

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

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

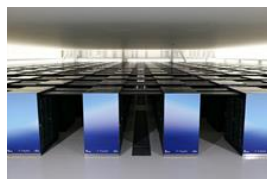
- ・放射光科学研究センター



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

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



バイオリソース



クライオ
電子顕微鏡

TRIP 事業

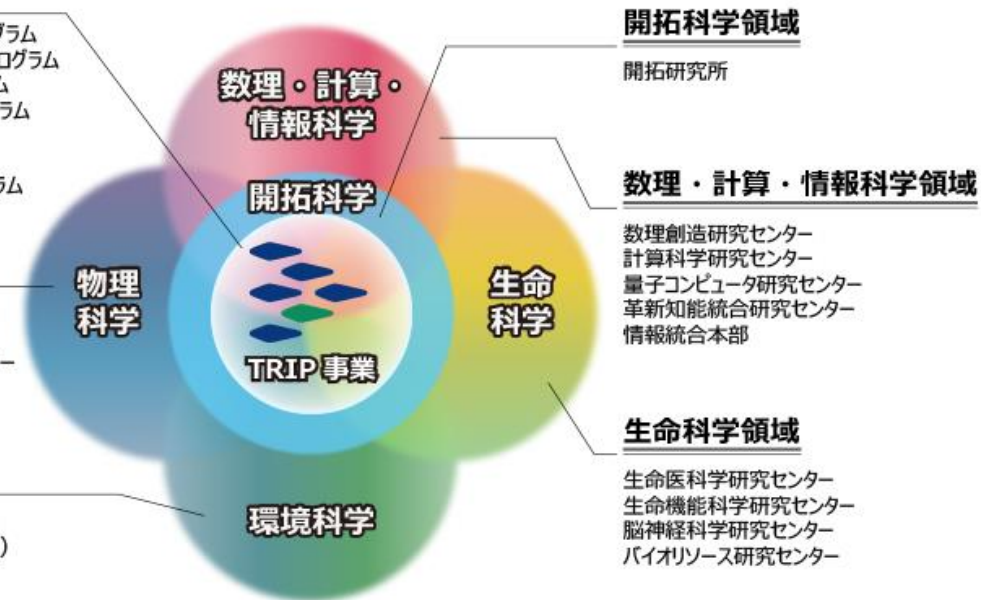
統合データ・計算科学プログラム
科学研究基盤モデル開発プログラム
基礎量子科学研究プログラム
創薬・医療技術基盤プログラム
先端半導体科学プログラム
理研産業協創プログラム
バトンゾーン研究推進プログラム

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)



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



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 市販メーカー製計測機器
 - 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
 - 保存／廃棄
- FAIR原則に基づいた公開・公共リポジトリへの登録



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



顕微鏡イメージング



メタボローム(代謝物)



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

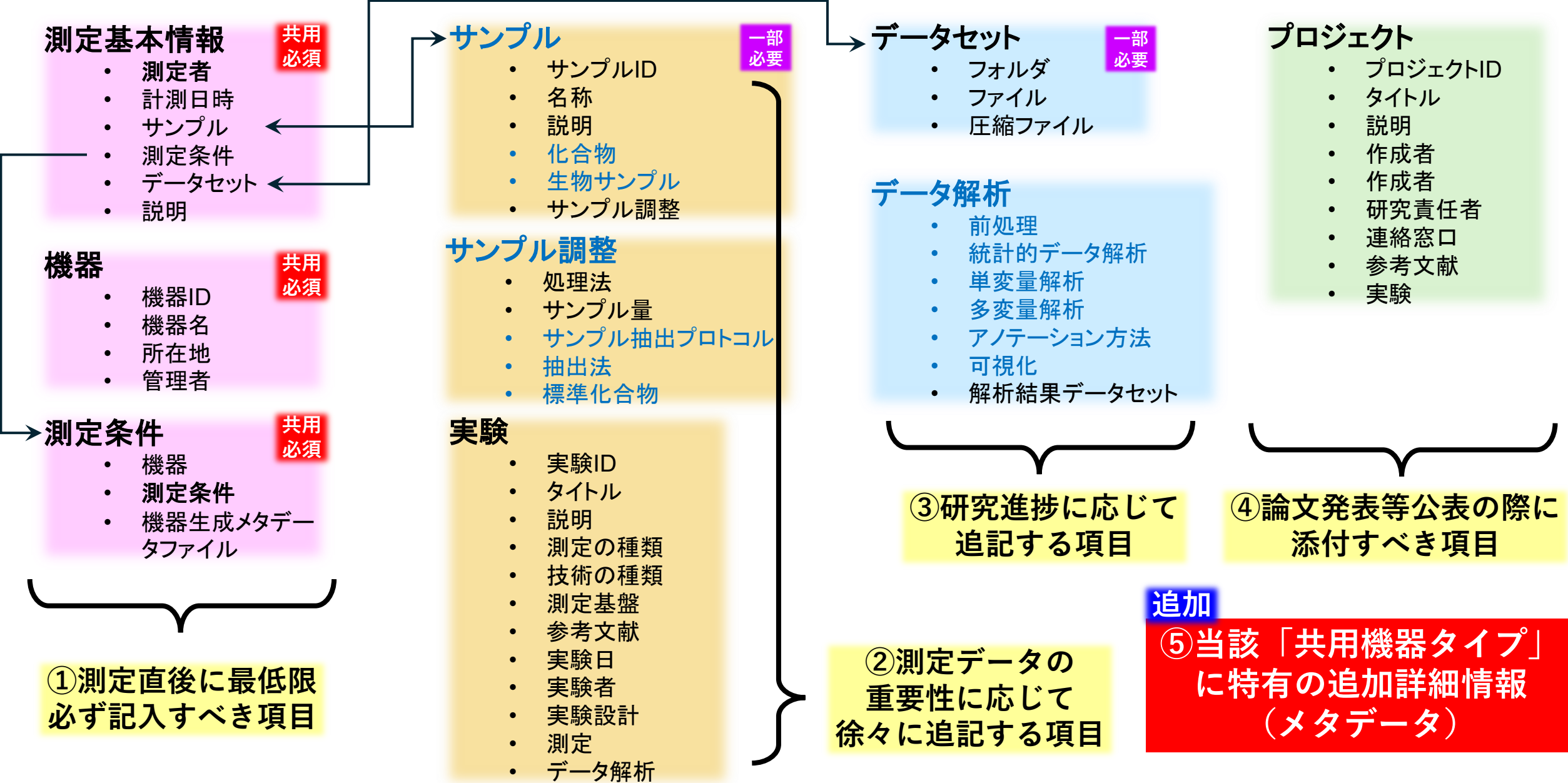
- サンプルID
- 名称
- 説明
- 化合物
- 生物サンプル
- サンプル調整

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

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



<https://metadb.riken.jp>

- 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**)



IMPC

A portal of
phenotype data
analyzed in Japan



JP phenome



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

RIKEN MetaDatabase

理研メタデータベースは、理研の成果をより広く研究者の方に活用していただくことを目的に、理研の研究者が公開するデータベースのメタデータを体系的に整理して公開するサービスです。

データベース
オントロジー編集
理研メタデータベースについて
関連メタデータベース

検索オプション

123件のデータベースが見つかりました

1 2 3 4 5 6 7 All

理研データベース電話帳 <http://metadb.riken.jp/db/Catalog>

▲ クラス一覧

データベース一覧、 組織/センター、 DB化対象項目、 データの種類、

データベース

FANTOM5 メタデータベース <http://metadb.riken.jp/db/Fantom5>

▲ クラス一覧

FANTOM5データ、

ライブラリ

CAGEライブラリ、 RNAseqライブラリ、 SRNaseqライブラリ、 CAGESCAN ライブラリ、

サンプル

サンプル、 製成とのサンプル、 系統、 品質管理、 時刻系列、

プロトコル

RNA塩化プロトコル、 RNAプロトコル認識ファイル、

FANTOM5 CAGE ピークアノテーション http://metadb.riken.jp/db/fantom5_cage_peak

▲ クラス一覧

マウス CAGE ピーク、 マウス アノテーション、 マウス 転写産物、 ヒト CAGE ピーク、 ヒト アノテーション

refTSS メタデータベース <http://metadb.riken.jp/db/refTSS>

▲ クラス一覧

プロジェクト

組織、 プロジェクト、 ファイル、 実験、

ヒト

ヒト refTSS 1、 ヒト refTSS 2、 ヒト refTSS 3、 ヒト 遺伝子アノテーション 1、 ヒト 遺伝子アノテーション 2、 ヒト 遺伝子アノテーション 3、 ヒト 対応転写物 1、 ヒト 対応転写物 2、 ヒト 対応転写物 3、 ヒト 対応転写物 4、 ヒト 対応転写物 5、 ヒト 対応転写物 6、 ヒト 対応転写物 7、 ヒト 対応転写物 8、 ヒト 対応転写物 9、 ヒト 対応転写物 10、 ヒト 対応転写物 11、 ヒト 対応転写物 12、 ヒト 対応転写物 13、 ヒト 対応転写物 14、 ヒト 対応転写物 15、 ヒト 対応転写物 16、 ヒト 対応転写物 17、 ヒト 対応転写物 18、 ヒト 対応転写物 19、 ヒト 対応転写物 20、 ヒト 対応転写物 21、 ヒト 対応転写物 22、 ヒト 対応転写物 23、 ヒト 対応転写物 24、 ヒト 対応転写物 25、 ヒト 対応転写物 26、 ヒト 対応転写物 27、 ヒト 対応転写物 28、 ヒト 対応転写物 29、 ヒト 対応転写物 30、 ヒト 対応転写物 31、 ヒト 対応転写物 32、 ヒト 対応転写物 33、 ヒト 対応転写物 34、 ヒト 対応転写物 35、 ヒト 対応転写物 36、 ヒト 対応転写物 37、 ヒト 対応転写物 38、 ヒト 対応転写物 39、 ヒト 対応転写物 40、 ヒト 対応転写物 41、 ヒト 対応転写物 42、 ヒト 対応転写物 43、 ヒト 対応転写物 44、 ヒト 対応転写物 45、 ヒト 対応転写物 46、 ヒト 対応転写物 47、 ヒト 対応転写物 48、 ヒト 対応転写物 49、 ヒト 対応転写物 50、 ヒト 対応転写物 51、 ヒト 対応転写物 52、 ヒト 対応転写物 53、 ヒト 対応転写物 54、 ヒト 対応転写物 55、 ヒト 対応転写物 56、 ヒト 対応転写物 57、 ヒト 対応転写物 58、 ヒト 対応転写物 59、 ヒト 対応転写物 60、 ヒト 対応転写物 61、 ヒト 対応転写物 62、 ヒト 対応転写物 63、 ヒト 対応転写物 64、 ヒト 対応転写物 65、 ヒト 対応転写物 66、 ヒト 対応転写物 67、 ヒト 対応転写物 68、 ヒト 対応転写物 69、 ヒト 対応転写物 70、 ヒト 対応転写物 71、 ヒト 対応転写物 72、 ヒト 対応転写物 73、 ヒト 対応転写物 74、 ヒト 対応転写物 75、 ヒト 対応転写物 76、 ヒト 対応転写物 77、 ヒト 対応転写物 78、 ヒト 対応転写物 79、 ヒト 対応転写物 80、 ヒト 対応転写物 81、 ヒト 対応転写物 82、 ヒト 対応転写物 83、 ヒト 対応転写物 84、 ヒト 対応転写物 85、 ヒト 対応転写物 86、 ヒト 対応転写物 87、 ヒト 対応転写物 88、 ヒト 対応転写物 89、 ヒト 対応転写物 90、 ヒト 対応転写物 91、 ヒト 対応転写物 92、 ヒト 対応転写物 93、 ヒト 対応転写物 94、 ヒト 対応転写物 95、 ヒト 対応転写物 96、 ヒト 対応転写物 97、 ヒト 対応転写物 98、 ヒト 対応転写物 99、 ヒト 対応転写物 100、 ヒト 対応転写物 101、 ヒト 対応転写物 102、 ヒト 対応転写物 103、 ヒト 対応転写物 104、 ヒト 対応転写物 105、 ヒト 対応転写物 106、 ヒト 対応転写物 107、 ヒト 対応転写物 108、 ヒト 対応転写物 109、 ヒト 対応転写物 110、 ヒト 対応転写物 111、 ヒト 対応転写物 112、 ヒト 対応転写物 113、 ヒト 対応転写物 114、 ヒト 対応転写物 115、 ヒト 対応転写物 116、 ヒト 対応転写物 117、 ヒト 対応転写物 118、 ヒト 対応転写物 119、 ヒト 対応転写物 120、 ヒト 対応転写物 121、 ヒト 対応転写物 122、 ヒト 対応転写物 123、 ヒト 対応転写物 124、 ヒト 対応転写物 125、 ヒト 対応転写物 126、 ヒト 対応転写物 127、 ヒト 対応転写物 128、 ヒト 対応転写物 129、 ヒト 対応転写物 130、 ヒト 対応転写物 131、 ヒト 対応転写物 132、 ヒト 対応転写物 133、 ヒト 対応転写物 134、 ヒト 対応転写物 135、 ヒト 対応転写物 136、 ヒト 対応転写物 137、 ヒト 対応転写物 138、 ヒト 対応転写物 139、 ヒト 対応転写物 140、 ヒト 対応転写物 141、 ヒト 対応転写物 142、 ヒト 対応転写物 143、 ヒト 対応転写物 144、 ヒト 対応転写物 145、 ヒト 対応転写物 146、 ヒト 対応転写物 147、 ヒト 対応転写物 148、 ヒト 対応転写物 149、 ヒト 対応転写物 150、 ヒト 対応転写物 151、 ヒト 対応転写物 152、 ヒト 対応転写物 153、 ヒト 対応転写物 154、 ヒト 対応転写物 155、 ヒト 対応転写物 156、 ヒト 対応転写物 157、 ヒト 対応転写物 158、 ヒト 対応転写物 159、 ヒト 対応転写物 160、 ヒト 対応転写物 161、 ヒト 対応転写物 162、 ヒト 対応転写物 163、 ヒト 対応転写物 164、 ヒト 対応転写物 165、 ヒト 対応転写物 166、 ヒト 対応転写物 167、 ヒト 対応転写物 168、 ヒト 対応転写物 169、 ヒト 対応転写物 170、 ヒト 対応転写物 171、 ヒト 対応転写物 172、 ヒト 対応転写物 173、 ヒト 対応転写物 174、 ヒト 対応転写物 175、 ヒト 対応転写物 176、 ヒト 対応転写物 177、 ヒト 対応転写物 178、 ヒト 対応転写物 179、 ヒト 対応転写物 180、 ヒト 対応転写物 181、 ヒト 対応転写物 182、 ヒト 対応転写物 183、 ヒト 対応転写物 184、 ヒト 対応転写物 185、 ヒト 対応転写物 186、 ヒト 対応転写物 187、 ヒト 対応転写物 188、 ヒト 対応転写物 189、 ヒト 対応転写物 190、 ヒト 対応転写物 191、 ヒト 対応転写物 192、 ヒト 対応転写物 193、 ヒト 対応転写物 194、 ヒト 対応転写物 195、 ヒト 対応転写物 196、 ヒト 対応転写物 197、 ヒト 対応転写物 198、 ヒト 対応転写物 199、 ヒト 対応転写物 200、 ヒト 対応転写物 201、 ヒト 対応転写物 202、 ヒト 対応転写物 203、 ヒト 対応転写物 204、 ヒト 対応転写物 205、 ヒト 対応転写物 206、 ヒト 対応転写物 207、 ヒト 対応転写物 208、 ヒト 対応転写物 209、 ヒト 対応転写物 210、 ヒト 対応転写物 211、 ヒト 対応転写物 212、 ヒト 対応転写物 213、 ヒト 対応転写物 214、 ヒト 対応転写物 215、 ヒト 対応転写物 216、 ヒト 対応転写物 217、 ヒト 対応転写物 218、 ヒト 対応転写物 219、 ヒト 対応転写物 220、 ヒト 対応転写物 221、 ヒト 対応転写物 222、 ヒト 対応転写物 223、 ヒト 対応転写物 224、 ヒト 対応転写物 225、 ヒト 対応転写物 226、 ヒト 対応転写物 227、 ヒト 対応転写物 228、 ヒト 対応転写物 229、 ヒト 対応転写物 230、 ヒト 対応転写物 231、 ヒト 対応転写物 232、 ヒト 対応転写物 233、 ヒト 対応転写物 234、 ヒト 対応転写物 235、 ヒト 対応転写物 236、 ヒト 対応転写物 237、 ヒト 対応転写物 238、 ヒト 対応転写物 239、 ヒト 対応転写物 240、 ヒト 対応転写物 241、 ヒト 対応転写物 242、 ヒト 対応転写物 243、 ヒト 対応転写物 244、 ヒト 対応転写物 245、 ヒト 対応転写物 246、 ヒト 対応転写物 247、 ヒト 対応転写物 248、 ヒト 対応転写物 249、 ヒト 対応転写物 250、 ヒト 対応転写物 251、 ヒト 対応転写物 252、 ヒト 対応転写物 253、 ヒト 対応転写物 254、 ヒト 対応転写物 255、 ヒト 対応転写物 256、 ヒト 対応転写物 257、 ヒト 対応転写物 258、 ヒト 対応転写物 259、 ヒト 対応転写物 260、 ヒト 対応転写物 261、 ヒト 対応転写物 262、 ヒト 対応転写物 263、 ヒト 対応転写物 264、 ヒト 対応転写物 265、 ヒト 対応転写物 266、 ヒト 対応転写物 267、 ヒト 対応転写物 268、 ヒト 対応転写物 269、 ヒト 対応転写物 270、 ヒト 対応転写物 271、 ヒト 対応転写物 272、 ヒト 対応転写物 273、 ヒト 対応転写物 274、 ヒト 対応転写物 275、 ヒト 対応転写物 276、 ヒト 対応転写物 277、 ヒト 対応転写物 278、 ヒト 対応転写物 279、 ヒト 対応転写物 280、 ヒト 対応転写物 281、 ヒト 対応転写物 282、 ヒト 対応転写物 283、 ヒト 対応転写物 284、 ヒト

理研メタデータベースチャットの開発

MetaDB Chat

Select a database

xsearch_cell

Chat History

Clear Chat History

Example questions:

What cell lines are associated with cancer

Which cell lines are characterized as stem cells?

Which cell lines are suitable for immunology research?

Which cell lines are suitable for neuroscience and cancer?

Provide Feedback

[English Feedback Form](#)

[Japanese Feedback Form](#)



What cell lines are associated with cancer

ユーザプロンプト



SPARQL Query

クエリ自動生成

```
PREFIX ns1: <http://metadb.riken.jp/db/rikenbrc_cell/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
...
```

Show full SPARQL query

Turtle Triples

```
@prefix ns1: <http://metadb.riken.jp/db/rikenbrc_cell/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
...
```

Show full Turtle Triples

SPARQL Results

メタデータ発見

?cell

http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB0307

http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB3710

http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB3711

Show all SPARQL results

Answer

結果を整理・表示

The following cell lines are associated with cancer:

- [cell_RCB0307](#)
- [cell_RCB3710](#)
- [cell_RCB3711](#)
- [cell_RCB3712](#)

<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)に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ェブツール**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



データの深い理解
スキーマ基本設計

図書館員、RA



データ流通・検索の知見
標準メタデータとの関連づけ

メタデータ専門家
(情報)



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

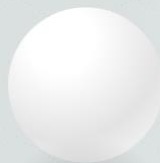
メタデータ設計・評価システムの実装

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



スキーマ可視化

③ スキーマ編集・管理・承認・公開

User

Users / Taro Riken

Name

Taro Riken

Affiliation

RIKEN

Expertise

Domain

Research fields

DataScience

Projects (6)

Test Project 001

Test Project ABC

Test Project 003

Test project 007

Test 008

Test Project 008

Role

Editor

Reviewer

Manager

Manager

Reviewer

Editor

① ユーザ管理

Edit

Project

UOsakaRikenMetadadataDesign

Research field

Analytical chemistry, Data science, Semantic web

Description

A joint research project regarding metadata design on core facilities between the University of Osaka

Schemata (3)

SampleSchema1

SampleSchema2

SampleSchema3

External schemata

RDFS

SKOS

Schema.org

DCMI Metadata Terms

qudt

OME-OWL (OME-Core-201606)

FOAF

Members (3)

Norio Kobayashi

Masako Nomoto

Furutani Hiroshi

② プロジェクト作成・管理

作成するスキーマ

既存のスキーマの取り込み

参加ユーザ

Edit

Schema

SampleSchema1

UOsakaRikenMetadadataDesign

Namespace URI

http://example.com/ss1/

Namespace prefix

ss1:

Description

A sample schema for instruments.

Classes (4)

Instrument

InstrumentOwner

InstrumentPart

InstrumentType

Properties (8)

hasInstrumentPart

hasOwner

hasPart

instrumentTypeID

isDescribedBy

isSharedInstrument

ownerPID

typeOfInstrument

Publication approvers (2)

Norio Kobayashi

Masako Nomoto

クラス(概念)

プロパティ(関係と値)

承認者

編集

承認

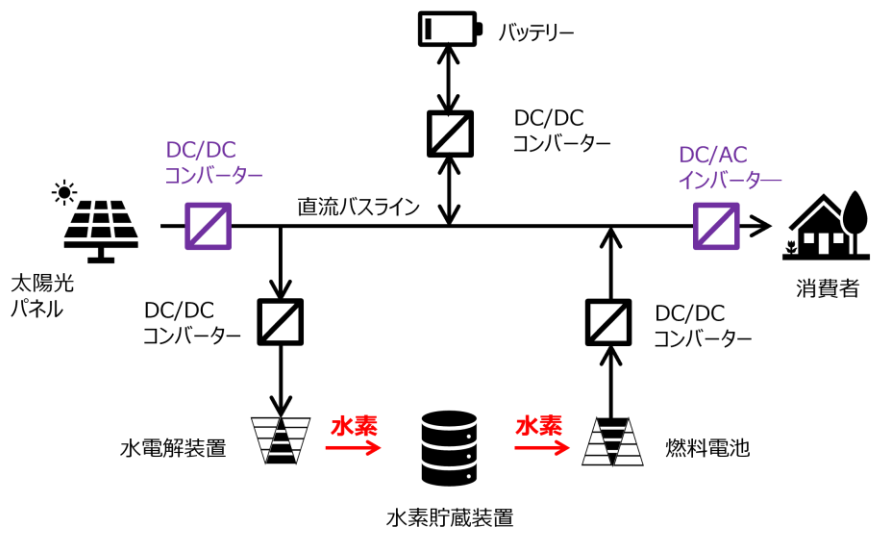
公開

フロー管理

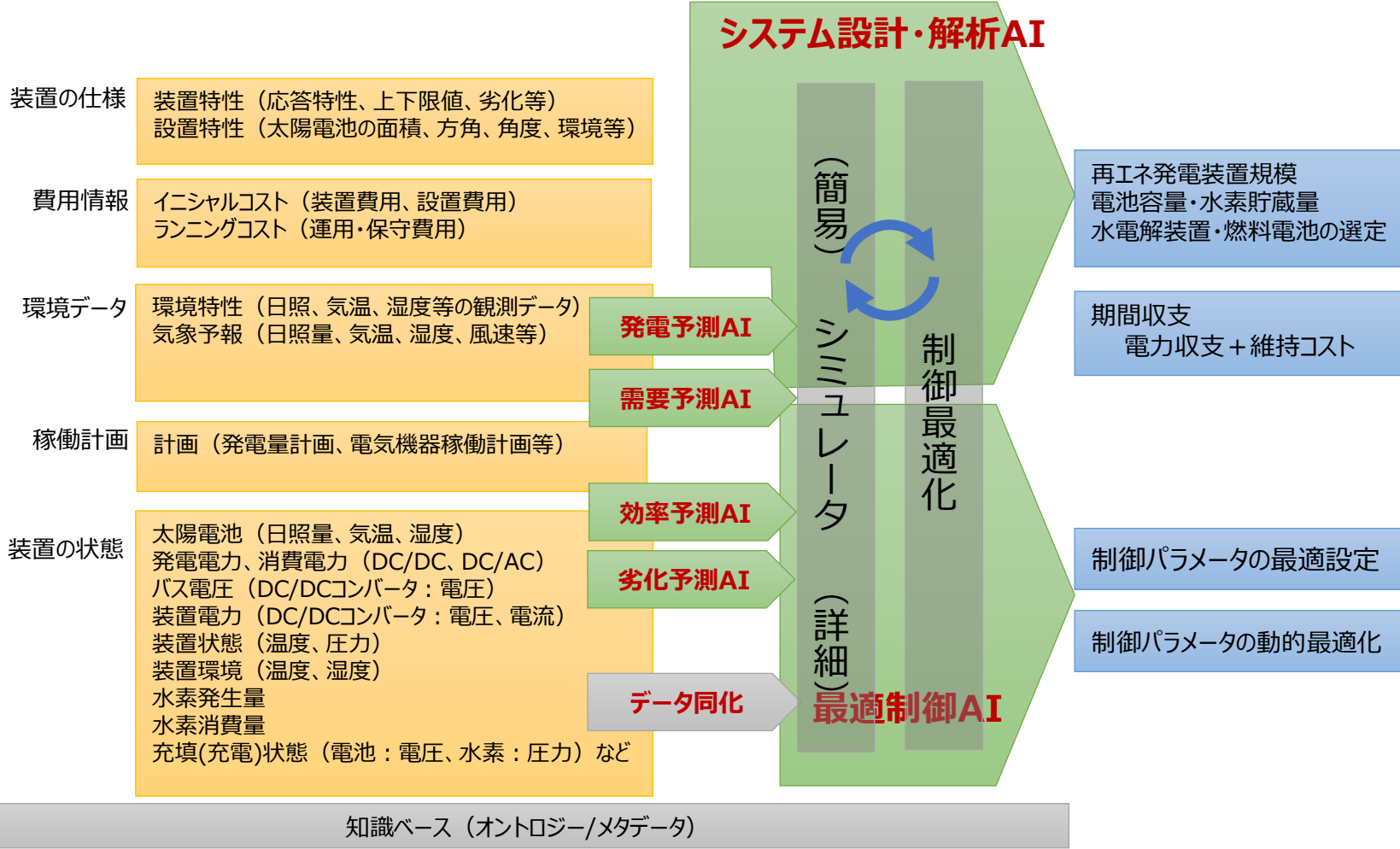
分散型水素エネルギーシステム

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



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



分散型水素エネルギーシステム 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

