented language models, underscoring its potential influence on the evolution of language models within the biomedical sector.</p>
16:40	総合討論 (10分)
セッション6 ハッカソン参加チーム紹介	
16:50	チーム代表者
17:15	<p>総括・閉会挨拶</p> <p>小林 紀郎 (理化学研究所 情報統合本部)</p>
17:20	閉会

[問い合わせ先]

国立研究開発法人理化学研究所 情報統合本部 基盤研究開発部門 データ知識化開発ユニット

isc-dko-secretary [at] ml.riken.jp

([at] を @ で置き換えてください)

[参加申し込み]

下記サイトにて、事前参加申し込みをお願いします。

https://krs2.riken.jp/m/hackathon_r5

参加申し込み数が定員に達した場合には、参加申し込みを締め切らせていただきます。

[個人情報に関する取扱について]

ご記入頂きました個人情報は理研ハッカソン運営管理の目的にのみ利用させていただきます。

また必要なセキュリティ対策を講じ、厳重に管理致します。