

研究のライフサイクルと研究データ基盤

提言

オープンサイエンスの深化と推進に向けて
(抜粋)



令和2年（2020年）5月28日

日 本 学 術 会 議

オープンサイエンスの深化と推進に関する検討委員会

要 旨

1 作成の背景

2009 年、4th paradigm なる書が公開された。コンピュータ界のノーベル賞であるチューリング賞受賞者 Jim Gray への追悼の書は、スーパーコンピュータによる計算科学（第3の科学）の次のステージとして、支配方程式のない科学の領域に向けて Data-Intensive Scientific Discovery、すなわちデータ駆動科学を提唱した。その後、2012 年に開催された画像認識コンペでは深層学習が従来手法に対し圧勝し、本格的な第4の科学の時代へと突入していく。若干遅れて、我が国は Society5.0（データ駆動社会、あるいは、データ駆動サイバーフィジカルシステム）を提案した。

オープンサイエンスとは、ICT を活用して科学を変容させることであり、現在研究データを含めた研究成果をインターネットの上で広く共有する科学の進め方に注目が集まっている。こうした新しい科学の在り方は、科学者の自発的な取り組みとして展開されるのが自然であるが、国内では政策的な観点からの議論が先行している。本課題別委員会では、これからの科学の作法をも変えうるオープンサイエンスを、各研究分野におけるデータ駆動科学の現状も加味した総合的な視点でとらえ、研究データ共有の促進と共有のためのプラットフォームの重要性を明らかにすることを目的として審議を進めた。

2 現状および問題点

新型コロナウイルス感染症（COVID-19）が猛威を振るう中、2020 年1月31日に NIH、AMED を含む国内外の研究資金配分機関や論文誌出版社が協力し、論文投稿前に実験データを WHO に提供することの合意形成がなされた。論文が出版されるまで待っている間に失われる命を救うための策である。これを受けて2月11、12日に WHO 本部で開催された COVID-19 緊急対策会議では「論文投稿時に WHO とデータを共有しないと論文を受理しない」ことが確認された。これは疫病と戦うための究極の国際的なデータ共有であり理想的な施策である。一方、この事象が意味する深層も理解する必要がある。すなわち、学術分野によってデータ共有の意識や実態が異なる中、これほどの危機的状況にならないと、データの共有が進展しないことも意味する。また、データが極めて重要な資産と見なされる時代において、その共有と公開に関しては、そのデータの特性に応じた、協調と競争のバランス取りが必要となる。データを国際的に共有するかは慎重に検討する必要があり、公開から産業展開を踏まえた非公開まで協調と競争のバランスをどのように取るかは深い課題ともいえる。

研究データの重要性が著しく高まる中で、データ利用の法制度も十分に整っているとはいえない。データは著作物ではないため、著作権法による直接的保護はない。データベースの権利は規定されているものの実施例は少ない。この状況に鑑み、我が国は、2019 年不正競争防止法に「限定提供データ」と名付けたフレームワークを導入した。これは法制度により、データを他者と共有するとともに、データの権利を保護する画期的なものである。しかしながら現時点では、我が国固有の法制度でしかないという課題もある。グーグルや

アマゾンの存在を念頭に、公正取引委員会では、デジタル社会における競争政策を検討している。旧来のビジネスと異なりデジタルプラットフォーマーにおいては、市場支配力が極めて強い事業者が相対的に出現しやすく、不当なデータ収集、不当なデータの囲い込みは独占禁止法上問題となりうる。加えて、データに個人情報が含まれる場合には、個人情報保護法が適用される。EUはGDPR（General Data Protection Regulation：一般データ保護規則）なる我が国より厳しい制度を有し、2020年米国はカルフォルニア州でCCPA（カリフォルニア州 消費者プライバシー法）なる同様の法制度が導入されつつある。

このような現状において、研究者が多層に関係する種々のルールに配慮することは大きな負担になり、また不適切なデータ利用ではないかという懸念により研究活動を萎縮させてしまいかねない。本提言では、データに対する現状の認識を多様な学術の専門家から聴取し整理することを試みたが、分野やセクターによる温度差は大きい。データ駆動科学を推進し、科学を変容するデータ共有の精神が進展すべく、日本学術会議で丁寧な議論を今後も重ねてゆくことが必須といえる。

3 提言の内容

(1)データが中心的役割を果たす時代のルール作りの必要性:データを扱う作法の混乱を避けるべく、政府は、不正競争防止法、個人情報保護法、著作権法等のデータに関連する法規制を集約・整理し、データを安心して活用できるルールを、国際的なコンセンサスを得ながら明確化する必要がある。研究データの特性と社会との関係性を踏まえた着実な取り組みを実践するための、適切なガイドラインの作成も必要である。

(2)データプラットフォームの構築・普及の必要性:膨大なデータを収集、キュレート、アノテート、メタデータ付与、保存等を推進すべく、学術界およびその研究活動を支える機関は、国の支援を得てプラットフォームの普及に努め、さらに機械学習や高次解析との接続を可能とする次世代機能の提供を実現していくべきである。このプラットフォームでは、公開から非公開までのバランスが取れたデータ共有を支援し、分野間やセクター間など多様なデータの潜在的融合を許容してさらに高い価値を生み出すことを指向すべきである。現時点においては、データを共有する意図があろうとも、経済的理由あるいはスキルの欠如等からデータ保存がなされない事態も散見される。職場の異動時や定年時を含め、データ散逸を防ぐために誰でも容易に利用できるプラットフォームが必須である。

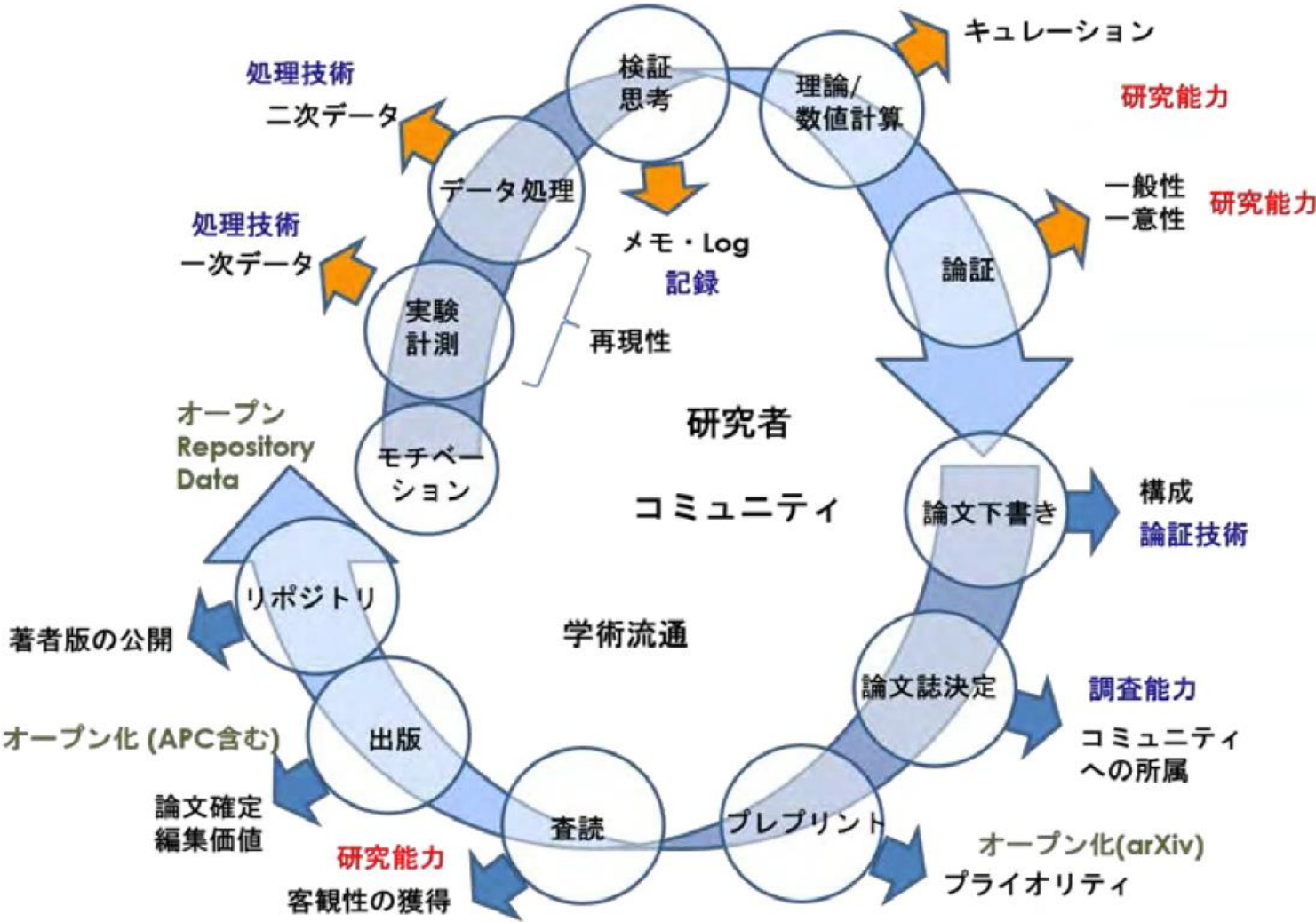
(3)第1次試料・資料の永久保存の必要性:研究成果を直接もたらした第1次試料（岩石、堆積物、土壌、流体、生物、物質、遺構、遺物など）の永久保存体制の構築および第1次資料（文書記録、書籍、景観、技術、生活様式、生産様式など）の維持保存体制を強化する必要がある。また、それらを抽出選択する背景となった第0次試料（未研究の採集試料・資料）の選択的保存について基本方針を確立する必要がある。保存のための設備、整理、管理運営と公開方針について国際・全国連携体制で進める必要がある。

(2) データプラットフォームの構築・普及の必要性

第3節でまとめたように、オープンサイエンスやデータ駆動科学が進んでいるとされる研究分野においても、より幅広い研究者に対し、膨大なデータを収集、キュレート、アノテート²²、メタデータ付与、保存等が容易に実現されるデータプラットフォームが必要とされている。NIIでは現在、研究データのライフサイクルに沿った管理・公開・検索を可能とする基盤の開発を進めているが、これを学术界およびその研究活動を支える機関と協調しながら多方面で普及していく必要がある。

＜付録図表 3＞研究のライフサイクル

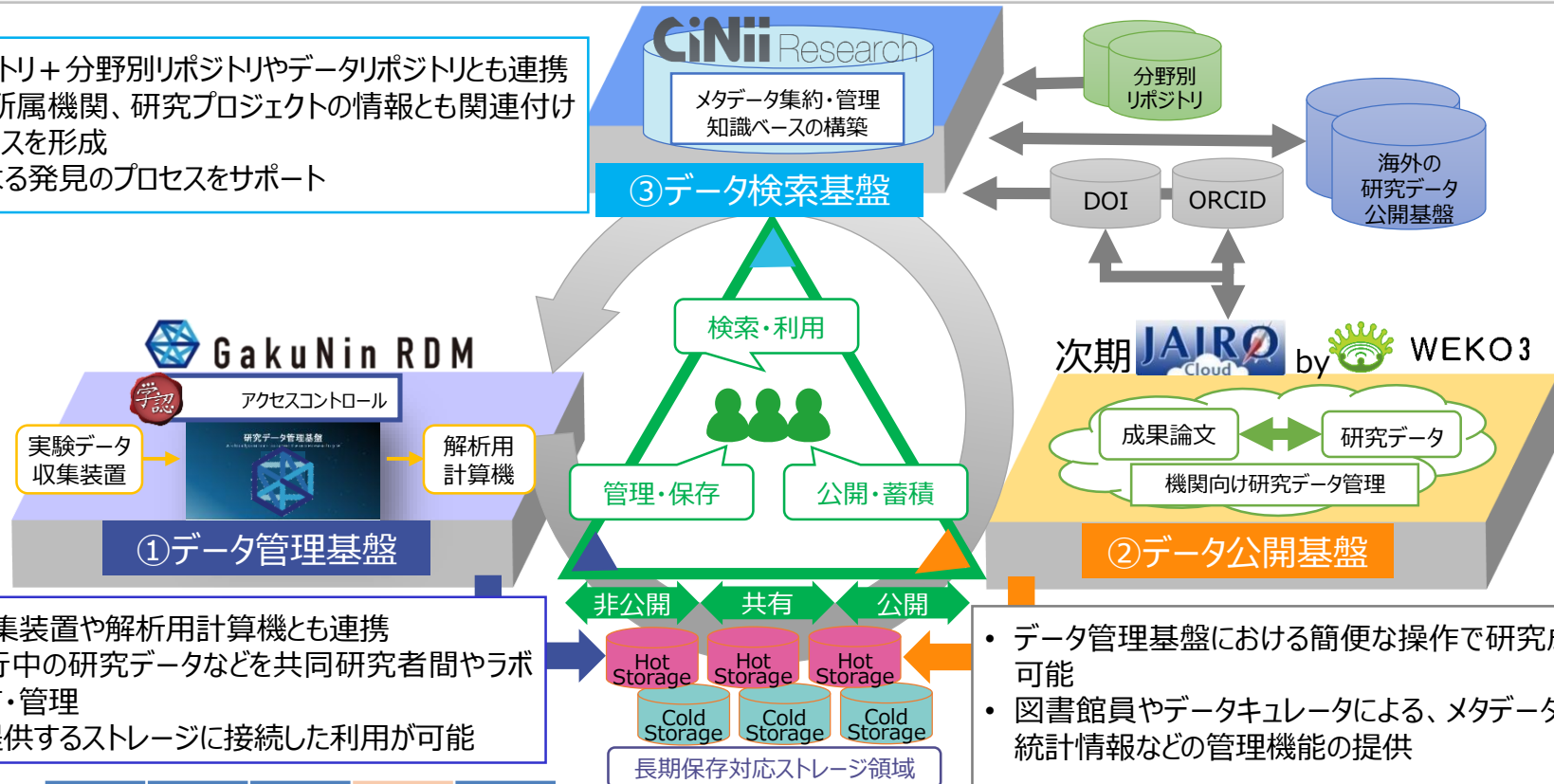
研究のライフサイクルとオープン化



オープンサイエンス推進のための研究データ基盤 (NII Research Data Cloud) の整備

- ◆ 国立情報学研究所において、研究データの平易な保存・管理、網羅的な検索等を実現するための共通システムとして研究データ基盤を開発中。2020年度に運用開始予定。
- ◆ 研究データの保存・管理を通じて、研究の再現性や透明性、継続した有用性を確保。また、研究データの利活用を促進することで、研究分野を超えた新たな知見の創出に寄与。
- ◆ ①データ管理基盤（研究者が簡単に自身の研究データをクラウド上のストレージに保存できるシステム）、②データ公開基盤（各大学で保存された研究データを他の研究者が利活用できるようリンクを提供するとともに、共有・公開するシステム）、③データ検索基盤（研究データ公開基盤や分野別リポジトリ等に登録された研究データを横断的に検索できるシステム）を一元的に整備し、オープンサイエンスを推進。

- ・ 機関リポジトリ+ 分野別リポジトリやデータリポジトリとも連携
- ・ 研究者や所属機関、研究プロジェクトの情報とも関連付けた知識ベースを形成
- ・ 研究者による発見のプロセスをサポート



- ・ データ収集装置や解析用計算機とも連携
- ・ 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- ・ 組織が提供するストレージに接続した利用が可能

- ・ データ管理基盤における簡便な操作で研究成果の公開が可能
- ・ 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

