

全国規模の材料データプラットフォームにおけるAI ベース検索システムの構築

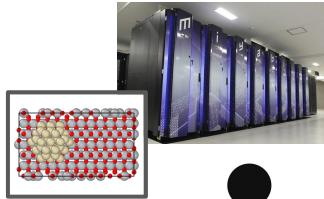
東京大学 情報基盤センター 華井雅俊

Outline

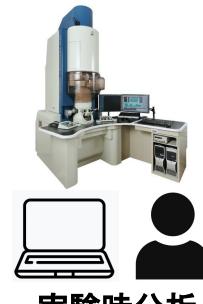
- 背景: 材料科学におけるデータ駆動型研究
- 課題: 材料データの公開と1次データの効率的収集
- 提案
 - ARIM-mdx Data System: Towards a Nationwide Data Platform for Materials Science
 - ARIM-mdxにおける材料実験データ収集
 - ARIM-mdx Sample Submission System
 - RxT 非ネットワーク機器向けデータ転送 IoT デバイス
- まとめ

背景: 材料科学におけるデータ駆動型研究

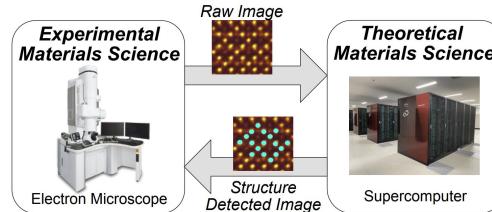
- **近年の研究手法の高度化により、多様な専門家が研究チームを形成**
 - 理論解析, 実験計測, 合成・製造・プロセス, etc.
- **大型実験施設・スパコン・クラウドをフル活用したデータ駆動型研究**
 - 大規模な理論シミュレーションとデータセット化・機械学習モデルの構築
 - 実験中のインタラクティブな分析
 - シミュレーションデータと実験データの融合分析
 - 自動実験における常時モニタリングとフィードバック



シミュレーション +
機械学習



実験時分析



融合分析



自動実験

課題: 材料データの公開と1次データの効率的収集

- 既存の大規模データプロジェクトなどでは、
整理されたデータの公開が主な目的
(論文・特許発表データや処理済2次データの公開)



- 国内では研究データ公開のルール化が進む
 - 科研費でのデータ公開の義務化
https://www.jsps.go.jp/j-grantsinaid/01_seido/10_datamanagement/index.html
 - ARIM 事業での材料実験データカタログ化 <https://nanonet.go.jp/>

- データ公開にいたるまでの **一次データの収集に課題**
 - どうやって大規模収集をするのか？どこに集めて管理するのか？
 - “頑張って集める”ではなくて “自然に集まる”サービスにしたい
 - 材料研究に典型的な多様な専門家チームにおいてデータは複雑にサイロ化
(理論研究者・スペコン・実験研究者・実験装置、等々)

ARIM-mdx Data System:

Towards a Nationwide Data Platform for Materials Science

M. Hanai, R. Ishikawa, M. Kawamura, M. Ohnishi, N. Takenaka, K. Nakamura, D. Matsumura, S. Fujikawa, H. Sakamoto, Y. Ochiai, T. Okane, S. Kuroki, A. Yamada, T. Suzumura, J. Shiomi, K. Taura, Y. Mita, N. Shibata, Y. Ikuhara
“**ARIM-mdx Data System: Towards a Nationwide Data Platform for Materials Science**” in IEEE BigData 2024

ARIM-mdx Data System



全国の材料実験装置とスーパーコンピュータをつなぐ大規模材料データインフラ

- 国内外これまで個々の研究施設・組織で実施された実験・計算融合インフラ⁴⁾から全国規模のインフラへ
- AI for Scienceに向け、材料研究向けデータインフラの高度化 + 大規模化を目指す

令和5年8月から一般利用を開始し、2年間で 現在1400以上の研究者ユーザーが利用 約1PBの材料データ(計算・実験)を保存

- mdx上に構築、各スパコン間にSINET 6で高速データ転送
- 50台以上の実験装置をIoTで接続し実験データを効率収集
- 東大、広大、SPring-8 (JAEA), 北大、電通大で利用、
九大、名大、産総研、NanoTerasu (東大物性研) で準備中

材料研究向けに誰でも無料で使えます

- <https://arim.mdx.jp>よりプロジェクトを登録
- 一口 30TBからの大容量クラウドストレージ+ Jupyterサービス

1) The Workshop on Extreme-Scale Experiment-in-the-Loop Computing in SC (XLOOP) <https://wordpress.cels.anl.gov/xloop-2025/>

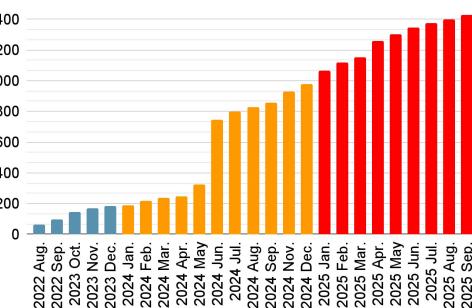
2) Argonne National Laboratory: スパコン×X線分析装置連携 (2019) "Balsam: Near Real-Time Experimental Data Analysis on Supercomputers" (XLOOP'19)

3) Lawrence Berkeley National Lab (2020) "Cross-facility science with the Superfacility Project at LBNL" (XLOOP'20)

4) 理研: 富岳 × SPring-8 連携 (2023) https://www.riken.jp/pr/closeup/2023/20231225_1/index.html

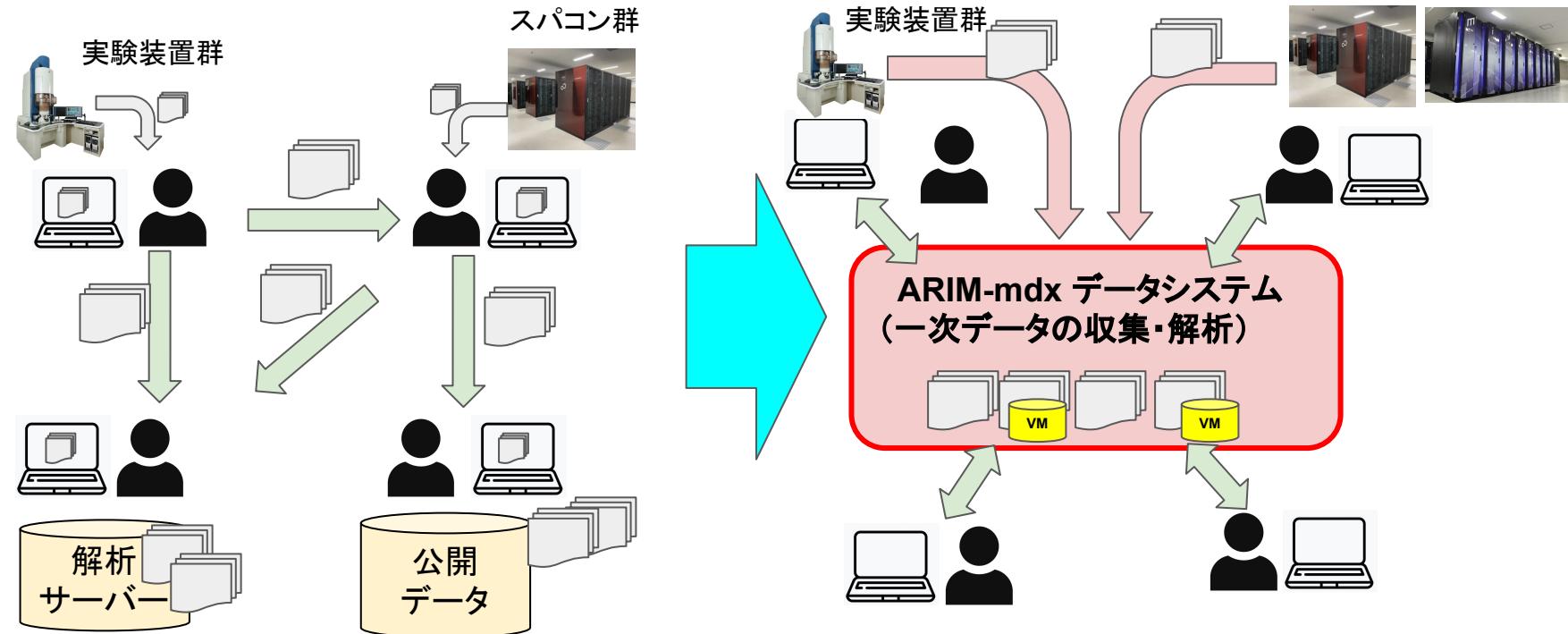


of Users in ARIM-mdx Data System



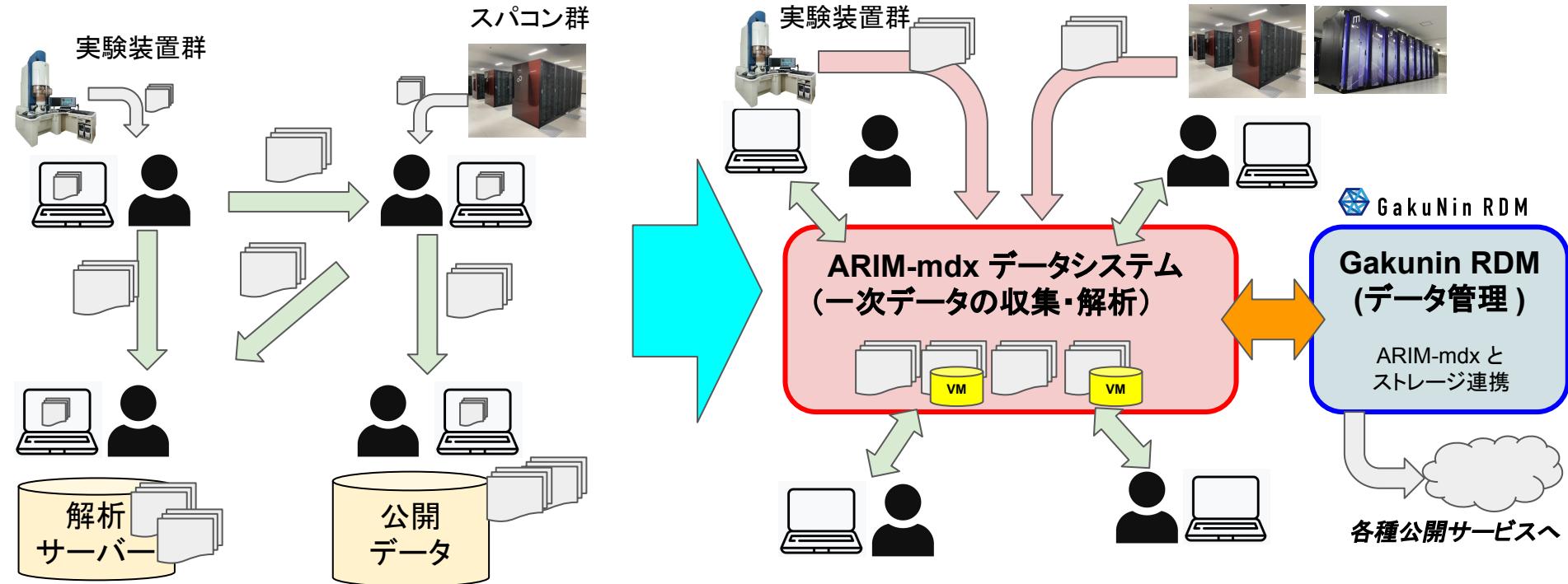
Key Concepts

- データ大規模化・メンバー増加によりデータがサイロ化、**基本的な操作もたちまち困難**に
 - 収集、蓄積、転送、共有、解析、etc
- **データをNextcloud (mdx I) に一元集約。Jupyter経由で解析サーバーを起動**



Key Concepts

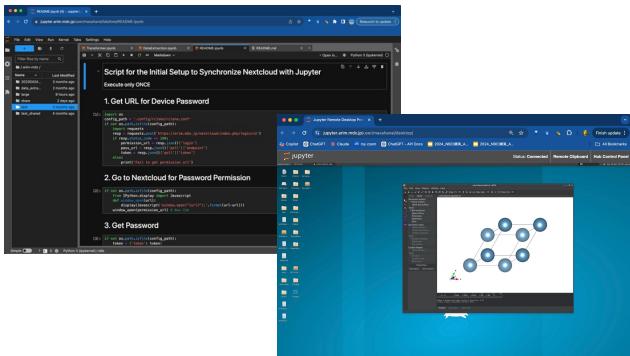
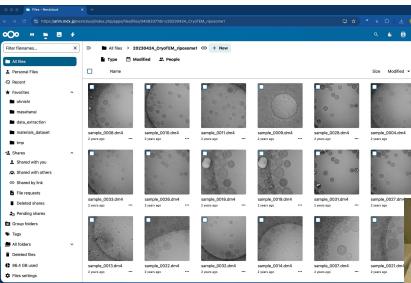
- データ大規模化・メンバー増加によりデータがサイロ化、**基本的な操作もたちまち困難**に
 - 収集、蓄積、転送、共有、解析、etc
- **データをNextcloud (mdx I) に一元集約。Jupyter経由で解析サーバーを起動**



提供のサービス

3つの基本機能をまずは提供
材料分野への特化 + 利便性 + 高性能 に注力

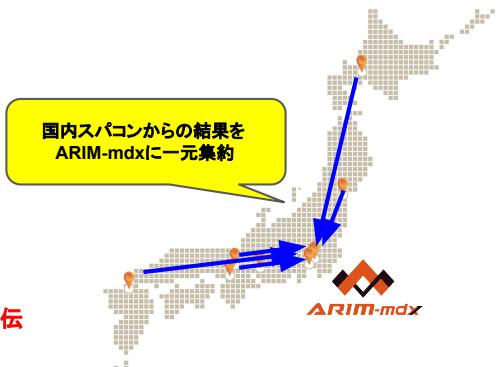
- **データ保存**
 - 合計3PBの高速クラウドストレージ (Nextcloud)
 - 各グループ毎に一口 30TBで無料提供
 - スパコン高速転送用のチューニング や
材料データ用に改造 (装置データの表示等)
 - <https://arim.mdx.jp/nextcloud>
- **データ収集**
 - IoTデバイスによる実験装置データの効率収集
 - <https://arim.mdx.jp/docs/ja/iot.html>
 - AIによる試料情報の詳細な収集
- **データ解析**
 - Jupyterサービス
 - MateriApps LIVE! で材料研究ツール群をAll-in-oneで提供
 - <https://jupyter.arim.mdx.jp>



Use Cases

- ARIM (文部科学省 マテリアル先端リサーチインフラ) における実験データ収集

- 実験設備の共用貸出事業におけるデータインフラとして利用
- 共用実験装置にIoTを設置し、全ユーザーの実験データをクラウドで管理
- NIMSでの材料データ公開システム（RDE）への効率的なデータ提供

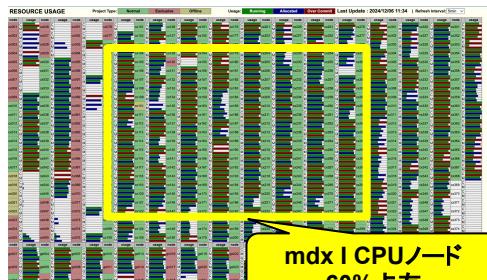


- 無機材料における熱伝導物性の DFTハイスループット計算

- M. Ohnishi, M. Hanai, et al. "Database and deep-learning scalability of anharmonic phonon properties by automated brute-force first-principles calculations" Arxiv 2025 <https://arxiv.org/abs/2504.21245>
- 既存オープンデータ 100材料 (Matbench) から 6000材料 ヘスケールへ
 - <https://matbench-discovery.materialsproject.org/>
 - DFT計算のベンチマーク (Meta / MS / DeepMind等も参加)
- 国内13のスパコン (HPCI + JHPCN) 合わせて **1億CPUコア時間以上、20,000材料をDFT計算し6,000材料の熱伝導データを生成** (圧縮後5TB程度)
- 各スパコンにサイロ化された計算結果を ARIM-mdxに集約、ARIM-mdx上でデータ分析・公開を実施

- 電池開発向け電解液物性の MDハイスループット計算および実験データ融合

- 中村公、華井雅俊、他 “ハイスループット分子動力学計算による電解液データベースの開発” 電池討論会 2025
- DxMT / DxGEM: 文部科学省 データ創出・活用型マテリアル研究開発プロジェクト（再生可能エネルギー最大導入に向けた電気化学材料研究拠点）での利用
 - <https://www.dx-gem.t.u-tokyo.ac.jp/wp/>
- **40万ノード時間で 9000 compositions を計算。400TB以上の計算結果**
 - mdx CPUノードの約60%を使った大規模計算実施
 - 300VM (46800 core) × 2週間 ~ 占有
- 実験データと合わせて ARIM-mdxにて収集



- 自動実験システム連携

- Digital Laboratory Consortium の自動実験システム @東大・東京科学大 (一杉研) との連携
- 自動実験のデータを ARIM-mdxに自動収集
- <https://digital-laboratory.jp/index.html>



ARIM-mdxにおける材料実験データの収集

材料実験データの収集の目的と課題

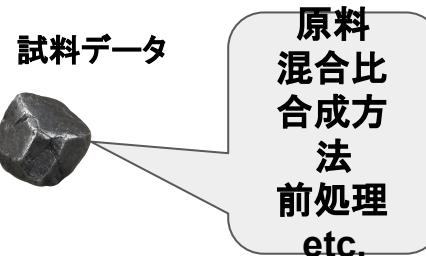
- ARIM-mdxにおいて、**実験データと計算データの一元的な収集を実現**
 - 計算データは中間データ・メタデータ含め詳細な収集が比較的容易(=転送するだけ)
 - **一方、実験データの効率的かつ詳細な収集は困難**
- ARIM (文部科学省 マテリアル先端リサーチインフラ) 事業においては、
共用実験装置の利用時に実験データの提供をユーザーにお願い している。
- **実験データは大きく試料データと装置データ**
 - 試料データ: 各研究者がもつ材料の情報・メタデータ
 - 作成・合成方法、加工方法、保存方法、購入経路、等々
 - **研究者しか知り得ない情報や構造化・テンプレート化の難しい詳細情報をどう収集するか？**
 - 装置データ: 計測装置 (電子顕微鏡など) での結果データや加工装置の利用ログ
 - **研究アクティビティ (実験 => データ分析 => 論文発表・社会実装) の中にいかにデータ収集を織り込むか？**



ARIM-mdxにおける材料実験データの収集

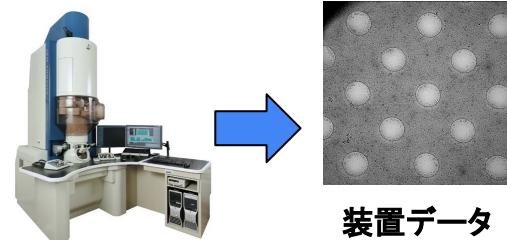
- 試料データ: *ARIM-mdx Sample Submission System*

- フォーマットフリーなテキストによる自由記述
 - 実験ノートや論文ドラフトのコピペで ok
 - これまで、Excelによる項目化やテンプレート化を実施した反省から
- LLMによる記述内容の評価システム
 - 特化分野での記述内容の妥当性やクオリティを AIで確認
 - 表記のズレ、英語 ⇄ 日本語、等を吸収



- 装置データ: *RxT データ転送デバイス*

- 実験装置 (スタンドアロン運用) からのデータ取り出しに着目
- USB DiskやDVDで取り出していた実験装置データを IoTによりリモート化



試料データ

ARIM-mdx Sample Submission System

- これまでの方法：項目の設定による Table化 (i.e., Excel ファイル)
 - ルール決めによって必要最低限の情報は収集可能
 - 詳細記述にむけた問題点
 - Scalability: 項目を増やして詳細さを追求すると頭打ち
 - 条件分岐の複雑化 (電池材料 => A, B, C 項目, 有機材料 => X, Y, Z 項目)
 - 真面目にやるとオントロジーが必要
 - 空項目だらけになる + 記入コスト増大
 - 材料全てに精通する専門家はいないので、得意な材料とそれ以外で項目の充実度がばらつく
 - Evaluation: 記述内容のクオリティーを一元的に評価出来ない
 - "足りないのでもう少し書いてください"をいいたい
 - 専門分野や目的が少し違うだけで注力する項目が大きく異なる
 - 表記のブレ、英語 ⇄ 日本語、専門用語の使い方、単位の違い、etc.

Tableによる最低限情報の収集

Autodesk

Aptos Narrow (Body)

Home Insert Draw Page Layout Formulas Data Review View Automate

Comments Share

D11

Cells Cell Styles Cells Editing Add-Ins Analyze Data

	A	B	C	D	E	F
1 試料名		リチウム、ニッケル、マンガン、コバルト酸化物、混合物				
2 試料分類		電池材料				
3 化学式		LNMO				
4 原料	Li2CO3	NiO	MnO2			
5 合成方法		電気炉による焼成				
6 混合比		Li: Ni: Mn = 1:0.5:1.5				
7 その他						
8						
9						
10						
11						
12						
13						
14						
15						
16						



図 1 「粒子」「粒径」の定義

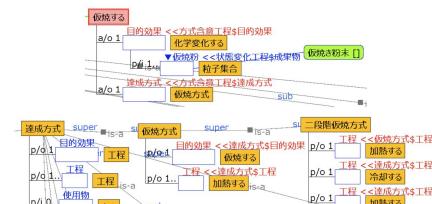
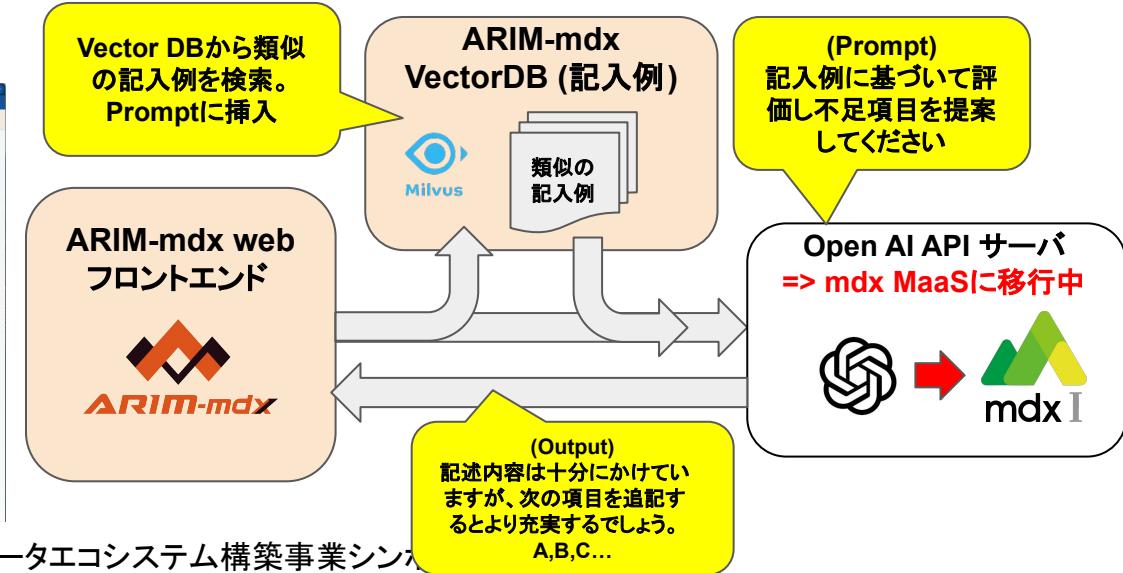
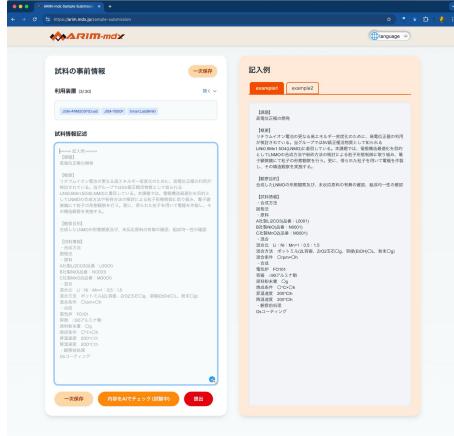
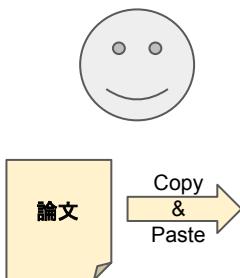


図 11 「仮焼する」と「仮焼方式」の定義

ARIM-mdx Sample Submission System

- テキスト入力によるシンプルな入力
 - フリーフォーマット(箇条書き、文章、英語 ⇄ 日本語、なんでもあり)
 - 論文・実験ノート・特許のコピペ、etc.
- LLM による評価システム
 - 記入例と照らし合わせて情報の充実度を評価 => 足りない情報の提案
 - 記入例を増やしていくことによるScalableな性能向上
 - 金属材料 => セラミックス => 有機材料 => etc.
 - 電池材料 => 半導体 => 生体材料 => etc.
 - 良くかけているユーザーの情報を記入例として増やしていく
- バックエンド API の mx 化を実施中
 - 機密データの利用などでメリット

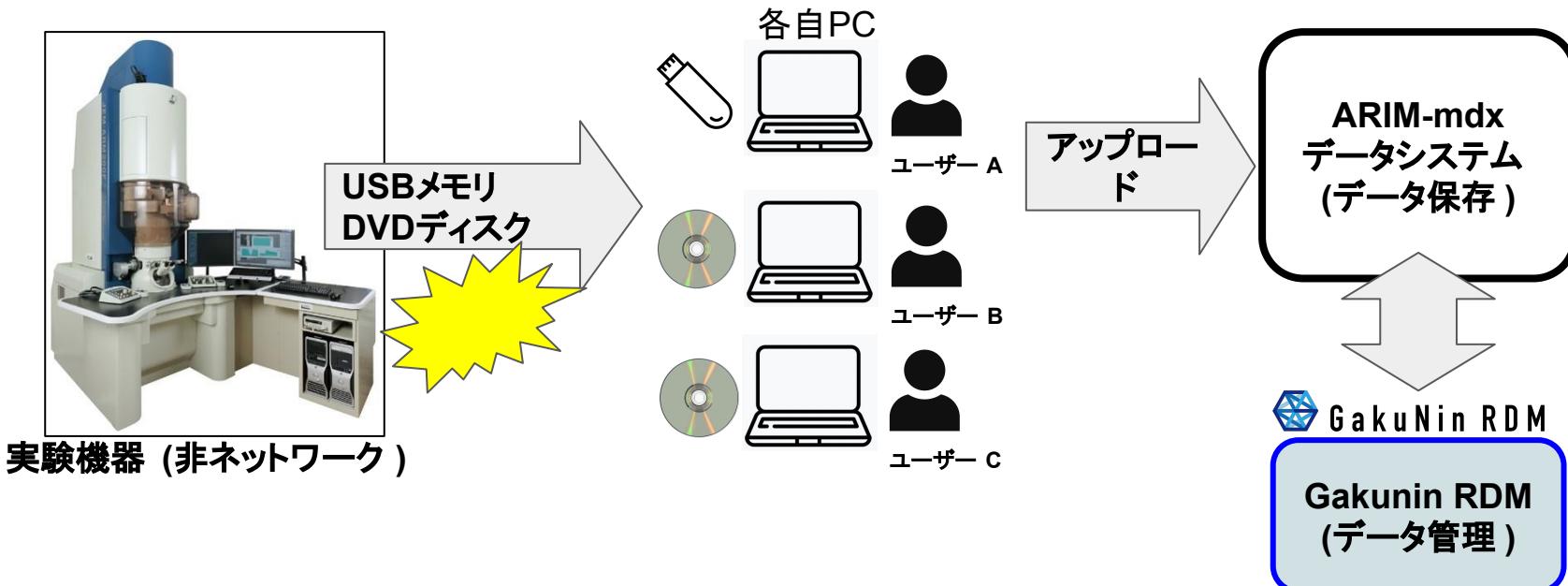


装置データ

RxT データ転送 IoTデバイス

課題: 非ネットワーク機器・共用 PCからのデータ取り出し

- USBメモリ・DVDディスクを用いた人手でのデータ取り出し
- 共用PCでのログアウト忘れ、パスワード漏洩 の多発



RxT データ転送 IoTデバイス

課題: 非ネットワーク機器・共用PCからのデータ取り出し

- USBメモリ・DVDディスクを用いた人手でのデータ取り出し
- 共用PCでのログアウト忘れ、パスワード漏洩の多発



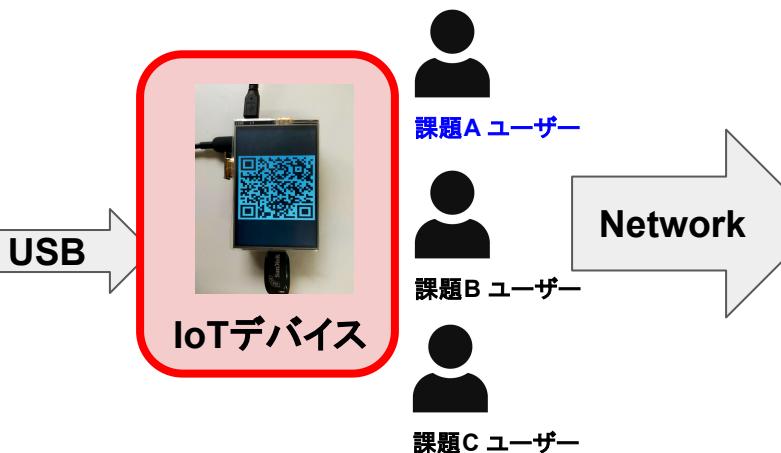
RxT データ転送 IoTデバイス

独自IoTデバイスを開発。IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



実験機器 (非ネットワーク)



M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

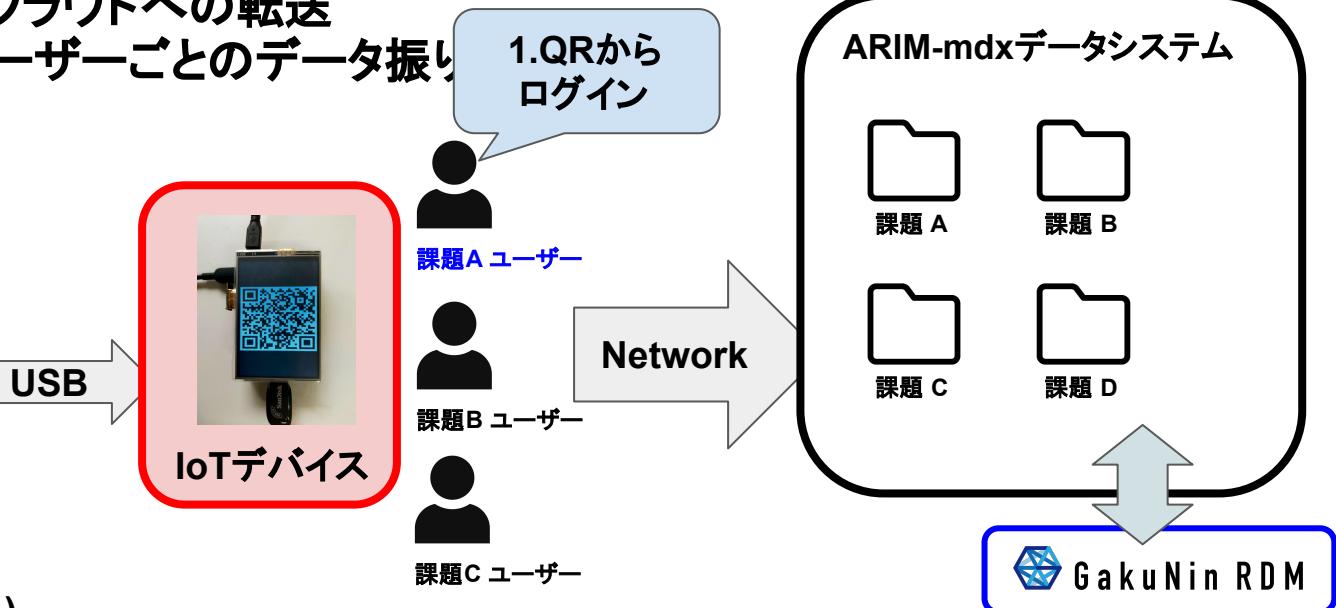
RxT データ転送 IoTデバイス

独自IoTデバイスを開発。IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



実験機器 (非ネットワーク)



M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

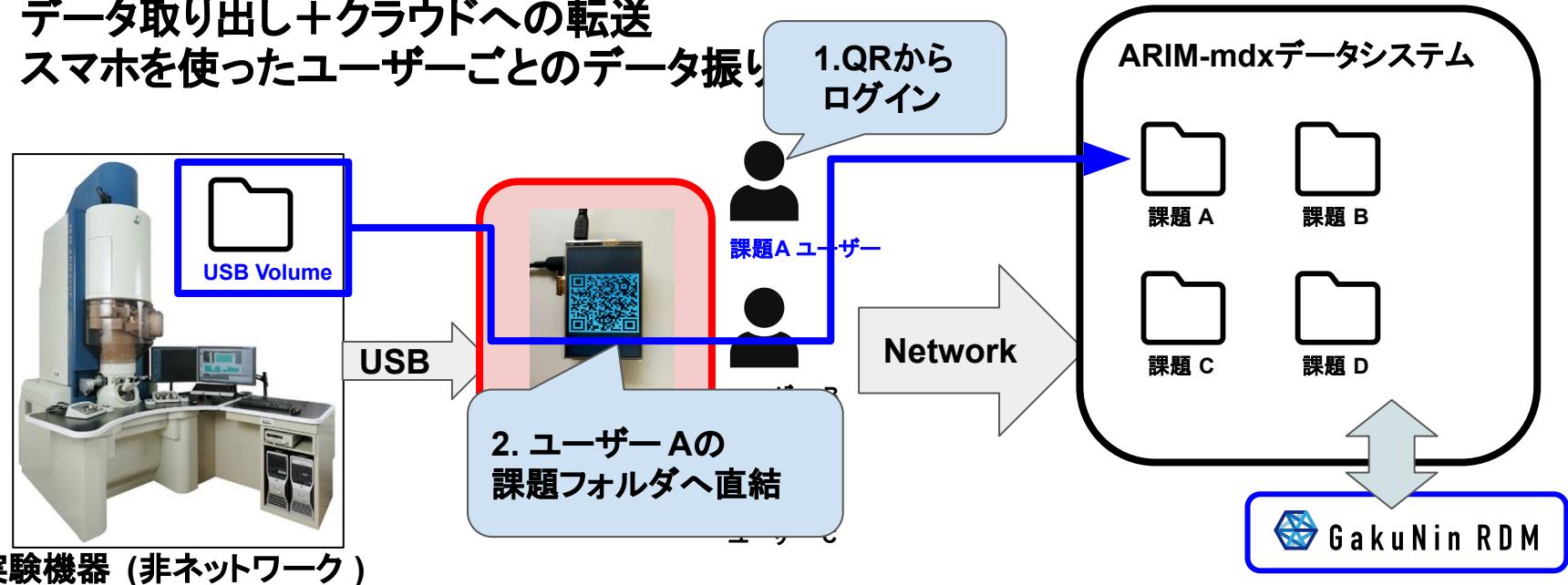
"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

RxT データ転送IoTデバイス

独自IoTデバイスを開発。IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

RxT データ転送IoTデバイス

独自IoTデバイスを開発。 IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



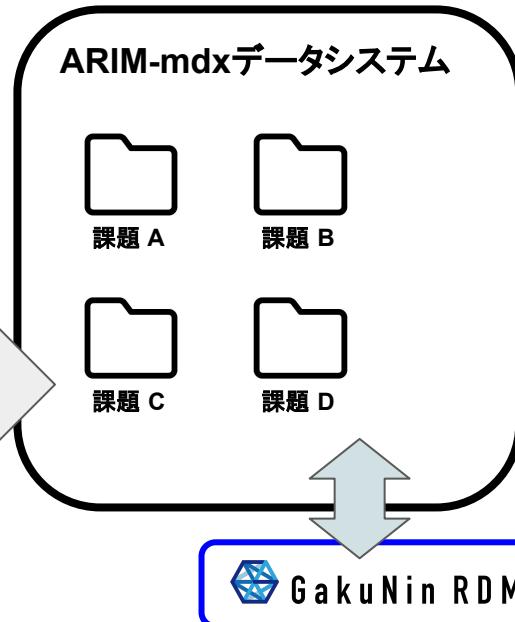
実験機器 (非ネットワーク)



IoTデバイス



Network



M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

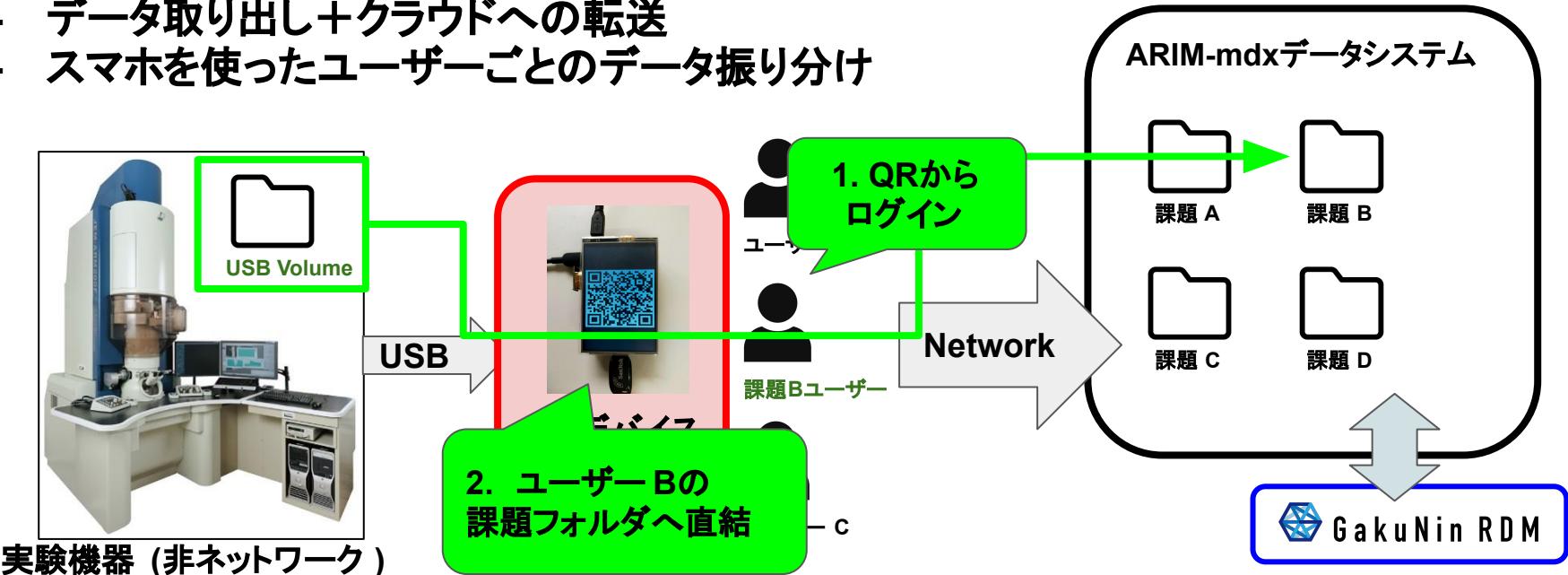
"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

RxT データ転送IoTデバイス

独自IoTデバイスを開発。IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

RxT データ転送 IoTデバイス

独自IoTデバイスを開発。IoTデバイスが以下を処理

- データ取り出し+クラウドへの転送
- スマホを使ったユーザーごとのデータ振り分け



実験機器 (非ネットワーク)

機器単位 FWによるセキュアネットワーク構築や NASによる多段構成は
装置側に手をいれる必要がありスケールしない
=> USB driverを使った汎用的な方法でスケール可能

1. QRから

c

2. ユーザー B の
課題フォルダへ直結

ARIM-mdxデータシステム

課題 A

課題 B

課題 C

課題 D

GakuNin RDM

M.Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura.

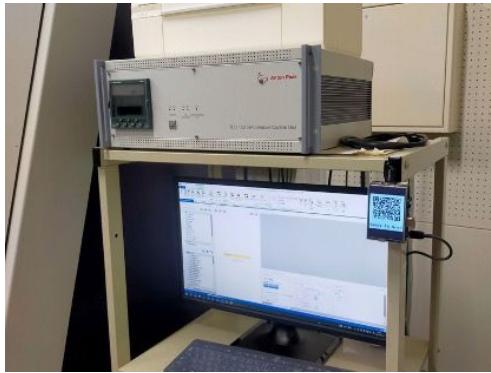
"Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System" IEEE/ACM UCC 2023

特願2023-156343 "IoT デバイス、データ転送システムおよびデータ転送方法"

デモ

RxTデータ転送デバイス.MOV

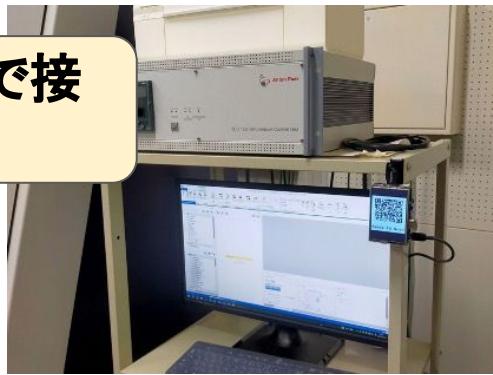
RxT データ転送IoTデバイス



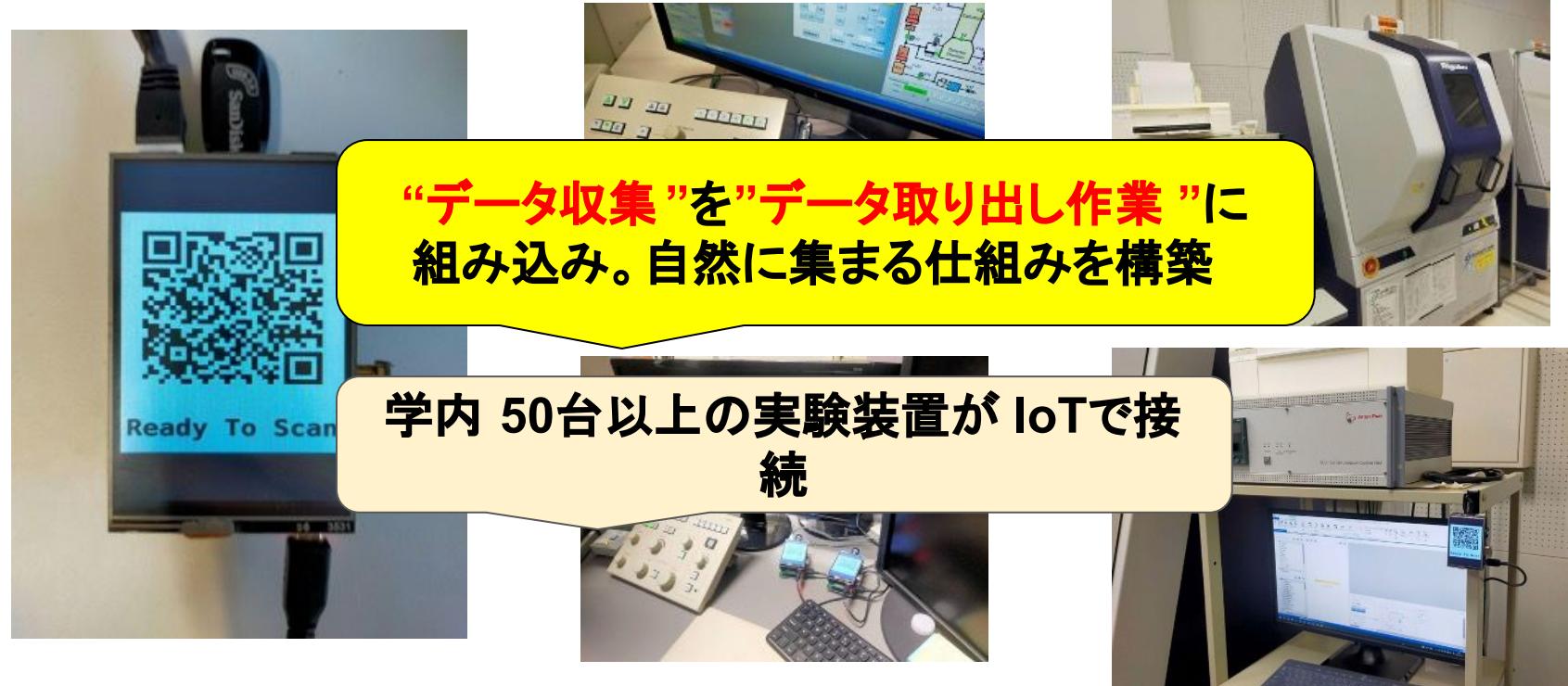
RxT データ転送 IoTデバイス



学内 50台以上の実験装置が IoTで接続



RxT データ転送 IoTデバイス



東京大学 x 株式会社ハウディとの共同で製品化 (RxT-01)

- 商用クラウドストレージへの対応 (Google, One Drive, Box)
- 耐故障・耐熱性能の強化
- 独自モジュール・専用ケースの開発
- 運用・サービス体制の整備
- 超高速版RxT-02制作中 (10x 高速)



通称 Hanaiデバイス
Ver 0.01



通称 Hanaiデバイス
Ver 0.02



通称 Hanaiデバイス
Ver 0.03



RxT-01

東京大学 x 株式会社ハウディとの共同で製品化 (RxT-01)

- 商用クラウドストレージへの対応 (Google, One Drive, Box)
- 耐故障・耐熱性能の強化
- 独自モジュール・専用ケース
- 運用・サービス体制の整備
- 超高速版RxT-02制作中 (1)

2月よりハウディHPで販売中

現在、アズワン社との協業で販売経路の拡大を
予定 (今年度中を予定)
=> 全国の研究室 から簡単に購入可能に



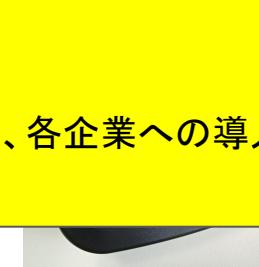
その他、各企業への導入や協業を実施中



通称 Hanaiデバイス
Ver 0.01



通称 Hanaiデバイス
Ver 0.02



通称 Hanaiデバイス
Ver 0.03

RxT-01

東京大学 x 株式会社ハウ

全国の研究室の材料実験データを簡単に mdx + GRDMへ !!

- 商用クラウドストレージ
- 耐故障・耐熱性能の強化
- 独自モジュール・専用ケース
- 運用・サービス体制の整備
- 超高速版RxT-02制作中 (1)

2月よりハウディHPで販売中

現在、アズワン社との協業で販売経路の拡大を
予定 (今年度中を予定)
=> 全国の研究室から簡単に購入可能に



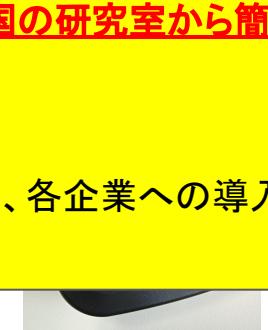
その他、各企業への導入や協業を実施中



通称 Hanaiデバイス
Ver 0.01



通称 Hanaiデバイス
Ver 0.02



通称 Hanaiデバイス
Ver 0.03

RxT-01

まとめ

ARIM-mdx Data System:

- 計算分野・実験分野・データ分野、全ての材料研究者のための大規模データプラットフォーム
- 運用2年間で**1400ユーザー・約1PBの材料データ**が蓄積
- GRDMとストレージ連携

ARIM-mdxにおける実験データの収集

- 試料データ: **ARIM-mdx Sample Submission**
 - 自由記述による簡単な方法、LLM + RAG による自動評価システム
- 実験データ: **RxT データ転送 IoTデバイス**
 - 非ネットワーク実験装置からのデータ取り出し+ クラウドupload

今後

- **検索ツール・AIツールの開発 (本プロジェクトのゴール)**
- GRDMとの結合、収集データの利活用

研究・開発・運用チーム

- 東京大学 情報基盤センター：
 - 華井雅俊 特任助教 (機械学習・グラフニューラルネットワーク・高性能計算),
 - 鈴村豊太郎 教授 (深層グラフ学習・グラフニューラルネットワーク・人工知能・高性能計算),
 - 田浦健次朗 教授 (並列分散処理・プログラミング言語・システムソフトウェア・高性能計算)
- 東京大学 工学系研究科：
 - 石川 亮 特任准教授 (光電子材料・二次元系物質・三次元電子顕微鏡法)
- 横浜国立大学
 - 河村 光晶 准教授 (マテリアルズ・インフォマティクス、物性計算)

Reference

- Web Page: <https://arim.mdx.jp/>
- 「材料研究DXを加速するARIM-mdxデータシステムを開発、全国の900名以上が利用開始」https://www.u-tokyo.ac.jp/focus/ja/press/z0310_00004.html
- 「ハウディ、東京大学との共同研究から生まれたデータ転送 IoTデバイス「RxT-01」を販売開始」<https://prtimes.jp/main/html/rd/p/000000012.000083225.html>
- 華井雅俊 “ARIM-mdxデータシステム:材料研究向け実験・シミュレーションの統合データプラットフォーム”, ARIM次世代ナノスケールマテリアル領域 研究会, 2025年3月
- 華井雅俊 “ARIM-mdxデータシステム:材料研究向け実験・シミュレーションの統合データプラットフォーム”, 第19回材料系ワークショップ, 2025年2月
- 華井雅俊 “東京大学ARIM データ基盤部門 活動報告” 「第3回革新的なエネルギー変換を可能とするマテリアル領域」シンポジウム 2024年12月
- 華井雅俊 “ARIM-mdxデータシステム:材料研究向け実験・シミュレーションの統合データプラットフォーム”, データ活用社会創成シンポジウム 2024, 2024年12月
- 華井雅俊 “ARIM-mdx データ収集・保存システムの紹介” 第2回 NBCI-ARIM 技術交流会, 2024年6月
- 華井雅俊 “エネルギー変換領域におけるデータ収集・保存システム” 「第2回革新的なエネルギー変換を可能とするマテリアル領域」シンポジウム, 2024年1月
- 華井雅俊 “ARIM-mdxデータシステム: 材料実験データの利活用に向けた実験施設のDX化”, 第一回UDAC-SRIS合同勉強会, 2023年12月
- M. Hanai + 19 authors “ARIM-mdx Data System: Towards a Nationwide Data Platform for Materials Science” IEEE BigData 2024
- M. Hanai, M. Kawamura, R. Ishikawa, T. Suzumura, and K. Taura. “Cloud Data Acquisition from Shared-Use Facilities in A University-Scale Laboratory Information Management System” IEEE/ACM UCC 2023
- 特願2023-156343 “IoT デバイス、データ転送システムおよびデータ転送方法”

協力 (敬称略・五十音順)

幾原 雄一 (東京大学), 飯盛 桂子 (東京大学), 井上 佳寿恵 (東京大学), 太田 悅子 (東京大学), 大西 正人 (統計数理研究所), 岡根 哲夫 (日本原子力研究開発機構), 沖津 康平 (東京大学), 押川 浩之 (東京大学), 落合 幸徳 (東京大学), 木村 鮎美 (東京大学), 黒木 伸一郎 (広島大学), 小西 邦昭 (東京大学), 近藤 兜之 (東京大学), 坂本 弘樹 (広島大学), 櫻井 治之 (東京大学), 佐藤 千津子 (東京大学), 澤村 智紀 (東京大学), 塩見 淳一郎 (東京大学), 柴田 直哉 (東京大学), 竹内 美由紀 (東京大学), 竹中 規雄 (東京大学), 寺西 亮佑 (東京大学), 豊倉 敦 (東京大学), 中村 公 (東京大学), 西尾 和記 (東京大学), 西村 芳治 (東京大学), 一杉 太郎 (東京大学), 府川 和弘 (東京大学), 福川 昌宏 (東京大学), 藤川 誠司 (日本原子力研究開発機構), 藤原 誠 (東京大学), 古川 恵 (東京大学), 松井 裕章 (東京大学), 松村 大樹 (日本原子力研究開発機構), 水島 彩子 (東京大学), 三田 吉郎 (東京大学), 森田 真理 (東京大学), 森山 和彦 (東京大学), 安永 竣 (東京大学), 山田 淳夫 (東京大学), 山田 真司 (広島大学), 山原 弘靖 (東京大学), レ デウック アイン (東京大学), 以上 43名

おわり

