資料4
AI for Scienceを支える研究
データの管理・利活用と流通
の在り方ワーキンググループ
(第５回)
令和8年5月22日

大型放射光施設SPring-8の研究データ基盤と AI for Scienceに向けた取り組み

初井宇記

理化学研究所・放射光科学研究センター

制御情報・データ創出基盤グループ

大型放射光施設SPring-8

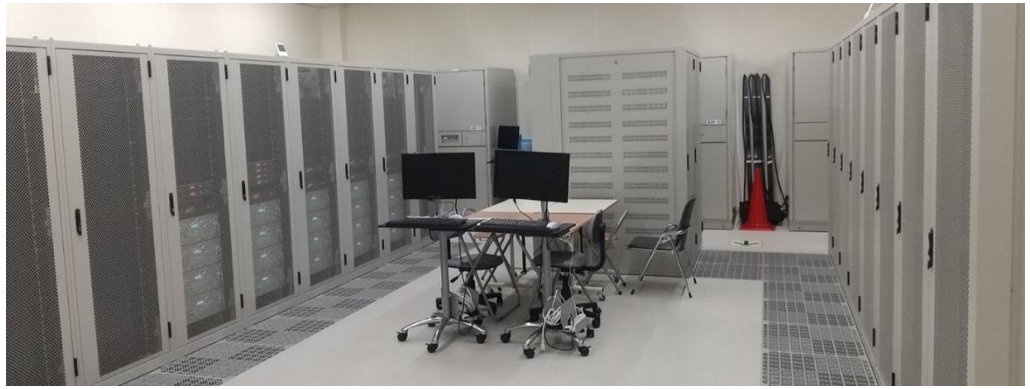
* 2024.9-2025.8

SPring-8概要

- 周長1.5 kmの加速器から高輝度X線を生成
- 延べ利用者数：約1万4千/年*（約2割が民間利用）
- 利用分野：生命科学(薬学)、化学、材料科学 、エネルギー、文化財、地球科学等
- 半導体戦略、国土強靱化、農学・食料安全保障などの国家課題解決に向けた戦略的利用を拡大
- 2029年度にSPring-8-IIへ高度化、100倍の高輝度化を達成見込み

データ基盤

- 分野ごとにデータ特性、オープン/クローズ方針、解析手法・ワークフローが多様
- 各分野を支える共通研究基盤が重要
- 文部科学省令和3年度補正予算により中核となるデータセンターを設置2023年9月から運用開始
  広帯域ビームライン5本分（データ帯域換算で約25%）について措置
- SPring-8データセンターはデータ流通・実験中データ解析に特徴をもつ基盤
- アーカイブ、大規模データ解析はHPCIリソース、パブリッククラウドと連携
- GakuNin RDM連携サービスも試行中



データセンターを活用した成果例：

河口彰吾ら(JASRI)

大規模計算とその場測定を用いて多元セシウム塩化物を効率的に探索

課題

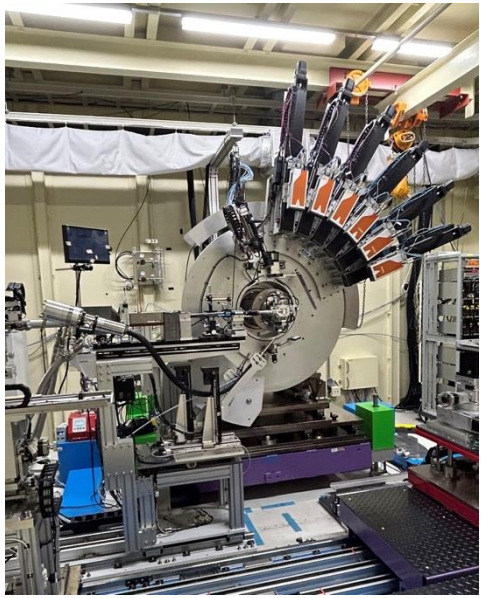
多元材料探索は組合せ爆発 、実験のみでは探索効率が低い。
大規模計算と実験統合が重要

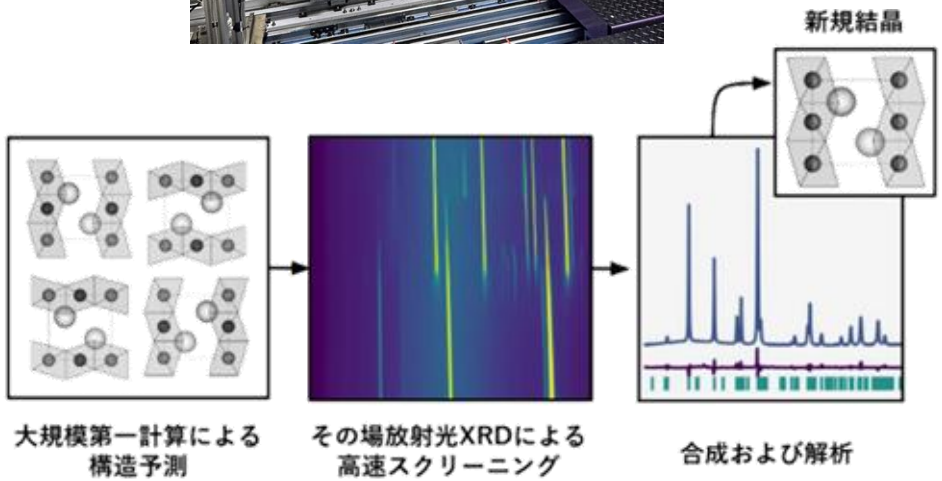
データ処理・データ量

画像データから1次元の回折プロファイルを自動処理
数百KB/データ、数千～数十万データ/日

成果

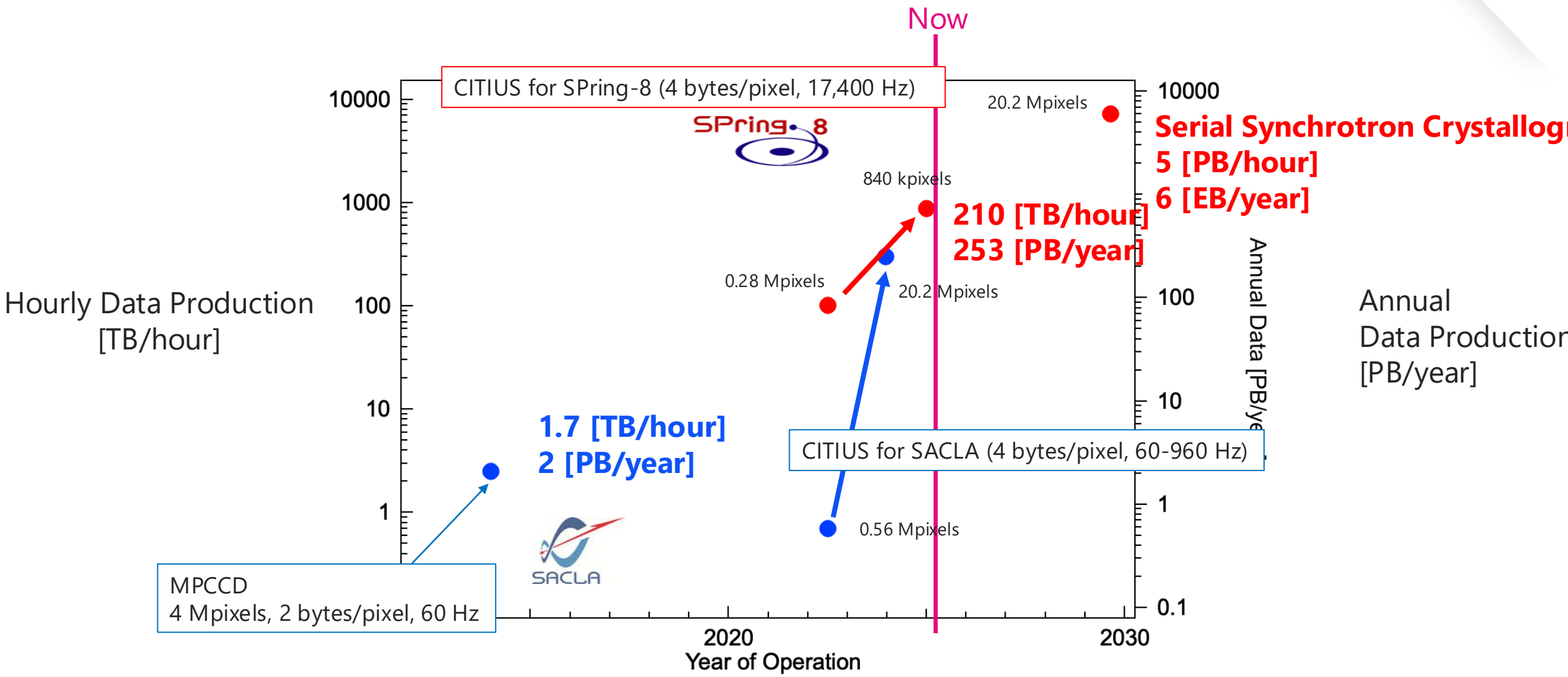
第一原理計算による大規模構造予測を用いて探索
放射光X線回折による高速スクリーニング
新規セシウム塩化物の合成に成功



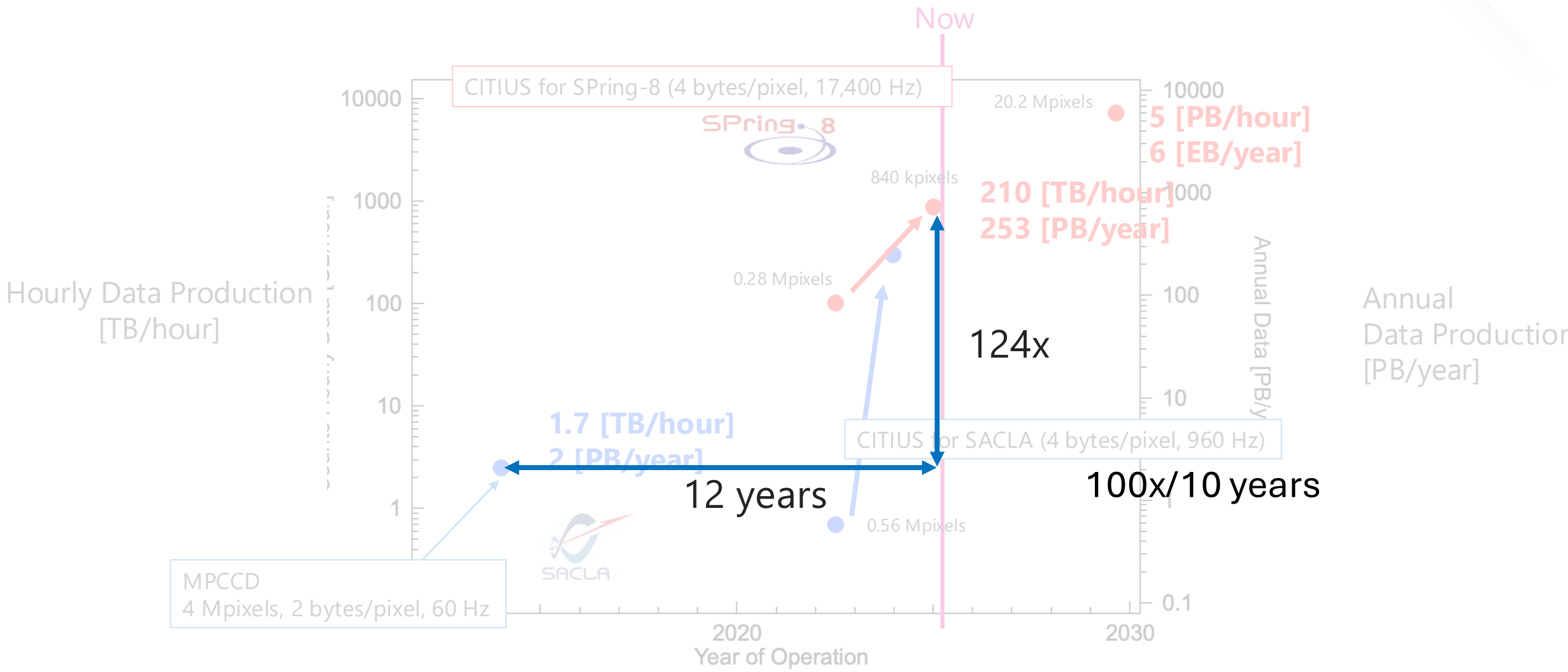


三浦章(北海道大学)、Ekin Dogus Cubuk(Google DeepMind) ら
A. Miura et al., J. Am. Chem. Soc. (2024) **146**, 29637

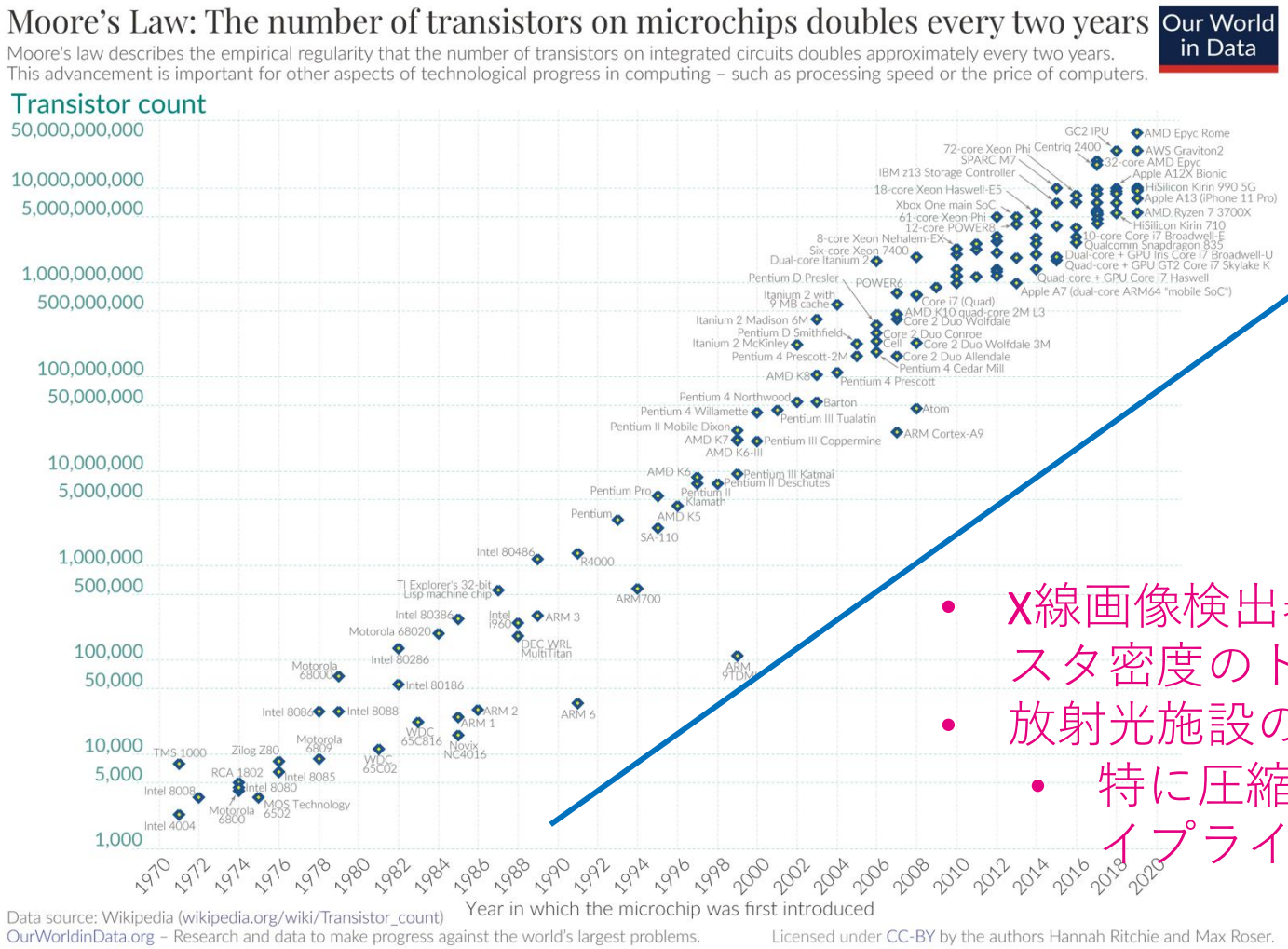
X線画像検出器の開発トレンド



X線画像検出器の開発トレンド



Moore’s law
Transistor density 2x in 1.5-2 years



40x in 10 years

- X線画像検出器の出力帯域トレンドはトランジスタ密度のトレンドを上回る速度で向上
- 放射光施設のデータ基盤R&Dが重要に
  - 特に圧縮技術を含む広帯域データ処理パイプライン

広帯域ビームライン
データ処理パイプラインの実装例

https://www.riken.jp/press/2026/20260403_1/index.html
M. Saito et al. Phys. Rev. Lett. 132, 256901 (2024)

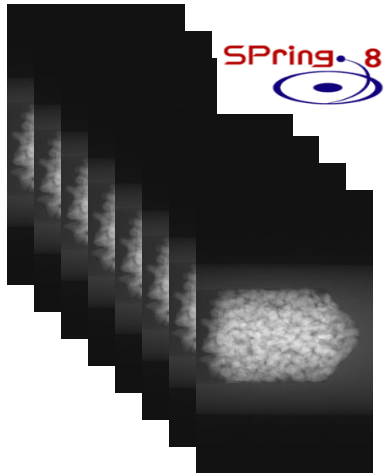
課題
- 理研が開発した次世代X線画像検出器CITIUSを利用した実験
- 大量データを生成(27 GB/s, 2.3 PB/day)、リアルタイムデータ処理基盤が必須

現状
- 独自開発のFPGA演算加速ボード等によるリアルタイム前処理・情報抽出・データ圧縮
- 自動データ転送とデータセンター内クラスタによる即時解析
- ブラウザベース解析環境（OpenOnDemand）
- 実験条件・解析履歴を含む構造化されたデータとして保存される

今後
- 米国Argonne研究所APS施設に本データ処理パイプラインの技術供与・立ち上げ中
- APS施設と共同で、AIによるリアルタイム解析・実験制御を見据えた基盤技術を開発中
- 米国Genesisミッションで策定されるデータ形式との相互運用を検討中

国家課題解決に向けたトップダウン型の戦略的利用
SPring-8(SPring-8-II) × 富岳 (富岳Next)

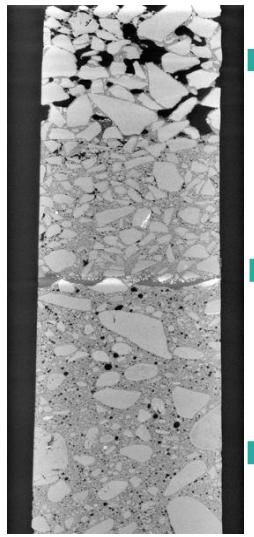
コンクリート(運用中)



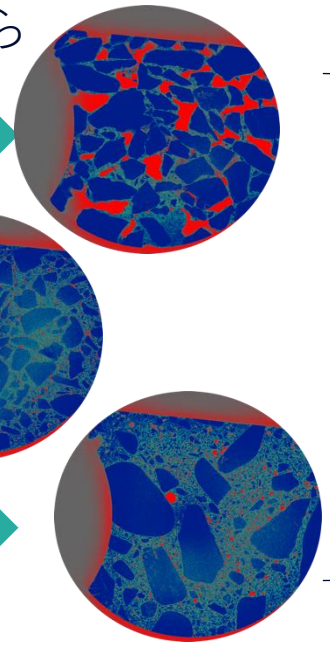
0.5 PB/day

像再構成

Mohamed Wahib (R-CCS, RIKEN)ら

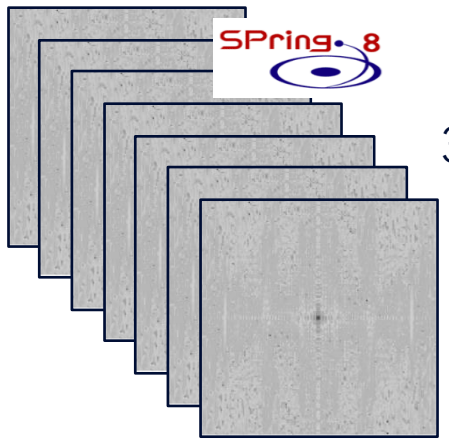


領域分割



コンクリートの寿命予測
長寿命化の知見創成

半導体(開発中)

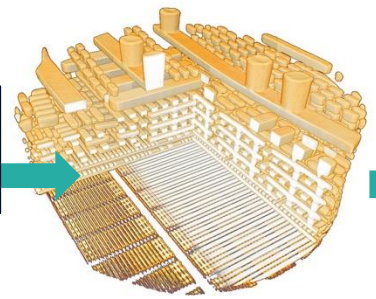


30 PB/day

像再構成

Yasumasa Joti (RIKEN, JASRI)ら

領域分割



先端半導体の
製造技術開発の
指針を与える構造情報

AI-Powered SPring-8に向けて：課題

データの移動

ビームライン、**SPring-8**データセンター、富岳と**3**箇所のストレージをデータが移動、ステージングがユーザの利便性を損ねている

セキュリティ

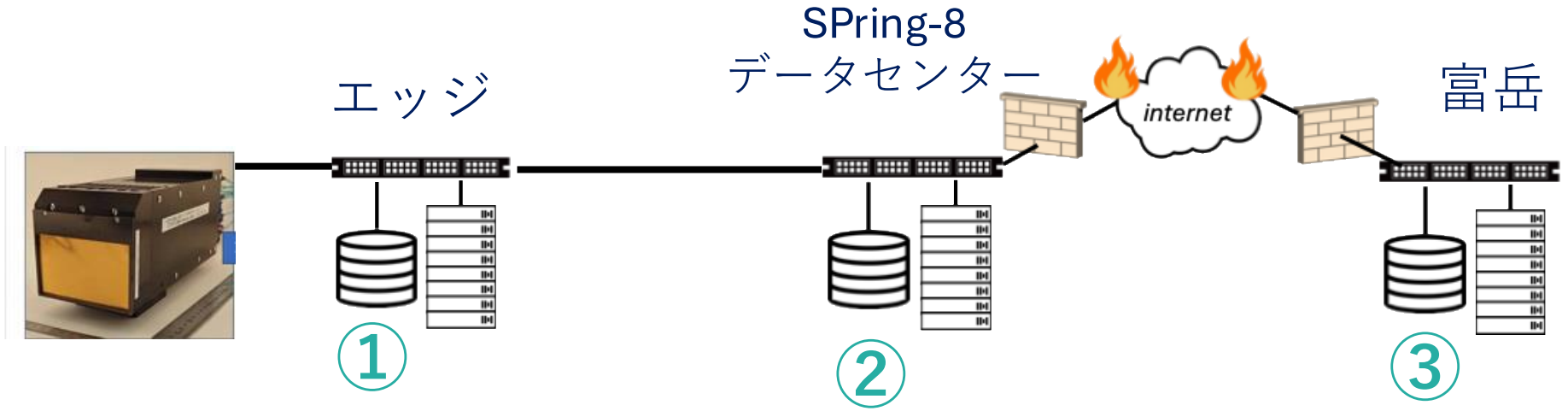
先端半導体や電池開発などの産業利用では高度なセキュリティが重要。既存インフラでは対応出来ない状況

民間クラウド事業者の**Best practice**と同等の運用・サービスが求めらている（サーバ設置室の警備員配置、監視システム、ユーザ領域の物理層での隔離オプション、監査、補償、事業継続措置等）

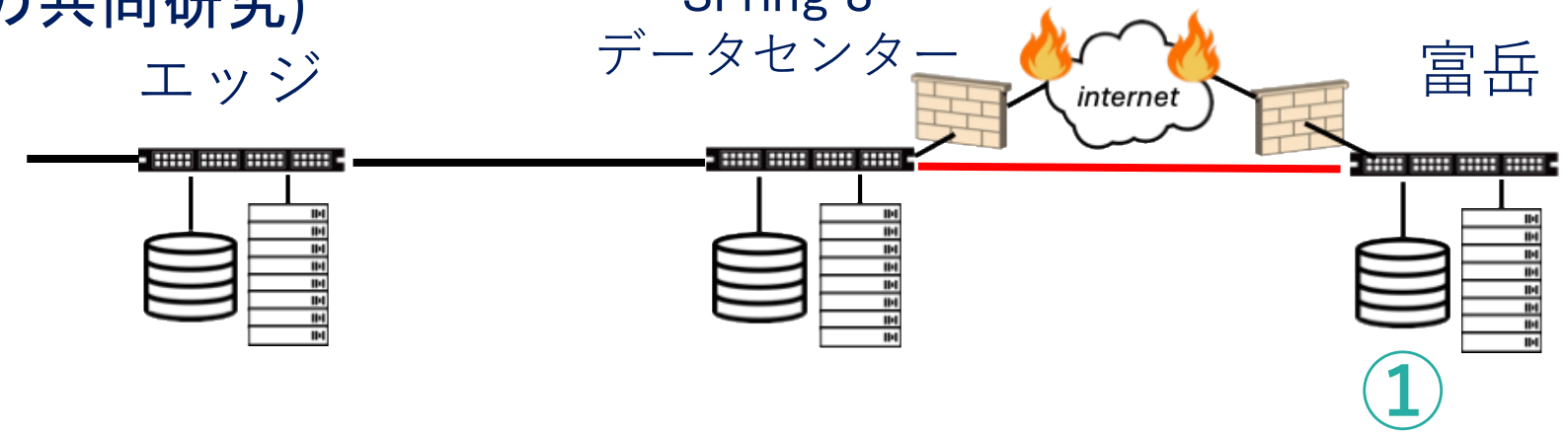
機密保持に対応出来る**LLM**が必要

セキュリティ強化：
SPring-8から富岳へのデータ転送

現状



開発中(NTTとの共同研究)



特定波長を独占的に利用できるため、専用線と同様に物理的に通信経路を隔離できる。
さらに、低レイテンシ特性を活かすことで、高性能で利便性の高いストレージ構築も原理的に可能。
（ステージングの解消）

SPring-8とデータ基盤：現状

データベース

- **SPring-8**は多くの分野で利用されている。
- タンパク質結晶構造解析など、**AI**時代に有用なデータの大量創成拠点
- ユーザコミュニティが主体となってデータベースを構築、**SPring-8**は支援
  データのオープン・クローズはユーザ(実験責任者)が判断できる運用

データ基盤

- **SPring-8**データセンターはデータ流通・実験中データ解析に特徴
- アーカイブ、大規模データ解析は、**HPCI**リソースおよびパブリッククラウドと連携
- データ管理について、**GakuNin RDM**連携サービスを試行中
- データ構造化への取組は重要。**AI**可読な形式でデータ保存。順次共用装置について対応中

AI for Scienceに向けた課題

利用者へのファーストタッチ

国内9施設に多数・多様な分析装置が分布。活用するためのコンシエルジュ機能が必須

認証

SPring-8では独自の認証基盤を運用。現在更新を検討している。民間ユーザ含む対応が必須。全施設で共通の認証基盤が必要

ネットワーク

SPring-8データセンター、富岳、HPCIストレージなどが活用されている。これらを接続するSINETの重要性は今後増大する。特にIOWNの低遅延・物理層分離ネットワークによるセキュリティ強化とステージング解消等の新規機能に期待している。

AI向け計算リソース

セグメンテーション向けのVision Foundation Modelの暫定見積もり(training 20M GPU/hours、inference 40 M GPU/hours)。大きなリソースが必要と見積もられる。

SPARE SLIDES