

日本学術会議 課題別委員会  
オープンサイエンスの取組に関する検討委員会  
(第23期・第5回)  
議事要旨

1. 日 時：平成27年9月14日(月) 15:00~17:00
2. 場 所：日本学術会議 5階 5-C(1) 会議室
3. 出席状況  
出席者：土井委員長、大杉幹事、岡委員、喜連川委員(4名)  
欠席者：杉田副委員長、戸田山委員、吉川委員、三成委員(4名)  
オブザーバー：  
村山泰啓 情報通信研究機構 統合データシステム研究開発室長  
真子博 内閣府政策統括官(科学技術・イノベーション担当) 付参事官補佐  
(国際統括)  
事務局：石井参事官、松宮補佐、大西専門職、鈴木専門職付、辻上席学術調査員
4. 配布資料：  
資料1 第4回議事要旨(案)  
資料2 日本学術会議「オープンサイエンスの取組に関する検討委員会」論点整理(案)  
資料3 第2-4回委員会における発表・意見交換のまとめ  
参考1 委員名簿

5. 議 事：

(1) 前回議事要旨案の確認

資料1に基づいて、前回議事要旨(案)が確認され、了承された。

(2) 論点まとめ

資料2に基づいて、土井委員長より、本委員会における論点整理の説明が行われた。以下の3点すなわち、1. オープンの意味、2. オープンサイエンスのコスト、3. オープンサイエンスの人材を中心にとりまとめる案が提示された。

具体的には、

1. オープンの意味として、「すべてオープンではない」という点を強調しながら、分野で異なるオープンクローズ戦略、オープン化のインセンティブを論じる。
2. オープンサイエンスのコストとしては、オープンサイエンスにはどのようなコストがかかるのかまだコンセンサスが得られていないことを踏まえ、運営コスト、標準化にかかる研究者の負担、保存方法について論じる。特に保存については、計算機的なものでは対応できないモノ保存のような、資材があること自体が重要な分野等も配慮する必要性を示す。
3. オープンサイエンスの人材としては、データを集める人、データをマイニングする人などいろいろな人が混在しているがこれらを整理し論じる。

についてとりまとめる方針案が示された。

その後、議論が行われた。本日の議論を踏まえて、土井委員長が論点整理を再度行い、次回の委員会で示すこととした。本委員会としては、報告書の形式でまとめることとし、内容はこれまでの発表などをベースにすることとした。

主な議論は次の通り。

【オープン化の方法論、コスト】

- ・オープン化の方法議論、どこでどのようにデータを蓄積していくのかについては、オープンサイエンスの意味に入るのかコストに入るのか。また、それにかかるコストは分野によって異なる。
- 今の案では、ひとつの場所に明示したようにはなっていない。どちらかと言えば、「2. オープンサイエンスのコスト」のところそれぞれ分野の、役割分担やコストについて事実をま

とめてはどうかと考えている。

→それぞれの分野でずいぶん違うという印象。それぞれの分野ごとに淡々と記載することとして、そのうえで、もし少しまとめられれば良いと思う。

#### 【学術におけるオープンサイエンスの対象：デジタル、モノ】

- ・学術会議として「オープンサイエンス」といったときに何を対象にするのか。デジタルデータのみを対象にするか、モノなども考慮する方が良いか。含まれないものが出てくるのはまずいのではないか。
- 内閣府でオープンサイエンスの議論を行った際には、デジタルデータを対象としており、web上でコントロールできるものを想定している。広く一般に利用できるという観点から、デジタルデータを考えている。定義をどこで線を引くのかということだろう。一方で科学界としてのとらえ方は異なるだろう。
- 学術としてオープンサイエンスを議論するので、デジタルデータ以外のものも含めて考えることが望ましい。オープンサイエンスの目的は、研究の加速、コストのセービング及び研究のエビデンスが目的である。一般の人が入る余地はあまりないのではないか。広く一般に利用できるというようなアウトリーチまで考えるときりがないと感ずる。

#### 【オープンサイエンスのオープンの意味】

- ・オープンサイエンスといった場合、すべてオープンにするということではない。天文学でも占有期間等がある。「オープンと言っても、すべてオープンではない。そのように運営しても良い。」ということ言うことが大切と考える。ひとつのデータベースを取り上げても、オープンの部分とそうでない部分が存在していることがヒアリングでも示されている。

#### 【国益・分野の違い、論文公表とオープンの対象】

- ・国際的にオープンサイエンスの議論が活発になる中で、日本としてそれらを認識したうえで分野別に判断し対応できるようになっておいた方が良好だろう。そうでなく、いつの間にか海外で標準化が決まってしまう不利になってしまうのは良くないだろう。日本の現状を踏まえて、現実に行えることを明らかにして強みを示していくことが重要と思われる。
- ・分野ごとにオープンの違いがあるということによくわかる。一方で国によって学問によって守らないとならない分野があるというのはどのような意味か。天文学などは国際協調であって出てきたデータが日本のものという意識はないだろう。これが日本固有で守らないとならない分野とはどのようなものか。ファンディングエージェンシーが、日本でファンディングして取ったデータを他に使わせないなどになると変なことになるのではないか。
- 産業と密着している分野では、守らないとならない部分はあるだろう。ただしそうした分野や範囲は狭いものであって、「国」という言葉の使い方は気を付けないとならないと思う。ファンディングエージェンシーが、オープン化に関して規制をかけるということはあまり望まれていない。
- ・原則として公開するデータとは、パブリッシュされたものということと理解している。学術的に価値があるとオーソライズされたデータをオープンにするということである。これは研究の成果であり、ファンディングエージェンシーとしては、ワールドワイドにオープンにしましょうということになるだろう。パブリッシュしたものは原則オープンにするということでは異論にならないのではないか。オープンにしたいくなければ、パブリッシュしなければ良い。特許などもその意味で取得しないことも往々にしてある。もう少しこの辺りは整理が必要だろう。天文学は論文にならないデータもオープンにしているが、これは経済的価値が少なく、そうしたデータをオープンにするかどうかの議論は意味がないのではないか。
- 測地学などでは、コミュニティとして人類の資産としてデータ整備することが不可欠であり使命であるが、それがリファアーされないという問題がある。データを成果物としてリファアーされることが求められる分野もある。
- 測地学は国益の塊。実際にはいろいろな人たちが関連分野に入っている、難しいところである。
- データをとっても論文を出さないということはないだろう。論文は出ても根拠のデータまでは示していない、あるいは、維持できなかつたなどの課題があるというシンプルな話なのではな

いか。

→ケミストリーだと論文とともにサプリメンタルデータがパブリッシュされる体制が整っているが、サプリメンタルデータのみが単体リファアされることはあまりない。しかし、データパブリケーションの流れからすると、それがリファアされ、研究成果として評価されるという流れがある。

#### 【オープンサイエンスの目的】

- 地球科学の分野などで、論文にまとめる元になった貴重なデータやモノを退官とともに消失してしまうという事例が多々ある。オープンデータが比較的ウェルカムな分野でもその状態なので、もう少し整えるように促す必要があるのではないか。
  - 何も言わなければ、手間や資金をかけてオープンにすることは行わないというのが現状だろう。あるいはどこかでレジストリを提供してもらえればやるという人はいるかもしれない。そのためにインセンティブをつけたり、ある程度強制したりしないと動かない可能性がある。
  - オープンサイエンスの話をしているのか、それとも、高い検証性（再現性）が求められる時代が変わったと啓蒙する話をしているのか、その2つを一度に論じるとわかりにくい。
- 本委員会の議論は、前者を行っていると思っている。効率性に関わる話である。ここでは、オープンサイエンスは、学問の加速化、学問のセービングが趣旨であるという議論だと考えている。その際に、研究者の人権を守るために、分野の差を考慮する必要がある。
- 天文学などでは、データを取得するために多大なコストがかかっている。こうしたところでは、きちんと保存しておかないとならない。オープンにするかどうかは、データ取得のコストにもよるだろう。

#### 【日本の取組】

- それぞれの分野で国際競争を行い、データをオープンにして世界の中で戦うといったときに、その人材が日本にいないかが違う。人がいないがためにオープンにしたいという分野もあるだろう。そのレベルはまちまちで、日本の弱点があるだろう。それをサポートできれば良いのではないか。
  - 既に走っている分野はグローバルに行っている。日本発でオープンサイエンスを始めるべき分野はあるか。例えば他の国が既に始めているものに乗っかるという方向で、問題はあろうか。
- オープンデータが進んでいったときに、日本では強いパブリッシャーがいない。海外のパブリッシャーが、保存場所を提供し、集めていくという状況がある。こうした状況が良いのかどうか。
- その流れは、IT的にはリネージをとったということであり、データに誰がクリエイトしたかのタグを付けたに過ぎない。そのうえで、ロケーションをどこにするかを考えることになる。国内のインフラで意味がないかという問いに答えると、国内のインフラでもある程度有効ではないかと考える。ポイントは、データの出所を付けることにより存在感を示すことで、そうすればスモールスタートでも価値があることができるのではないか。そして、データの価値はデータをとった人が一番わかっており、分野のコミュニティの中で、ある程度「データの大切さ観」は醸成されている。ただ、シェアリングのユーザビリティが良くない部分があるので、知を横ぐしして共有化できるようにするという話は必要だろう。
- そうした場合に、頼りになるリポジトリがないという問題がある。またラベルを書くことができるのはその分野に精通している人でないと難しい。その分野から離れてデータを扱う人も必要であるが、やはりその手前にそのデータに適した処理をすることができる専門家が必要である。そうした人のためのコストを手当てしないと難しいというメッセージを出す必要があるのではないか。

#### 【分野による違い、資金・雇用の状況の違い】

- 大型の研究を行うような分野の状況と、競争的資金のような時限による研究を行うような分野の状況では、状況が異なる。後者においては任期付き雇用が多く、人材育成が困難である。分野の違いと言うよりは、資金や雇用の違いで生じているという面があるだろう。真理の追求のための研究と産業につなげるための研究では、かなり違いがあり、そうした違いをどう取り扱

うか、整理しておかないと難しいかもしれない。

- ・リテラシーは、その分野・データを維持していこうという意識やパッションが基本ではないか。大学でも人事ポストについては分野が考える話なのであるが、現実にはそれが行われていない。そうした現状の問題点について学術会議がメッセージを出す価値がある。また、データを気軽に置いておく場所を確保して、そのデータをキュレートする人もきちんと処遇すべきであるといったことは王道のメッセージとして書くことが良いのではないか。

### 【人材育成】

- ・人材についてデータはあるのか。人材について論じる場合は、どれぐらいの人材がいるなどの資料・エビデンスがないと、論じにくいのではないか。
  - 日本については難しいが、海外ならキュレータというポジションがあるので、少しは取れるか。
  - 人数もだが処遇も大切である。そうしたデータが必要である。
  - 日本でも既にデータベースを持っているところはあるので、そうしたところではどれぐらいの人数でどのぐらい時間をかけて行っているかわかると良いのだが。
  - ライフサイエンスについては問い合わせをして数字が出せるかもしれない。
- ・『オープンデータを実現するための広めの人材』を対象として考える方が良いのではないか。データキュレータだけ言うよりは、システムに関わると思われる。
  - キュレータ、マネジメント、セキュリティができる人が必要である。今までは下支えしていた人を今後はもっと前面に出していかないとならないだろう。

- ・少子化と言われる中、こうした人材を増やしていくという話ばかりでなかなか増えない。単にキュレータを増やすと言っても難しい。まずデータキュレーションを既に行っている人を認めてキャリアパスを認める仕組みを作る必要がある。
  - そのために、EUの議論では、研究キャリアパスとしてデータパブリケーションの評価を位置付けている。
  - 研究キャリアパスの他に、純然と研究を扱う人たちも考えないとならない。
  - ライフサイエンスの分野では同じような議論を行い、結局のところポストの話になった。そのポストのために研究ポストを減らすかというところで議論が止まっている状況ではないか。そこまで立ち入って、ポストを確保することを他の分野の人が認めるかどうかということが今回のポイントなのではないか。

### 【データキュレータ】

- ・データキュレータは何をする人なのか。
  - データを二次利用できる形に加工する人をデータキュレータとしている。
  - そうした人材の処遇については、URAのポストを増やす程度で良いという話にならないとも限らない。実際には、プロフェッサーを目指している人が行っているのが現状である。欧米では、ライブラリアンやデータキュレータの社会的地位も高くそうした専門職の人たちが行っている。大学にデータ管理と研究というポストがあって、若手のキャリアパスとして示せば良いのではないか。
  - ・データをリファーされる人は、キュレータかデータを提供した人か。
    - 現在は後者が多いが、キュレータが中心的な役割を果たす場合も想定される。固まっていない。
    - 論文のエビデンスとしてのデータと、データジャーナルの示すようなデータベースとしてウェルキュレートされたデータとでは、データの次元が違う。そこを混同して議論するとわからなくなる。キュレータは後者の方で、データベースのエコシステムをデザインする人とイメージすると良いのではないか。
    - それは、研究者なのではないか。研究がわからないとできない。素粒子の分野では、パーティクル・データベース・グループ (PDG) という 30 人ぐらいのグループがあり、そして、パブリッシュされたデータを集めて、整理している。彼らは高く評価され、業界をリードするようになっている。研究者として必要だと考えて、ボランティア的に行っている。そのために自分の研究が進まないということはあるかもしれないが、それが嫌な人はやらない。このデータの引

用数（グループとして）はとても多い。

#### 【ローデータとキュレーション、キュレータ】

- ・提言を出す場合、オープンサイエンスで取り扱うのは、論文のエビデンスでなく、それ以外の元のデータ（ローデータ）を対象とした場合どのような問題があるのか。
- 論文のエビデンスは論文に紐づいている、そのため書きぶりを考えると、論文からデータにたどり着かないとならないので、書きづらいかもしい。
- ・パーティクル・データベース・グループ（PDG）のような取組が評価され、業界をリードしているときに、こうしたことが今行われていない分野で、どうしたら行われるようになるのか。ヘルスケアの分野などでもいざタグを付けようとするとなかなか難しいということがある。データにタグを付けて共有することを海外を巻き込みつつ実現していくことがどうしたらできるのか。
- ・地球環境のグローバルなデータベースを扱う際の紹介をすると、次のようになっている。現地のデータは現地の人しかわからないので、ツールを用意し、現地の人にキュレートしてもらうようお願いする。そしてキュレートされたメタデータを付けながらアップロードされていく。ややこしいのはメタデータは一度つけたとしても収束するものではない。フィックスなものではなく、使い方によってどんどん変化し、ダイナミックなものであり、変わっていく。分野によるダイバーシティは大きい。
- ・キュレートする、データベースとしてまとめるということは、ある意味では評価をするということである。評価をしながら確立していく、研究作業である。
- ・データを利活用するためには、保存するだけでなく使える形にキュレーションがないとまらない。両者が必要であるが、現在ではキュレータがいなくて片手落ちになってしまう。
- ・分野ごとにキュレータの定義をしていかないと意味がないかもしれない。
- ・またキュレータの仕事の中身が、サイエンスの中身の変化に伴いどんどん変化している。
- ・図書館員が評価されると良いのではないかと。一方で定員削減が行われているので、難しい。
- ・図書館はオープンジャーナル対策で大変で、オープンデータまで手が回っていないのが現状だろう。

#### 【その他】

- ・研究の健全性に関連してはじまっている、研究データの保存の義務との関係についても本当は考える必要がある。
- ・企業的にみると、IP（知財）の管理の一環だと考えれば納得する。ライセンスの際に、何をどここの製品に紐づけされているか理解する。
- ・ベースラインを、天文にするのか、それとももっとビジネスに近いところにするのかで議論が異なるだろう。
- ・内閣府では、ファンディングエージェンシーの調査を行う予定である。日本学術会議には、多様な学協会に向けてどのようなメッセージを伝えるのか、気になっている。
- ・学協会に対しては、商業ベースだと意にそわないものになると思う。本委員会として、「その分野のあるべき姿を検討すべき」というウォーニングを言う必要がある。

#### (3) その他

##### 【次回以降の開催予定】

次回（第6回）は10月6日（火）10:00～12:00、次々回（第7回）は11月4日（水）10:00～12:00に開催する。

**【次回の内容等】**

次回（第6回）は、科学技術・学術政策研究所科学技術動向研究センターの林和弘上席研究官を参考人として招きヒアリングを行う。

また、提言（あるいは報告、記録）のとりまとめについても議論を行う。

加えて、11月12日に開催予定の内閣府総合科学技術・イノベーション会議オープンサイエンス推進に関するフォローアップ検討会において、土井委員長が本委員会の審議内容を紹介する予定であるが、その際の内容についても、次回委員会で相談することとした。

以上