ジャパンリンクセンター「対話・共創の場」
「AI時代におけるメタデータの再考」
2026年2月2日

メタデータ技術に基づいた研究データ知識化の実現

理化学研究所 情報統合本部 基盤研究開発部門

小林 紀郎

An Introduction to RIKEN

和光

- 最先端研究プラットフォーム連携(TRIP)事業本部
- 開拓研究所
- 数理創造研究センター
- 計算科学研究センター
- 量子コンピュータ研究センター
- 情報統合本部
- 脳神経科学研究センター
- 環境資源科学研究センター
- 創発物性科学研究センター
- 光量子工学研究センター
- 仁科加速器科学研究センター
- 放射光科学研究センター

量子コンピュータ


重イオンビーム

けいはんな

- バイオリソース研究センター
- 革新的知能統合研究センター
- 情報統合本部

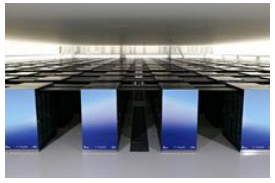
播磨

- 放射光科学研究センター

放射光施設

神戸

- 生命機能科学研究センター
- 数理創造研究センター
- 開拓研究所
- 計算科学研究センター

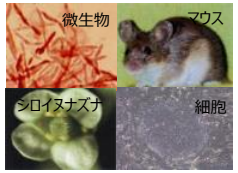

スパコン「富岳」

仙台

- 光量子工学研究センター

筑波

- バイオリソース研究センター
- 環境資源科学研究センター



バイオリソース

東京

- 革新的知能統合研究センター
- 数理創造研究センター
- 計算科学研究センター

横浜

- 開拓研究所
- 数理創造研究センター
- 計算科学研究センター
- 生命医科学研究センター
- 環境資源科学研究センター


クライオ
電子顕微鏡

TRIP 事業

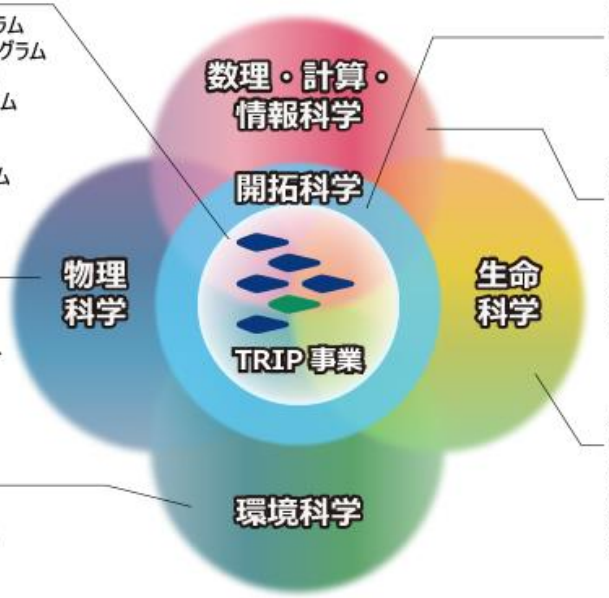
統合データ・計算科学プログラム
科学研究基盤モデル開発プログラム
基礎量子科学研究プログラム
創薬・医療技術基盤プログラム
先端半導体科学プログラム
理研産業協創プログラム
バトンゾーン研究推進プログラム

物理科学領域

創発物性科学研究センター
光量子工学研究センター
仁科加速器科学研究センター
放射光科学研究センター

環境科学領域

環境資源科学研究センター
(バイオリソース研究センター)



開拓科学領域

開拓研究所

数理・計算・情報科学領域

数理創造研究センター
計算科学研究センター
量子コンピュータ研究センター
革新知能統合研究センター
情報統合本部

生命科学領域

生命医科学研究センター
生命機能科学研究センター
脳神経科学研究センター
バイオリソース研究センター

つなぐ科学　TRIP構想

RIKEN TRIP
Transformative Research Innovation Platform of RIKEN platforms

https://trip.riken.jp/

Transformative Research Innovation Platform of RIKEN platforms

データ・予測アルゴリズム・先端計算による科学研究の革新

経済成長・社会課題解決へ繋げる

良質なデータとは

研究データのライフサイクルにおける次の利用者が、
推測に頼らず、
メタデータとして明示された情報を読むことで、
データの理解と再利用に必要な事項を確認できるデータ

研究データのライフサイクル

- 計画
- 生成
- 集約
- 解析／統合
- 共有／公開
- 保存／廃棄

メタデータと知識化

メタデータ

- データ生成の背景やデータの意味（語彙／概念）を記述する情報

知識化

- スキーマと語彙（オントロジ）の厳密な定義と共有により、統語的（構造）および意味的相互運用性を担保
- 機械処理を前提とした形式で実装、AIとの接続を可能とする
- 目的または課題に対応できるように、既存のメタデータとオントロジーを体系的に集約・整理・構造化し、得られる知見（知識グラフ）を利用可能な形で扱う

良質なデータとは

研究データのライフサイクルにおける次の利用者が、推測に頼らず、**メタデータ**として明示された情報を読むことで、データの理解と再利用に必要な事項を確認できるデータ

研究データのライフサイクル

- 計画　　データ管理計画（DMP）
- 生成　　市販メーカ製計測機器
- 集約　　研究情報管理基盤・データリポジトリの整備
- 解析／統合
- 共有／公開　　FAIR原則に基づいた公開・公共リポジトリへの登録
- 保存／廃棄

公開に用いるメタデータから
遡って
研究メタデータを考える

＋

AIの支援を得て、メタデータ管理を低コストで実現したい

生命科学共通メタデータスキーマの抽出

生命科学の国際リポジトリ・コンソーシアム・公共リポジトリのメタデータを比較して
生命科学研究の最小限のメタデータスキーマを抽出

次世代シーケンサー

DDBJ DNA Data Bank of Japan

顕微鏡イメージング

OME

QUAREP-LiMi

メタボローム（代謝物）

Plant Metabolome MetaDatabase

プロジェクト
- プロジェクトID
- タイトル
- 説明
- 作成者
- 作成者
- 研究責任者
- 連絡窓口
- 参考文献
- 実験

実験
- 実験ID
- タイトル
- 説明
- 測定の種類
- 技術の種類
- 測定基盤
- 参考文献
- 実験日
- 実験者
- 実験設計
- 測定
- データ解析

測定
- サンプル
- 測定条件
- データセット
- 計測日時
- 説明

データセット
- フォルダ
- ファイル
- 圧縮ファイル

測定条件
- 機器
- 機器生成メタデータファイル

機器
- 機器ID
- 機器名
- 所在地
- 管理者

データ解析
- 前処理
- 統計的データ解析
- 単変量解析
- 多変量解析
- アノテーション方法
- 可視化
- 解析結果データセット

サンプル
- サンプルID
- 名称
- 説明
- 化合物
- 生物サンプル
- サンプル調整

サンプル調整
- 処理法
- サンプル量
- サンプル抽出プロトコル
- 抽出法
- 標準化合物

60クラス、152項目を定義　青色部分は研究ごとに定義が必要

生命科学実験研究メタデータを「共用機器測定メタデータ用」に整理

大阪大学 コアファシリティ機構 古谷 浩志先生よりご提供

測定基本情報 共用必須

- 測定者
- 計測日時
- サンプル
- 測定条件
- データセット
- 説明

機器 共用必須

- 機器ID
- 機器名
- 所在地
- 管理者

測定条件 共用必須

- 機器
- 測定条件
- 機器生成メタデータファイル

①測定直後に最低限必ず記入すべき項目

サンプル 一部必要

- サンプルID
- 名称
- 説明
- 化合物
- 生物サンプル
- サンプル調整

サンプル調整

- 処理法
- サンプル量
- サンプル抽出プロトコル
- 抽出法
- 標準化合物

実験

- 実験ID
- タイトル
- 説明
- 測定の種類
- 技術の種類
- 測定基盤
- 参考文献
- 実験日
- 実験者
- 実験設計
- 測定
- データ解析

②測定データの重要性に応じて徐々に追記する項目

データセット 一部必要

- フォルダ
- ファイル
- 圧縮ファイル

データ解析

- 前処理
- 統計的データ解析
- 単変量解析
- 多変量解析
- アノテーション方法
- 可視化
- 解析結果データセット

③研究進捗に応じて追記する項目

プロジェクト

- プロジェクトID
- タイトル
- 説明
- 作成者
- 作成者
- 研究責任者
- 連絡窓口
- 参考文献
- 実験

④論文発表等公表の際に添付すべき項目

追加

⑤当該「共用機器タイプ」に特有の追加詳細情報（メタデータ）

理研メタデーベース

https://metadb.riken.jp

- 生命科学系の、理研が生成した知識グラフ（メタデータ）を中心に集約、統合、公開

- 研究者には扱いにくいグラフ構造を、表形式等のわかりやすい方法で表示

32の公共オントロジによるデータ記述
57の理研DBと74の外部DBを統合
(計 **131DB**)

国内外の研究コミュニティで活用

IMPC

Brain/MINDS

ViBrism-DB New Platform ViBrism DB

A portal of phenotype data analyzed in Japan J-phenome

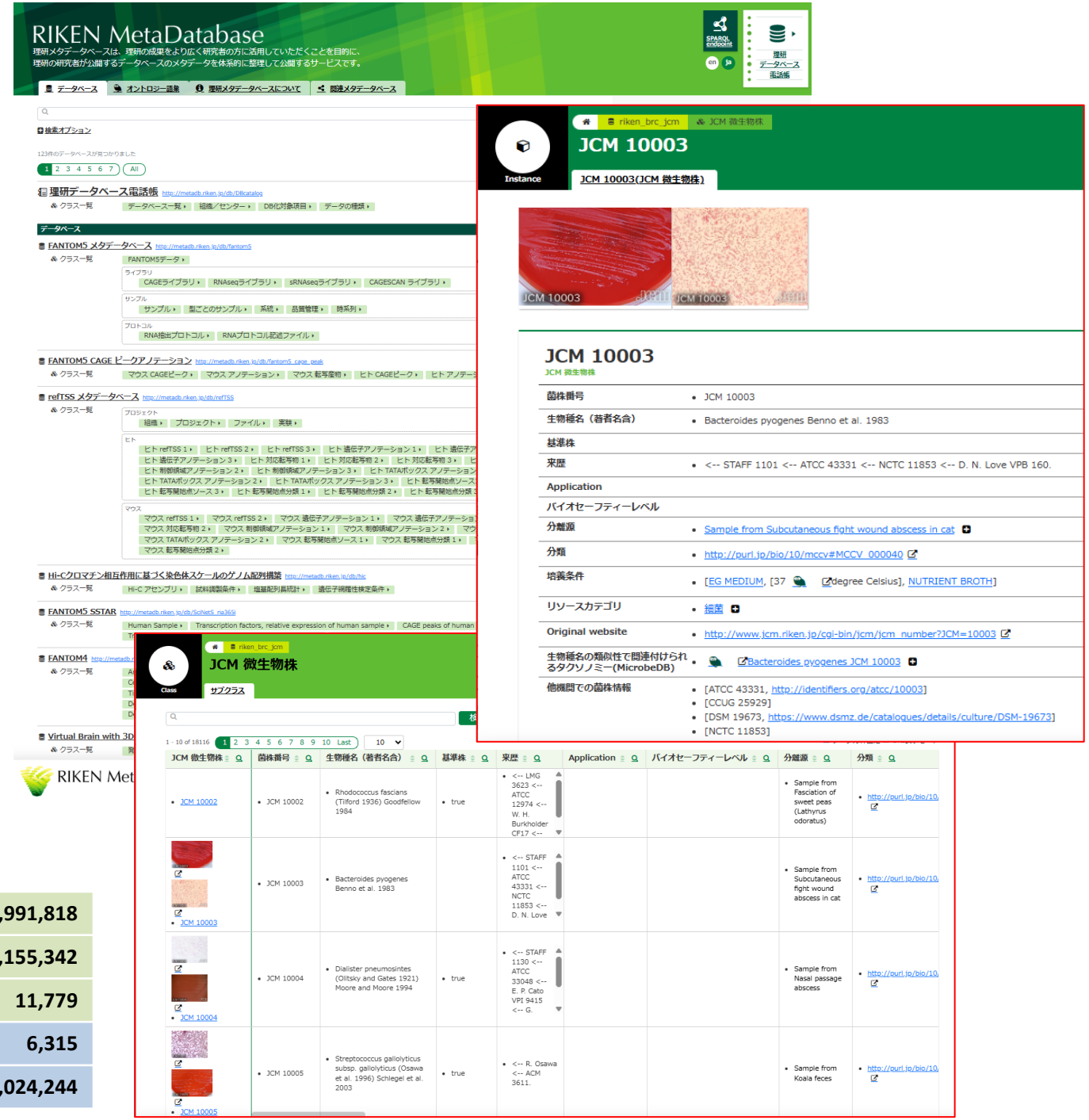
BrainTx

RIKEN Microstructural ImagingMetadatabase

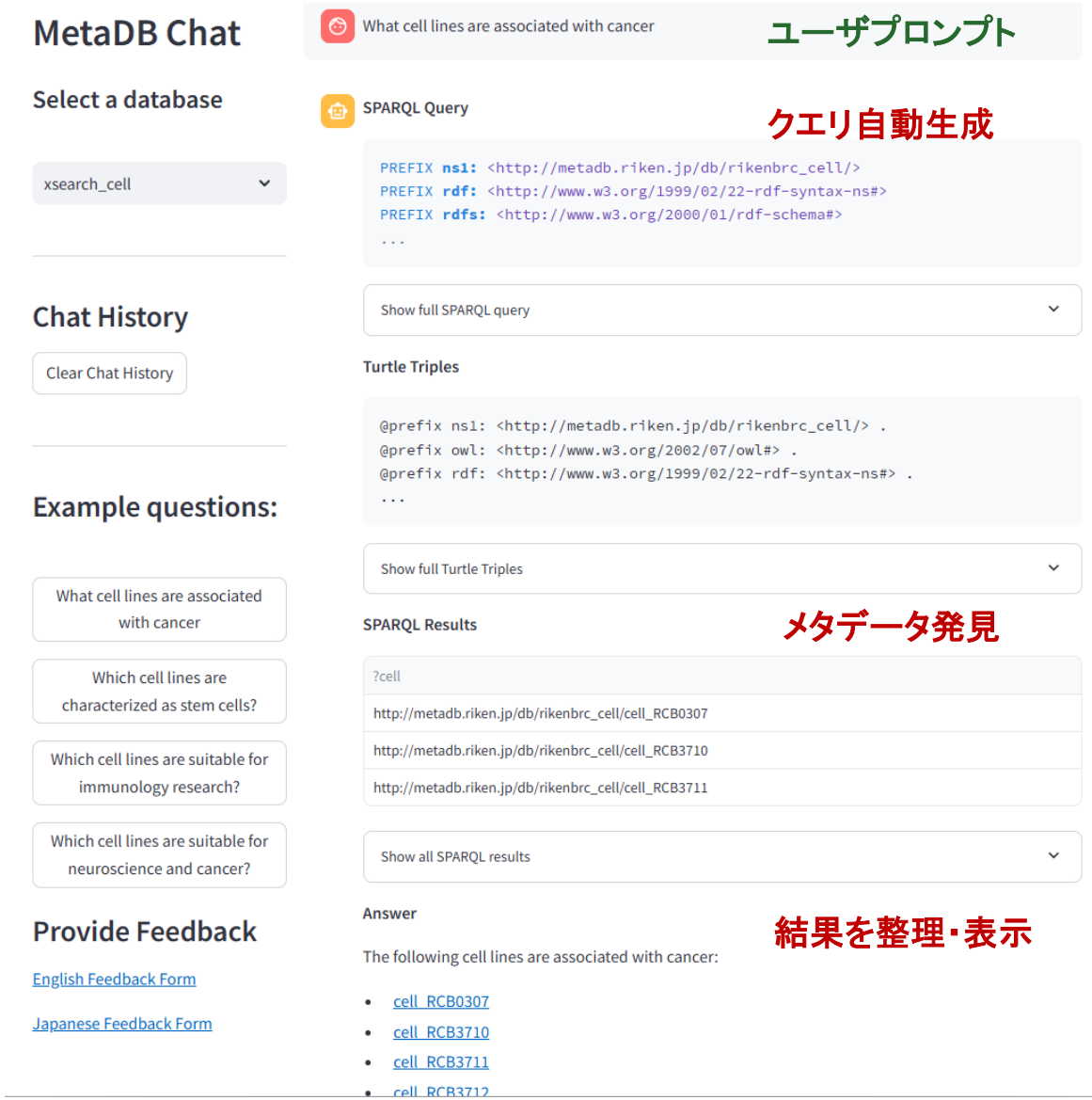
RIKEN Plant Metabolome MetaDatabase

OME
Open Microscopy Environment

概念数	3,991,818
トリプル数	1,449,155,342
関係数	11,779
月ユーザ数	6,315
月アクセス数	1,024,244



理研メタデーベースチャットの開発



https://metadb.riken.jp/chat/

自然言語の質問をSPARQLに自動変換して、メタデータを検索

生命科学公開メタデータ（RIKEN MetaDatabase）を用いて大規模言語モデル(LLM)よる検索枠組みを試作・検証し、その成果を実際のデータ管理基盤における研究データ検索へと展開できる要素技術を確立する。

今後の展開

- メタデータの自動生成に応用
- 構造化されていないメタデータを許容するデータ検索
- 論文・文献と組み合わせたメタデータ検索・抽出

メタデータ設計・評価システム

メタデータ定義には、統制語彙の利用や構造化においてメタデータの専門知識が必要であるが、
組織や職制（立場）を超えて、複数人で協力してメタデータを設計できるウエブツールを整備して課題解決を試る。

プロトタイプ版を公開中　https://metadb.riken.jp/metadataDesign/

大学・研究機関向けの認証「学認」に対応し、全国の大学・研究機関等で利用可

システム上でプロジェクトを作成し、プロジェクト内で共同作業

研究者	図書館員、RA	メタデータ専門家（情報）
		
データの深い理解 スキーマ基本設計	データ流通・検索の知見 標準メタデータとの関連づけ	メタデータ技術 オントロジー・既存スキーマを利用した意味情報の付与

ユーザフィードバックを得て、特にGUIの改良を進めている

メタデータ設計・評価システムの実装

Meta-data Toolkit

① ユーザ管理

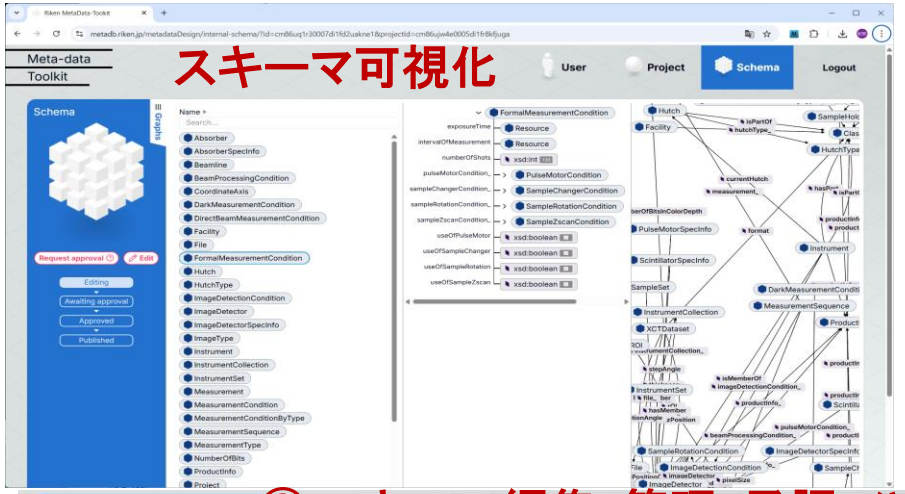
② プロジェクト作成・管理

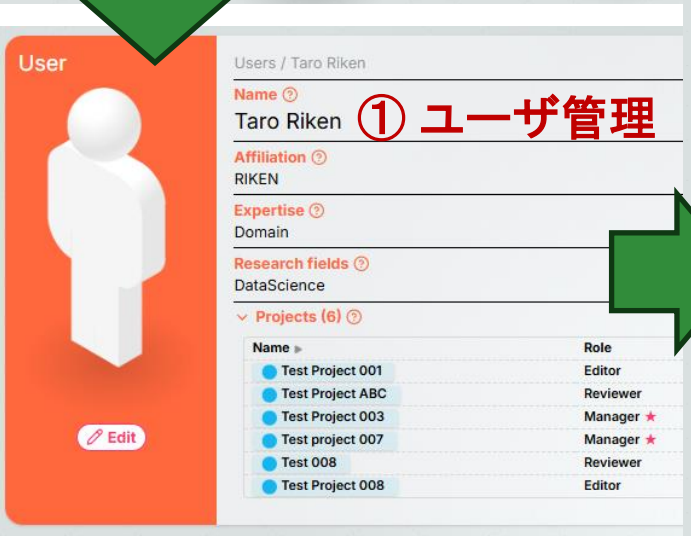
③ スキーマ編集・管理・承認・公開

Manage Users

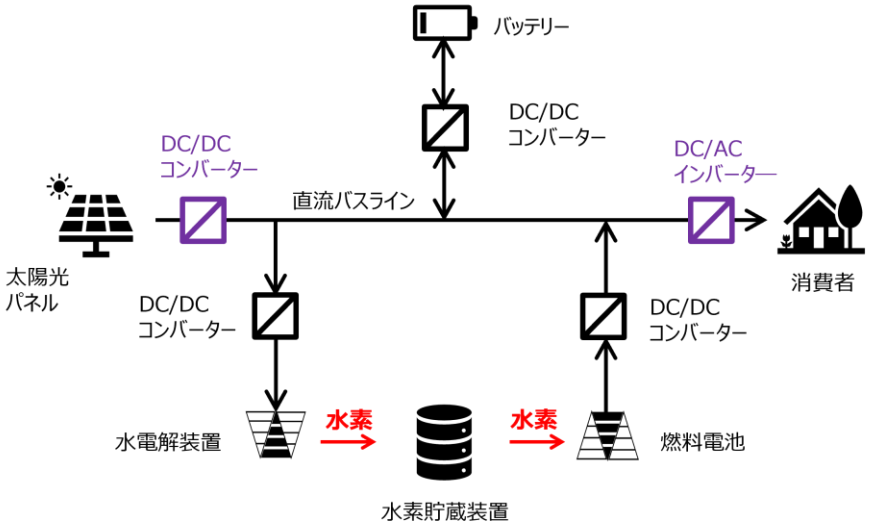
Manage Projects

Manage Schemata





分散型水素エネルギーシステム



[1] D. Yamashita, et al., Int. J. Hydrogen Energy 44 (2019) 27542.
[2] K. Tsuno et al., IFAC PapersOnLine 56-2 (2023) 9098.

高度な知識AI基盤を備えたエネルギーシステム

集約された知識をAI・シミュレータで利用

システム設計・解析AI

装置の仕様	装置特性（応答特性、上下限値、劣化等） 設置特性（太陽電池の面積、方角、角度、環境等）
費用情報	イニシャルコスト（装置費用、設置費用） ランニングコスト（運用・保守費用）
環境データ	環境特性（日照、気温、湿度等の観測データ） 気象予測（日照量、気温、湿度、風速等）
稼働計画	計画（発電量計画、電気機器稼働計画等）
装置の状態	太陽電池（日照量、気温、湿度） 発電電力、消費電力（DC/DC、DC/AC） バス電圧（DC/DCコンバータ：電圧） 装置電力（DC/DCコンバータ：電圧、電流） 装置状態（温度、圧力） 装置環境（温度、湿度） 水素発生量 水素消費量 充填(充電)状態（電池：電圧、水素：圧力）など

発電予測AI
需要予測AI
効率予測AI
劣化予測AI
データ同化

（簡易）シミュレータ（詳細）

制御最適化

最適制御AI

- 再エネ発電装置規模
電池容量・水素貯蔵量
水電解装置・燃料電池の選定
- 期間収支
電力収支＋維持コスト
- 制御パラメータの最適設定
- 制御パラメータの動的最適化

知識ベース（オントロジー/メタデータ）

分散型水素エネルギシステム での知識化

多種多様なデータをメタデータで体系化して制御や予測に活用