

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3回)
議事要旨(案)

1. 日 時：平成27年6月15日(月) 16:00~18:00
2. 場 所：日本学術会議 6階 6-C(1) 会議室
3. 出席状況
出席者：土井委員長、大杉幹事、喜連川委員、三成委員(4名)
欠席者：杉田副委員長、戸田山委員、吉川委員、岡委員(4名)
参考人：大石雅寿 国立天文台天文データセンターセンター長
黒川顕 東京工業大学地球生命研究所副所長
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：盛田参事官、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2回議事要旨(案)
資料2 天文学における世界規模のデータ共有
資料3 微生物・環境データの統合化と今後の展開
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) 先例紹介：天文学における世界規模のデータ共有

大石雅寿 国立天文台天文データセンターセンター長より、資料2に基づいて、「天文学における世界規模のデータ共有」として、観測データの活用方法、天文学データアーカイブ利用の歴史、天文アーカイブ利用に関わる新たな課題(Virtual Observatory:VOとその国際アライアンス、データサービスの公開と利用の仕組み、標準策定作業、VO論文、欧米と日本の違い)、将来に向けて(大規模データを生成する望遠鏡計画、データが爆発する時代のデータマネジメント・解析、ICSUのWorld Data System、行政への期待)などについて説明が行われた。

その後、議論が行われた。主な発言は次の通り。

【データの占有期間】

・オープン化されるまでの観測データの提案者による占有期間は、どのように決まっているのか。
→観測所がユーザーコミュニティとの駆け引きで決める。提案者にとってはそのデータをなるべく長い間占有したいという希望があるが、一方で観測所としては、観測所のデータを使った論文がたくさん公表されるようにしたいという希望がある。なお論文を書くか書かないかに関係なく、占有期間が過ぎれば、データはオープンになる。

【公表される対象(データ、プログラム)】

・分析のプログラムは公開されているのか。そうしたプログラムもデータアーカイブに載っているのか。
→分析のプログラムについては、公開されているもの(ユーザーコミュニティが作成したもの。手弁当)とされていないもの(特定の目的のために作成されたり、秘密の処理がしてあり、使いにくいものなど)がある。基本的には公開している方が多いと思う。

【研究資金の調達方法】

・LSST(Large Synoptic Survey Telescope)にGoogle社が参加しているということであるが、そ

の場合の資金はNSF (The National Science Foundation) から出るのかそれとも Google 社が自前で出すのか。

→詳細な情報はわからないが、LSST に関しては、NSF が全面的に資金を提供しているということではなく、LSST が自前で資金を集めている部分がかかなりある。おそらく Google 社にも資金を提供してもらっているのではないかと。

【天文学分野のオープンデータによる論文、研究の質】

・観測をしてオリジナルのデータをもとにして書かれた論文と、オープンデータをマイニングして書かれた論文の割合はどのようになっているのか。

→最近では、われわれが「サーベイ」と呼んでいる全天サーベイすることを目的としているプロジェクトがあり、そうしたプロジェクトでは全天データを処理したうえで天文コミュニティに公表している。「自分では観測をしないけれども、公開されているサーベイを使って研究する」という人は増えており、そうした論文は多い。

・天文学全体の論文の中で、VO (Virtual Observatory) を使って生み出された論文はどの程度か。VO 以外の論文は、オリジナルのデータによるものと理解して良いのか。

→VO の論文の割合は、全体の中では多くない。現時点では従来のやり方 (オリジナルのデータ) の方が多い。オリジナルデータにアーカイブデータを加えて論文にする例も多い。VO、オリジナルのデータ以外に、サーベイによる論文も多い。VO は伸びていく傾向にある。

・質的な話として、VO によって良い成果につながるのが多いのか。

→例えば、銀河の中心に巨大ブラックホールがあることは、天文学の常識であるが、その成長については謎のままである。このなぞに関連する VO を活用した研究として、7000 万銀河のデータについて VO を使って集め統計処理をして、重たいブラックホールの周りの銀河の集中度を示したものがある。この研究のように、昔からやりたいと思っけていてもできなかったような研究が、できつつある。量が質に変わりつつある。

【天文学分野の研究方法】

・観測時間の獲得競争が 10 倍というのは、具体的にはどういうことか。

→例えば、サーベイとは異なり、アルマのような狭い範囲を指定した高感度で見るとような望遠鏡は、どこをどのように観測するのかについて、あらかじめ提案されたものから審査を経て選ばれたものについて行っている。提案数と採択数の比率が 10 倍ということである。一方でサーベイの場合は、広い範囲を一定の感度で均一に観測する。アルマとサーベイでは観測の仕方が異なっている。

・観測チームの人数や役割分担はどうなっているのか。

→研究室レベルでは、5 人程度のチームで行い、オリジナルの計画の策定から観測後の処理まで行う。一方、全天のサーベイになると 100 人単位で、ハードウェアの担当も含めると数 100 人になる。データ処理・ソフトの開発だけでも 100 人ぐらいいる。そのぐらいの人数で責任をもってデータを作り、公表している。データのクレジットはチーム全体になる。

【日米の慣習の違い】

・日本と欧米では、目指すもの、成果の使い方が異なるのか。

→あまり違いはないと思うが、ただし、天文学の研究者の数が欧米は日本の 10 倍いる。このことにより、装置を作らないで研究成果で勝負する人たちが欧米にはたくさんいる。研究の提案や、オープンになったデータの活用について常に考えている人たちがたくさんいる。質も量も高い。また欧米ではユーザーサイドからの観測所への要求も厳しい。観測所としてはその要求にこたえながら多くの研究論文が書かれれば「よくやった」という評価につながっている。

→文化的な違いはそのままよいのか。

→それは問題で、科学的成果をたくさん出すこと・きちんと論文にすることについて、研究者がもっと意識する必要がある。そうした機運を高めるために、雑誌の特集を行ったり、デモンス

トレーションを行ったりしている。

【人材にかかわる課題】

- ・IT リテラシーの高い人材にデータ処理をしてもらわないとならないが、そうした人材が不足し、その人たちのキャリアパスも見えてこないということが、多くの分野で生じているようであるが、この点について天文学分野の場合はどうか。
- 同じように困っている。オープンデータ・オープンサイエンスの流れもあり、人を付けるように国立天文台に対しても働きかけているところであるが、持続的な人材（シニアと若い人の組み合わせ）で人を配置することが難しい。特に若手のポストは、助教クラスの場合は任期付きになってしまい、10年計画が立てられない。
- あらゆる分野で同じ状況が生じているので、分野ごとに人材を考えるよりは、もう少し大きなメタなところで整理した方が良い。また、分野ごとのデータセンターというよりは、もう少し大きなまとまりなのかなという考えを持っている。例えばMicrosoftのWorld Wide Telescopeのようなものを活用できるような仕組み、エコシステムを考える必要があるのではないかという感想を持った。

(3) 先例紹介：微生物・環境データの統合化と今後の展開

黒川 東京工業大学地球生命研究所副所長より、資料3に基づいて、「微生物・環境データの統合化と今後の展開」として、微生物研究を取り巻く現状、メタゲノム解析、MicrobeDB とそれを活用した解析、データ公開に対する問題点、GenBank や MG-RAST (metagenomics Rapid Annotation using Subsystem Technology) 等諸外国のデータベース、ビジネス展開、今後の展開・問題点等について説明が行われた。

その後、議論が行われた。主な発言は次の通り。

【MicrobeDB について】

- ・遺伝子や微生物に関するデータは農水省関係の研究開発法人などでも収集しているがそれと、MicrobeDB の関係はどうなっているのか。また、MicrobeDB を一般向けにしたのはなぜか。
- MicrobeDB については、オープンされている情報をすべて集めている。したがってそうしたところのものでもクローズドのものは集められていない。また、MicrobeDB を一般向けにしたのは、微生物は地球上至る所に生息し、ありとあらゆる分野・研究と関係があるため、他分野の人々が使うということを想定して、この形になった。

【オープン化のドライバー】

- ・FASTA や BLAST というソフトが GenBank の発展のドライバーになったのはなぜか。FASTA や BLAST がオープンならば自分で解析もできるのではないか。
- 当時は DNA の解析にお金がかかっていた。自分で解読できる遺伝子情報がわずかであったため、それを集めて集合的に検索できるということが大きなメリットであった。FASTA や BLAST によって解析できるようになったため、皆がデータを GenBank に入れたがるようになった。
- MG-RAST も同じようなことか。
- 今、ヒトゲノム、マウスゲノムなどについてはほとんどオープンになっており、研究者にとっては自分が再解析したマウスのゲノム情報をデータベースに登録することはやりたがらない。しかし、「メタゲノム」については状況が全く異なり、環境中にどんな微生物や遺伝子があるのかがやっとわかった段階である。したがって、データ登録により解析して、どんな種類の細菌がいるのか知りたいという気持ちが、ドライビングフォースとなっているように思う。
- ・解析手法がアトラクティブポイントなのか、それとも他の人のデータが使えることがアトラクティブポイントなのか。
- GenBank の場合は他の人のデータと比較したいということが理由であった。一方で MG-RAST の場合は、そうではない。メタゲノムという膨大な情報量を持つものを対象にしているため、自

分の PC で計算させると 1 年間以上かかるが、MG-RAST にデータをアップロードすると使えるパソコンと統計処理ソフトにより、1 か月程度で結果が得られるというメリットがある。

【他のデータのリンク、応用イメージ】

- ・他のデータとのリンクはどのようなものか。
- アメリカではシカゴ川のメタゲノムを定期的に解析している。他の環境データとの関連性を解析はもちろん、コンクリートを打ったことの影響、排水の経路と影響なども見ており、その結果を都市計画に生かそうとしている。

【オープンクローズ戦略】

- ・情報がビジネス界の大きなところにどんどん利用される面があれば、公的なものに利用される面もあるだろう。データベース等を利用したビジネス展開があり得る場合に、何をどこまでオープンにするべきと考えているのか。どんどんオープンにすれば良いのか。
- MicrobeDB では企業のプライベート利用が開始されていて、彼らは自分たちのデータをオープンにすることはしない。ただし、最初に協力依頼をする際に、何をどこまで出して良いかについて（配列情報、アノテーション情報・メタ情報）確認している。公表される情報は、段階的になっている。検索をした際に、「配列情報のみ」など限定された情報であってもヒットすることが他の研究者にも役に立つ。農業をやっている人に話を聞くと、農法や農地のメタゲノム情報について、「自分はオープンにしたいくないが、他人のものは見たい」という意見を聞くことがある。こうした点から考えて、公的な研究費によって行われた研究は、徹底的にオープンにする必要があるのではないか。
- ・Google 社が来る中でどうやって国益を守るのか。日本はどこで生業をつくっていくのか。オープン・クローズ戦略をどうするかが問題である。データが国のアセットになりつつある。その点を考えなければならない。その戦略をインパラレルで行わないとならない。どう守るかと言うことをきちんと行わないとデータの循環が行われないのではないか。
- 日本人は長寿なので、食べ物のデータなど、魅力があるデータがあり、流出を危惧しているが、対応についてはノーアイデアである。
- そのあたりは JST の COI（センター・オブ・イノベーションプログラム）で議論しているが、ビジネスに対してどこまで守るかと言った点についてはそこまで話が進んでいないのが現状である。
- 「オープン化」というと、「データを出せ」という話からはじまりがちであるが、実は、その前段階の語彙の統一やメタデータの整備など前作業を各分野が合理的にするかということが非常に問題となっている。重要なのは、オープンが可能なような段階までデータをアセット（資産）として整えておくことである。まずデータマネジメントの作法がきちんと出来て、アセット（資産）として確立するプロセスがあって、その先にオープンにするかどうかということになる。その中で守るべきは守り、出すべきものは出すというように使い分ければ良いのではないか。
- 構造化および標準化のところが本当に難しい。各国で概念が異なり、例えば、海の砂浜の波打ち際の微生物は、sea なのか soil なのか何なのか、すりあわせをしていかないとならない。そうした作業は、多くの場合は業績として評価されない。そのあたりも含めて、業績になるような仕組みを考える必要がある。

（４）その他

【次回の開催について】

次回は、7月21日(火)13時より開催し、北村行伸一橋大学経済研究所教授、青柳正規文化庁長官の2名を参考人として招きヒアリングを行う。

【今後の審議の進め方について】

次回の委員会は9月以降に開催する。その際に、これまで出てきた論点などを土井委員長が整理し、それに対する意見や追加ヒアリングの必要性等について議論を行う予定とした。

以上