研究データ管理・利活用の事例

ムーンショット型研究開発制度

事例紹介（JST･目標1）

石黒プロジェクトで収集されたデータの有効活用事例

石黒プロジェクトでは、CAシステムを通して得られる多様なデータの再利用に取り組んでおり、計２４件の体系的なデータ収集と整備を行っている。１４件がプロジェクト内で共有されており、このうち２件がGitHubを通して広く公開されている。

日本語日常対話コーパス（日本語の日常的な対話データ）

日常生活、学校、旅行、健康、娯楽の5つのトピックに関する日常会話を収録した高品質なマルチターン対話データセット。すべての対話は基本的な語彙と語順で標準的な日本語で書かれている。

公開場所：GitHub（https://github.com/jqk09a/japanese-daily-dialogue)
公開時期：２０２３年５月１７日
アクセス数：２６５件/２週間（8月末～9月上旬に集計）

BPersonaChat（日本語と英語の対訳がセットとなった対話データ）

英語多言語チャットコーパスPersona-chatと日本語多言語チャットコーパスJPersona-chatに基づく評価データセット。各チャットは、人工的なペルソナを想定した2人のクラウドワーカーの間で行われ、発言者は、自己紹介、趣味、その他を含むがこれらに限定されていない。

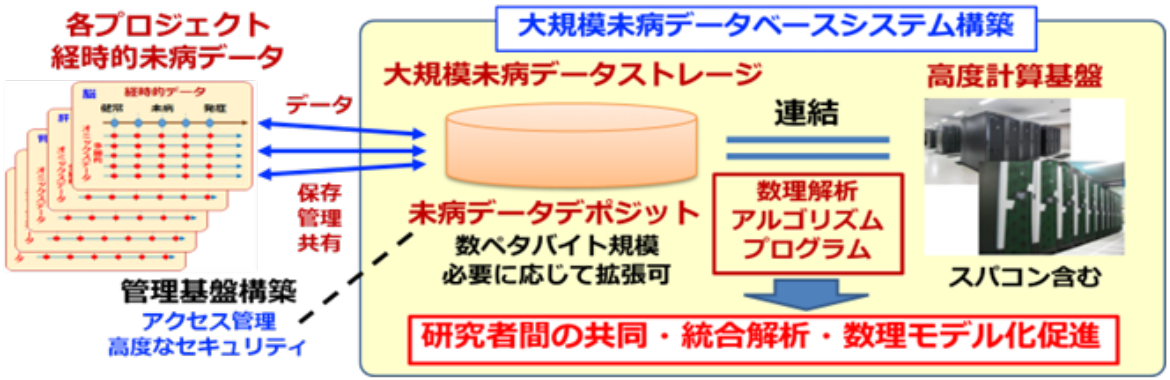
公開場所：GitHub（https://github.com/cl-tohoku/BPersona-chat）
公開時期：２０２３年１月１２日
アクセス数：６件

いずれも公開からの期間が短いため、具体的な活用事例についての情報は、確認できていない。日本語日常対話コーパスについては、１００件弱のcloneがあると推定されており、水面下で活用されている可能性が高い。

出典：JST提供

事例紹介（JST・目標2）
2050年までに、超早期に疾患の予測・予防をすることができる社会を実現

- 【目的】各プロジェクトで創出されたデータを共有し、円滑・高速な統合解析、数理モデル化を行うため、大規模データベースシステムを構築中。

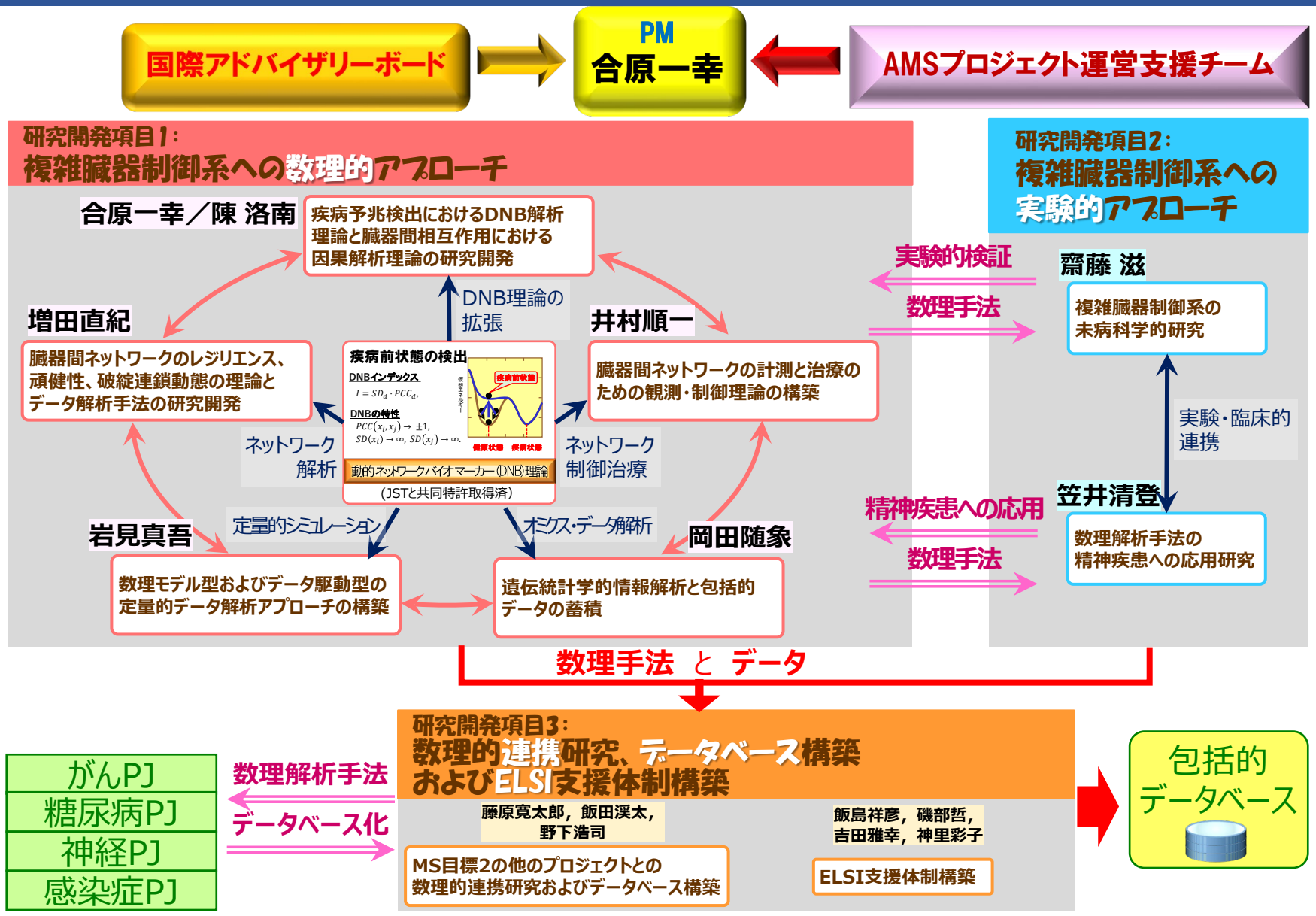


NII（国立情報学研究所）GakuNin RDM

- 【DBシステム】NII（国立情報学研究所）のGakuNin RDMを利用。
- 【体制】全プロジェクト横断で、数理データ連絡会議、データベース作業部会を構築し、最適なDB運用に向けて議論し、データフォーマット、メタデータ設計、規程整備、データベース利活用規程等を策定し推進。DBマネジメントやELSIの専門的な側面支援を目的に、データベースマネージメント支援チーム、ELSI対応チームを構築。
- 【管理データ】各プロジェクト経時的未病データ（Bulk RNA-seq、一細胞RNA-seq、空間的遺伝子発現、ゲノム・エピゲノム等）
- 【共有範囲／将来構想】：当初は目標２内の共有を実施。将来的には、本格的な未病社会の構築に向けた統合的な超早期の疾病の予測・予防の研究において、国際的な未病データ基盤の礎となることを想定。

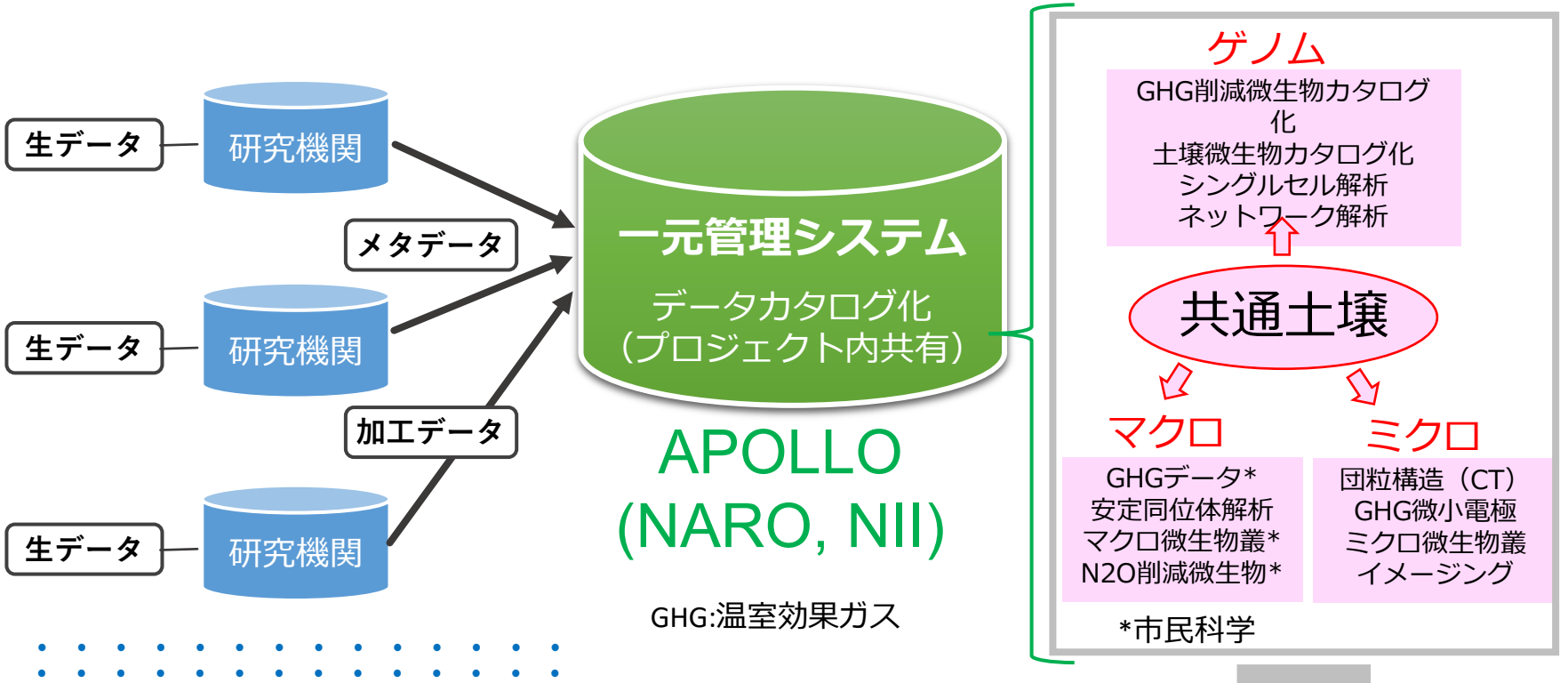
出典：JST提供

事例紹介（JST・目標2）
合原一幸PM「複雑臓器制御系の数理的包括理解と超早期精密早期医療への挑戦」



事例紹介（NEDO・目標4）
南澤 究PM　「資源循環の最適化による農地由来の温室効果ガスの排出削減」

取組事項
- プロジェクト内のGHG削減土壌メタデータを一元管理するシステムを構築
- 蓄積されたデータをプロジェクト内で活用する取り組みを実施



- 管理対象データの範囲：共有
- 公開、共有、非公開・非共有の区分の基準：共有

事例紹介（BRAIN･目標5）
竹山PM「土壌微生物叢アトラスに基づいた環境制御による循環型協生農業プラットフォーム構築」

【取組内容】

- 研究データは、研究機関ごとの管理からプロジェクト全体での共有管理への移行を推進中。
- **「土壌微生物アトラス」**と**「農業デジタルツイン」**を基軸とするデータマネジメントによって、**地域・日本・地球の健康**に貢献する研究成果を追求。

＜土壌微生物アトラス＞

土壌微生物の特徴(系統･遺伝子)を土壌の性質ごとに体系化し、土壌微生物を中心にした地図帳（アトラス）を構築

データ収集

- 細菌叢データ
- 細菌ゲノムデータ
- ラマン分光データ
- 土壌･環境･作物データ etc.

データ分析

- 有用微生物の共起ネットワーク解析
- RNA-seq 解析
- 代謝物解析 etc.

データ活用

- 土壌健康度の指標
- 微生物資材
- 食味と生産性を兼ね備えたダイズ育種 etc.

プロジェクト完了時の全データは、～１PB（千兆バイト）規模のデータ量を想定

＜農業デジタルツイン＞

土壌微生物を含めた農業のデジタル管理を中心にワンヘルスを実現

データ収集	･収集システム ･データ蓄積･加工	
データ分析	･モデル構築 ･予測結果導出	
データ活用	･分析ツール提供 ･API提供	
土壌分析関連機能	収量予測	品質予測 / **環境負荷予測**
サプライチェーン関連機能	品種別需給管理	直接取引市場（国内/海外） / スマートフード促進コミュニティ管理

大規模圃場試験（北海道～九州の6地点）と全国圃場（33道府県59圃場）で取得した、土壌微生物アトラスデータを含む数万を超える測定項目のマルチモーダルデータ（マルチオミクス）で学習された統合モデルを搭載

- **管理対象データ**： 論文のバックデータを必須としつつ、可能な範囲でデータの共有をプロジェクト内に求めている。
- **公開、共有、非共有・非公開の区分**： 業界を発展させるために必要なデータは、可能な限り公開・共有。
  知財の保護等に係るものは、非共有・非公開。

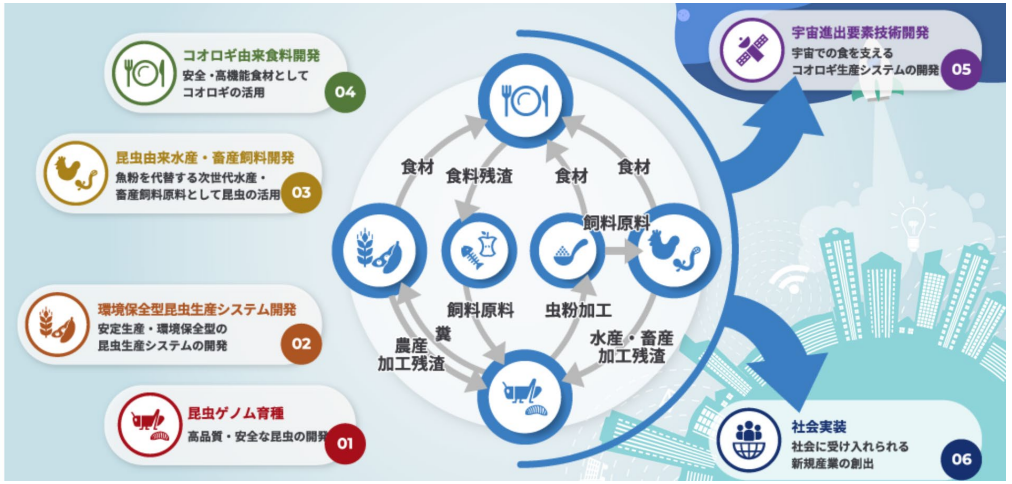
出典：BRAIN提供

事例紹介（BRAIN･目標5）

由良 敬PM　「地球規模の食料問題の解決と人類の宇宙進出に向けた昆虫が支える循環型食料生産システムの開発」

【取組内容】

- 研究データは研究者が所属する各研究機関で管理しているが、プロジェクト内で研究データを参照（共有）できるシステムを開発中。
- バイオインフォマティクスの観点から、多様な研究データを活用・分析することで研究成果を生み出すことを一つの目標としている。



出典：「ムーンショット型農林水産研究開発事業『地球規模の食料問題の解決と人類の宇宙進出に向けた昆虫が支える循環型食料生産システムの開発』」（https://if3-moonshot.org/rd/subproject/）

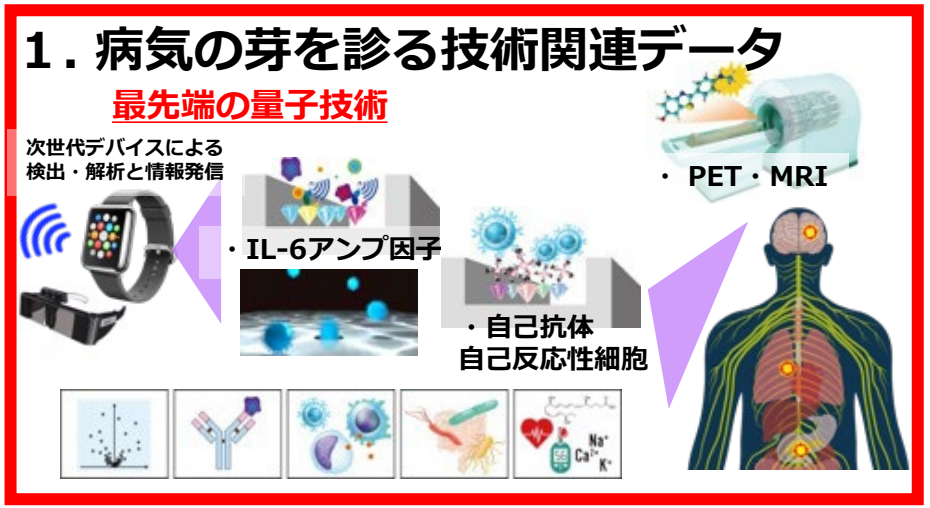
- 管理対象データ：論文のバックデータを必須としつつ、可能な範囲でデータの共有をプロジェクト内に求めている。
- 公開、共有、非共有・非公開の区分：業界を発展させるために必要なデータは可能な限り公開・共有。また、知財の保護等に係るものは非共有・非公開。

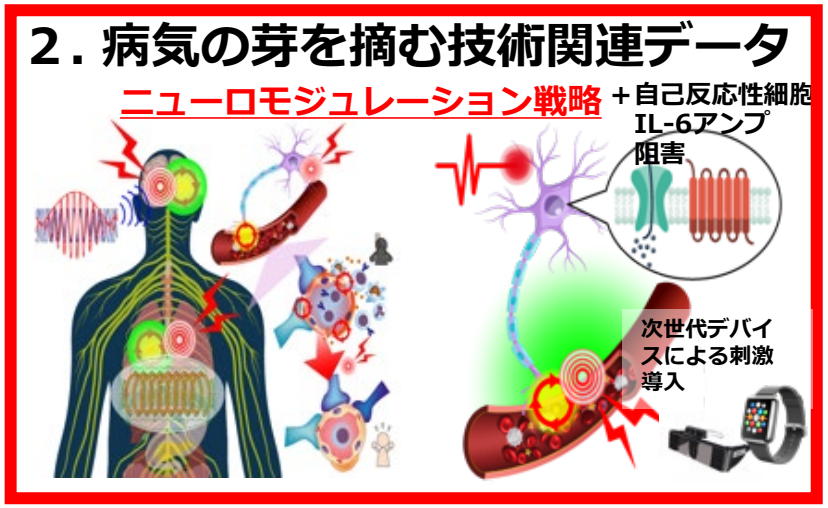
事例紹介（BRAIN･目標7）

村上 正晃PM　「病気につながる血管周囲の微小炎症を標的とする量子技術、ニューロモデュレーション医療による未病時治療法の開発」

取組事項

- プロジェクト内の共同研究データを一元管理するシステムを構築中。
  （村上PMらが管理するデータサーバーを利活用）
- 研究データの共有により、データ解析を行うことで、量子技術による超高感度解析、ニューロモジュレーション医療を実現することを目標に研究を実施。





- 管理対象データの範囲
  個人情報は匿名化を行うことを徹底
  量子計測デバイス関連データ 、大容量画像データ、遺伝子発現関連データ
  生理•行動情報に関するデータなどが対象
- 公開、共有、非公開・非共有の区分の基準
  - 実験データの取得から解析までが一つのグループで完結しない場合はデータの公開•共有を進める