

「研究データの公開・共有という用語について」

令和2年11月2日

国立遺伝学研究所 有田正規

研究データは公開という原則

生命科学論文が扱う研究データは、公開が前提である。塩基配列、立体構造、オミックス情報等はそれぞれの公的リポジトリから公開しないと論文も受理されない。

データ公開に関する背景

米国では1980年代にデータ公開 (data sharing) に関する議論を経て、国際ヒトゲノム計画が発足。ヒトゲノム情報は論文に関わらずデータ取得後即時公開を原則とした (バミューダ原則 1996)。

日本は2000年から科学技術会議でデータベース戦略が練られ、2008年に内閣府と4省で統合データベース構築のロードマップを策定。2011年にJSTのNBDCが発足。

データ利活用における用語の問題点

データ共有という日本語 (非公開を意味) は、英語の data sharing (公開) と対応しない。AMEDのゲノム制限共有データベース AGD は、データアクセスの可否をデータ提供者が判断。

公開形式の分類

無償で公開	パブリックドメイン	どんな利用形態も許す。著作権なし
	フリー	著作者が全ての権利を有し、利用者の権利不明
	オープン	利用者の権利が明記される
	制限アクセス	第三者 (DAC) がアクセスを制御
有償で公開	売買契約が必要	第三者による制限アクセスに近い (例: 学術出版)

国内の「制限共有」は英語でいうと research collaboration が近い。

要点

データアクセスの管理はデータ生産者・利用者から独立した第三者が担い、利用履歴公開がよい。DACやICも統一基準の実施がよい。(例: 科研費 先進ゲノム <https://www.genome-sci.jp/ethic>)

参考となる資料

- RIETI 「日本のバイオテクノロジー分野の研究開発の現状と3つの課題」 (2002)
- 学術会議バイオインフォマティクス分科会からの提言要旨「持続可能な生命科学のデータ基盤の整備に向けて」 <http://www.scj.go.jp/ja/info/kohyo/kohyo-24-t279-1-abstract.html>
- NBDC グループ共有データベースのページおよびヒトデータグループ共有ガイドライン別表 <https://gr-sharingdbs.biosciencedbc.jp/agd-guidelines>

