

AI for Life Science 基盤としての 統合データベースの在り方について

有田 正規
国立遺伝学研究所

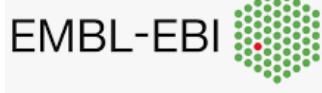
現在のオープン＆クローズ戦略

誰にでも公開
(非制限公開)

審査つき公開
(制限公開)



国際連携



学術出版による
要請が駆動力。

研究者は登録に
協力的ではない。

どう継続？

アカデミアも企業も、
殆どのデータは非公開

バイオバンク
など個人情報

クローズド

ヒト情報以外もアクセス制限

データ主権 (data sovereignty)： 生物多様性条約 (CBD) 、
パンデミック協定において、**生物資源は国の財産**

CBDでは、各國が生物遺伝資源の窓口とするクリアリングハウスの設置を義務付け。提供国の明示は必須。

ABSCH ACCESS AND BENEFIT-SHARING CLEARING-HOUSE

HOME ABOUT SEARCH SUBMIT COUNTRY PROFILES HELP FORUMS NAGOYA PROTOCOL

CBD / ABSCH / Country profiles / Japan

Russian Federation

Japan

Party to the Nagoya Protocol

1	ABS National Focal Point
0	Competent National Author
2	Legislative, Administrative
0	ABS Procedure
0	National Model Contractua
0	Internationally Recognized
1	National Websites or Datab
1	Checkpoint
27	Checkpoint Communiqué
0	1st National Reports on the
1	Interim National Reports or

<https://absch.cbd.int/en/countries/JP?status>All>

日本の情報
(CBD)

公式なウェブ
サイトは環境
省のABSページ

同様の議論が、パンデミック協定 (PA) 、公海条約 (BBNJ) 、作物ゲノムの条約 (PGRFA) でも進行中

統合DBに求められる条件

目標：誰もがデータを利活用できる社会

- 無償で簡単に目的データを探せる (FAIR)
 - 検索用インデックスがないと駄目
 - メタデータ記載の統一が必要
 - 検索やキュレーションはAIが支援
- 提供者や利用条件等の明確化 (未対応)
 - 提供国や明記と利用条件は、今後必須
 - 例外処理が多く、自動化も難しい
 - データベース間の連携も必要
- AI時代に向けた人材育成

ナショナルと呼べる組織の必要性

諸外国から求められるもの：

- ・日本が提供する生物資源リスト（一次データ）
- ・利用条件等の「担保」

企業がフリーソフトを使えないように、フリーゲノムも使えない時代に突入。

国内で必要なもの：

- ・新データに対応できる基盤作り
- ・データベース間の人材流動と育成

個人ゲノムは社会保障の枠組みに入っていく。

データ提供・整理に対するインセンティブの設計。

リポジトリの重要性

研究データの保存と公開（ライセンシング？）

- 研究倫理、再現性
- AI時代には「学習可能データ」が生命線

データ出版時代に見合うインセンティブ

- データセンター、認定ブローカー
例：先進ゲノム支援、TRIP-AGIS
- 登録作業の分業、体系化
例：wwPDB、司書

議論から抜けている「図書館」

- データ基盤・登録における主役のはず
文献とデータをつなぐ必要あり

求められる統合化

一次データを AI-ready な形で蓄積

- 原産国、提供者などの正確な情報は必須
- インターフェースは AI

全世界の研究者に対応

- 海外機関と協力して網羅することが重要
- 登録のインセンティブやキュレータ資格の設計

ELSI（倫理、法、社会問題）の考慮

- 国によって異なる事情にも対応
- 日本語に対応できれば他の母語もOK