

生命科学系データベースアーカイブのインフラ・バックアップ体制の強化

八塚茂¹、栗原英輔²、加藤健弘²、井上圭介³、畠中秀樹⁴

1 国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター

2 株式会社日立製作所

3 株式会社日立社会情報サービス

4 大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター

生命科学データベースとは

<https://dbarchive.biosciencedbc.jp/index.html>

生命科学系データベースアーカイブは、国内のライフサイエンス研究者が生み出したデータセットをわが国の公共財としてまとめて長期間安定に維持保管し、データ説明(メタデータ)を統一して検索を容易にすると共に、利用許諾条件などの明示を行うことで、多くの人が容易にデータへアクセスしダウンロードを行えるようにしている。データを長期にわたり保全し、データベース作成者のクレジットを明示する一方、公的機関や民間等様々なユーザが利用しやすい形にすることで、それぞれの研究の生命科学へのいっそうの貢献を支援する。当アーカイブはNature publishingのScientific Data誌の推奨レポジトリに認定されている。

主な特長

- NBDC「統合化推進プログラム」で産生されたデータベースも含め、145以上のデータベースをダウンロード可能
- 各データベースのメタデータが付与されており、データの再利用が容易
- 各データベースの利用許諾は、明確かつシンプル(CC BY-SA ライセンス等)
- 各データベース及びそのサブセットには DOI (Digital Object Identifier) が付与されており、論文等で簡単に同定・引用できる
- 多くのデータがRDF化対応済み
- Scientific Data誌の推奨レポジトリに認定

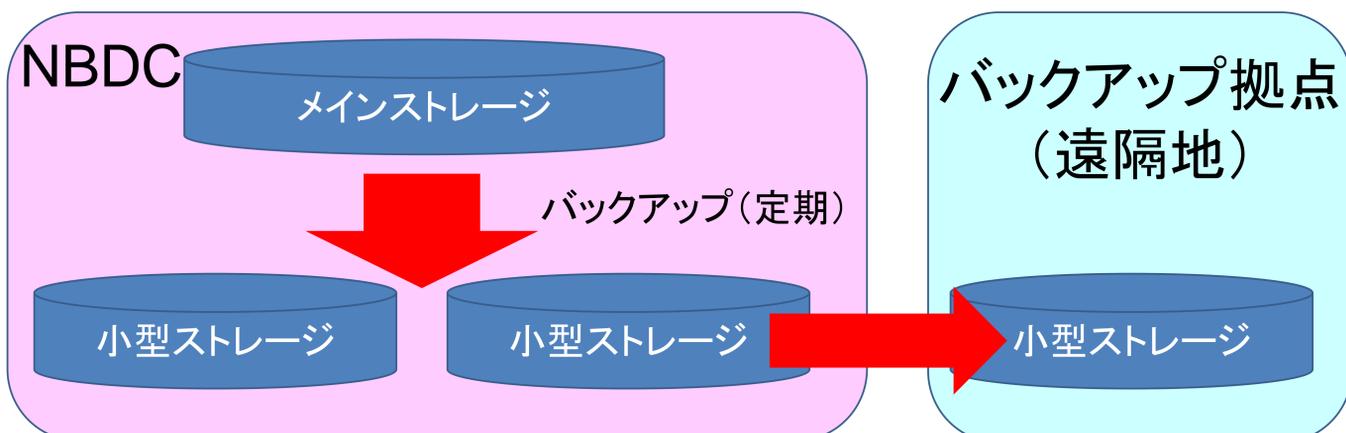


インフラの強化



- 最大約500TBのデータを格納できる大型ストレージ(RAID5)をメインストレージとして導入
- 予備(データ加工作業、メインストレージの補完等)用のストレージも準備

バックアップ体制の強化



- JST/NBDC(東京)が地震等により被災した場合を想定
- データを定期的に運搬可能な小型ストレージにバックアップ
- 小型ストレージは遠隔地にあるバックアップ拠点で保管