

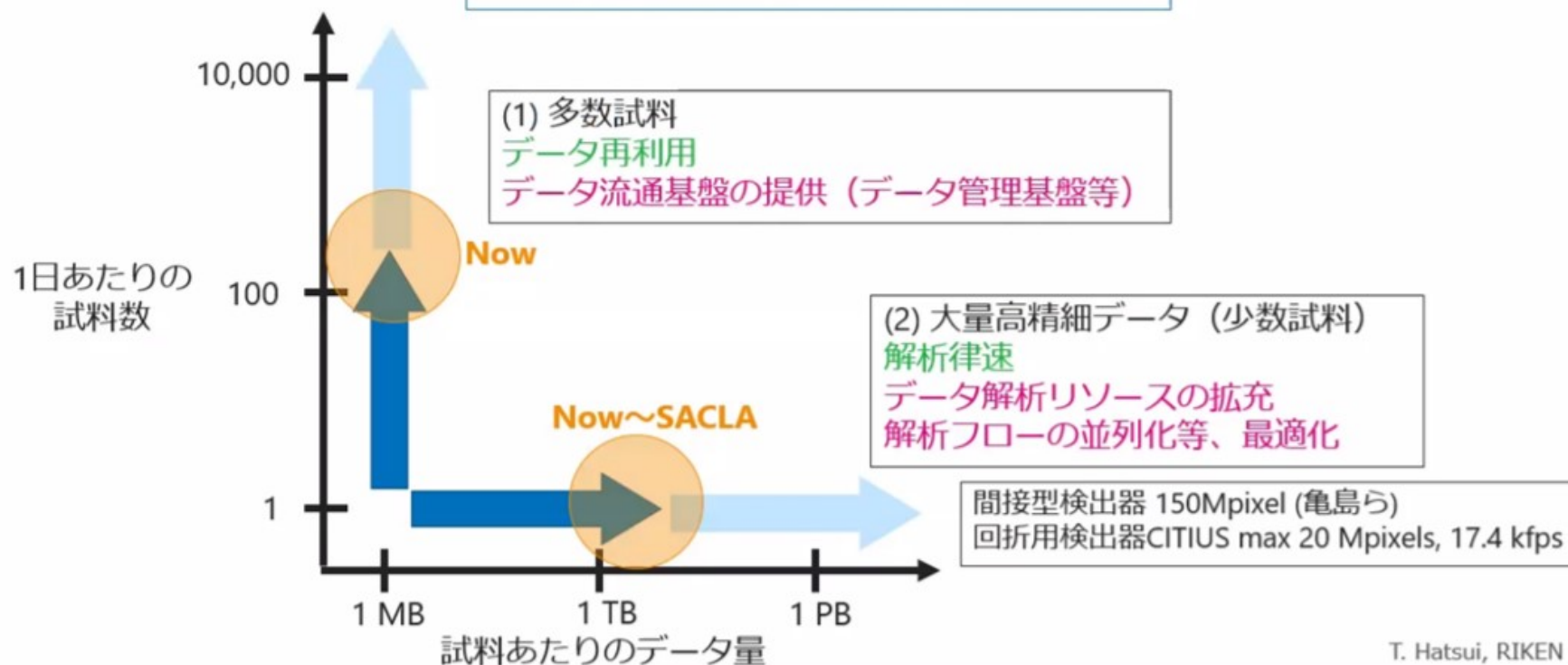
SPRUC 第4回 BLsアップグレード検討ワークショップ
2022年03月14日(月) 09時00分から16時10分

13:45-14:10 <データ> (座長 SPRUC幹事 藤原秀紀)
SPring-8 データセンター構想 (理研/JASRI 初井宇記)

データ増大に関する類型化と課題、検討中の機能

SPRING-8・放射光技術の進歩

- 光源
- 要素技術（X線光学、検出器、ロボット...）
- エンドステーションの自動化...



2) 大量高精細データに対する トータルソリューション

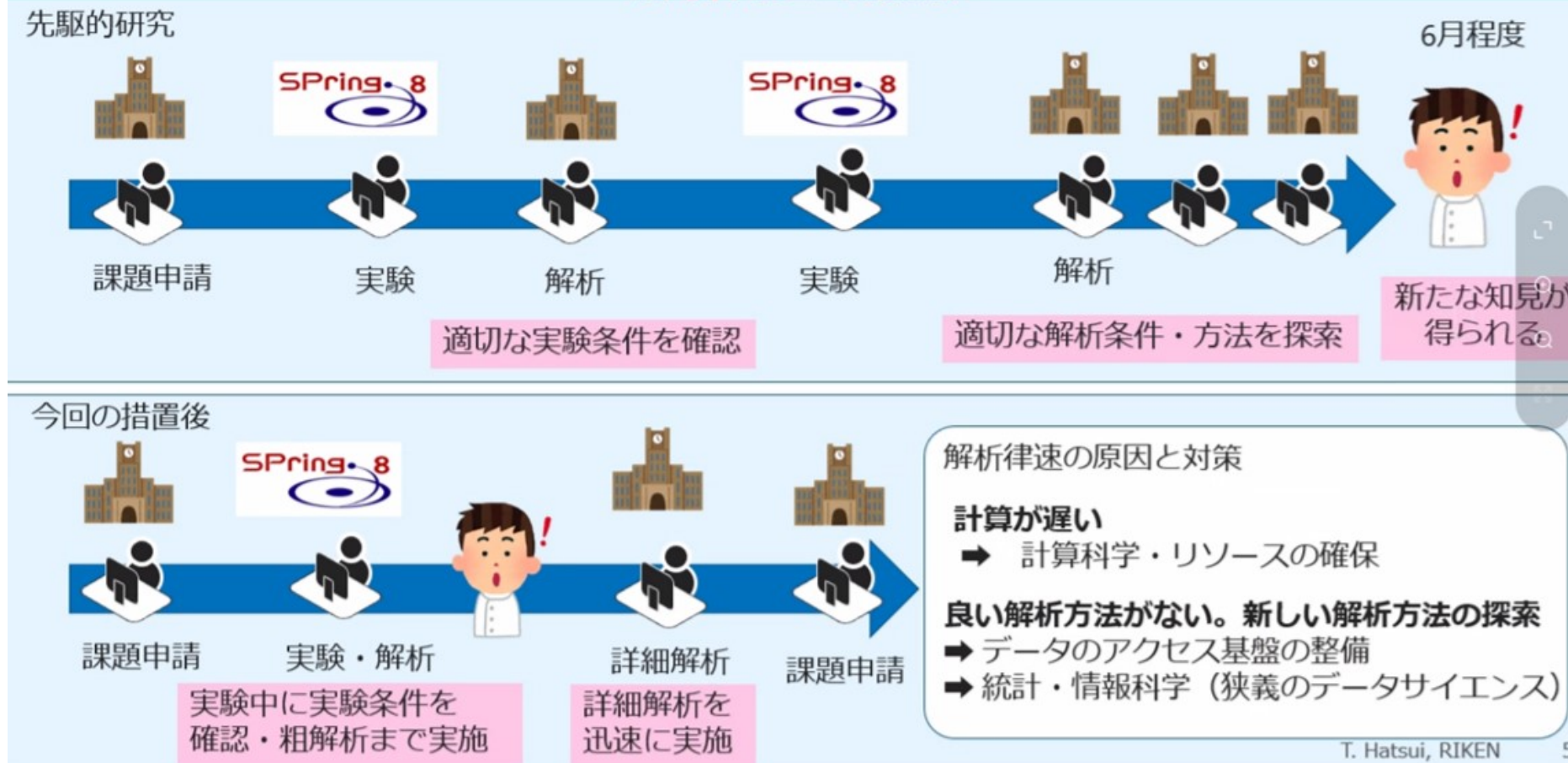
第1回データワークショップ「SPring-8データセンター構想によるデータ解析力の強化」

開催日時: 2021年12月13日(月) 14:30~17:00

開催形式: Zoomによるオンライン開催

<https://dncom.spring8.or.jp/workshop/data-workshop/index.html>

2) 大量高精細データ 解析律速の解消



先駆的研究



今回の措置後



先行研究の蓄積をもとに候補となる解析条件をいっぺんに探索

- 並列計算基盤KMRを活用(SACLAデータセンターで利用実績あり)
- スパコン「京」時代に開発された技術。簡単に大規模並列化可能

非専門家にも利用できる解析フローへ

GUIを整備し、非専門家への敷居を下げる

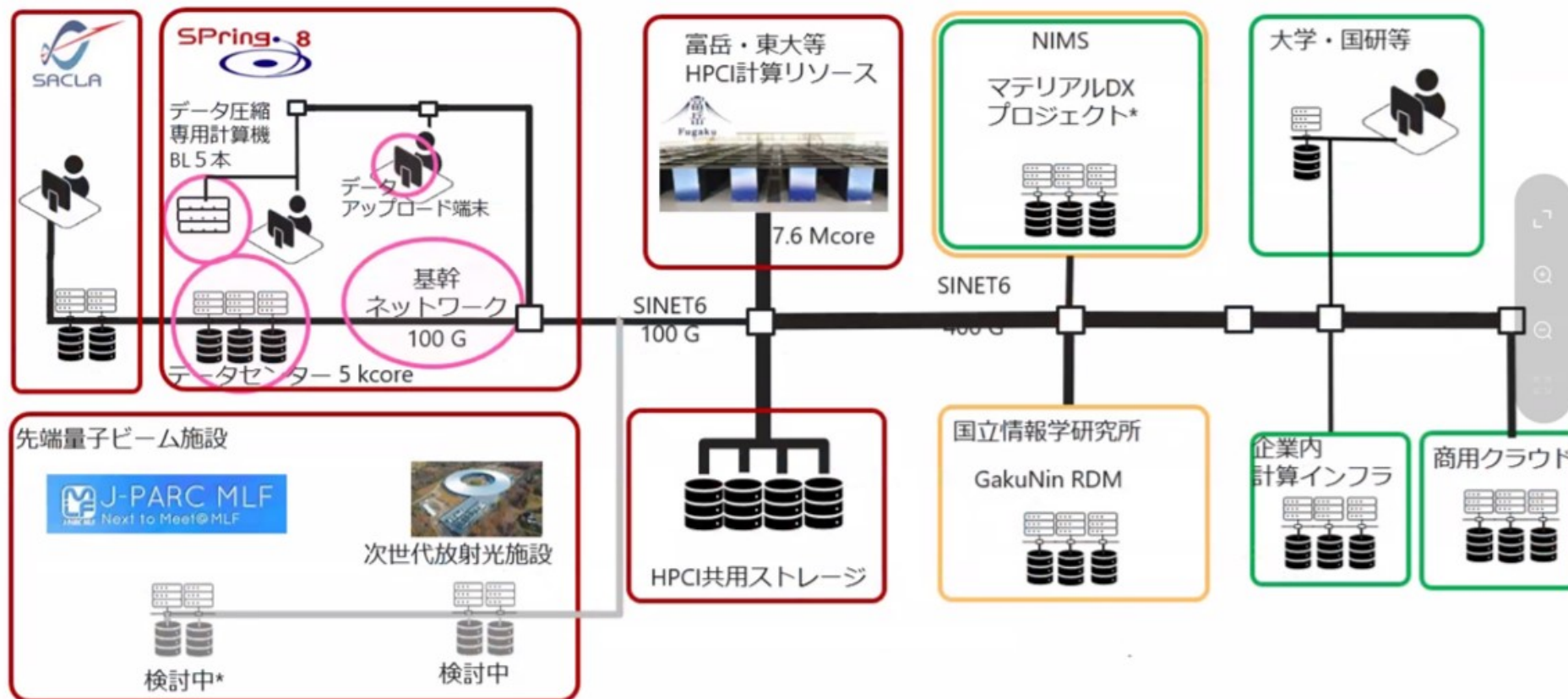
データ共有

情報科学研究者とのデータ共有を促進、共同研究を支援

Run ID	Status	Total Size	Processed Size	Date	Index
000001	Completed	1.0GB	1.0GB	2014-01-01	1
000002	Completed	2.0GB	2.0GB	2014-01-02	2
000003	Completed	3.0GB	3.0GB	2014-01-03	3
000004	Completed	4.0GB	4.0GB	2014-01-04	4
000005	Completed	5.0GB	5.0GB	2014-01-05	5
000006	Completed	6.0GB	6.0GB	2014-01-06	6
000007	Completed	7.0GB	7.0GB	2014-01-07	7
000008	Completed	8.0GB	8.0GB	2014-01-08	8
000009	Completed	9.0GB	9.0GB	2014-01-09	9
000010	Completed	10.0GB	10.0GB	2014-01-10	10

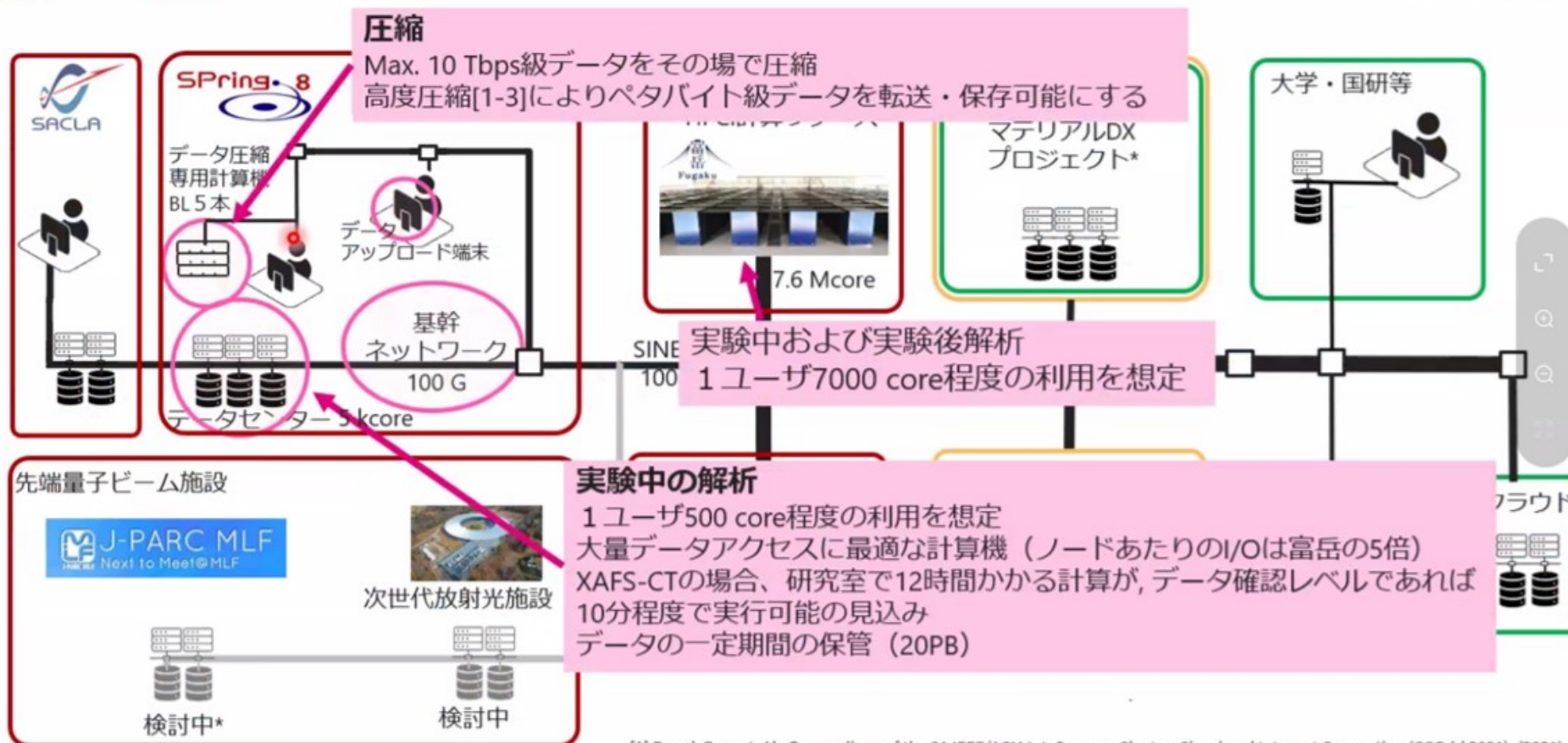
SACLAのバイオ系ユーザが利用しているGUIの例。非スパコンユーザでも100万枚規模の画像解析が可能

SPring-8データセンター構想: インフラ



* クラウド上に設置を予定

SPring-8データセンター構想: インフラ



* クラウド上に設置を予定

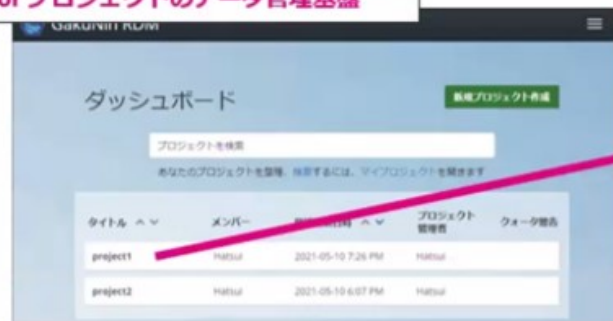
[1] Rupak Roy, et. Al., Proceedings of the 21 IEEE/ACM Int. Sym. on Cluster, Cloud and Internet Computing (CCGrid 2021), (2021).
[2] 平木俊幸、初井宇記、特願2021-175897
[3] T. Hiraki et.al., in preparation

T. Hatsui, RIKEN

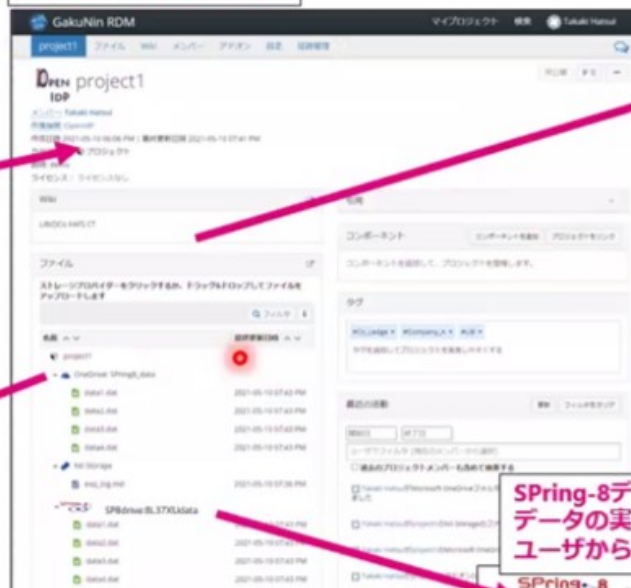
8

1A-3) NIIのGakuNin RDMでデータを管理

NII上に構築されているGakuNin RDM
or プロジェクトのデータ管理基盤



研究テーマごとのページ



SPring-8データセンター or
HPCI共用ストレージ上（新規）
実験中の解析を最先端のアルゴリズムで
実施できる（ユーザのブラウザ上）

フローベース：一般ユーザ向け



コマンドベース：専門家向け



所属機関サーバ
プロジェクトサーバ
商用クラウド

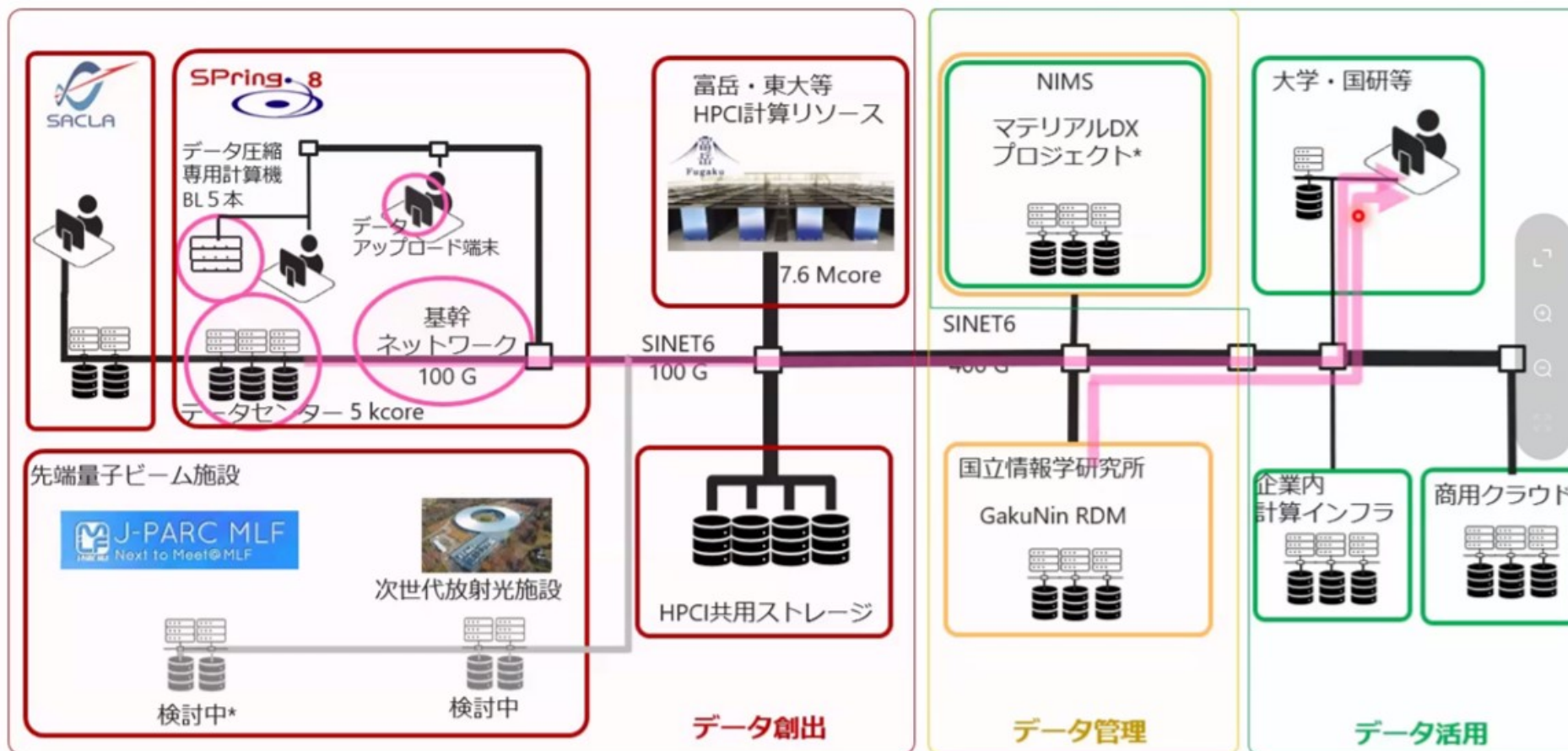
SPring-8データセンター or HPCI共用ストレージ上（新規）
データの実体はSPring-8が管理するサーバにあるが
ユーザからは手元にあるように見える

SPring-8 SP8drive:BL37XUdata

data1.dat	2021-05-10 07:43 PM
data2.dat	2021-05-10 07:43 PM
data3.dat	2021-05-10 07:43 PM
data4.dat	2021-05-10 07:43 PM

研究室の他の装置のデータと一括管理が可能
SPring-8のデータは実体はSPring-8にあるが参照先が分かる。
解析もここからアクセス可能とする予定
データ公開機能もあり

SPRING-8データセンター構想: インフラ



* クラウド上に設置を予定

1) 多数試料系

第2回データワークショップ「SPring-8データセンター構想とMDXプロジェクトとの連携」

開催日時: 2022年2月25日(金) 13:00-17:00

開催形式: Zoomによるオンライン開催

<https://dncom.spring8.or.jp/workshop/data-workshop/index.html>

第3回データワークショップ「SPring-8データセンター構想におけるデータ流通基盤」

開催日時: 2022年4月18日(金) 13:00-17:00

開催形式: Zoomによるオンライン開催

1)多数試料：材料の基礎特性の精密データ

- 多くの試料の高品質データ
 - データの収集段階が大変
 - 個別に集めなくてはならない
 - データの使用許諾をとってもデータを集める手間が大変
 - データもそれぞれ違った形式で保存されている
- 今回の措置
 - データの流通基盤を提供
 - web経由でデータアクセスが可能 (Amazon S3互換サービスを想定)
 - メタデータ (ワークショップのwebでたたき台を公開中)
 - 試料：記述方法についてはユーザコミュニティと連携 (NIMS、放射光学会、先端量子ビーム施設)
 - 実験条件：施設側で整理
 - データ公開
 - 国立情報学研究所(NII)が進めている研究データ基盤、NIMSと連携

SPring-8データセンター構想の動機・現状および検討中のサービス

データの創出基盤・管理基盤・活用基盤の整備

利用シーン

- 1A-1) NIMSのMDXプロジェクトへデータを登録
- 1A-2) SPring-8データセンターへデータを保存
- 1A-3) NIIのGakuNin RDMでデータを管理
- 1A-4) 企業内サーバーでデータ活用
- 1A-5) 商用クラウドでデータ活用
- 1A-6) データベース作成・公開へのサポート

1B) 複数量子ビーム施設のデータをアクセス

新設

1A-1) NIMSのMDXプロジェクトヘデータを登録



データアップロード専用の
端末(DUC)を設置

NIMSのアップロード用webサーバーの画面が表示される。

SPring-8のビームラインにデータアップロード専用の端末(DUC, 新設)を設置。この端末からNIMSの運営するサーバー（商用クラウド上）にアップロードする。

SPring-8 Data Flow Systemのユーザー画面（検討中の案）

Login

SPring-8 Data Portal

User

Password

Login

OR

Login with XXX

SPring-8データセンターのID

富岳IDでログイン
J-PARC, 次世代放射光IDでも
ログインできるように拡張予定

Upload

SPring-8 Data Portal

Upload

Proposal ID
XXXXX

Sequence ID
XXXXX

measurement metadata
XXXXX.json

sample metadata
XXXXX.json

Sequence Data

Drop your file(s) here, or browse file(s)

Upload

xxx MB

データの認証・認可
先端大型施設間で連携

メタデータとデータをま
とめてアップロード

Download

SPring-8 Data Portal

Sequence ID	Date Created	Size (MB)	Owner
XXXXXXXX	2022-01-01	10	User-A
XXXXXXXX	2022-01-01	11	User-B
XXXXXXXX	2022-01-01	15	User-C
XXXXXXXX	2022-01-01	21	User-D
XXXXXXXX	2022-01-01	21	User-E

Download

Metadata

メタデータを閲覧

Access Controlへのリンク

データへのリンク(REST API)

データの表示(jupyter notebook)

データをまとめてダウンロード

Access Control

SPring-8 Data Portal

Sharing Settings

Proposal ID
XXXXXXXX

Sequence ID
YYYYYYYY

Sharing

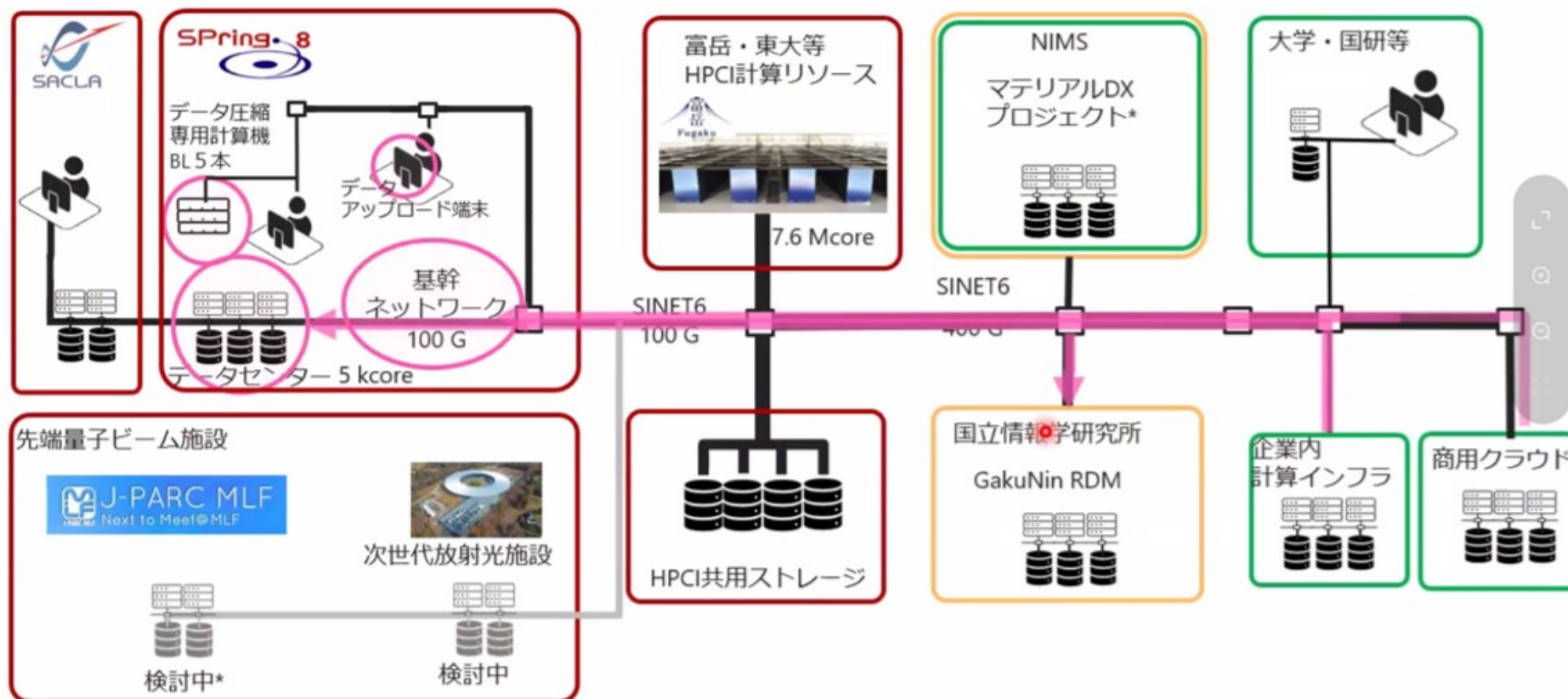
Everyone Limited

User	Access	Action
User-A	Manage	Share
User-B	Manage	Share
User-C	Manage	Share

Close

実験責任者はデータのアクセス
権限を設定することができる。
詳細はユーザーコミュニティ・
放射光学会と協議

SPRING-8データセンター構想: インフラ

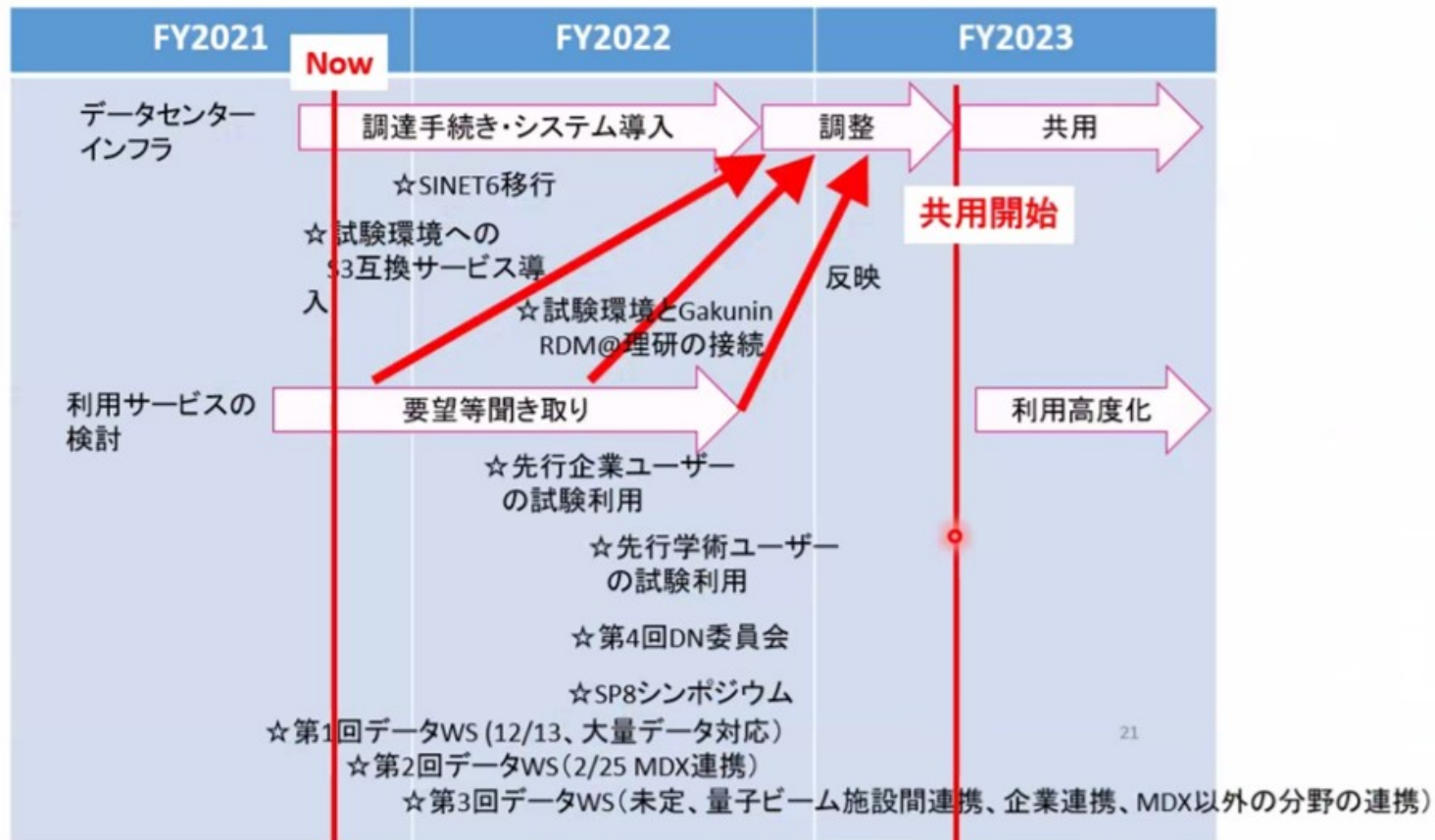


* クラウド上に設置を予定

新設

2021年度補正予算

スケジュール



まとめ

- 類型化と課題、施設側で検討中の機能
 - (2) 大量高精細データ（少数試料）
 - オンサイトのデータセンター（新設）と富岳等スパコン
 - 使いやすいツール
 - 5ビームラインについて整備
 - (1) 多数試料
 - 国が進めている研究DXと歩調を合わせて整備
 - 国立情報学研究所NIIのGakuNinRDMとの接続
 - 物質・材料研究機構NIMSの材料DX拠点との連携
 - 量子ビーム施設間の連携
 - 共通
 - 情報科学研究者との共同研究の促進と成果の共有
 - ← データの流通基盤
 - 企業ユーザのニーズへの対応
 - ← データの流通基盤、データ前処理
 - スケジュール
 - 2023年夏から共用を開始
- お願い：是非ご意見をお寄せください。
 - データセンターへのご要望
 - 試験利用への参加をいただける方
 - データセンター機能の利用を想定される専用BL
 - MDXへ参画されるBLは必ず

- ご連絡先
dncs@spring8.or.jp
- データ・ネットワーク関連の情報
<https://dncom.spring8.or.jp/>
- データワークショップ
<https://dncom.spring8.or.jp/workshop/data-workshop/>
- 第3回データワークショップ
「SPRING-8データセンター構想におけるデータ流通基盤」
2022年4月18日(金) 13:00-17:00