

# わが国におけるデータシェアリングの あり方に関する提言

科学技術情報委員会

平成27年4月



## はじめに

近年、ICT技術革新が、科学の研究手法や研究成果の公開、共有方法を大きく変革しつつある。特に、この技術革新は研究データへのアクセスを容易にし、新たな研究手法としての「機械判読可能なデータ形式による二次利用可能な研究データの公開(オープンデータ)」や、「第三者とのデータ共有、交換(研究データシェアリング)」を可能とした。

欧米をはじめとする諸外国の政府関係機関では、この新しい可能性を後押しする取り組みを始めている。公的資金を用いた研究成果(論文、研究データ等)を広く社会に公開するオープンサイエンスの取り組みは、科学界だけではなく産業界を含む社会全般に対し、研究成果への広く容易なアクセスとその利用を可能とし、効果的な科学研究の推進とその活用によるイノベーションを視野に入れている。

研究コミュニティにおいてもオープンデータ、研究データシェアリングの取り組みが、科学研究のやり方を変え、研究活動に大きく貢献する可能性を認識し、議論に取り組み始めている。

一方で、産業界ではビッグデータが注目を集め、大量データの収集、蓄積、分析によって新しい価値を創出する機会を手に入れることが、競争優位の鍵となりつつある。これは、科学界においてデータ駆動型科学研究手法による革新的なイノベーション創出の期待を高める効果を産んでいる。

このように世界規模で研究環境が大きく転換しつつある状況において、諸外国の政府関係機関、研究コミュニティ、産業界、社会といった様々なステークホルダーは、研究データシェアリング、オープンデータ、データ駆動型科学研究手法に関する枠組みの議論や具体的取り組みを急速に活性化しつつある。そうした中、わが国はまだ国としての取り組み姿勢が明確でなく、世界情勢に対し大きな遅れをとっている。

わが国の研究者にとって研究環境を世界最高水準のものとしていくことは、他国の研究者との競争に勝つための必須条件であり、惹いては、産業競争力の維持にも欠かせない。同時に他国の研究者と共同してグローバルな課題にチャレンジする為にも、研究データシェアリングという新たな研究プロセス、手法を導入することが必要である。その上で、対象研究データが持つ特性と状況に応じ、データの公開制限、非公開を可能とするよう、わが国としてのデータ公開判断基準をしっかりと議論することが重要である。

その為、政府関係機関は海外の先行事例に学びつつ、研究データシェアリングを活用した科学技術政策を立案、実行していくことが急務である。この政策は、研究者や研究コミュニティが自ら進んでデータシェアリングに参加できる意識の醸成、必要な人材の育成・確保、インフラ整備や公的資金の使途の変革等の様々な課題に柔軟に対応することが求められる。

本書では、わが国の研究環境を整備し、科学技術研究の更なる発展と研究開発力向上に貢献するため、データシェアリングに関連する議論の活性化を促し、研究者のデータ共有意識を高めていくために、提言を行うものである。

## 提言の要旨

科学技術情報委員会は、わが国における科学技術研究の更なる発展及びイノベーション創出につながる新たな知の発見のため、研究手法としてのデータシェアリングのあり方について 以下の提言を行う。

### 提言1: データシェアリングポリシーを早急に策定すべき

#### 【データシェアリングポリシー策定】

- ・組織的にデータシェアリングの仕組みを稼働させるため、内閣府総合科学技術・イノベーション会議（CSTI）が策定する方針に基づき、各研究機関等が早期にデータシェアリングポリシーを策定すべき
- ・国際的な情報戦略を意識し、「データ収集基準」、「シェアリング範囲」、「品質保証方法」、「データ格納方法」、「データ整備ルール」などの規定が必要

#### 【公的研究資金配分機関等におけるデータシェアリングのあり方】

- ・公的研究資金配分機関は、公的資金による研究データの所在、維持管理を把握するため、各研究プロジェクト等におけるデータの取り扱いを定める「データ管理計画書」作成の方策を導入すべき
- ・データ管理計画の実行に必要な基盤構築には、各研究プロジェクト、研究組織の各レベルにおいて必要とされる費用を、研究費の直接、間接経費から充当できるような仕組みを導入すべき

#### 【データシェアリングを行うべき分野、領域の選定】

- ・国際的に研究開発をリードしていくため、政策的観点から実現可能な重点的分野を順次選定し、リソースを集中的に投下すべき
- ・新領域、新分野を対象とする研究助成制度の立ち上げや、新たな研究資金配分機関等の設置の際、その当初からデータシェアリングの仕組みを導入することが重要

### 提言2: わが国としての持続的な研究データ基盤を構築すべき

- ・研究データの共有、再利用を目的とするデータ基盤構築は、データを産出する個々の研究プロジェクトよりも長い持続性を持ち、データの品質、ライフサイクル管理が可能な体制を構築すべき
- ・多様な研究データの利用促進のため、最新技術を活用したデータ構造化研究を推進すべき

### **提言3: 研究データに係る人材の確保と育成を政策的に推進すべき**

- ・データシェアリング、データ駆動型研究手法による研究活動を支える、インフォマティクスと専門研究分野の知識を併せ持つ専門人材の持続的な確保と育成を、政策的に推進すべき
- ・専門人材の確保と育成には、大学等高等教育機関における育成コース、カリキュラムの制定と同時に、専門人材が適材適所で活躍できるようにテニユアに至る明確なキャリアパスを形成するプログラム制度を改革すべき

### **提言4: 研究データに係る研究業績評価の仕組みを見直すべき**

- ・研究者や研究組織のデータ共有意識を高揚させるため、従来の論文発表による評価と同様に、研究データ作成と提供を成果として評価対象とし、研究データシェアリングが研究プログラムや研究組織の業績として評価されるよう、内閣府及び各省の研究開発評価指針を改定すべき
- ・改定にあたっては、データ研究、データ基盤構築に係る研究業績の新たな評価軸、研究費の一定額をデータ基盤整備等に投入できる枠組みの導入などが必要

### **提言5: 国はデータシェアリングを推進する研究コミュニティに対し、充実した支援サービスを行うべき**

- ・公的機関による支援サービスとして、持続的なデータ基盤の構築と運用、データの利便性向上、データ品質保証、専門人材の確保と育成、データ統合に係る研究開発等を実施すべき
- ・同時に、研究データシェアリングで得られるアウトカムを把握し、評価する技術、方法の開発が重要
- ・公的機関は、データシェアリングを機会に、国内学協会によるデータの収容確保など学術活動の機能を戦略的に見直し、我が国の科学技術研究のレベルに比肩する学術情報保全を図るための支援をすべき

## 目次

1. データシェアリング推進の位置づけと目的
  - 1.1 科学技術研究の更なる発展
  - 1.2 イノベーション創出のための新たな知の発見への期待
2. わが国におけるデータシェアリングの課題
  - 2.1 わが国のデータシェアリングポリシーの策定
  - 2.2 公的研究資金配分機関における検討課題
  - 2.3 データシェアリング対象分野、領域の検討
  - 2.4 持続的なデータ基盤構築
  - 2.5 専門人材の不足、育成
  - 2.6 データ研究、基盤構築に係る研究業績評価
  - 2.7 研究者、研究コミュニティのデータ共有意識
3. データシェアリングのあり方に対する提言
  - 提言 1 データシェアリングポリシーを早急に策定すべき
  - 提言 2 わが国としての持続的な研究データ基盤を構築すべき
  - 提言 3 研究データに係る人材の確保と育成を政策的に推進すべき
  - 提言 4 研究データに係る研究業績評価の仕組みを見直すべき
  - 提言 5 国はデータシェアリングを推進する研究コミュニティに対し、充実した支援サービスを行うべき
4. おわりに

### ■別添

- 資料 1 ステークホルダーにおけるデータシェアリングの動向
- 資料 2 研究データ共有ポリシーに関する調査

### ■附則資料

- 1)委員会名簿
- 2)委員会スケジュール

## 1. データシェアリング推進の位置づけと目的

### 1.1 科学技術研究の更なる発展

ICT 技術の革新による研究環境の変化の中、第三者が作成した研究データを容易かつ効果的に再利用する研究データシェアリングの仕組みは、研究者のデータ取得、解析にかかる時間を大幅に削減するとともに、重複研究を回避するなど研究リソースを最大化させ、研究活動を効率化し、科学技術研究の更なる発展を加速する。

同時に、研究データを共有し、交換可能とすることは、データを作成・提供する研究者にとって、データサイテーション等による新たな研究業績評価につながる可能性がある。

他方で、研究データが多数の第三者に相互に検証、再利用されることにより、科学研究の透明性と再現性を担保し、社会の科学に対する信頼確保につながる。

### 1.2 イノベーション創出のための新たな知の発見への期待

近年の ICT 技術の急激な発展は、膨大なデータの産出とともに、その連携、高度処理、データの意味抽出を軸としたデータ集約型科学(第四パラダイム)という概念を産み出した。実験科学(第一パラダイム)、理論科学(第二パラダイム)、計算科学(第三パラダイム)に続く、この新たな科学的研究手法は、今後本格化するであろうデータ活用社会において、新たな知の発見と価値の創造、社会的課題解決に有効な手法として注目を浴びている。

膨大なデータの中から新たな発想を見出すデータ駆動型科学研究手法は、将来の科学技術研究による革新的イノベーション創出の一角を担っていくことが期待されており、新たな研究プロセス、手法による科学研究の革新が、研究を目覚しく進歩させる可能性がある。

## 2. わが国におけるデータシェアリングの課題

わが国がデータシェアリングという新たな研究手法を有効活用することで、科学技術研究を発展させイノベーション創出につなげるためには、研究者や研究コミュニティのデータ共有意識の醸成、必要な人材の確保・育成、業績評価、インフラ及び資金面等の様々な課題が存在する。

以下に、わが国におけるデータシェアリング議論の検討課題を列挙する。

### 2.1 わが国のデータシェアリングポリシーの策定

わが国の研究環境を世界最高水準にするためには、これまで研究者が個別に行っていたデータ共有の仕組みを、組織の仕組みとして稼働させ、組織としてのデータシェアリングポリシーを策定することが求められる。

国際的なデータシェアリングの動きを見据え、わが国における政府関係機関、研究コミュニティ、産業界などのステークホルダー間で、世界情勢から遅れることの危機意識を共有し、科学技術政策、産業政策の両側面から、わが国としてあるべきデータシェアリング施策について検討していくことが必要である。



## 2.2 公的研究資金配分機関における検討課題

欧米をはじめとする諸外国の主要な研究資金配分機関の多くは、公的研究投資の影響力や責任を最大化することを目的に、公的研究プログラムに応募する研究者に対して研究データの保存、共有・公開の有無等について記載した「データ管理計画書」の提出を義務化している。

わが国においても、公的研究資金を含む研究成果の論拠となるデータについては、それらをしっかりと再利用可能な形で保管、蓄積し研究の透明性を確保することが、公的研究資金配分機関としての責務と考えられる。

この点において、わが国としてのポリシーメイキング、データ共有意識の醸成、必要な支援方策等の具体的なデータシェアリング施策を検討する必要がある。

## 2.3 データシェアリング対象分野、領域の検討

データシェアリングの必要性、有効性、方法論等については研究分野毎に、特徴的な課題があり、データシェアリングの仕組みが良く機能する分野とそうではない分野が存在する。

例えば、バイオサイエンスや地球観測、天文学の分野などは、公にデータを共有することで研究が促進される分野であり、既に国際的にデータの共有、オープン化の取組みは進められている。

逆に、応用研究に近い技術開発や、わが国が世界に先駆けて高い競争力を持つ分野、領域については、産業への悪影響を避けるため、データを公開しない等の配慮をすべきである。

データシェアリングはまだ新しい取り組みであり、政策的観点から重点的な取り組み分野を選定し、リソースを集中することが賢明であろう。分野毎の特徴的な課題を念頭に置き、データシェアリングを推進すべき分野における意義目的を明確にする戦略を、スピード感を持ってしっかりと検討する必要がある。

## 2.4 持続的なデータ基盤構築

データシェアリング手法の導入には、産出された研究データをグローバルかつ分野横断的に幅広く共有し、再利用するため、データ利用者のニーズに応えられる情報流通のデータ基盤が重要である。

しかしながら、わが国では研究データの情報流通の仕組みが確立しておらず、共有すべき貴重なデータでも研究者個人にその管理を依存している状況である。貴重なデータを散逸させず、持続的にマネジメント可能な基盤的仕組みの構築が必要である。

持続的なデータ基盤の構築にあたっては、第三者が安全かつ容易に再利用可能となる相互運用性、データの品質保証とデータを確実に再利用できる利便性向上の仕組みが求められる。