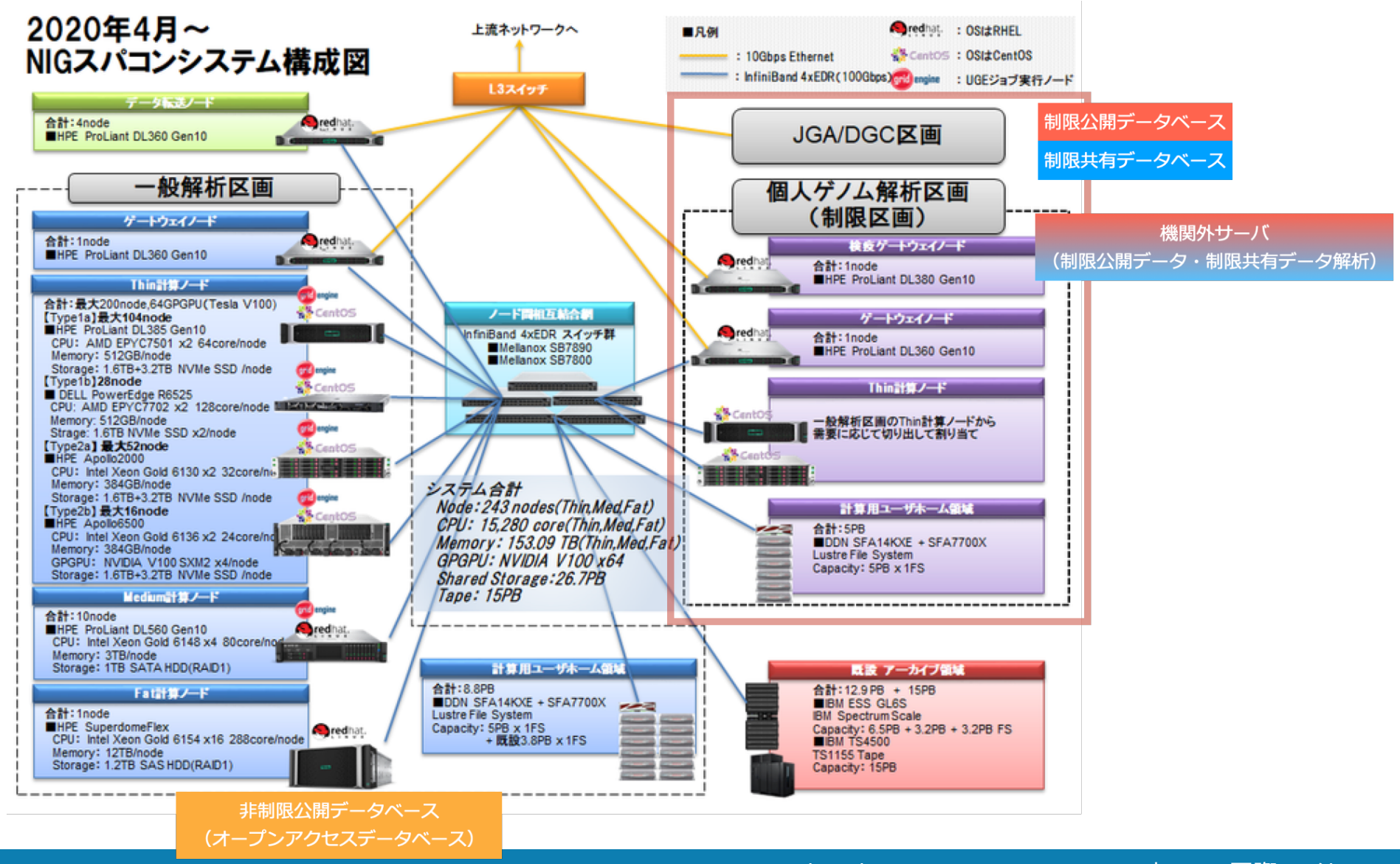
3 DDBJサービスの統合化を目指して

○藤澤貴智、小笠原理、中村保一、有田正規

国立遺伝学研究所　生命情報・DDBJ センター

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

DDBJ：サービスする解析プラットフォーム



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

DDBJ：サービスするデータベース

www.ddbj.nig.ac.jp/services/index.html?tag=DDBJ,database,submission

DDBJ　サービス　スパコン　統計　活動　センターについて　DDBJ Web Sites　利用規約　問合せ　English

サービス

ホーム > サービス > サービス

News カスタム検索

タグ

- database 10
- search 14
- submission 10
- analysis 5
- annotation 1
- DDBJ 19
- DBCLS 8
- BINDS 1
- NBDC 2

非制限公開データベース（オープンアクセスデータベース）

BioProject
database　submission　DDBJ
BioProject はプロジェクトに由来する生物学的データをとりまとめ、関連するデータへのリンクを提供します。

BioSample
database　submission　DDBJ
BioSample は実験で使われた生物学的サンプルに関する情報を収集しています

DDBJ
database　submission　DDBJ
アノテーションが付与されたゲノム、遺伝子、転写産物の塩基配列

DRA
database　submission　DDBJ
DRA は高性能シークエンス機器からの生シークエンスデータとアライメント情報を格納しています。

GEA
database　submission　DDBJ
GEA は MIAME に準拠したアレイとシークエンスベースの機能ゲノミクスデータを受付けています。

JGA
database　submission　DDBJ
JGA は個人の遺伝型と表現型に関する情報を共有するためのアクセス制限データベースです
制限公開データベース　NBDC

JVar
database　submission　DDBJ
ヒトゲノムバリアントのためのアーカイブ

MetaboBank
メタボロームデータおよび解析データのためのレポジトリ

AGD (AMED)　**SHD (SIP)**
制限共有データベース　NBDC

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

DDBJがかかえる課題とアプローチ

マイクロサービス化

- 異なる仕様・実装・システムで構成されたDBが多すぎる
  - アカウント（スパコン、データベースサービス利用者）
  - システム（レポジトリ・検索）
  - メタデータ・データ種別（塩基配列・配列リード・functional genomics・メタボローム）
  - アクセス制限（非制限公開、制限公開、制限共有）とその組合せ

アカウント統合

DDBJ Search

広域登録システム

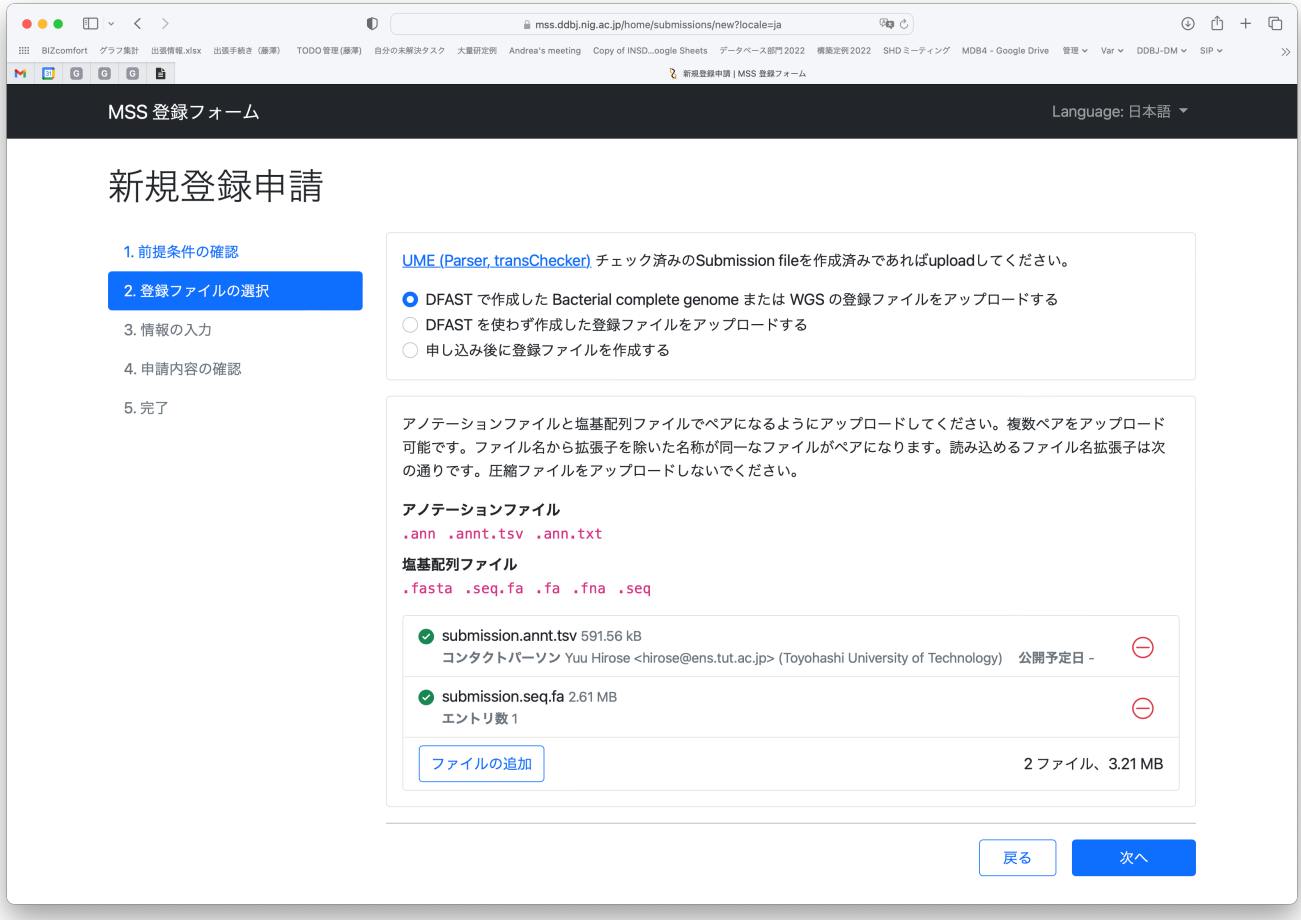
データ種別・入力形式・バリデーション

ネットワーク区画

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

ゲノム等の大規模配列登録システム

- 2022年6月に正式リリース
- 基本操作は、DDBJサービス利用ユーザ（D-way）アカウントでサインインし、前提条件の確認後に登録形式ファイルをドラッグ＆ドロップするだけ
- メールでのデータ受領を廃止し、登録者・DDBJキュレーター両者のコミュニケーションコストを軽減



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

データベース登録ナビゲーション

- Q&A方式で生物種やデータ種別を絞り込み選択してもらうことで、登録したいデータはどのデータベースへ登録が必要かをナビゲーション
- DDBJホームページに組み込み、URL参照可能なようにパーマリンクを生成

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター

今後の展望

- ゲノム等の大規模配列登録システムを発展させて、他のデータベース登録に拡張、既存システムの段階的にリプレイス
- データベースやデータ種別、生物種に連動して登録に必須項目が変わるため、まずは、メタデータ登録形式ファイル生成のための定義情報の再構築
- 大規模データのアップロード用のscpやsftpや高速転送プロトコルによるインターフェースとの連動、スパコンアカウント連携
- データ・メタデータのバリデーションツールや査定内容の公開

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　生命情報・DDBJセンター