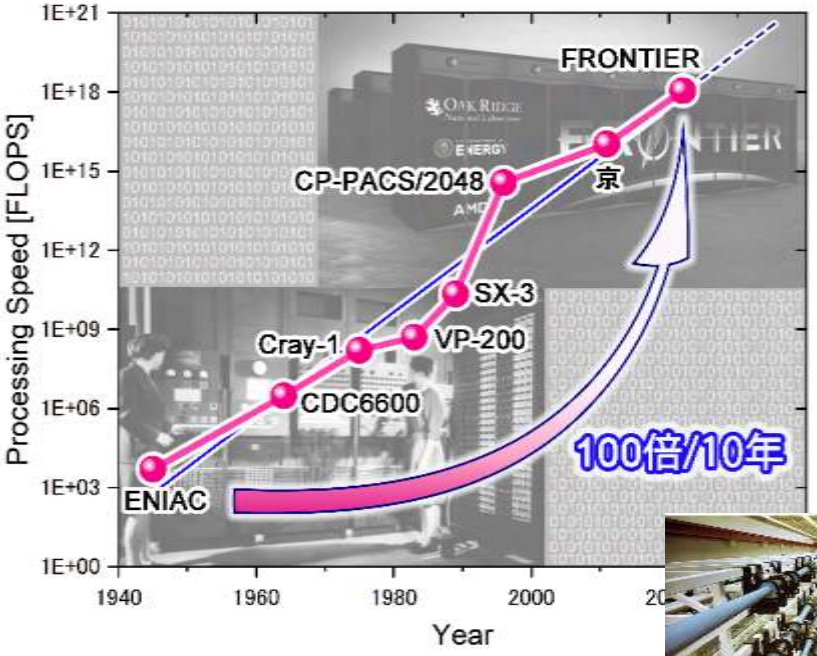
物質・材料分野における
オープンデータ・サイエンス

国立研究開発法人物質・材料研究機構
石井真史

エクサスケールの時代を迎えて
オープンデータの原動力

計算機の歴史
(1945→2022)

ENIAC (1945)
CDC6600 (1964)
CRAY 1 (1975)
VP-200 (1983)
SX-3 (1989)
CP-PACS/2048 (1996)
京 (2011)
富岳 (2020)
Frontier (2022)



膨大なデータ処理が可能に

- レーザー開発⇔レーザー応用科学の発展
- データが使える⇔データ駆動科学の発展

データ集中科学（データ駆動科学）
第四のパラダイム(2000年代)

第1:「経験科学」

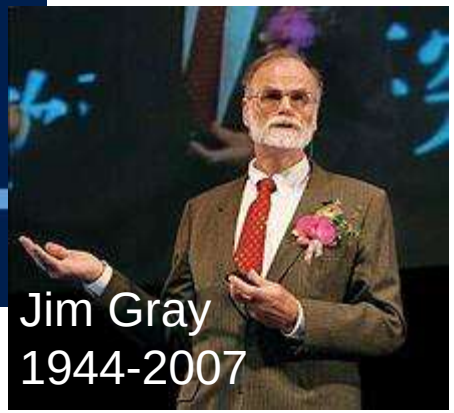
第2:「理論科学」

$$D = \epsilon_0 E + P$$
$$H = \mu_0^{-1} B - M$$

第3:「計算科学（シミュレーション）」

第4:「データ集中科学」
Jim Gray
第四のパラダイム

Tony Hey,Stewart Tansley,Kristin Tolle
Published by Microsoft Research, October 2009
http://research.microsoft.com/en-us/collaboration/fourthparadigm/



- ビッグデータの重要性
- オープンデータの利活用
- データマイニングによる新発見

国内のデータ利活用の取り組み
データ管理インフラの充実

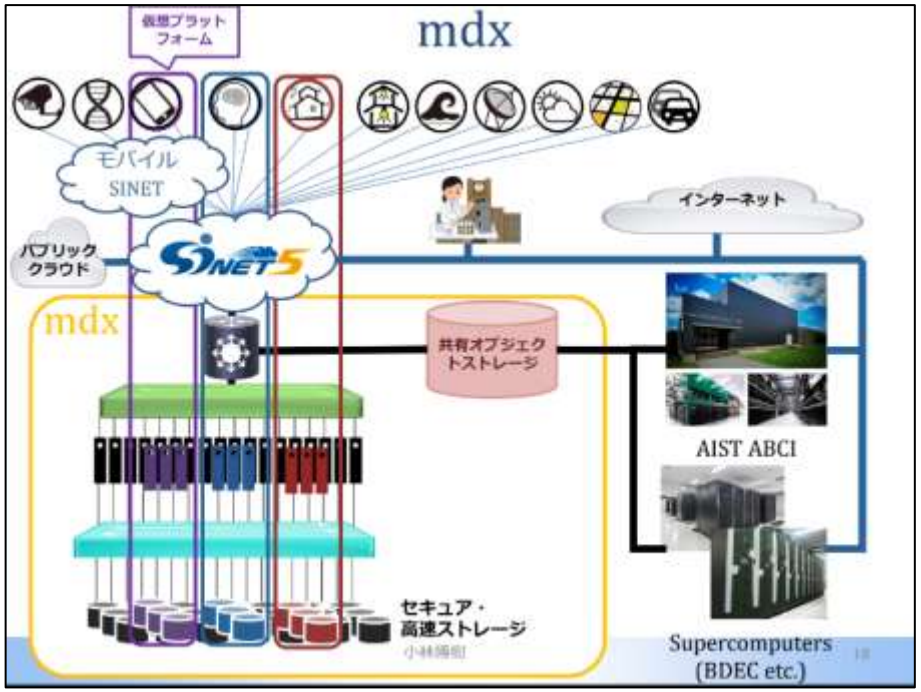
東京大学MDX

仮想プラットフォーム

- 仮想マシンとVPNを使った「疑似占有環境」
- 柔軟性とセキュリティ

「mdx:データ活用社会創成プラットフォームの紹介」

https://dss.csis.u-tokyo.ac.jp/documents/1640-kobayashi.pdf

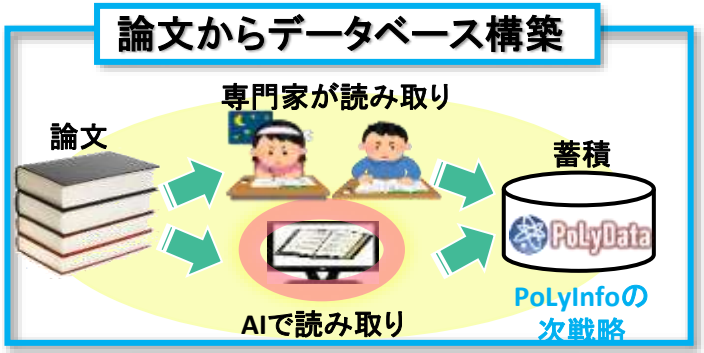


- 高度なシステム
- データポリシー

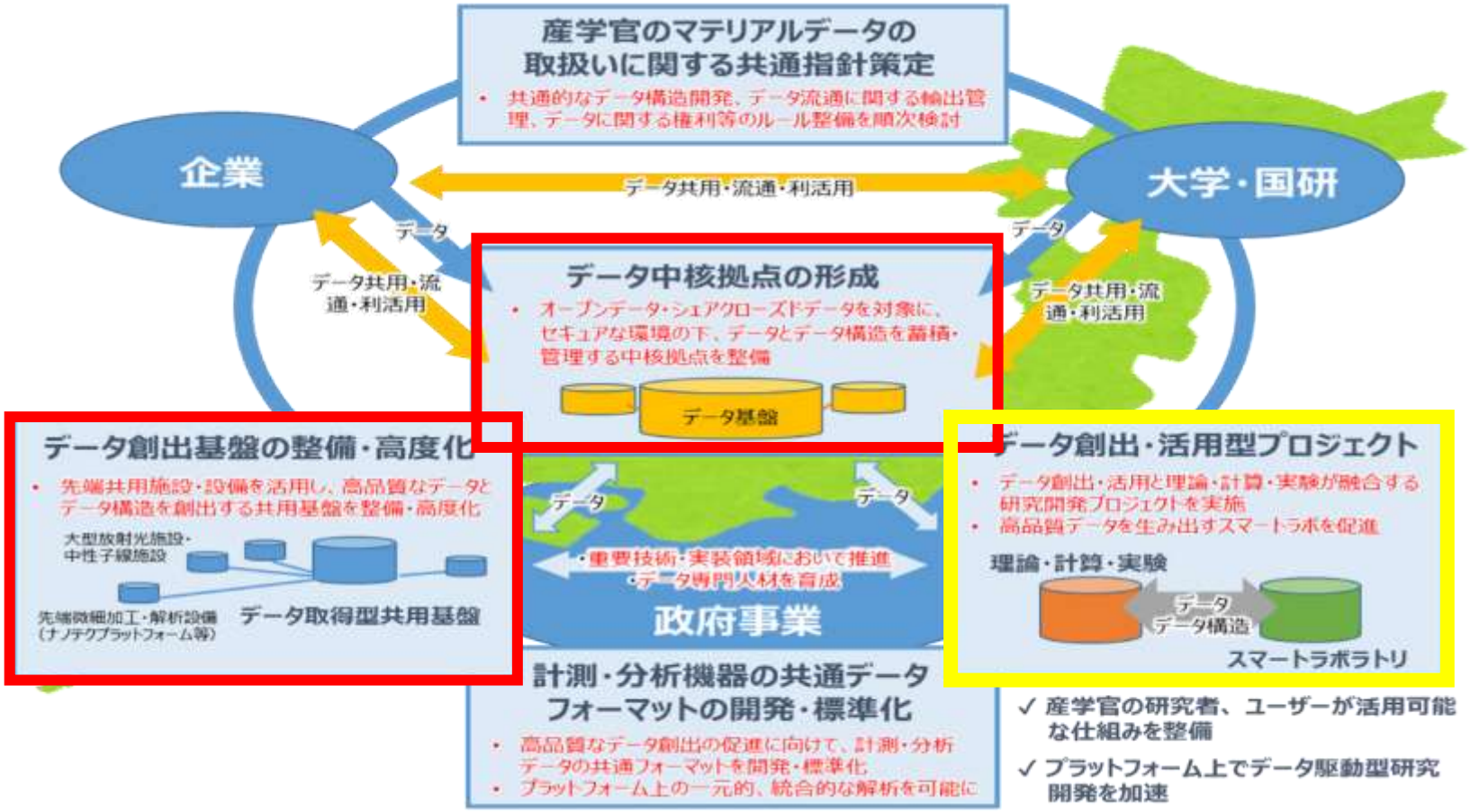
基盤の充実（データ管理に必要なインフラ）

NIMSの取り組み 2017~
Data platform (DPF1.0)

データをつくり、使い、公開する



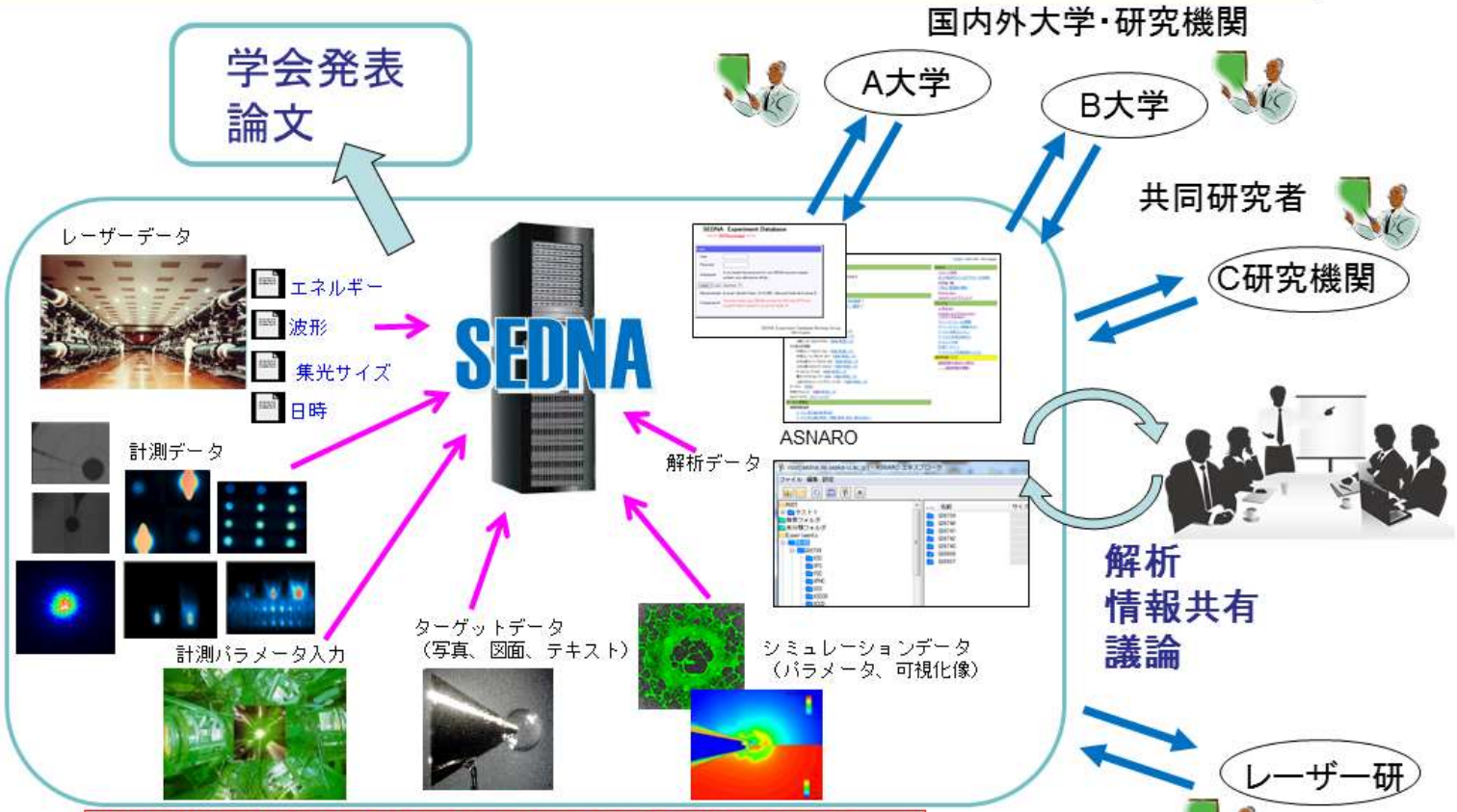
「マテリアルDXプラットフォーム」 2022～
次のデータクラウド DPF2.0



セキュアなデータ共有システム(SEDNA)

https://www.ile.osaka-u.ac.jp/sedna/

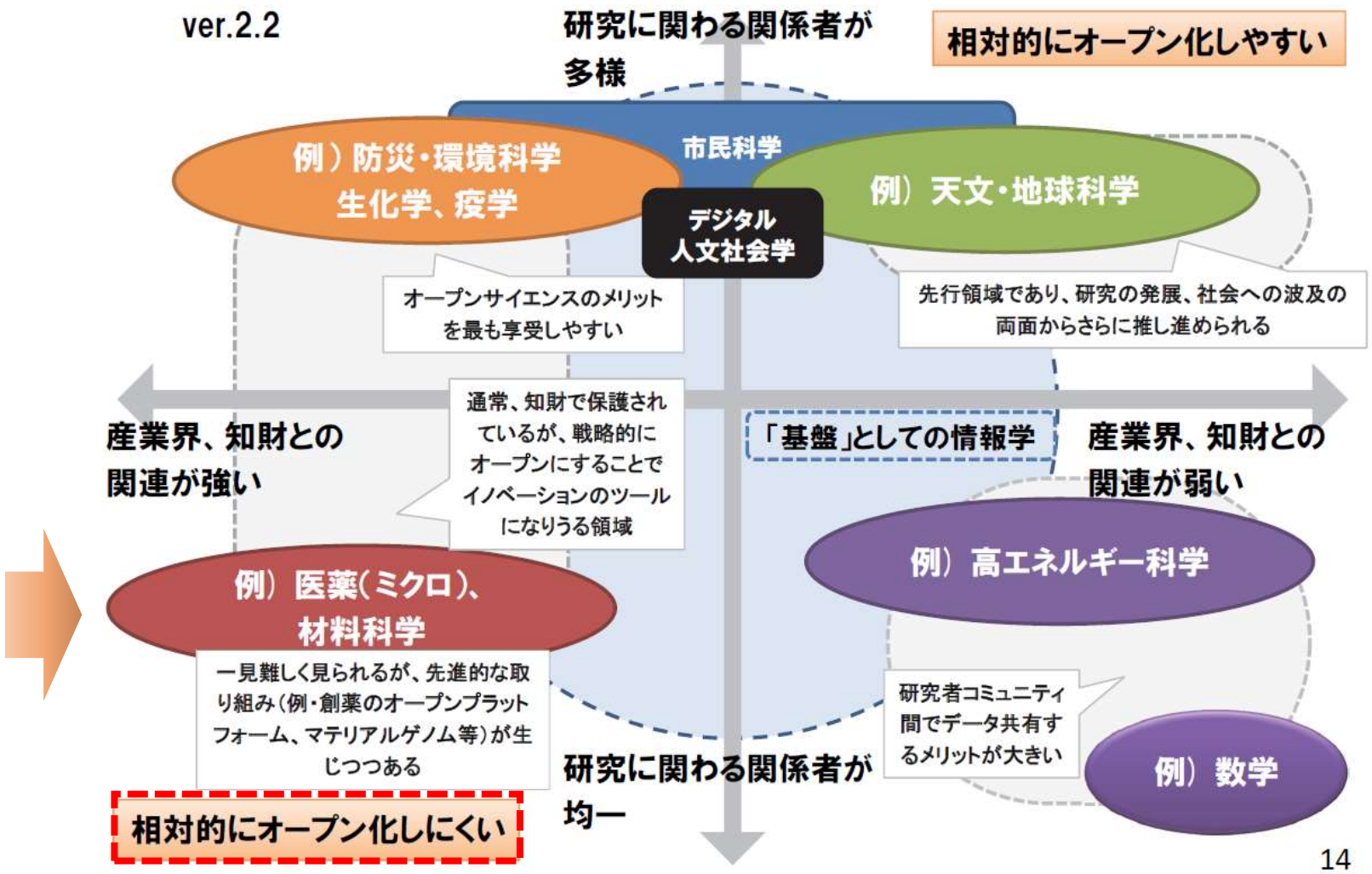
実験・レーザーデータベース(SEDNA)をハブとして複数の研究組織が参加したバーチャルな研究環境を提供している



様々な共同研究に対応するため、データアクセス権限を細かく設定できる。
データの不正改ざん防止のため、修正履歴をすべて記録できる。

オープンサイエンスに係る分野別相対マッピング

https://www.nistep.go.jp/wp/wp-content/uploads/review2018_presentation_6.pdf



オープンサイエンスの潮流

施設：

- データ共有リポジトリの増加とメタデータ設計

出版：

- オープンアクセス（OA）ジャーナルの増加

制度・要請：

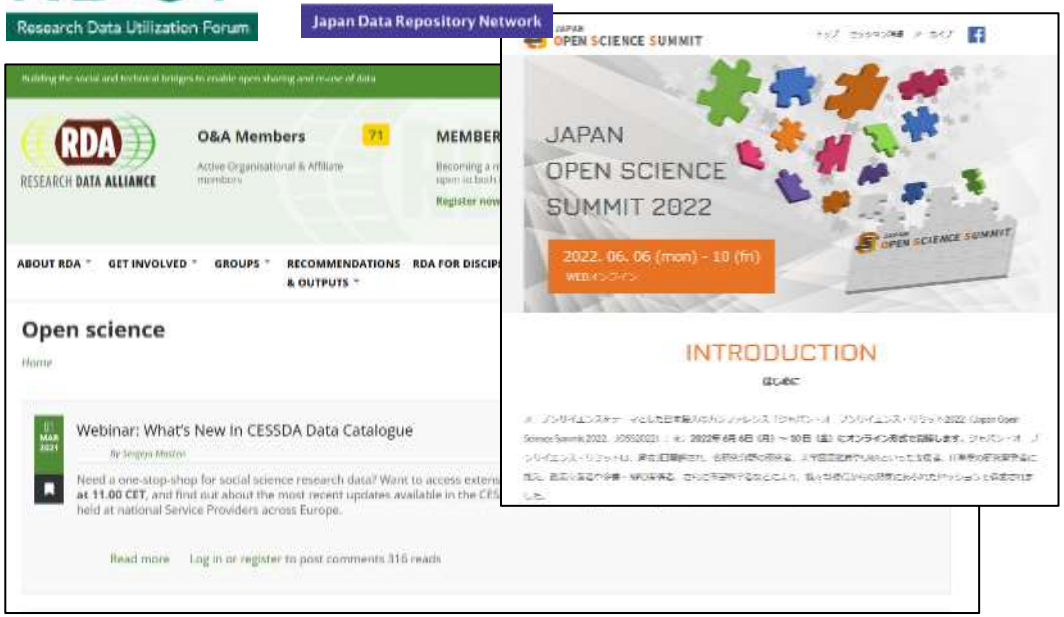
- データマネージメントプラン（DMP）の義務化
- 再現性や信頼性の公開

組織：

- CODATA（the Committee on Data for Science and Technology：科学技術データ委員会(1966)
- JaLC(Japan Link Center)

研究データ利活用協議会 RDUF Research Data Utilization Forum

ジャパン・データリポジトリ・ネットワーク JDARN Japan Data Repository Network



Japan Link Center

国立国会図書館 National Diet Library, Japan

NII 国立情報学研究所 National Institute of Informatics

JST 国立研究開発法人 科学技術振興機構 Japan Science and Technology Agency

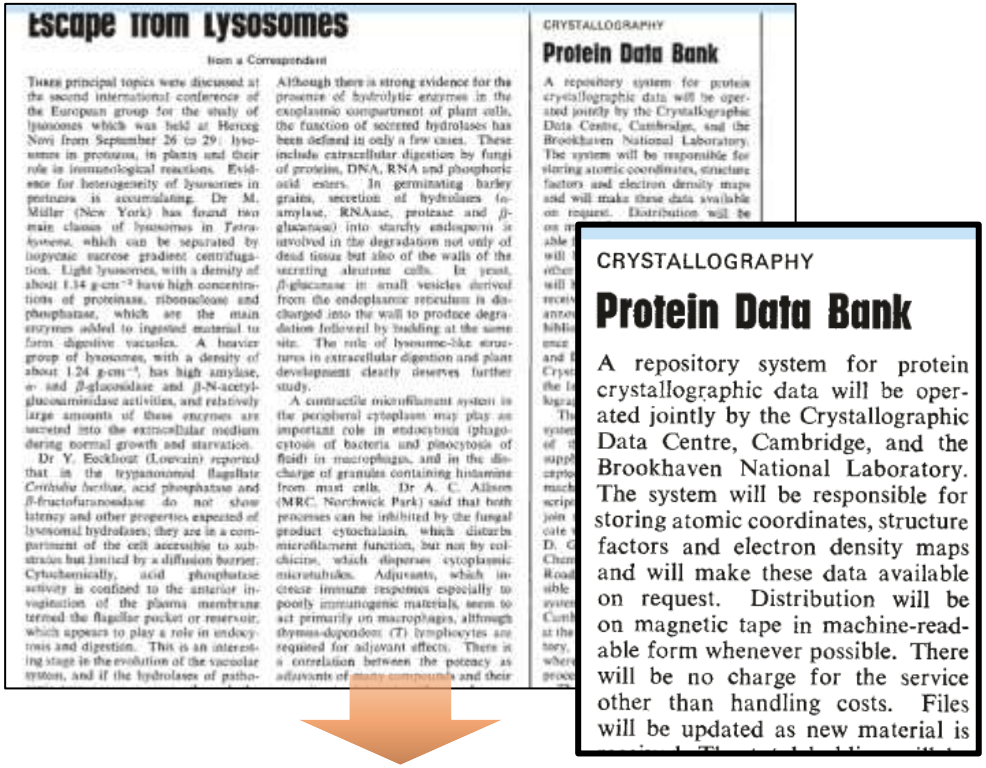
NIMS 国立研究開発法人 物質・材料研究機構 National Institute for Materials Science

オープンサイエンスの歴史ある取り組み

Open data repository since 1971

Selected Historical Publications (All Publications | Citation Information)

- Open-access data: A cornerstone for artificial intelligence approaches to protein structure prediction
  (2021) Structure 29: 515-520 doi: 10.1016/j.str.2021.04.010
- PDB 50th Anniversary: celebrating the future of structural biology
  (2021) Nature Collection
- Synergies between the Protein Data Bank and the community
  (2021) Nature Structural & Molecular Biology 28: 400–401 doi: 10.1038/s41594-021-00586-6
- wwPDB biocuration: on the front line of structural biology
  (2021) Nature Methods 18: 431–432 doi: 10.1038/s41592-021-01137-z
- How the Protein Data Bank changed biology
  (2021) Journal of Biological Chemistry Thematic Series
- Virtual Issue: The PDB in JBC
  (2021) Journal of Biological Chemistry (Issue | Structures)
- Creating a community resource for protein science
  (2012) Protein Science 21: 1587-1596 doi: 10.1002/pro.2154
- The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future
  (2012) Structure 20: 391-396 doi:10.1016/j.str.2012.01.010
- The evolution of the Protein Data Bank (part of special essay collection on the 40th anniversary of the PDB)
  (2011) Nature Structural & Molecular Biology 18: 1310 doi: 10.1038/nsmb1211-1304
- The Protein Data Bank: a historical perspective
  (2008) Acta Cryst. A64: 88-95. doi: 10.1107/S0108767307035623
- The PDB: A case study in management of community data
  (2004) Current Proteomics 1: 49-57. doi: 10.2174/1570164043488252
- Announcing the worldwide Protein Data Bank
  (2003) Nature Structural Biology, 10:. 980. doi: 10.1038/nsb1203-980
- The Protein Data Bank
  (2000) Nucleic Acids Research 28: 235-242. doi: 10.1093/nar/28.1.235
- **Protein Data Bank**
  (1971) Nature New Biology 233:223 doi: 10.1038/newbio233223b0



Escape from Lysosomes

from a Correspondent

Three principal topics were discussed at the second international conference of the European group for the study of lysosomes which was held at Herceg Novi from September 26 to 29: lysosomes in protozoa, in plants and their role in immunological reactions. Evidence for heterogeneity of lysosomes in protozoa is accumulating. Dr M. Müller (New York) has found two main classes of lysosomes in *Tetrahymena*, which can be separated by isopycnic sucrose gradient centrifugation. Light lysosomes, with a density of about 1.14 g·cm^{-3} have high concentrations of proteinase, ribonuclease and phosphatase, which are the main enzymes added to ingested material to form digestive vacuoles. A heavier group of lysosomes, with a density of about 1.24 g·cm^{-3}, has high amylase, α- and β-glucosidase and β-N-acetylglucosaminidase activities, and relatively large amounts of these enzymes are secreted into the extracellular medium during normal growth and starvation.

Dr Y. Eeckhout (Louvain) reported that in the trypanosomid flagellate *Crithidia luciae*, acid phosphatase and β-fructofuranosidase do not show latency and other properties expected of lysosomal hydrolases; they are in a compartment of the cell accessible to substrates but limited by a diffusion barrier. Cytochemically, acid phosphatase activity is confined to the anterior invagination of the plasma membrane termed the flagellar pocket or reservoir, which appears to play a role in endocytosis and digestion. This is an interesting stage in the evolution of the vacuolar system, and if the hydrolases of patho-

Although there is strong evidence for the presence of hydrolytic enzymes in the exoplasmic compartment of plant cells, the function of secreted hydrolases has been defined in only a few cases. These include extracellular digestion by fungi of proteins, DNA, RNA and phosphoric acid esters. In germinating barley grains, secretion of hydrolases (α-amylase, RNAase, protease and β-glucanase) into starchy endosperm is involved in the degradation not only of dead tissue but also of the walls of the secreting aleurone cells. In yeast, β-glucanase in small vesicles derived from the endoplasmic reticulum is discharged into the wall to produce degradation followed by budding at the same site. The role of lysosome-like structures in extracellular digestion and plant development clearly deserves further study.

A contractile microfilament system in the peripheral cytoplasm may play an important role in endocytosis (phagocytosis of bacteria and pinocytosis of fluid) in macrophages, and in the discharge of granules containing histamine from mast cells. Dr A. C. Allison (MRC, Northwick Park) said that both processes can be inhibited by the fungal product cytochalasin, which disturbs microfilament function, but not by colchicine, which disperses cytoplasmic microtubules. Adjuvants, which increase immune responses especially to poorly immunogenic materials, seem to act primarily on macrophages, although thymus-dependent (T) lymphocytes are required for adjuvant effects. There is a correlation between the potency as adjuvants of many compounds and their

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is

- バイオロジーに見られるデータ共有の源泉
- 50年のオープンサイエンスの歴史

https://www.rcsb.org/pages/about-us/history

Biologyにおけるデータ管理と国際的共有

Protein database (PDB)

生体高分子の3次元構造の原子座標を蓄積

2003年 Worldwide Protein Data Bank (wwPDB)発足

完全に同一のデータを全世界で公開

- 日本：日本蛋白質構造データバンク (PDBj)
  https://pdbj.org/
- 米国：構造バイオインフォマティクス研究共同体 (RCSB PDB)
- 英国：欧州蛋白質構造データバンク (PDBe)
- 米国：生体分子磁気共鳴データバンク（BMRB）

2018年　データ捏造事件をきっかけに検証レポートを導入
https://pdbj.org/news/20180510

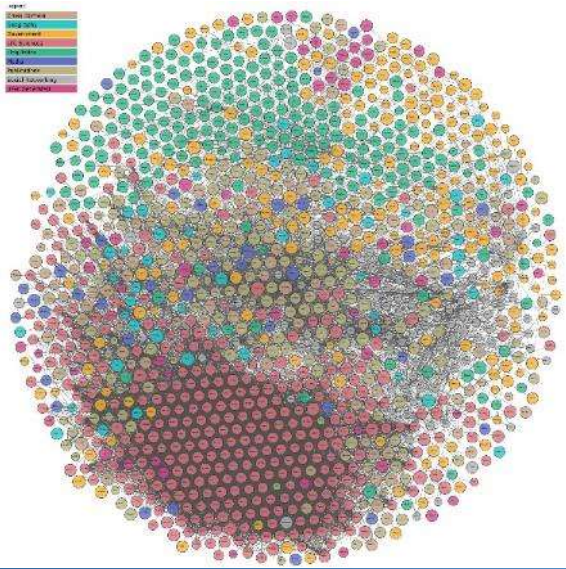
https://pdbj.org/

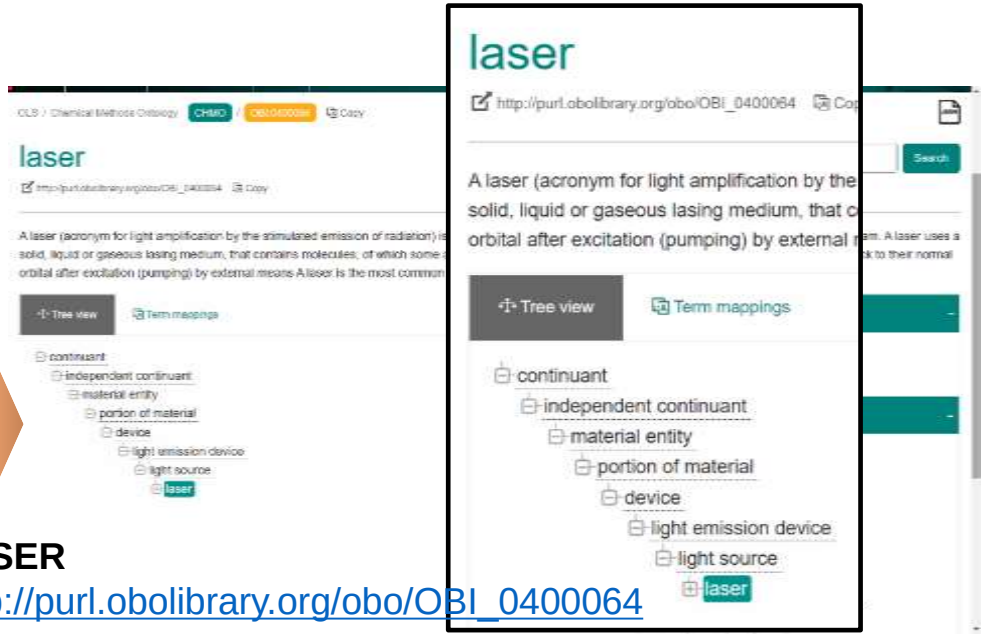
- PDBの登録番号と検証レポートが論文投稿に必須
- 検証レポートが登録データの品質向上

Biology周辺のオープンデータ技術
物質・材料との連携と協調発展のターゲット

LOD（Linked Open Data）技術

- セマンティックウェブ技術
- 機械可読メタデータ
- タキソノミー（分類学）
- オントロジー





LASER
http://purl.obolibrary.org/obo/OBI_0400064

- Resource Description Framework (RDF) によるデータベース間の接続
- データの階層構造化による類似・近接データへのアクセス
- 機械可読な学術体系の記述

The Linked Open Data Cloud from lod-cloud.net, https://lod-cloud.net/

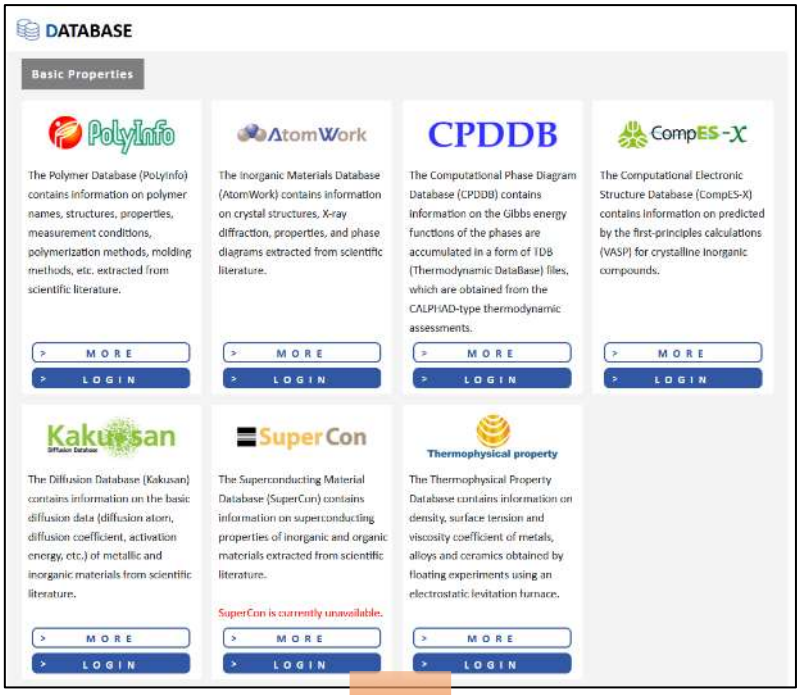
NIMS MatNavi
NIMS　材料データベース群

MatNavi

https://mits.nims.go.jp/

- 高分子データベース
- 無機材料データベース
- 金属データベース
- 計算状態図データベース
- 電子状態データベース
- 拡散データベース
- 超伝導データベース
- 高温熱物性データベース

他データシートなども



- **論文からのデータ収集 >20年**
- **NIMS一機関のクローズ運用**

NIMS MDR (Materials Data Repository)
NIMSにおける材料オープンサイエンスの試み

材料研究データや文献等を保存・公開しているオープンデータリポジトリ

論文や学会発表資料＋サポートデータ

データそのものを公開

マテリアルズ・インフォマティクスなどで利活用

メタデータ項目検索

閲覧・ダウンロード

認証なし

Publications　Datasets　Collections

https://mdr.nims.go.jp/

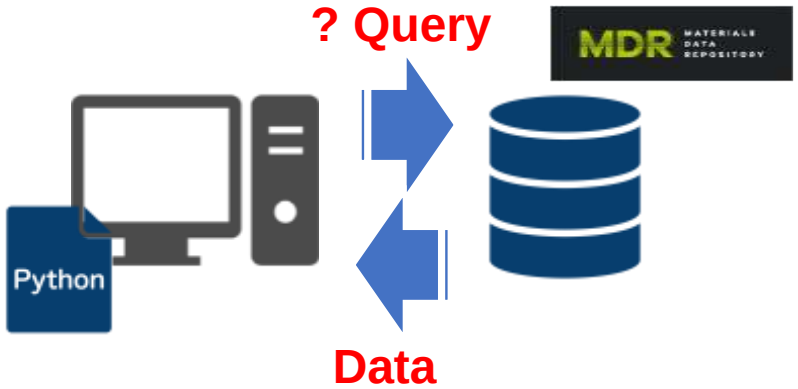
DOIをつけて
オープンアクセス

MDR XAFS DB
NIMS MDRの機能を最大限活用

- 人可読
  GUIベース
  https://mdr.nims.go.jp/collections/qz20st57x?locale=en

- 機械可読
  APIベース
  Application Programming Interface (API)
  - メタデータの取得
  - 全収録データの取得

X線吸収分光スペクトルデータの共有基盤

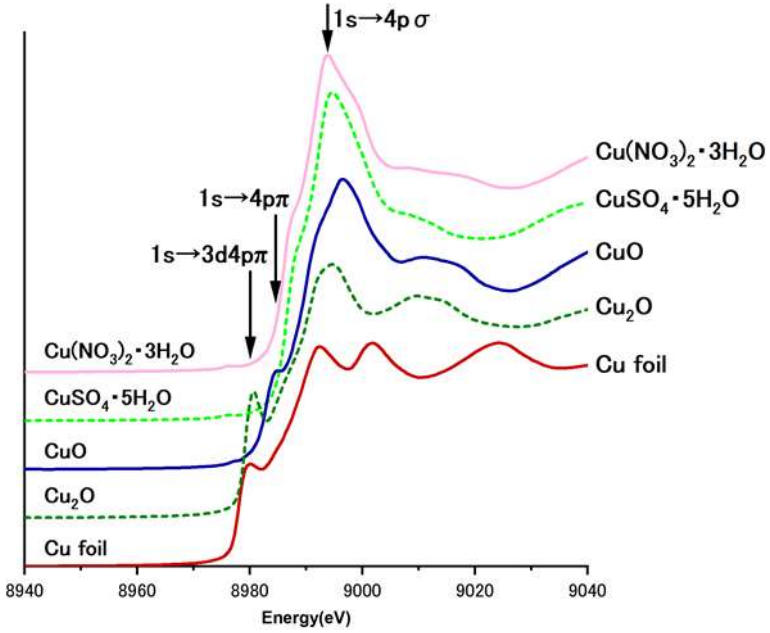


みんなで作り、みんなで使うデータベース

データ共有とXAFSの相性
需要のあるユースケース　X線吸収分光 XAFS

XAFSの場合

- 分野間の差が小さい＝「原子の周辺」の誰もが共有できるサイズ感
- スペクトル単位で再利用できる＝表記が単純でデータ統合しやすい
- 比較・参照が解析精度を向上＝他の研究者の解釈の参照→新しい知見
- 特殊性が少ない（適用性が高い）＝突出したデータは発展性が乏しい



データを登録するのは手段＝登録した後の使い道の設定が必要

データ統合による研究活用のストーリー

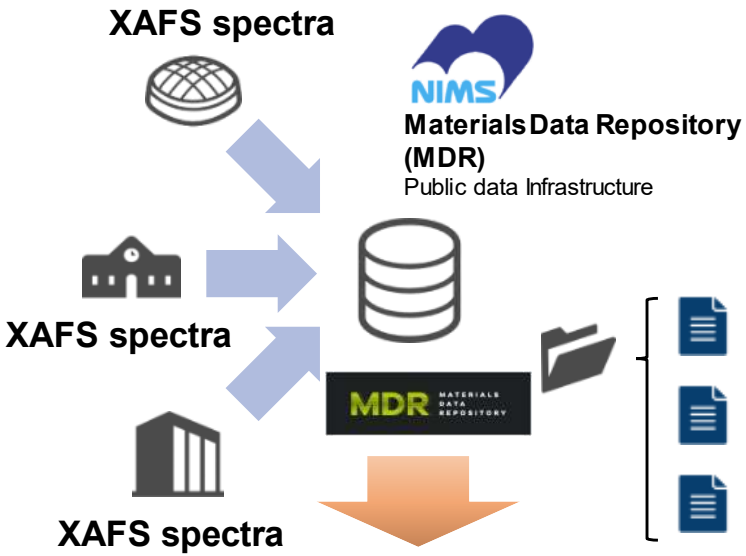
https://mdr.nims.go.jp/collections/qz20st57x?locale=en
https://doi.org/10.48505/nims.1447

MDR XAFS DB
多機関連携型データベース

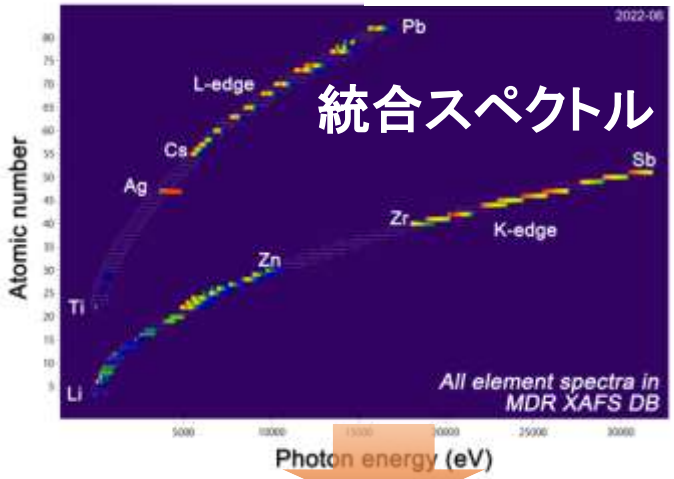
多機関からデータを持ち寄り一つの大きなデータベースを作る

SPring-8 立命館 北大 KEK…

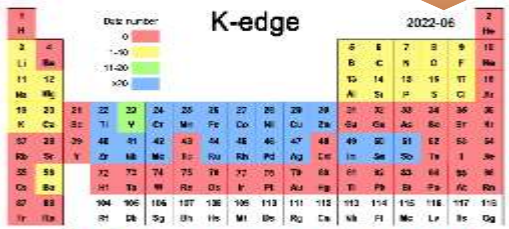
軽元素から重元素まで
（Li から Pbまで, 2022/6）

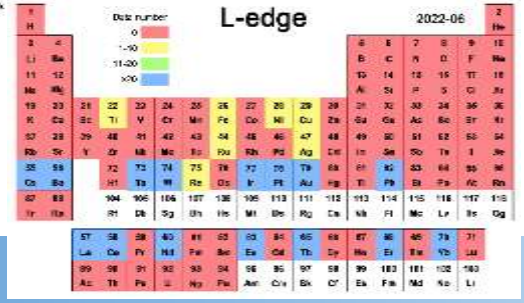


スペクトルデータの共有



2022-06

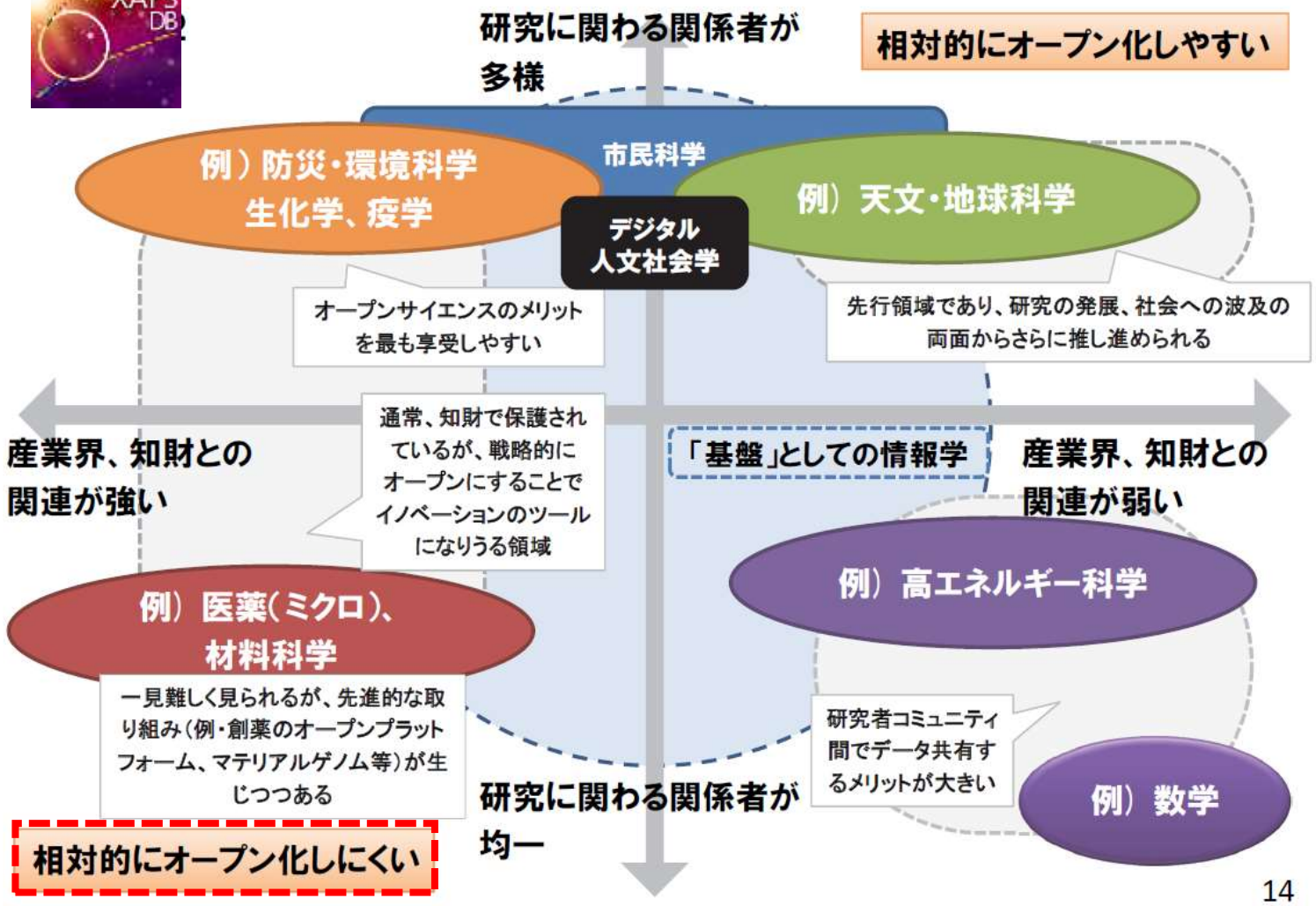




オープンサイエンスに係る分野別相対マッピング

https://www.nistep.go.jp/wp/wp-content/uploads/review2018_presentation_6.pdf

材料に特異な
オープンデータ
コア



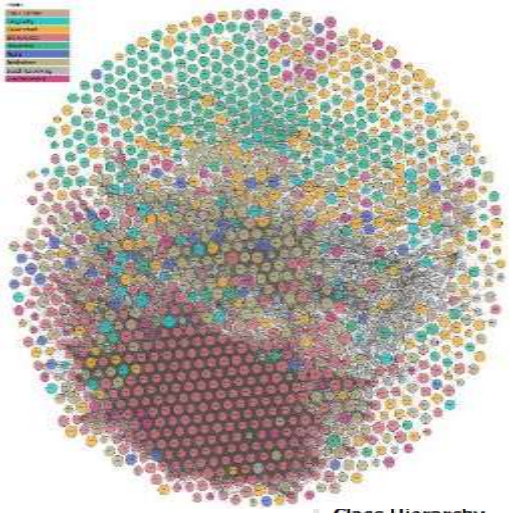
材料データ連携スキーマ
材料におけるセマンティックウェブ技術

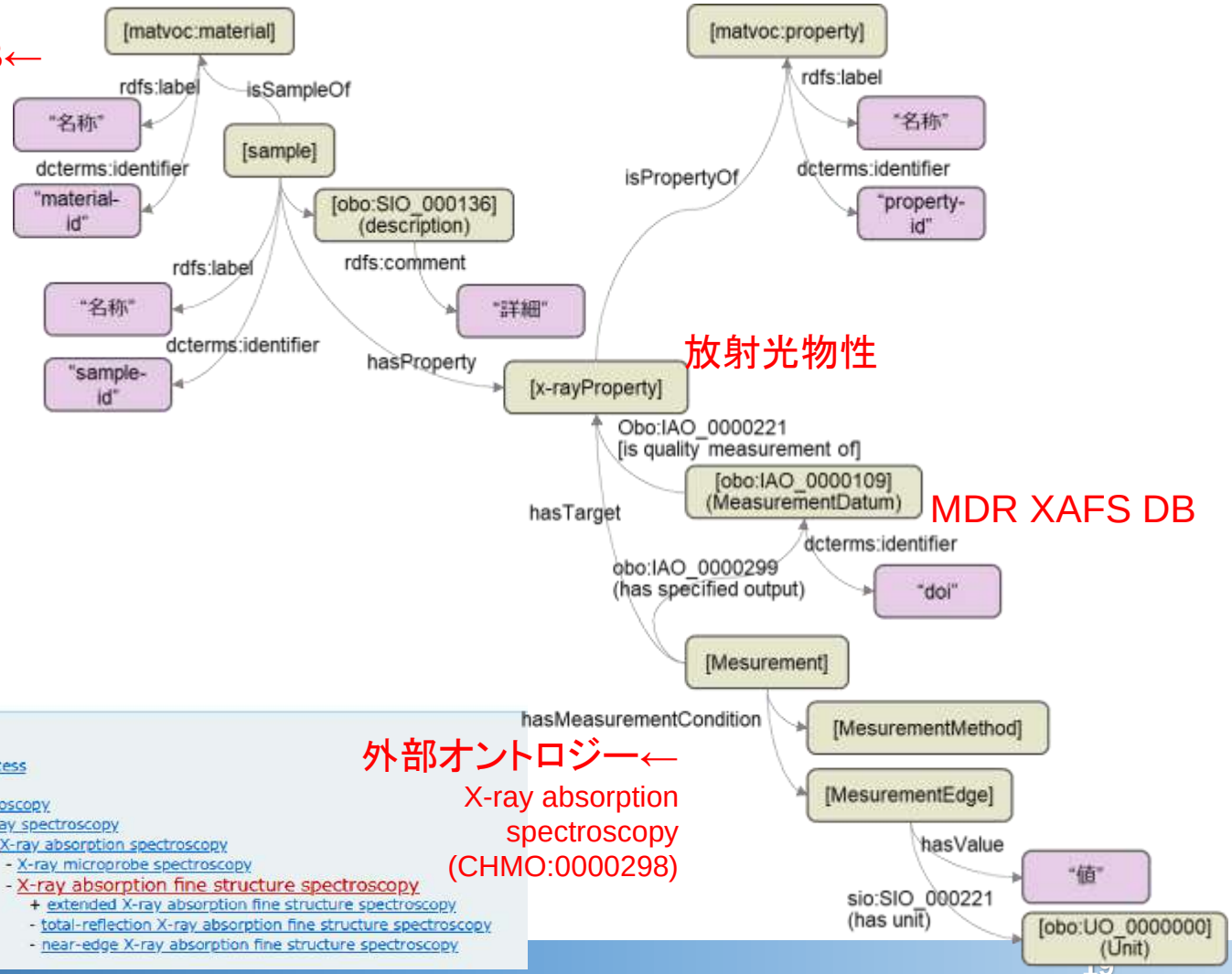
材料辞書

物性辞書

NIMS DB←
MatNavi

The Linked Open Data Cloud from lod-cloud.net, https://lod-cloud.net/





Class Hierarchy

Thing
+ process
+ planned process
+ assay
+ spectroscopy
+ X-ray spectroscopy
+ X-ray absorption spectroscopy
- X-ray microprobe spectroscopy
- X-ray absorption fine structure spectroscopy
+ extended X-ray absorption fine structure spectroscopy
- total-reflection X-ray absorption fine structure spectroscopy
- near-edge X-ray absorption fine structure spectroscopy

外部オントロジー←
X-ray absorption spectroscopy
(CHMO:0000298)

材料のオープン化の試み

- インフラ構築の急進展
- バイオ関係の先端的取り組み
- 権利・知財/競争
  ⇔オープンサイエンス
- 協調のメリットが多いコアの設定
  - MDR XAFS DB
- セマンティックウェブ技術による連携拡張

物質・材料分野における オープンデータ・サイエンス

国立研究開発法人物質・材料研究機構
石井真史