

学術情報委員会におけるオープンサイエンスの検討について（案）

1 議論の枠組みについて

○研究成果（論文・研究データ）の公開についての基本的考え方等を検討し、指針として示す。

- ・研究成果の公開を推進する意義・基本方針
- ・「論文のエビデンスとしてのデータ」及び「研究過程で生じたその他のデータ」の定義・範囲
- ・その他下記2に関係する事項 など

○海外の動向及び民間の動向を踏まえた戦略的な公開の在り方を検討

○指針には、大学・研究機関及び研究資金配分機関が取り組むべき内容も示す。

- ①データ公開についての実施計画等（公的研究資金申請者のデータ管理計画の提出を含む）の策定
- ②データの管理体制（管理方針の策定等を含む）の構築
- ③データの保存・公開における図書館等の役割・機能 など

※オープンアクセス（主として論文に係る）については、これまでの審議の蓄積があることから現状についてレビューし、必要な検討を加える。

2 オープンデータに係る取組の具体化について

○研究データの公開と共有及び保存の体制をどう考えるか。

- ①データリポジトリやプラットフォーム及び検索システム構築の在り方
- ②データの質や正確性を担保する仕組み及び再利用可能な形での保存・公開（例）フォーマットの標準化、メタデータの付け方、データセットの識別子付与、保存期間の設定 など
- ③データの利活用に係る許諾ルール（教育での利用を含む）の明示
- ④データへのアクセス及び利用に係る制限事項（プライバシーの保護、財産的価値のある成果物の保護等）
- ⑤データを作成・公開する研究者のインセンティブ及び業績評価への反映
- ⑥学協会等によるデータジャーナル出版の可能性及びその支援方策

○学会や日本学術会議及び各省との連携

3 各研究分野の取り扱いについて

○具体的な事例に則して検討を進めるため、ライフサイエンス分野を例として取り上げ検討。その上で、各分野で共有、参照すべき事項について検討。

○人文・社会科学分野におけるオープン化の在り方

1 議論の枠組みについて

○研究成果（論文・研究データ）の公開についての基本的考え方等を検討し、指針として示す。

・研究成果の公開を推進する意義・基本方針

（目的・意義）

公的研究資金による研究は、新たな発見や洞察を提示し、それにより、健康、エネルギー、環境、農業といったあらゆる分野の進歩を強力に牽引してきた。こうした公的研究資金による研究成果として得られた論文や研究データへのアクセスを可能とすることで、研究成果の理解促進と同時に、成果の再利用による新たな発見や、新たな研究概念とイノベーションを加速し、新たな産業の創出、競争力の強化、地球規模での研究の促進、経済成長等に貢献するものである。

教育に関する観点等についても言及する必要があるのではないか。

（基本姿勢）

公的研究資金による研究成果（論文、研究データ等）の利活用促進を拡大することを我が国のオープンサイエンス推進の基本姿勢とする。

（公開の範囲）

公的研究資金による研究成果のうち、論文及び論文のエビデンスデータとしての研究データは、原則公開とし、その他研究開発成果としての研究データについても可能な範囲で公開することが望ましい。

研究データのオープン化においては、科学研究活動上の便益・損失や研究途上の情報の取扱及び機微の判断など、関与する専門家、研究者又は科学コミュニティが判断すべきものが含まれる。

・「論文のエビデンスとしてのデータ」及び「研究過程で生じたその他のデータ」の定義・範囲

研究データは原著論文と同等又はそれ以上の研究開発成果として位置づけられ、公開の対象となる研究データは、メタデータ、数値データ、テキストレコード、イメージ、ビジュアルデータなど多様なデータが含まれる。これには原則として以下のものは含まれない。研究ノート、予備的分析、論文の草稿、将来の研究計画、同僚との私信、物理的対象物（研究室の試料、バクテリアの菌株、実験動物など）等及び所属機関との契約により公開に条件が付されているもの。

公開の対象となるデータの種類の例示されているが、「論文のエビデンスとしてのデータ」等については定義されていないため、規定する必要があるのではないか。

○海外の動向及び民間の動向を踏まえた戦略的な公開の在り方

オープンアクセスのための新たなビジネスモデルの開発や新しいデータキュレーションと保存サービスの提供に取り組んでいる学術出版会への対応（出版社や第三者機関に預けるのか自前（日本）で管理するのかという公的な知的財産の保有ルール）についても整理しておく必要がある。

出版社では、研究論文投稿時に利用データ公開を促す例が増加している。データ置き場については、海外機関を推奨するケースが多い。我が国の研究者が投稿する場合、最先端研究データが海外機関へ保管されることになる。国際的には、データリポジトリの継続的運用モデルについても議論が活発に行われている。この場合、保存データ利用時に料金を徴収するモデルも見受けられ、日本の研究データへのアクセスに海外への料金支払いが必要になることも想定される。

日本学術会議情報学委員会国際サイエンスデータ分科会の報告によると、米国、中国、インドでは、各分野に数十人から数百人のデータベースセンターが存在しており、多くのデータベースを開発している。日本には、このような大規模なデータベースセンターは存在しないが、研究、教育、産業、行政、医療など各分野で、データベースが生み出されている。しかし、国としてのデータ戦略がないことから、それらは、有機的に繋がることなく、バラバラに作られ、日本全体のデータ活動は把握されていないと指摘している。

複数の大手出版社においては、オープンアクセスデータジャーナルを刊行するなど、研究データセットをジャーナルにする動きも活発になってきている。

例えば、ネイチャー・パブリッシング・グループが、2014年5月にオリジナル論文との連携が容易な Science Data を創刊しており、学术界に大きなインパクトを与えている。

海外の出版社からのデータジャーナルの刊行や論文投稿時に出版社から求められるデータの置き場が海外の機関を推奨される動きがある中で、戦略的なデータの公開についてどのように考えるべきか。具体的には、データの保管場所や公開の仕方などについての考え方を示す必要があるのではないか。

○大学・研究機関及び研究資金配分機関が取り組むべき内容

①データ公開についての実施計画等（公的研究資金申請者のデータ管理計画の提出を含む）の策定

（実施方針・実施計画）

関係機関が定めるオープンサイエンスに関する実施方針及び実施計画には、次の要素が含まれていることが必要であると考えられる。

- ・イノベーションや競争力を強化する取組
- ・ステークホルダーにとって透明なプロセス
- ・公的研究資金による研究成果に係るデジタル研究データの所在を特定しアクセスできる方法

- ・研究成果の長期的管理を保証した上で、アクセスし易さや相互運用を奨励するような検索、保管及び普及機能を最適化するためのアプローチ方法
- ・機関の計画等の遵守に関する評価とともに、必要に応じて、計画等に実行徹底を図るための戦略
- ・計画等を実行するための運用経費等の確保
- ・計画等を実行するための優先度を考慮したロードマップ策定
- ・データインフラ（リポジトリ等）の整備計画
- ・公的研究資金を受けている研究者に対して義務を周知するためのガイドラインの策定、関連する規則の改正等

（データ管理計画）

- ・公的研究資金による支援を受ける研究者が、必要に応じ、データ管理計画を作成するようにすること。同計画は、データの長期保存及びアクセスが困難な場合において理由を説明することにも活用される
- ・研究者から提出されたデータ管理計画が適切に評価されるようにすること
- ・研究者自ら提出したデータ管理に従っていることを確実にするための仕組みを盛り込むこと
- ・適当な場合には、ユーザーがアクセス可能なデータベースにデータを預けることを奨励すること
- ・計画の下で利用可能となる研究データセットに対し、適切な属性を特定し、提供するための基盤を開発すること
- ・研究データ管理、分析、保管、保存及び共有技術等に関する訓練、教育及び人材養成を支援すること

②データの管理体制（管理方針策定等を含む）の構築

公的研究資金を用いた研究を行うための設備、ルール、人材等を具備している機関においては、論文、研究データ等の研究成果の管理に係る規則を定め、特に、研究成果の散逸、消滅、損壊を防止するための具体的施策を講ずる必要がある。

- ・研究終了後も研究データの保存・整備が必要
- ・データ保存・整備のしくみ（データリポジトリ等）が必要
- ・論文や研究データの永続的、長期的保存を担保するためには、国立国会図書館、大学図書館、情報基盤センターをはじめ、科学技術・学術文書・書籍を扱える図書館や情報基盤センター等の協力を得ることが有効
- ・公開されたデータや情報の権利関係（ライセンス）を明確にし、また永続的保存や提供サービスに関しての制度づくり・責任体制の議論が必要
- ・保存に関わる経費は、コンピュータなどの経費だけではなく、データを管理・維持するための人的コスト等が重要であることを考慮すべき。計算資源としてはクラウドを活用するなど、コストの圧縮を考えておくことが必要
- ・将来に向けて持続的なオープンサイエンス制度として自立できるよう、組織的なデータ利用をライセンス化するなどのビジネスモデルの創出（著作権制度などのイメージ）

大学等関係機関に求められる事項が例示されているが、オープンアクセスに係る大学等の現状や研究者の理解促進の観点も踏まえて、特に規定すべきと考えられる事項は何か。

③データの保存・公開における図書館等の役割・機能 など

大学・研究機関等においては、技術職員、URA、大学図書館職員等を中心としたデータ管理体制を整備できるように、データサイエンティストやデータキュレーターなどを研究支援人材として位置づけられるよう、包括的な育成システムを検討し、推進することが必要である。

※オープンアクセスについては、これまでの審議の蓄積があることから現状についてレビューし、必要な検討を加える。

2 オープンデータに係る取組の具体化について

○研究データの公開と共有及び保存の体制をどう考えるか。

①データリポジトリやプラットフォーム及び検索システム構築の在り方

○効率、効果的な基盤整備の在り方についてどう考えるか。

- ・大学等の機関リポジトリの活用
- ・NIIの共用リポジトリ（JAIR Cloud）の活用
- ・プラットフォームとして、JSTのJ-STAGEやNIIの学術機関リポジトリポータル（JAIR）活用の可能性
- ・JSTの科学技術情報統合検索システム（J-GLOBAL）の活用可能性
- ・分野別データベースの活用
- ・その他のシステム構築の必要性 など

②データの質や正確性を担保する仕組み及び再利用可能な形での保存・公開 (例)フォーマットの標準化、メタデータの付け方、データセットの識別子付与、保存期間の設定 など

本原則は、G8の下に置かれている政府高官グループ（GSO）において合意されたものである。

○現在広く用いられている検索手法を使って、必要とするデータ（もしくは、データセット）が容易に見つけられることである。

具体的な技術的手段としては、

- ・適切で持続的な識別子フレームワークの実現
- ・記述メタデータ標準の採用
- ・適切なデータフォーマット
- ・データの分類
- ・メタデータ検索とエキスパートAPI（活用インターフェース）

○倫理的、法的、商業的利益の保護などの制約は認めつつも、研究データはアクセス可能でなければならない。

アクセス可能となったデータの利用者は、論文などの引用と同じく引用元を明らか

にする義務がある。この引用により、研究データの貢献が記録される。
併せて、研究データの引用として、DOI (Digital Object Identifier) のような世界標準仕様に沿ったデータの同定と所在地 (Landing Page) を特定する用意が必要である。

○数値、テキスト、イメージ、動画などは、コンテキスト、セマンティクス、データ分析ツール、アルゴリズムなど一体となって提供されることで、理解可能なものとなる。

観測・実験データは、データが生成された際の条件（中略）に関するメタ情報によって理解可能となる。データ品質については、データの完全性と由来を確実にするために、生データから公開データまでに至るまでの加工プロセスについて、当該データが属する研究分野で広く承認された一定の規約が確立されなければならない。

○研究データが効率的・永続的かつ効果方法で管理されるためには、データ管理方針と計画が、研究プロジェクト及び組織のレベルで全てのデータについて定められていなければならない。

データ管理方針と計画は、誰がデータの有効性を維持する役割を果たすか、そして、キュレーション、保管と利用サービスの維持に必要な経費がどのように支払われることになっているかという運営と、利用者の負担を明らかにしなければならない。

研究データの技術的な品質の評価、またそれぞれの研究分野の固有に基準に沿った評価について合意しておく必要がある。論文のピアレビューに相当するような評価の仕組みが必要。

基本的な考え方は示されているが、ガイドラインとする場合、ブレイクダウンした内容を示す必要があるのではないか。

③データの利活用に係る許諾ルール（教育での利用を含む）の明示

提供するデータを自由に利用できることを示す利用ルールを付す必要がある。

・研究データは著作物ではないため、CC0 を採用することが望ましい。また、データの集積・整理や品質管理などの点で専門的研究者の労力や高度なノウハウが入ったデータベースなど、知的生産物として著作権が発生する場合には CC-BY を採用することが望ましい。

研究成果利活用促進の上で避けては通れない著作権の課題について取り組む必要がある。

例えば、学協会著作権データベースによると、登録学協会 2,609 件のうち、著作権ポリシーを有していない学協会が、1,355 件となっている。

研究成果の公開と利活用を促進する観点から許諾ルールをどのように規定するか。

④データへのアクセス及び利用に係る制限事項（プライバシーの保護、財産的価値のある成果物の保護等） など

ここでいう研究成果のうち、機密保持、企業秘密、国益及び国家安全保障に係るもの、研究成果の商用化・産業化を目的として収集されたデータ、又は民間企業が保有するデータ並びに共同研究契約などで研究成果の公開に制限がある場合などは、公開適用対象外として考えるべきである。

そのほか、データへのアクセスやデータの利用には、個人のプライバシー保護、財産的価値のある成果物の保護の観点から制限事項を設ける必要がある。

⑤データを作成・公開する研究者のインセンティブ及び業績評価への反映

政策誘導として、研究者及び科学コミュニティに対するインセンティブを高め、オープン化に対する努力を評価することが重要である。

アクセス可能となったデータの利用者は、論文などの引用と同じく引用元を明らかにする義務がある。この引用により、研究データの貢献が記録される。【再掲】

研究データの技術的な品質の評価、またそれぞれの研究分野の固有に基準に沿った評価について合意しておく必要がある。論文のピアレビューに相当するような評価の仕組みが必要。【再掲】

高品質なデータを提供した研究者に適切な報酬（助成金や昇進機会）を与えるなど、成果に見合う処遇の仕組みを設けるための検討が必要。

データ公開に対する研究者及び科学コミュニティに対するインセンティブを高める方策として、どのような事項を規定すべきか。

⑥学協会等によるデータジャーナル出版の可能性及びその支援方策

複数の大手出版社においては、オープンアクセスデータジャーナルを刊行するなど、研究データセットをジャーナルにする動きも活発になってきている。

例えば、ネイチャー・パブリッシング・グループが、2014年5月にオリジナル論文との連携が容易な Science Data を創刊しており、学术界に大きなインパクトを与えている。【再掲】

JST が提供する J-STAGE や JSPS が行う科学研究費助成事業（研究成果公開促進費（国際情報発信強化））の支援によるデータ出版の可能性

○学会や日本学術会議及び各省との連携

日本学術会議においては、「オープンサイエンスの取組に関する検討会」で方策について議論していることから、これらの検討内容に留意することが必要。

また、データ公開に係るガイドラインの共有やリポジトリ等に関して、各省との連携が必要。

3 各研究分野の取り扱いについて

○具体的な事例に則して検討を進めるため、ライフサイエンス分野を例として取り上げ検討。その上で、各分野で共有、参照すべき事項を検討。

○人文・社会科学分野におけるオープン化の在り方

科学技術と社会の関係を巡る課題に対し、社会のための科学技術、社会の中の科学技術という観点に立った人文・社会科学研究を推進し、その成果を踏まえ媒介的活動が活発に行われることが期待。

人文・社会科学の学術情報を、他の分野との連携や新たな研究の方法論も見据えて、どのような形でオープン化の中に位置づけるか。