

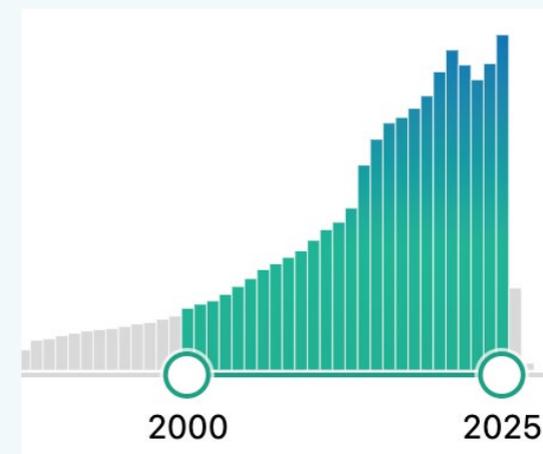
第3回 基礎・横断研究戦略作業部会

分野横断的に利活用可能なデータベースの構築の必要性 —AI for Scienceに向けて—

基礎・横断研究戦略作業部会
委員 早稲田大学 坂内博子

現状

- 国際的な論文数の増加
- 論文に付随するデータ量の増加
- さまざまなモダリティのデータ
(遺伝子、イメージング、化合物、医療データ)



Pubmed
keyword:biology

サステナブルにデータを整備し、利活用できる体制の構築、
知的財産・生物遺伝資源の保護が求められている

課題

- データベースの維持・管理が個々の研究機関
- 数年ごとの競争的資金により運営されている

不安定、疲弊
スピードについていけない

- モダリティごとにデータの登録法、フォーマットが異なる
- AI-readableでない
- データ・メタデータ登録がデータ取得者にとって容易ではない

データベースの拡充・
統合が困難
分野融合の壁

- 生物資源保護、個人情報保護

何を保護し、何をオープンにするかの戦略が必要

次期NBRP、NLDP、BINDSに必要なと思われる要素

データベースは国策として安定に運営

- データベースを運営する研究機関が連携する
- 「DDBJ, PDBj, jPOST, SSBD, DBCLS等を統合的・戦略的に運営する組織を構築」
- NBRPのバイオリソースのデータを理想的な形で格納

AIを活用した、データ登録・検索インターフェースの構築

- LLMを用いた対話型インターフェースにより、データ・メタデータ・研究キーワード等を収集
- AI-readable な、国際基準に適合するフォーマットでデータを格納
- デジタル実験ノートにより、実験時にメタデータを登録
- キーワード検索により、関連すれば異なる分野の研究データも閲覧できるように

知的財産、生物遺伝資源、個人情報保護

- 何を保護し、何をどのような形でオープンにするか早急に議論を
- 医療データ匿名化