

新規自動化X線CTビームラインにおける データセンターとの連携の検討状況

平木俊幸

理化学研究所

<目次>

- 解析環境に関するニーズ
- SP8データセンターとの連携の将来像
- X線CTビームラインの自動化
- 自動X線CTビームラインとSP8データセンターの連携
- 自動X線CTビームラインとSP8 DCとの連携のスケジュール
- まとめ

第1回SP8データWSでのビームラインセッションおよびサイエンスケースセッションでの講演を基に整理。

- イメージングBLから（JASRI 上杉様）
 - **BL28B2でX線CT測定の自動化**→1 TB/day, 400 TB/yearでのデータ生成の見込（再構成結果含む）
 - 現状、一部のユーザーにとってデータ処理・解析環境が不十分
 - 計測からデータ処理まで一貫して行えるインフラ整備（＝データセンター）が今後必須
- CT-XAFS@BL37XUから（JASRI 関口様）
 - 時分割CT-XAFSの開発→10 TB/day以上のデータ生成の見込
 - データ解析がヘビー。環境準備が新規参入にとって敷居が高い→データセンターでの解析に期待
- 名古屋大学 唯様から
 - 研究スタイルの変革→リアルタイムで異分野研究者が討議、測定にフィードバック
 - 解析速度の向上→物質開発サイドによる測定データの迅速診断につながる
 - イメージング計測の高度化による、これまで手に入らなかった物質ビッグデータの提供と活用を期待
- 豊田中央研究所 長井様から
 - Operando CT測定では1000枚程度/数秒、数分～数十分/条件の撮像、3日程度で20条件→1M画素検出器を仮定すると12 TB/3日でデータ生成。
 - 測定してみないとわからない。施設ごとの過去の類似実験があれば測定などの計画が立てやすい→類似実験が参照できる仕組み
 - 現場（豊田BL）での解析律速
 - データサイズが大きいため詳細解析環境（長久手）へデータ転送ができず、持ち帰ってから詳細解析
 - 現場と詳細解析環境（長久手）の物理的距離
 - 測定・解析の途中経過をいつでもだれでも各々の観点で見れば実験確度が向上すると期待→外部からアクセスできるデータ共有・解析資源（データセンターやクラウド環境）
 - SP8/J-PARCのデータがバラバラ→統一的なフォーマットでデータを管理・検索できる仕組み

オンサイトなデータ転送・解析環境、リモートアクセス

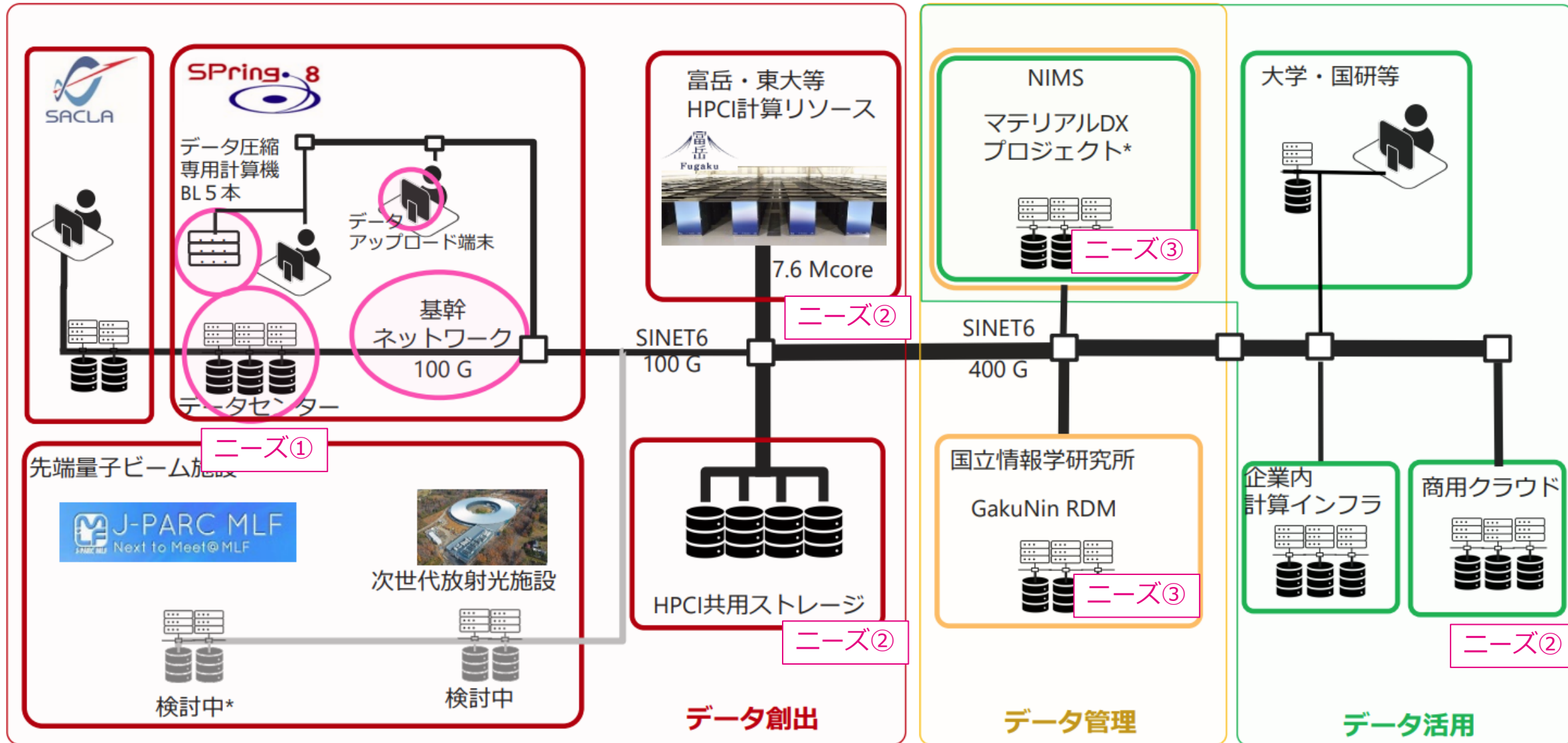
第1回SP8データWSでのビームラインセッションおよびサイエンスケースセッションでの講演を基に整理。

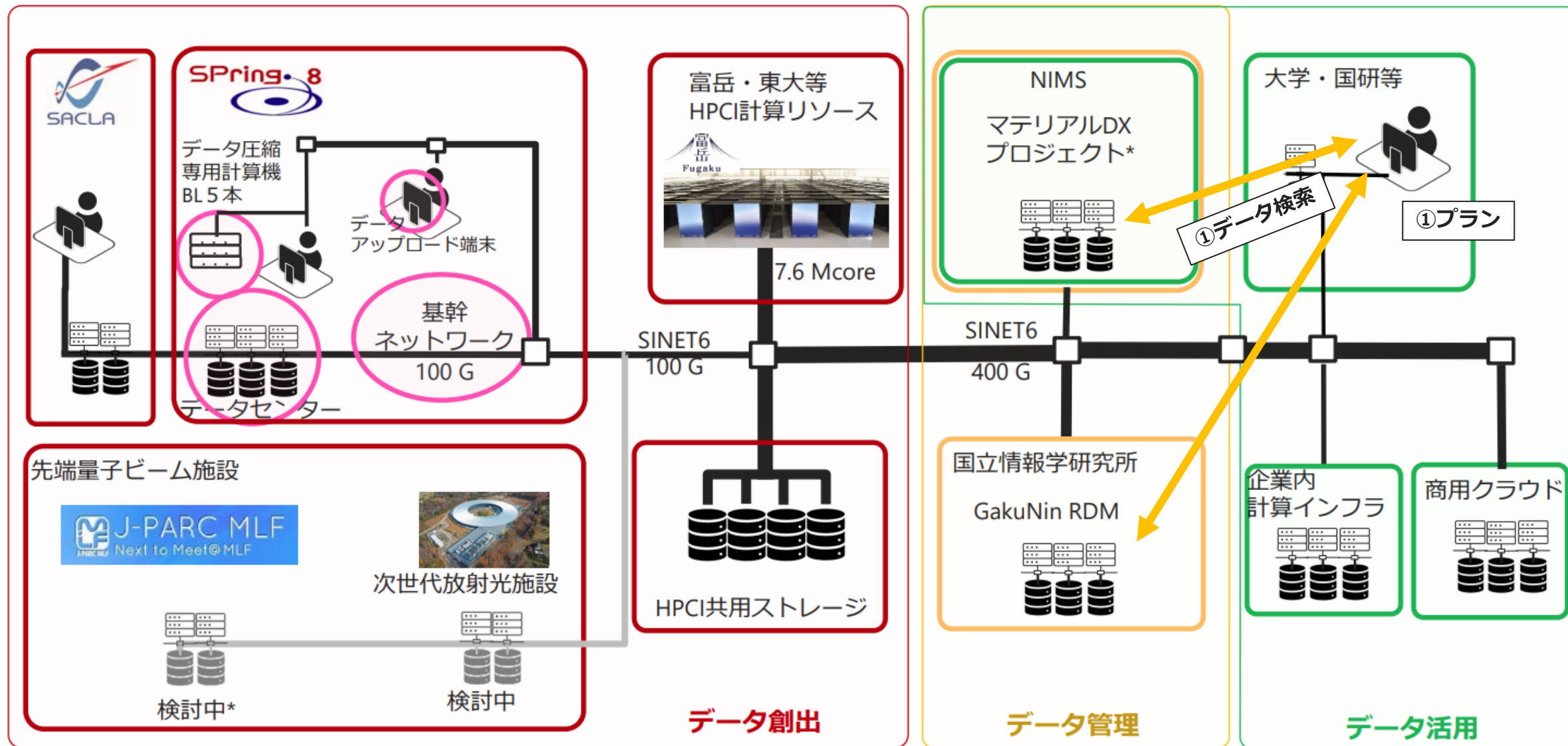
- イメージングBLから（JASRI 上杉様）
 - **BL28B2でX線CT測定の自動化**→1 TB/day, 400 TB/yearでのデータ生成の見込（再構成結果含む）
 - 現状、一部のユーザーにとってデータ処理・解析環境が不十分
 - 計測からデータ処理まで一貫して行えるインフラ整備（＝データセンター）が今後必須
- CT-XAFS@BL37XUから（JASRI 関口様）
 - 時分割CT-XAFSの開発→10 TB/day以上のデータ生成の見込
 - データ解析がヘビー。環境準備が新規参入にとって敷居が高い→データセンターでの解析に期待
- 名古屋大学 唯様から
 - 研究スタイルの変革→リアルタイムで異分野研究者が討議、測定にフィードバック
 - 解析速度の向上→物質開発サイドによる測定データの迅速診断につながる
 - イメージング計測の高度化による、これまで手に入らなかった物質ビッグデータの提供と活用を期待
- 豊田中央研究所 長井様から
 - Operando CT測定では1000枚程度/数秒、数分～数十分/条件の撮像、3日程度で20条件→1M画素検出器を仮定すると12 TB/3日でデータ生成。
 - 測定してみないとわからない。施設などの過去の類似実験があれば測定などの計画が立てやすい→類似実験が参照できる仕組み
 - 現場（豊田BL）での解析律速
 - データサイズが大きいため詳細
 - 現場と詳細解析環境（長久手）
 - 測定・解析の途中経過をいつで
 - 源（データセンターやクラウド）
 - SP8/J-PARCのデータがバラバラ

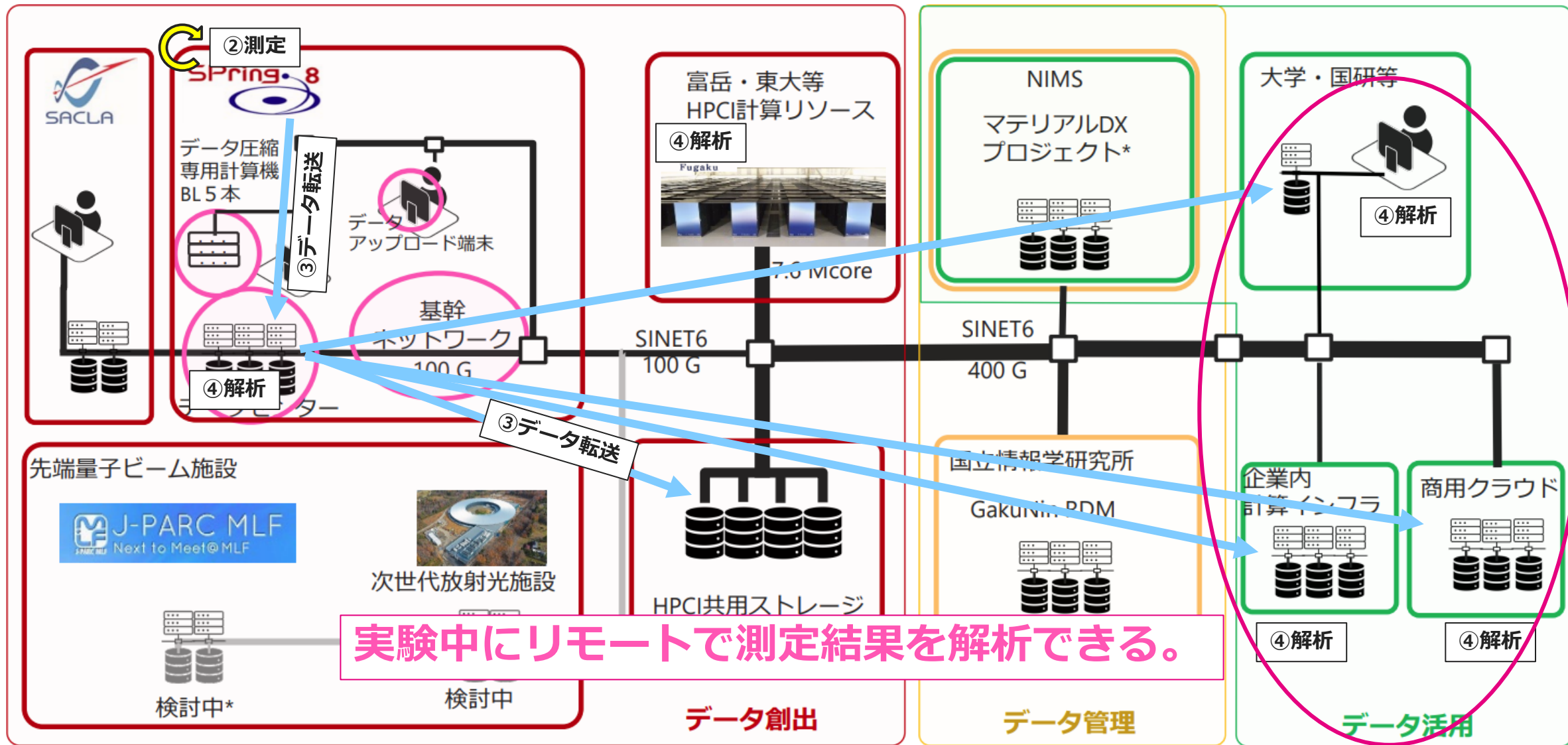
ニーズ①：オンサイト解析インフラ
 ニーズ②：外部資源との連携
 ニーズ③：先行研究を参照・比較する仕組み
 ニーズ④：測定へのフィードバック
 ニーズ⑤：ビッグデータの利活用

オンサイトなデータ転送・解析環境、
 リモートアクセス
 アクセスできるデータ共有・解析資

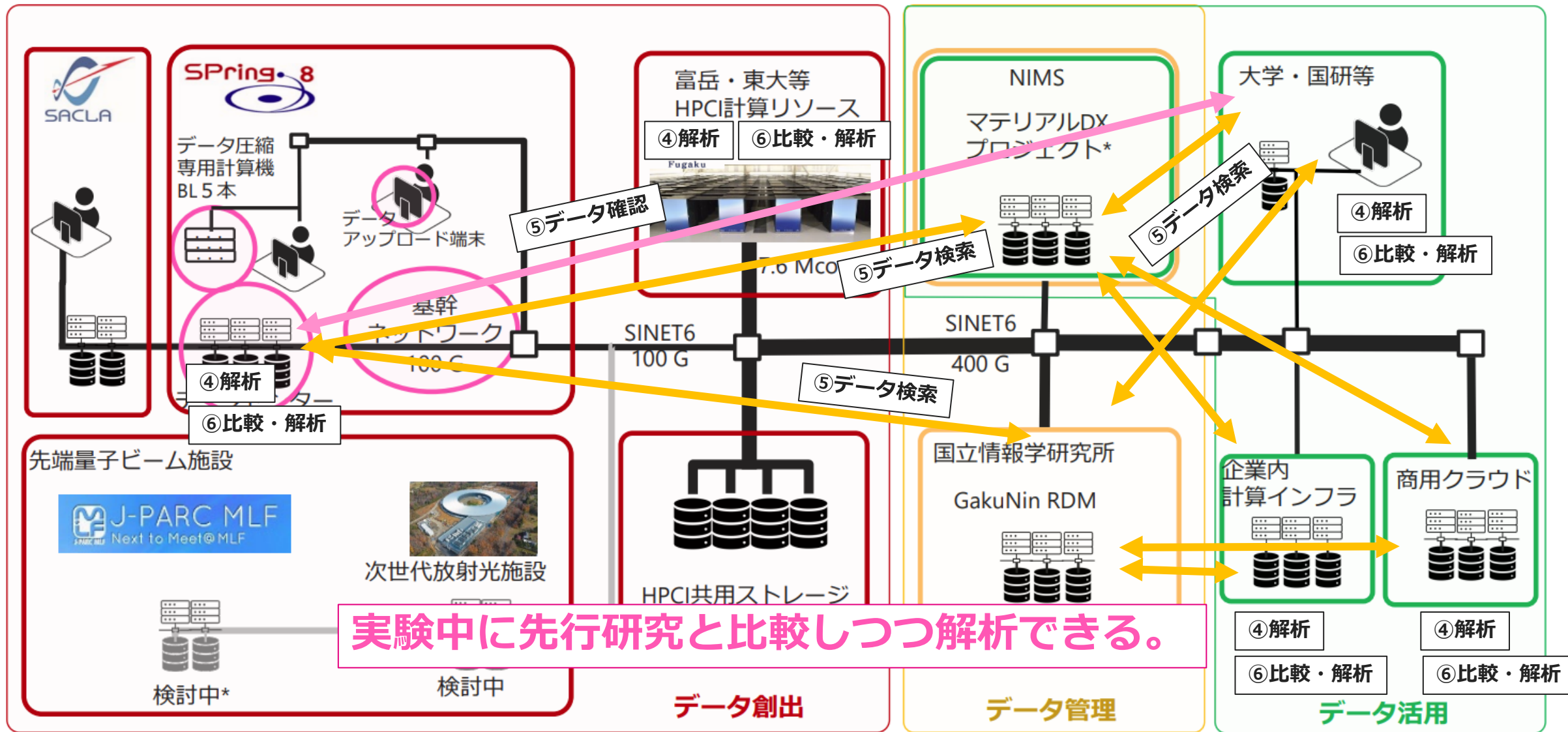
第2回SP8データWSでの初井様（理研）「SPring-8のデータ流通基盤とMDXへの接続」p.6より抜粋



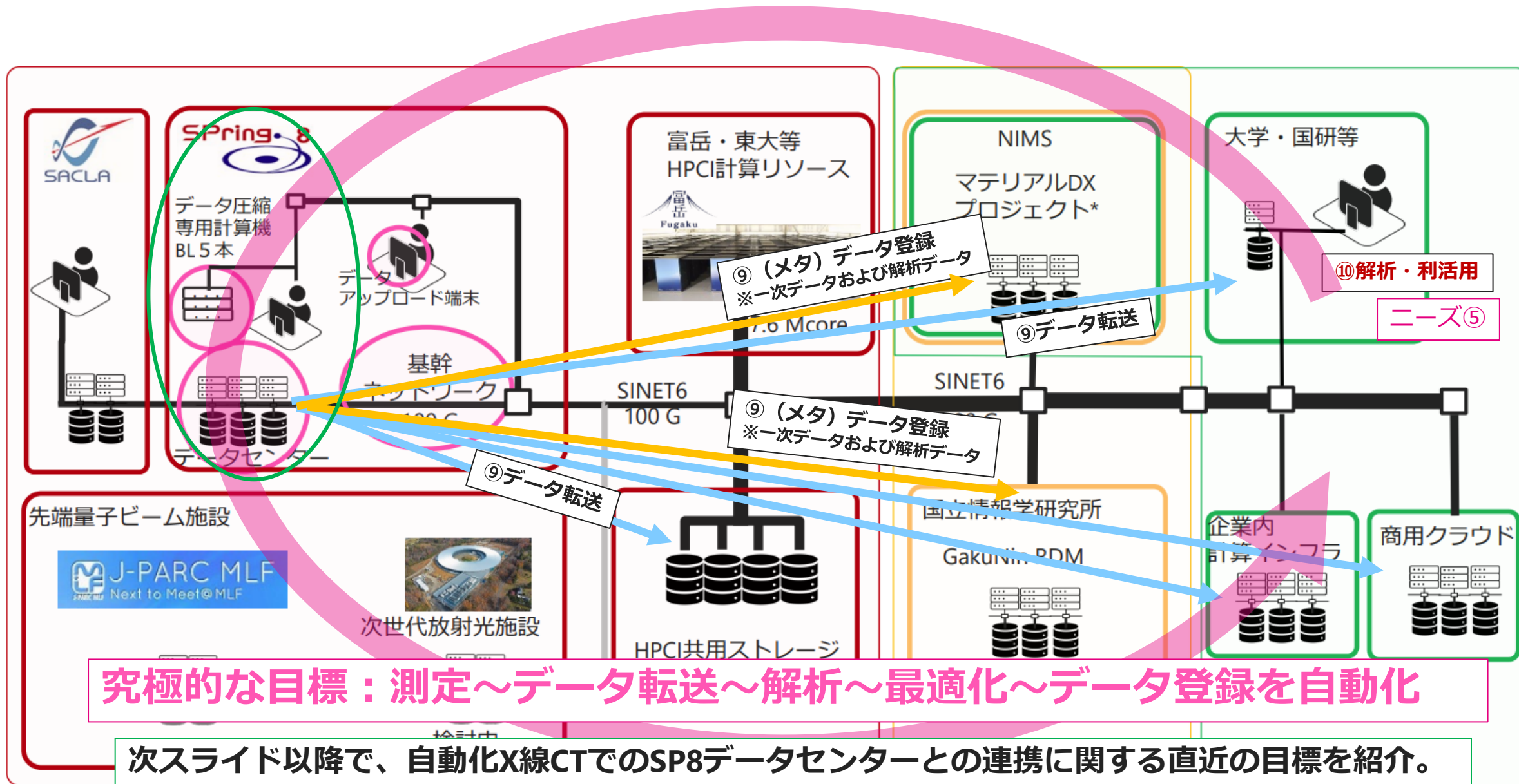


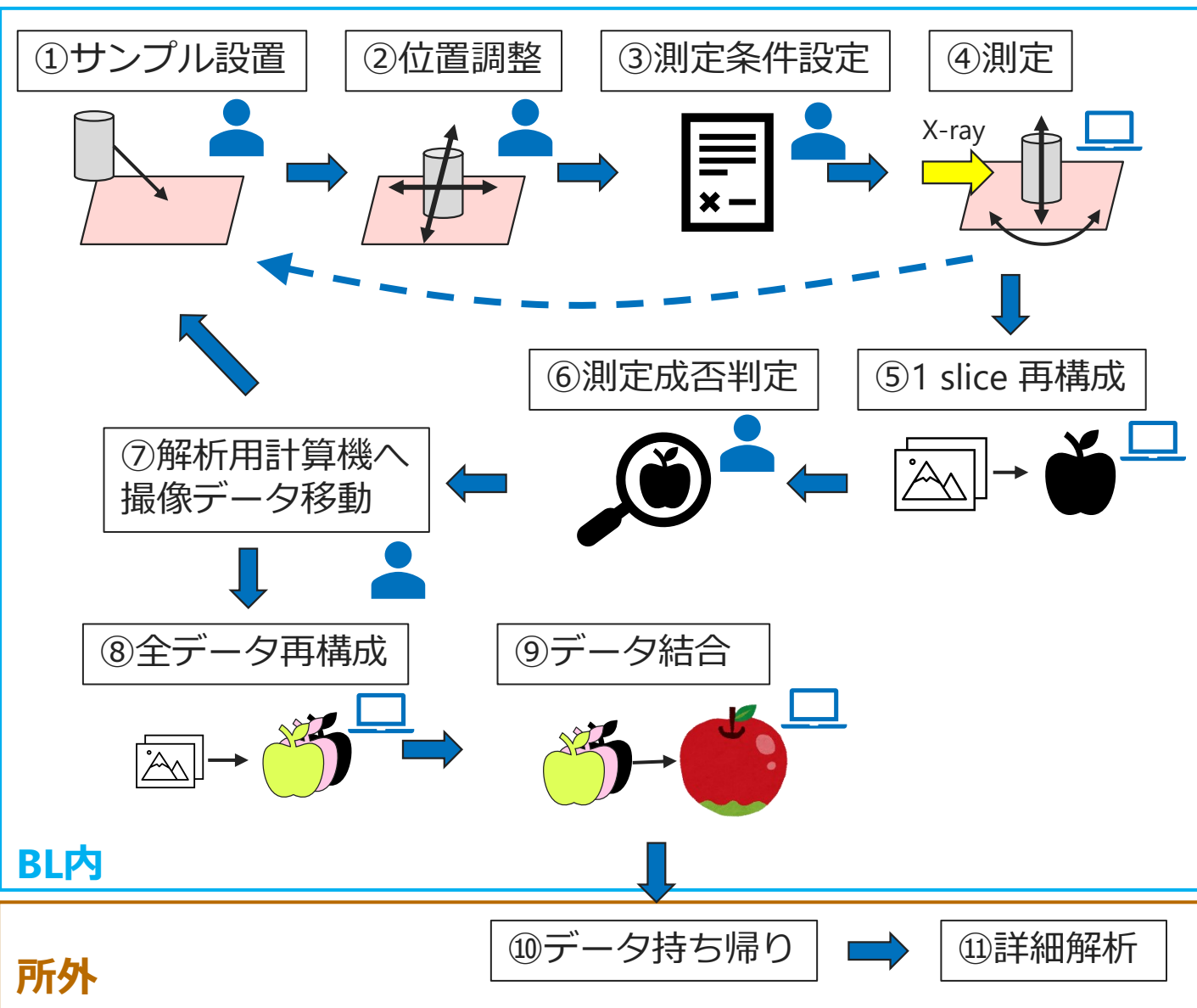


実験中にリモートで測定結果を解析できる。





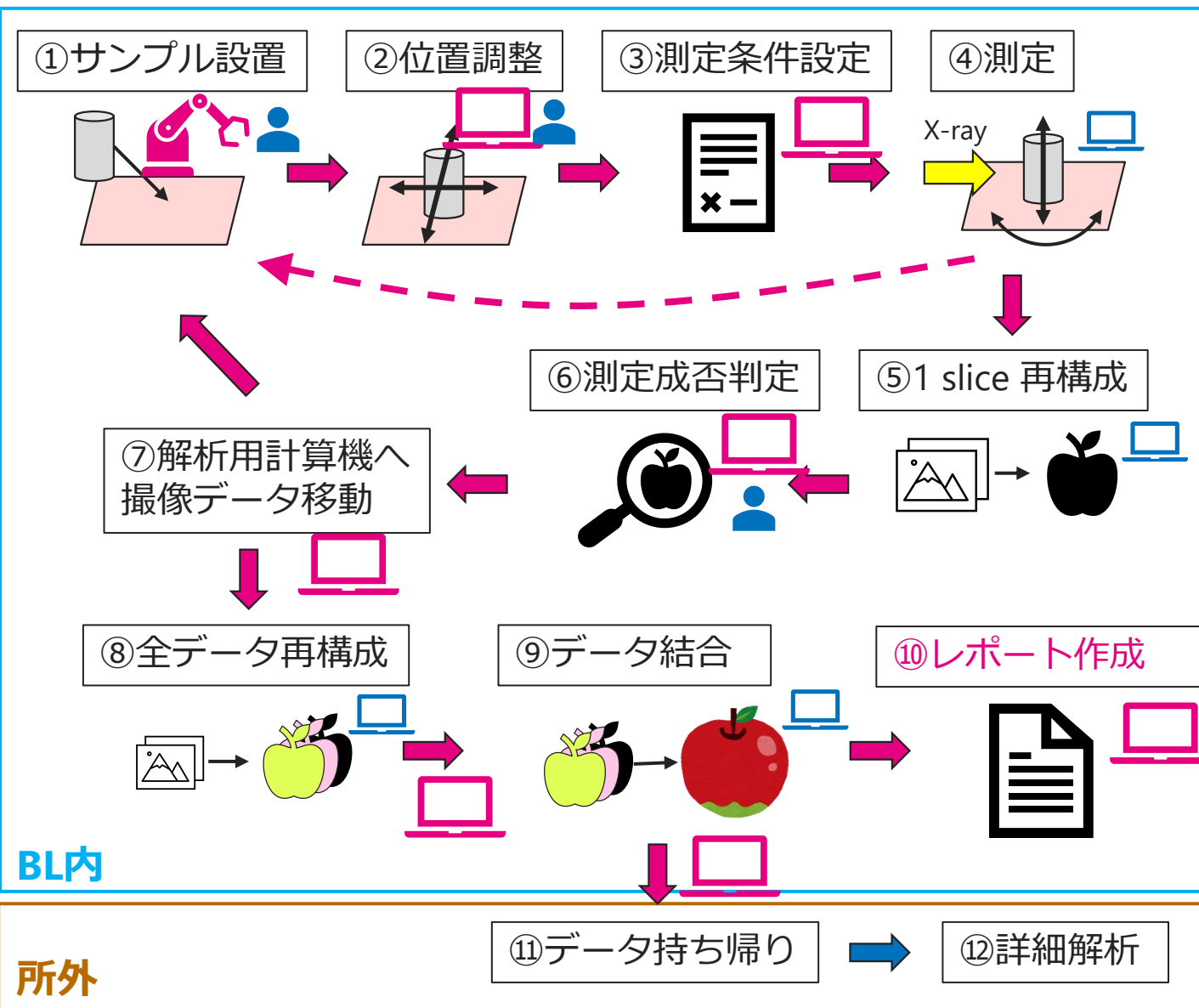




＜現システム＞

- 測定から再構成まで各BL内で可能。
- 操作および設定は**手動**。
- システム上、サンプルとデータを紐づけるのは、**測定開始に設定される保存先ディレクトリ名のみ**。
- 測定条件は**ユーザー自身が記録する**。
- 詳細解析は**ユーザーが持ち帰って実施**。

測定サンプル数（スライス数）は、**測定に慣れたユーザーでも 30~70/BT**。

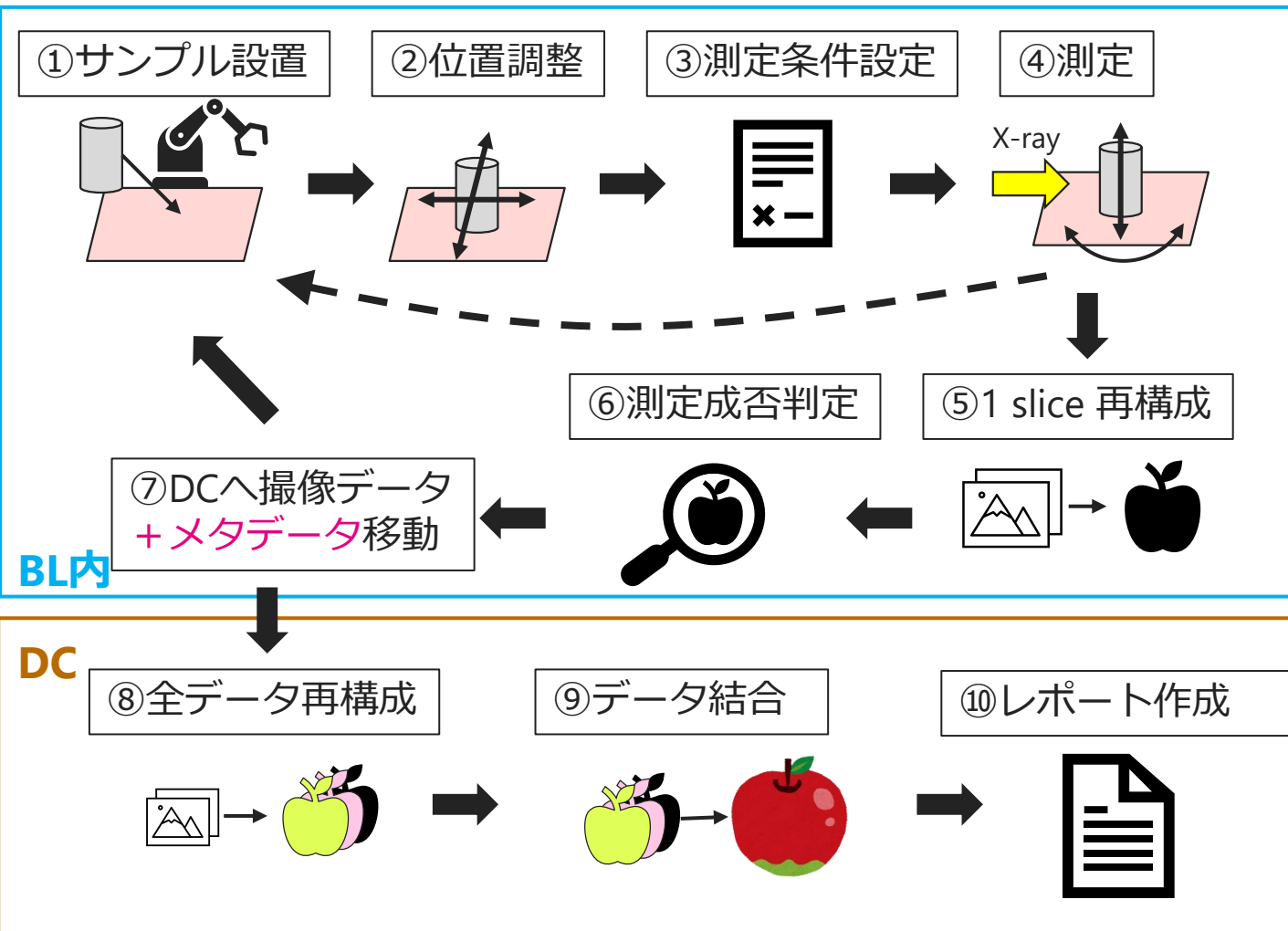


<自動化> ※最初は BL28B2 で実装。

- ・ **測定から再構成まで** (ほぼ) **自動**。
 - ・ 定形ホルダ利用時、サンプル交換はロボットが行う。
- ・ **メタデータ** (サンプル情報と測定条件) を付与。
 - ・ より管理・再現しやすくする。
 - ・ 申請内容と測定結果の整合性の取り方については協議中。
- ・ **測定・再構成結果のレポート**を自動作成。

定形ホルダ利用時、測定サンプル数は**ユーザーによらず** 30~70/BT を目指す。

(1 TB/day のデータ生成の見込み。)



撮像データはSP8データセンター（DC）に自動転送される。

⇒撮像データも再構成データも実験後に持ち帰る必要はない。

- DCでのデータ保持ポリシーにより“Hot”に利用できる期限はあるかも。

以下の処理をDC上で自動的に行う。

- 全再構成
 - フィルター処理・規格化を含むかも
- 再構成データ結合
- レポート作成

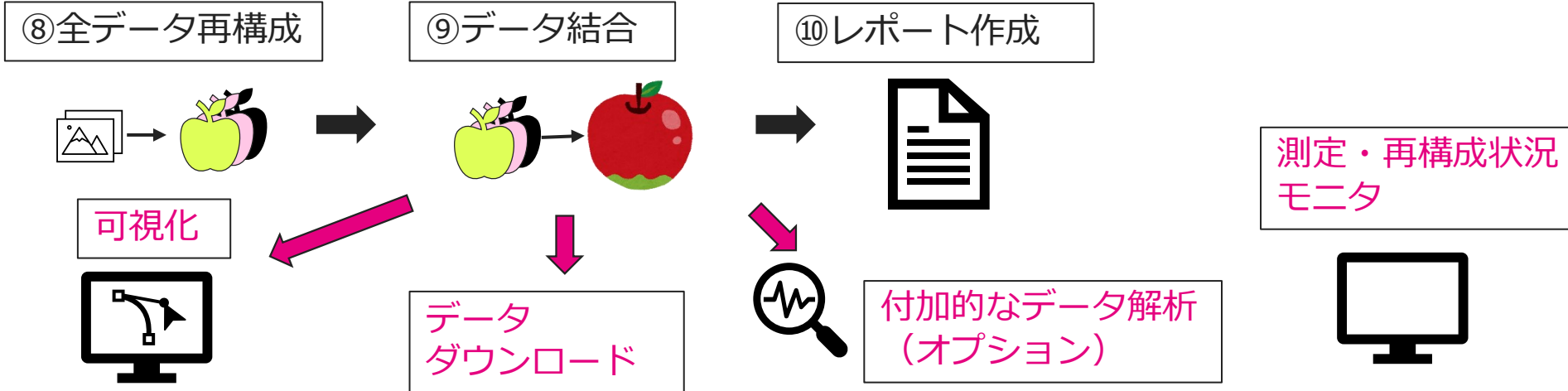
データ処理に関して、再構成まではユーザーのスキルに依存せずスムーズに行われるようになる。

どのユーザーもDCのクラスタ計算機を用いて詳細解析することが可能。

以下の機能もDC上で提供予定。

- **再構成結果のビューアー**
 - メタデータとの関連付けて検索可能に。
 - 撮像データ用ビューアーの必要性は別途検討。
- **再構成データダウンロード**
 - 縮小、トリミング、フォーマット変換なども利用可能。
- **測定・再構成状況モニタ**
- 付加的なデータ解析（オプション）の提供も検討。
 - セグメンテーションなど。

DC



	2022A	2022B	2023A	2023B	2024A
SP8 DC		システム導入	調整	供用開始	
自動X線CT	定形ホルダ 利用開始	B版システム リリース	Ver1.0 リリース		
メタデータ	構想中	A版@BL内	B版@BL内	Ver1.0@BL内	Ver1.0@DC
レポート	構想中	A版@BL内	B版@BL内	Ver1.0@BL内	Ver1.0@DC
他ソフト開発、 SP8 DCとの連携	構想中	Proto機で テスト	A版	B版	Ver1.0

- 解析環境に関するニーズ：
 - 実験環境との相互作用の強化
 - ユーザーの環境、外部の計算環境、データ共有環境とのスムーズな連携
 - ユーザーの持つスキル・資源に依存しないデータ処理・解析環境
- SP8の各BLとSP8 データセンター（DC）との連携により上記ニーズに応えていく。

- 自動X線CT BLでの当面の目標：
 - 撮像データがメタデータ付きでDCに保持される。
 - 再構成がDC上で自動実行。結果はDCからダウンロード可能。
 - 測定・再構成状況を所外から確認可能。
 - レポートが自動生成され、DCからダウンロード可能。
 - 解析環境としてDCの計算資源をどのユーザーも所外から利用可能。

