The State of Open Dataレポート 2023にみる

研究データ公開の国際的状況

宮入暢子
学術情報コンサルタント
nobuko@miyairi.info
http://orcid.org/0000-0002-3229-5662

2023年11月30日 / J-STAGE Data ユーザーミーティング

Agenda

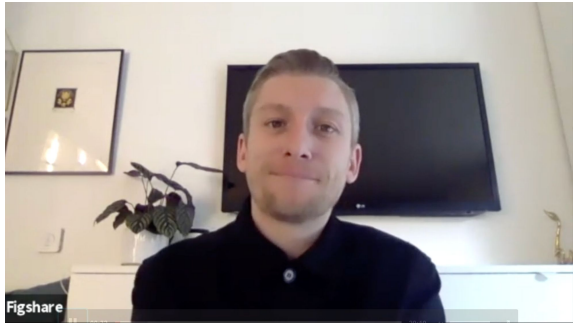
- ★ J-STAGE Dataがプラットフォームとして利用するデータリポジトリFigshareの紹介
- ★ Figshareが毎年発表するThe State of Open Dataレポートに見られる研究データ公開に関する国際的な状況：研究者の視点から

自己紹介

1991-1999	東京外国語大学附属図書館
2000-2002	ハワイ大学マノア校図書館情報学修士
2003-2012	Thomson Reuters (現Clarivate)
2012-2015	Nature Publishing Group (現Springer Nature)
2015-2018	ORCID, Inc. (regional director, APAC)
2018-	学術情報コンサルタント (フリーランス)
2019-	情報通信研究機構ナレッジハブ (招聘専門員)
2019-	筑波大学図書館情報メディア系 (非常勤講師)
2020-	DataCite (APAC expert group member)
2022-	Research Data Alliance (PID IG co-chair; member since 2016)
2022-	青山学院大学コミュニティ人間科学部 (非常勤講師)

【参考】Hahnel, Mark; Miyairi, Nobuko; Nomura, Norimasa (2021): How to make your data repository successful and sustainable: learnings from figshare-powered repositories. figshare. Presentation & Video. https://doi.org/10.6084/m9.figshare.14815311.v1

Figshareとは



- Mark Hahnelが博士課程在籍中に自身の研究データを広く公開することを目指して2011年に創業、2012年よりDigital Science社の傘下に
- 査読済みのジャーナル論文「以外」のさまざまな研究成果の公開と、研究に関わる全ての個人の「研究貢献」を明確にすることが目的
- 研究者個人による利用は無償（容量20GBまで）
- 各アイテムにはDOIを付与し、被引用数やオルトメトリクスを表示

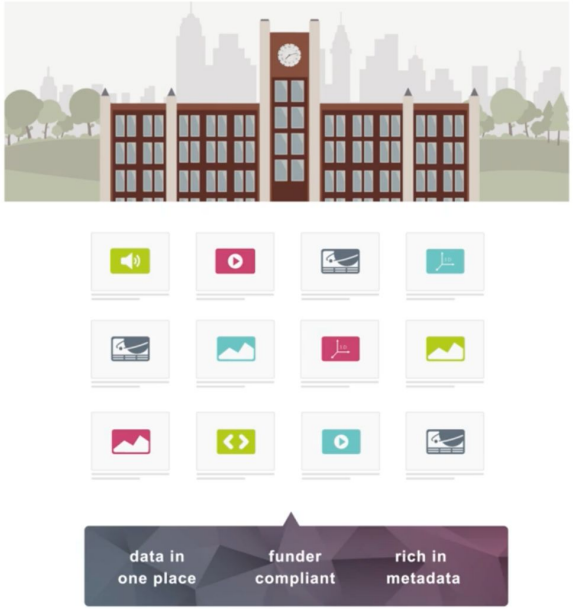
Figshare機関版

- 研究機関や学会、出版社などが利用できるクラウドソリューション
- Figshare.comが提供するリポジトリ機能をカスタマイズ可能に
- メンテナンスや国際標準への準拠、DOIの付与、SSOなど多くの共通課題に対応

An infrastructure that you have to ~~build and maintain~~ simply configure and let run

The Figshare platform is ready to run and constantly maintained, enabling your researchers/authors to store and publish their data with ease. Integrations are available, allowing to use institution IDs and Single Sign On (SSO).

The workflows are made for research data, and customisations and extra services such as curation can be added on.

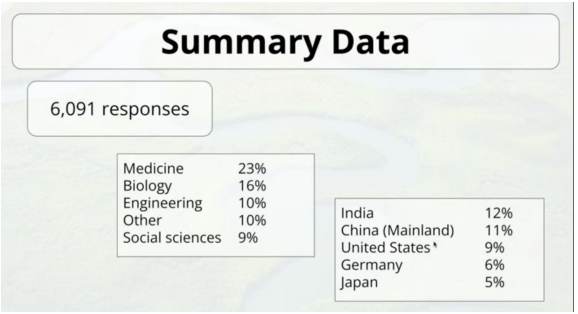


Figshare機関版

https://knowledge.figshare.com/working-with

The State of Open Data

- 全世界の研究者を対象にアンケートを実施した結果をまとめたThe State of Open Dataレポートを毎年公開
- 2016年のアンケート調査の回答者は約2000人→最新の2023年調査の回答者は約6000人
- これまでに延べ約3万3000人が回答しており、研究データに関する調査では最も大規模で長く続いているものの一つ
- 日本からの回答者は313人（インド、中国、米国、ドイツに続く第5位）



Key takeaways

1. **研究者が必要とする支援を受けられていない**
   回答者の約4分の3が、研究データ公開のための支援を受けたことがないと回答。
2. **一様な支援では対応できない**
   専門分野や地域によって回答にばらつきが見られ、研究データ管理には状況に応じた適切な支援が必要である。
3. **固定観念に対する疑問**
   経験の長い研究者は新しいことに反対しているのか？今回の調査では、オープンデータに対する認識や支援のレベルにおいて、研究者としての経験が重要な要因ではないことがわかった。
4. **クレジットに関する継続的な問題**
   本調査を実施して以来8年間、一貫して研究者は研究データの公開について十分な評価を受けていない、という懸念をもっている。
5. **AIに関する認識と行動のズレ**
   今回初めて、ChatGPTや類似のAIツールをデータの収集、処理、メタデータ作成などに利用しているかどうかを尋ねた。

1. 研究者が必要とする支援を受けられていない

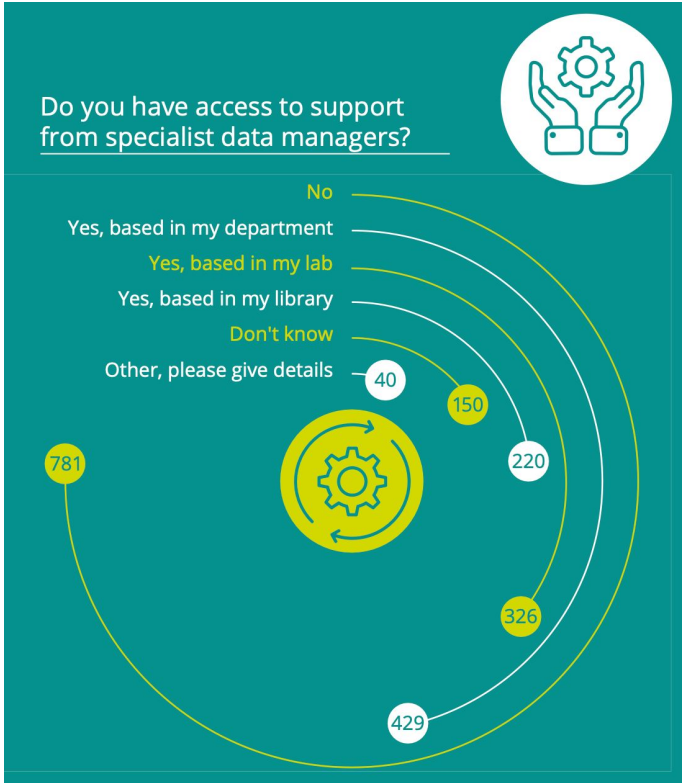
「データマネジメントプラン (DMP)について知っているか？」との質問に41%が「Yes」と回答。そのうち約半数は DMP作成にあたり何らか の支援が受けられる環境にあるが、約 4分の3は支援を受けたことがない、と回答。

支援を受けたことがあると回答した者は、以下の支援を受けている（複数回答）

- 研究室の同僚や上司から 61%
- 図書館から 31%
- 研究支援室など 26%

日本の回答者のうち、83%は実際にDMPを作成したことがない

- そのうち、作成したことがある回答者は、研究室 (16)や図書館 (4)から支援を受けている



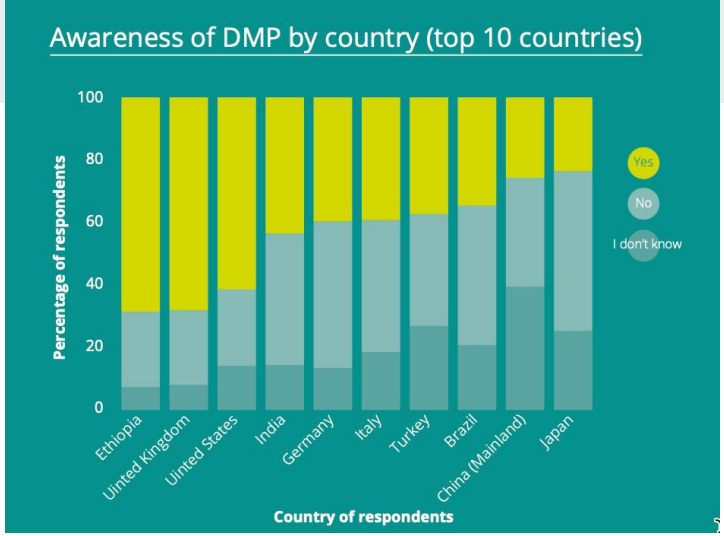
2. 一様な支援では対応できない

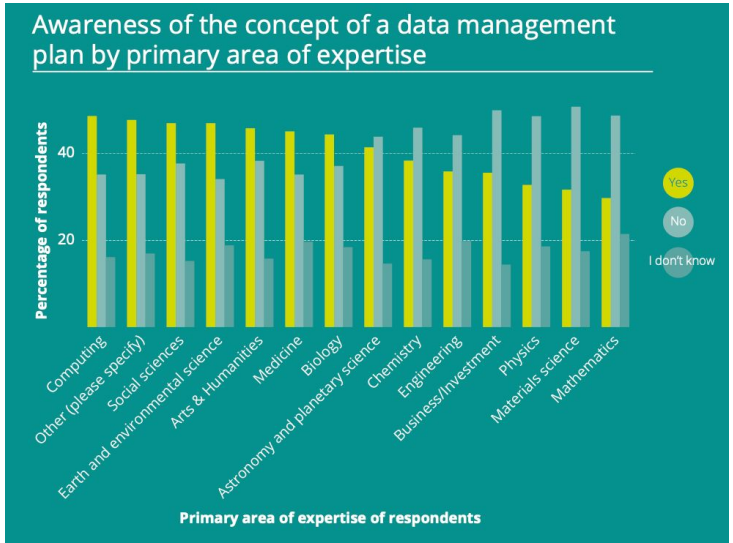
「データマネジメントプラン(DMP)について知っているか？」との質問への回答は地域によって大きな差が見られる

- 最もDMPの認知度が高いのはエチオピア、英国、米国、インド、ドイツなど
- 日本の回答：Yes (24%) No (51%) I don't know (25%)

同様に、分野によっても差がある

- 最もDMPの認知度が高いのは計算機科学、その他、社会科学、地球科学、人文学など
- 化学、工学、物理学、材料科学などの方がむしろ認知度は低い



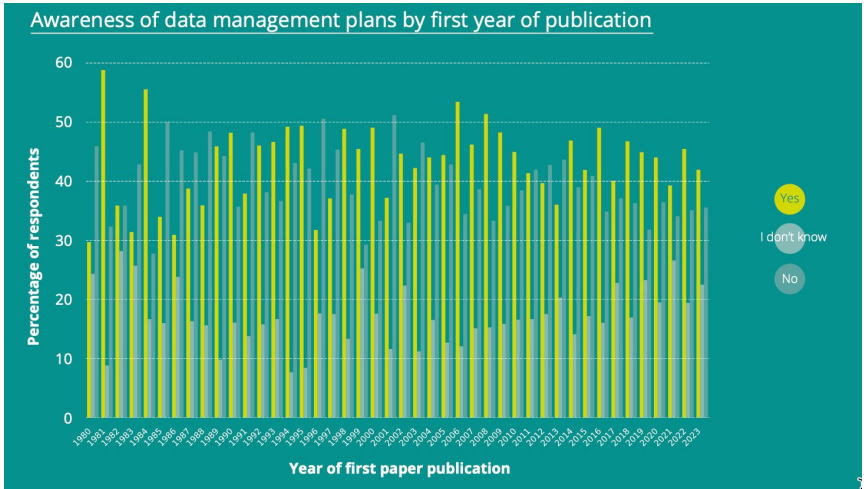


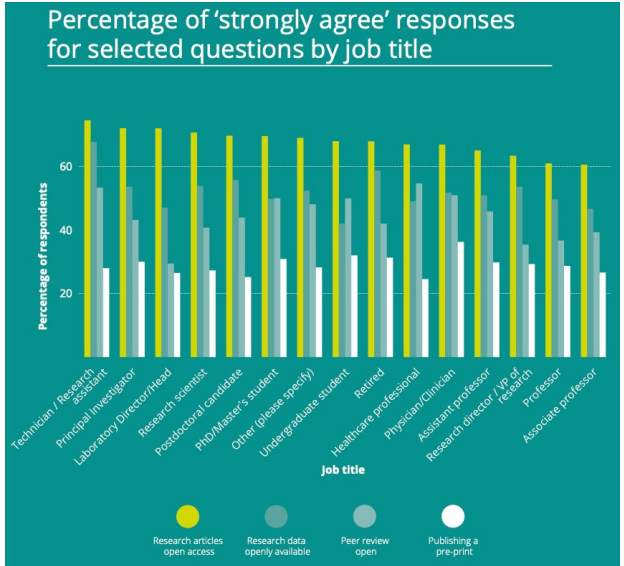
3. 固定観念に対する疑問

オープンデータに対する認識において、研究者としての経験は重要な要因ではない。

- 最初の研究成果から何年経過しているか、を研究者の経験と考えた場合、DMPの認知度とは相関しない。

「若い研究者ほど新しい概念に順応している」という固定観念は当てはまらない。





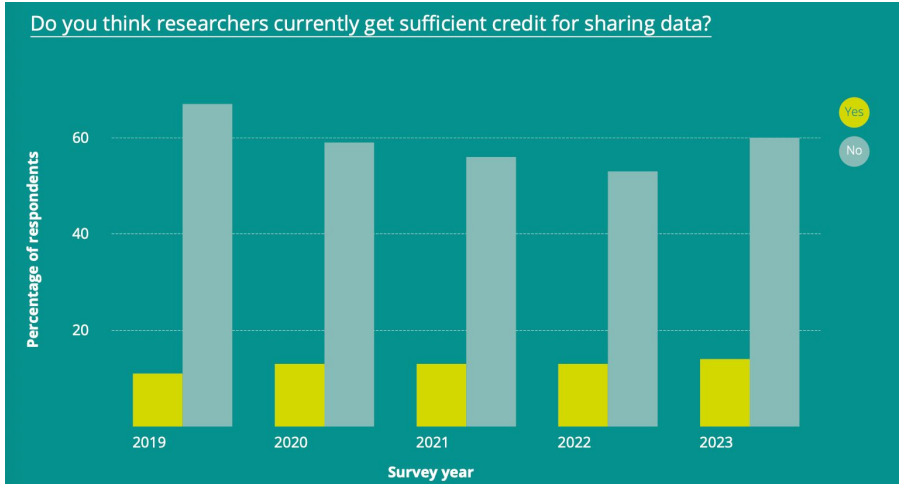
4. クレジットに関する継続的な問題

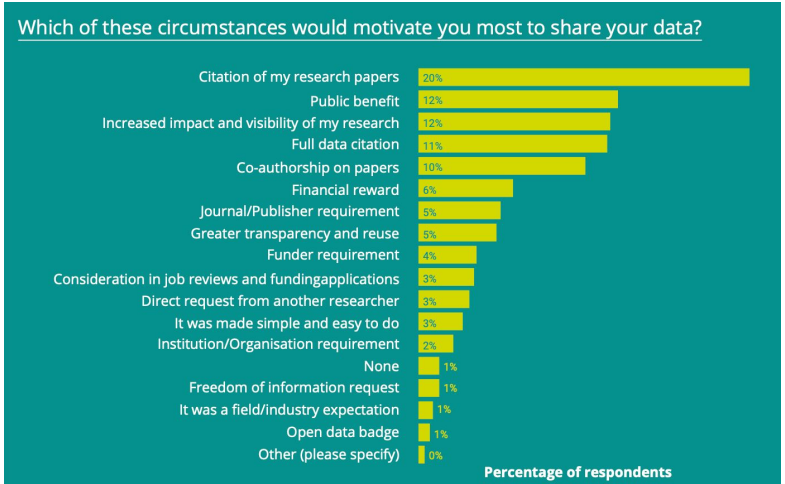
本調査を実施して以来8年間、一貫して研究者は研究データの公開について十分な評価を受けていない、という懸念をもっている。

- 調査全体では「Yes」13%
- 日本の回答者は「Yes」9%

データ共有の動機は、これまでの調査と同様、「論文への引用を促進」がトップ、続いて「社会貢献」、「論文のインパクトや可視性の向上」など。

- 日本の回答者は「引用」「共著者としてクレジット」「社会貢献」など





5. AIに関する認識と行動のズレ

今回の調査で初めて、ChatGPTや類似のAIツールを研究データの収集、処理、メタデータ作成などに利用しているかどうかを尋ねた。

- 最も多かった回答は、「ツールの存在は知っているが、利用しようと思わない」
- ただし、すでに利用し始めている研究者もいる（収集 12%、処理 14%、メタデータ作成 10%）
- 日本の回答者の59%は「利用しようと思わない」が、すでに利用し始めている研究者も（収集 9%、処理 9%、メタデータ作成 6%）

研究コミュニティへの示唆

Understanding the 'state of open data' in our specific research setting

研究データに画一的なアプローチは存在しない。対象コミュニティのニーズと関心レベルを見極めて、真に必要なサポートを提供すべき。

Credit where credit's due

一貫して研究者は十分な評価を得られていないと感じている。データ自体や論文への引用から採用・昇進条件にいたるまで、データ公開のためのインセンティブを高めるためにさまざまなシステム改革が必要。

Help and guidance for the greater good

NIHが助成する[GREIイニシアチブ](#)のように、研究コミュニティ全体に波及するような研究データのグッドプラクティスを追求し、汎用的なトレーニングとよりきめ細かいガイダンスの提供を考えるべき。

Making outreach inclusive

研究データ公開に熱心なのは若い研究者だけではない。全ての年代の研究者が一定以上のモチベーションや課題を抱えており、アウトリーチは区別なく、計画的に進めるべき。

学会としての取り組みの一例

欧州地球科学連合（EGU: European Geosciences Union）は2023年の年次大会で**データヘルプデスク**を設置し、研究データ公開について参加者が気軽に質問できるようにした。

ヘルプデスクはEGUや米国地球物理学連合（AGU: American Geophysical Union）などの関係者によって企画され、米国国立科学財団（NSF）や欧州連合（EU）などによるオープンサイエンス関連の助成金により、研究データ公開の経験を持つポスドクや研究者、図書館員などを配置した。

AGUでは今月開催する2023年次大会でも同様のヘルプデスクを設置する予定。

https://x.com/ShelleyStall/status/1650170885117034500?s=20