SPARC Japan NewsLetter
ISSN 1883-826X
NO.31 2017 年 1 月

SPARC Japan ニュースレターでは、各回セミナーの報告に講演やパネルディスカッションを書き起こしたドキュメントを加え、さらにそのほかの SPARC Japan の活動をご紹介します。今回は 2015 年以降の海外動向調査についてご報告します。

CONTENTS

■ SPARC Japan 活動報告

SPARC Japan 運営委員会

SPARC Japan 運営委員会の会議資料をウェブサイトで公開しています。
http://www.nii.ac.jp/sparc/about/committee/

海外動向調査

第 4 期及び第 5 期 SPARC 事業の達成目標の一つとして「オープンアクセスに関する基礎的情報の把握」を掲げており、その具体的活動として海外動向調査を行っています。ここでは国際会議参加及び積極的にオープンアクセスを推進している海外関係機関への訪問調査等の報告をご紹介します。

- 研究データ管理調査報告 : IDCC、DCC ワークショップ、サウザンプトン大学実践調査
  岡山大学　大園 隼彦、鹿児島大学 西薗 由依（以上、機関リポジトリ推進委員会協力員）
  2015 年 2 月 8～17 日
  http://id.nii.ac.jp/1280/00000083/
- COAR-SPARC Conference 2015 参加およびミーニョ大学訪問調査報告書
  千葉大学　三角 太郎、鹿児島大学　西薗 由依（以上、機関リポジトリ推進委員会協力員）
  2015 年 4 月 14～17 日
  http://id.nii.ac.jp/1280/00000107/

- OpenAIRE のメタデータマネジメント調査出張報告書
  九州大学　林 豊、北海道大学　三隅 健一（以上、機関リポジトリ推進委員会協力員）
  2016 年 2 月 20～27 日
  http://id.nii.ac.jp/1280/00000204/
- 11th International Digital Curation Conference 参加報告書
  国立極地研究所　南山 泰之（機関リポジトリ推進委員会協力員）
  2016 年 2 月 22～25 日
  http://id.nii.ac.jp/1280/00000203/
- Research Data Alliance 第 7 回総会参加報告
  千葉大学　三角 太郎、国立極地研究所　南山 泰之、鈴鹿工業高等専門学校　青山 俊弘、
  お茶の水女子大学　香川 朋子（以上、機関リポジトリ推進委員会協力員）
  2016 年 3 月 1～3 日
  http://id.nii.ac.jp/1280/00000202/
- CRIS2016＆OR2016 参加報告
  九州大学　林 豊（機関リポジトリ推進委員会協力員）
  2016 年 6 月 8～16 日
  http://id.nii.ac.jp/1280/00000205/

■ SPARC Japan セミナー報告

第 2 回 SPARC Japan セミナー 2016 (オープンアクセス・サミット 2016)

「研究データオープン化推進に向けて
：インセンティブとデータマネジメント」

2016 年 10 月 26 日（水）　国立情報学研究所 12F 会議室　参加者：112 名

本セミナーでは、研究データのオープン化について、「図書館員・研究者の協同」という観点から、オープン化を進めるインセンティブやデータマネジメントについて取り上げます。研究や業務のフローにデータマネジメントを組み込むための議論を、実際の取組みを踏まえつつ展開しました。

次ページ以降に、当日参加者のコメント（抜粋）、企画後記およびドキュメント全文（再掲）を掲載しています。その他の情報は SPARC Japan の Web サイトをご覧ください。（http://www.nii.ac.jp/sparc/event/2016/20161026.html）

概要

日本におけるオープンサイエンス推進のあり方については、2015 年 3 月に内閣府から報告書が公表された。それによると、国としての基本姿勢・基本方針は、公的研究資金による研究成果の利活用促進を拡大することとされており、ここで言う研究成果には研究の過程で得られた「デジタル化された研究データ」も含まれている。

研究データ・サイエンスデータのオープン化に当たっては、大学・公的研究機関、データを生み出した研究者の積極的な役割が期待されるところであるが、昨今の厳しい研究環境を背景として研究者の内発的動機づけに至っていないうえ、ある程度ワークフロー化した研究データマネジメントシステムが確立されていないのが現状である。こうした現状の指摘は、この半年ほどの間に「戦略的創造研究推進事業におけるデータマネジメント方針(JST)」、「G7 茨城・つくば科学技術大臣会合『つくばコミュニケ』(内閣府)」、「オープンイノベーションに資するオープンサイエンスのあり方に関する提言(日本学術会議)」などに相次いで見られ、その解決は焦眉の急といった様相を呈している。

研究者及び科学コミュニティに対しては、研究データのオープン化を進めることにより新たな知見や価値が生み出せるというインセンティブに加え、オープン化の成果に見合った処遇を与えるといったインセンティブを高めることも重要であると考えられる。研究データマネジメントに関しては、データの長期保存・管理・公開において図書館・機関リポジトリ・データセンターが果たす基盤的な役割は大きく、こうした機関の構成員と研究者との協同をワークフローに組み込むことはこうした問題を解決する可能性を秘めている。

以上を背景として、本セミナーでは、自然科学分野で実際に行われている図書館と研究グループ連携の取り組みや機関リポジトリの現状などの話題提供を通して、日本における研究データ・サイエンスデータのオープン化を「図書館員・研究者の協同」という観点から今後どのように推進していくことができるかを考えてみたい。

参加者から

（大学/図書館関係）

・各リポジトリで公開されているデータをどのように探すのか。本当に目録作業の延長なのか。図書等と違って、研究データは誰が見てもそうとわかるものではなく個別的というか分野で特性があったりするのではないか。リポジトリを運営する大学などのインセンティブは？

・セミナー全体の話の要旨は、研究者が自身の研究データ等を公開していく取組に煩わしさを感じており、本来の研究や教育に忙殺されて、研究成果や取得データの公開に対しモチベーションが下がっている。そのためにどんな方略が有効か、みんなで考えていきましょうという話が多かった。当初聴講をするつもりだった内容とずれていたが、研究者（大学教員等）の本音

が多く聞けてこれはこれで有意義な時間だった。
（大学/研究者）
・図書館員の役割について理解が深まった。
（大学/その他）
・自分が扱う分野と異なるデータの話も興味深くきけた。図書館のリソース(人)は限られており、メタデータ作成を手伝う部署の設置が必要に思った。
（その他/その他）
・研究者、学芸員、図書館員などそれぞれの役割・それぞれのインセンティブを考えるきっかけとなった。
・機関としてどう取り組むか、どのような課題があるかの視点が得られた。

企画後記

今回、SPARC Japan セミナーの企画に初めて主査として携わりました。普段は、大学教員として研究・教育を行うことに加えて、研究データのデータセンターを運営することを業務としています。そういった背景から、「オープンサイエンスを推進するためには、データを取得している研究者やデータを管理している研究者に対して、インセンティブを示す必要があるのではないか」と常々考えていました。第 2 回セミナーの開催趣旨は、この考えに端を発しています。企画を進めていく上でのブレインストーミングで、研究者へインセンティブを付与していく上で、図書館・機関リポジトリ構成員の果たす役割が明確になっていったり、当日のセミナーで、インセンティブにもいろいろな定義がありうるということが理解できたり、と私にとっていろいろな発見がありました。今後、日本でオープンサイエンスを推進していく上で、理念や理想だけではなくインセンティブという視点での議論も盛り込んでいければと考えています。多くの皆様にセミナーへご参加をいただきましたことを御礼申し上げます。

能勢　正仁
（京都大学大学院理学研究科）

研究分野の流行り廃りを超えて長い年月を費やし研究データを作成管理してきた研究者にとって、研究データ共有へのインセンティブは関心の一大事であるということだ。研究の遂行に一体化されたデータ作成と管理は労働集約型で、観察から得られる知見を検証せしめるにはなければならないものである。にもかかわらず、研究データ共有への新たなインセンティブは眼前に現れてこない。研究者にとってのインセンティブは学術の文化として埋め込まれ、制度として確立されて、もはや動かしようのないシステムのようだ。いま、我々は、地政をもってこれに改革を試みようとしている只中にいる。

蔵川　圭
（国立情報学研究所）

今回は twitter 係として、セミナーの内容の情報発信を担当した。会場では、オープンアクセスウィークということもあってか、インセンティブとデータマネジメントの観点からの研究データオープン化推進に関する議論が盛り上がりにみせたのに、それを十分に伝えきれなかったのが残念でした。せっかく公開可能なコンテンツなのだから、リアルタイム配信が今後も実現するよう次回以降に期待したい。

坊農　秀雅
（情報・システム研究機構
ライフサイエンス統合データベースセンター）

今回のセミナーでは企画側からの登壇となりました。企画段階では研究者と図書館員の行動規範の違い、という話題で盛り上がっていましたが、セミナー当日にはあまりご紹介できなかったのが少し心残りです。考え方も目的も異なる（かもしれない）両者が協働することで、科学がより活性化し、よりオープンな方向性に向かう可能性に期待しています。

南山　泰之
（国立極地研究所）

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

趣旨説明

能勢　正仁
（京都大学大学院理学研究科）

能勢　正仁

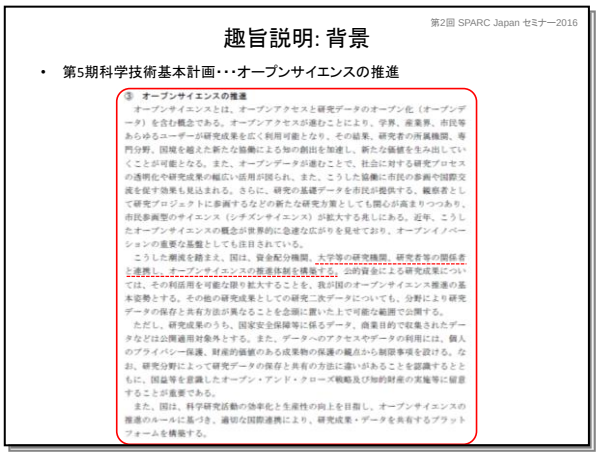
1998年に京都大学理学研究科で博士（理学）取得後、米国ジョンズホプキンス大学でポストドクトラルフェローとして3年間研究を行う。2001年帰国、現職。専門は、超高層物理学、地球電磁気学。主な研究テーマは、地磁気変動・脈動、内部磁気圏の高エネルギー粒子ダイナミクス、サブストーム、地磁気指数など。最近は、科学データへデジタルオブジェクト識別子を付与する活動にも積極的に関わっている。

最近、科学技術政策に関する文書で、「オープンサイエンス」という言葉が目に付くようになりました。図 1 は一例として「第 5 期科学技術基本計画」から引用してきた文章です。オープンサイエンスに関して、このように一つの節を設けて記述されています。赤線部で、「国は大学等の研究機関、研究者等の関係者と連携し、オープンサイエンスの推進体制を構築する」と述べています。これは、オープンサイエンスに興味を持つわれわれとしてはありがたいことなのですが、その推進を具体的にどのように進めていくかに関してはまだ手探りの状況で、われわれが考えていかなければいけない問題です。

推進にはインセンティブが必要

オープンサイエンス推進に当たって重要な役割を果たすプレーヤーとしては、データを持っている大学の研究機関や研究者が挙げられます。しかし、データをオープンにすることに興味を持つ研究者はまだ少数にとどまっているのが現状です。従って、推進に当たっては、インセンティブが必要だと考えられます。研究者は研究を行い、その研究成果を論文として発表しますが、なぜそのようなことをするかというと、もちろん研究自体が楽しくて、自然がどのような仕組みになっているかを究明していきたいという理由もありますし、論文を書いて世に問うという理由もありますが、論文を書かなければ自分の研究成果として認められないという理由もあります。要するにインセンティブが働いているのです。

同じようなことがデータのオープン化に対しても言えるのではないでしょうか。大局的もしくは内的なイ



（図 1）

ンセンティブとしては、データをオープンにすることによって新たな知見や価値が生み出せる、データをシェアすることによって科学技術の進展に貢献できるということがあります。

局所的もしくは外発的なインセンティブとしては、成果に見合った処遇や研究費が得られるといった、馬にニンジンを与えるような（図 2）ことがあります。このようなインセンティブを考えていく必要があると思います。

研究者と専門家集団の対話・協同

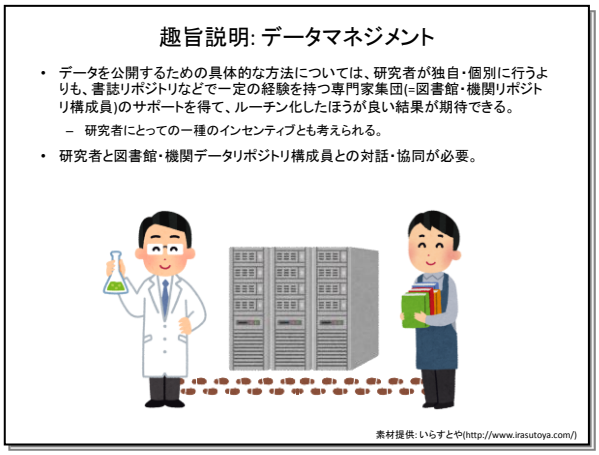
仮に、このようなインセンティブの仕組みができたとしても、研究者はデータをオープンにすることに慣れておらず、どのような形でオープンにしていけばいいのか分からないので、データを公開するための具体的な方法に関しては、研究者が一人で悩んでいるよりも、リポジトリなどで一定の経験を持つ専門家集団（図書館や機関リポジトリの構成員）のサポートを得てルーチン化した方が良い結果が得られることが期待できます。これは研究者にとっても、データをすぐ簡単にオープンにできるといった一種のインセンティブとして捉えることができます（図 3）。

従って、研究者と図書館・機関データリポジトリ構成員との対話・協同が今後は必要になってくるのではないでしょうか。

本セミナーの狙い

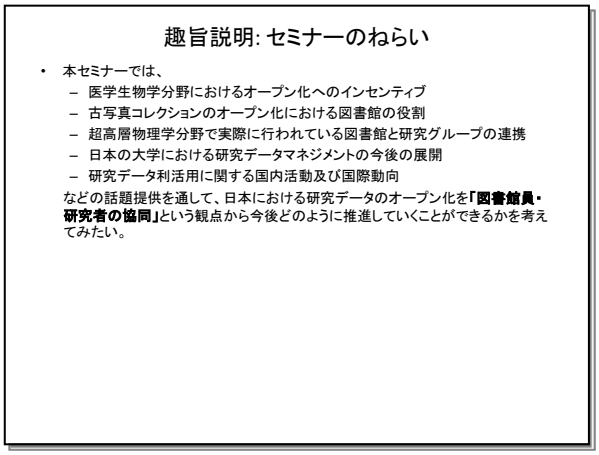
以上のことを受けて、このセミナーでは図 4 の五つのテーマについて各先生方から話題提供いただき、日本における研究データのオープン化を「図書館員・研究者の協同」という観点から、今後どのように推進していくことができるかを考えてみたいと思います。

最後のパネルディスカッションでは、皆さまに、活発なご議論を通して、このセミナーの狙いを達成していただければと存じます。



（図 3）

（図 2）



（図 4）

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

医学生物学分野における データのオープン化とそのインセンティブ

仲里　猛留

（情報・システム研究機構ライフサイエンス統合データベースセンター）

講演要旨

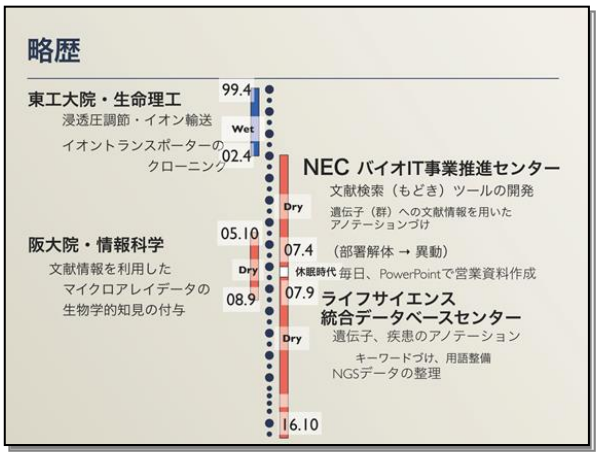
医学生物学分野では、文献情報が MEDLINE に、 塩基配列情報が GenBank にと 1970 年代からデータベースとして保存されており、インターネットの普及とともに誰でも広く利用可能となった。現在は米国 NCBI、欧州 EBI、日本の DDBJ によって、これらに加え遺伝子発現や化合物情報なども公共データベースとして蓄積・運用がされている。このように自然に生命科学データが集積されたのは、出版社が投稿規定内に実験データを公共データベースに登録することを義務化した背景がある。近年は、機器の発展やコストの低下により、診断などヒトを対象にした研究も広くなされるようになり、場合によってはデータをオープンにすることにより個人情報を侵害するのではないかとの懸念も持たれている。

仲里　猛留

東京工業大学大学院生命理工学研究科修士課程修了。在籍中はウナギの海水適応機構をテーマに分子生物学実験の日々を過ごす。修了後、NECに入社。バイオインフォマティクスの部署で文献情報を活用した大規模データの解析のためのソフト開発に従事。2007年より情報・システム研究機構（ROIS）ライフサイエンス統合データベースセンター（DBCLS）に勤務。DDBJと連携して公共データベース中の次世代シーケンサー（NGS）データを検索するサイトの運用や、それら大規模データの生物学的な解釈を行うための基盤整備を行う。2011年より現職。大阪大学大学院情報科学研究科博士後期課程修了。博士（情報科学）。

私が所属するライフサイエンス統合データベースセンター（DBCLS）はあまり聞き慣れないと思いますが、親組織が情報・システム研究機構で、国立情報学研究所、国立遺伝学研究所、統計数理研究所、国立極地研究所と一緒です。

私の年表は図 1 のようにまとめられます。今日は研究者の立場からデータのオープン化についてお話ししたいと思います。大学院、マスター（修士課程）までは魚から遺伝子を抽出して研究していました。例えば、ウナギを買ってきて 30 匹ほど解剖し、そこから血圧調節に関係する遺伝子を薬品を調合して抽出しました。ウナギはサケやマスと一緒で、淡水と海水の両方に行ける魚です。淡水のときは周りが真水なので、自分のイオンが抜けて水が入ってきます。海に行くと逆にな



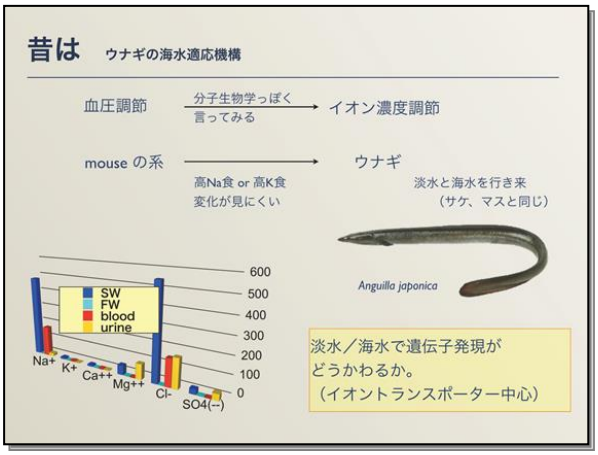
（図 1）

って、周りが塩辛いので、水が抜けてイオンが入ってきます。ちょうど逆になるので、淡水だけ出てくる遺伝子、海水だけ出てくる遺伝子を探していたのです（図 2）。

その後、私は実験に挫折し、IT バブルに乗って、NEC バイオ IT 事業推進センターに就職しました。そこも IT バブルがはじけて 5 年で部署がなくなり、その後、縁があって、DBCLS というアカデミックな機関に所属しています。

国立遺伝学研究所の中にある、日本 DNA データバンク（DDBJ）という機関は、DNA の A・C・G・T などの並び順をたくさん読める次世代シーケンサーから取ったデータを集めていますが、その検索サイトをつくる仕事が今の本業です（図 3）。

DBCLS の目的は、たくさん出てくる生命科学系のツールやデータをうまくリサイクルして研究者の人に使ってもらうことです。



（図 2）

（図 3）

生命科学分野におけるデータベース

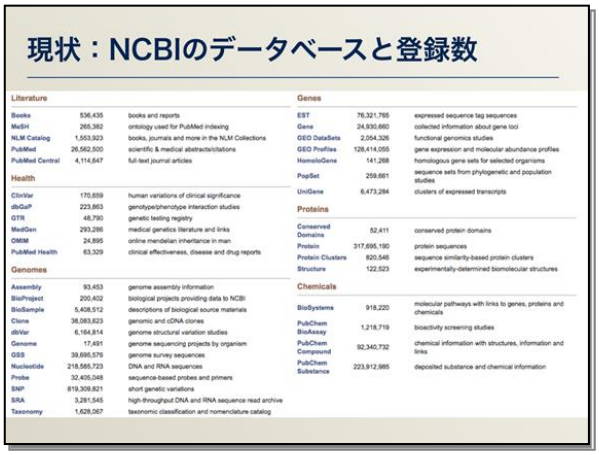
私は生命科学分野にいるので、そのデータベースの実情を最初にお話ししたいと思います。

バイオ系のデータは、アメリカ・ヨーロッパ・日本の三つでコラボレーションして集めています。私が学生のころは、アメリカが一番集めていて有名だったので、アメリカの国立生物工学情報センター（NCBI）のデータベースを使うように言われていました。NCBI のデータベースがどのようになっているかを図 4 に列挙しています。

文献関係、ヘルスケア関係、DNA 関係、化合物関係などいろいろなデータベースがあるのですが、左の下から 5 番目の「Nucleotide」にデータが約 2 億 1,800 万件入っています。文献関係では「PubMed」が上から 4 番目にあり、約 2,650 万件入っています。

私が研究室に入ったのは学部の 4 年生ですが、そのころから PubMed や Nucleotide などを見て、いいデータを探すような仕事をしていました。

図 5 は、コラボレーションしている国立遺伝学研究所のデータです。赤線が右の縦軸で Nucleotide の件数を表しています。約 2 億件です。水色のバーは左の縦軸でデータの量を表しています。bp（base pair）というのは塩基の数で、A や C などが何個入っているかだと思えばいいのです。単位は billion で 10 億ですから、2,000 億ぐらいとたくさん入っているのです。す



（図 4）

ごくたくさんのデータが入っていて、私が学部 4 年生だったのは 1999 年なので、ほぼ地を這うようなデータです。今は 2016 年なので、もう一つ棒が立つかどうかのところにいて、その何倍だろうかというぐらい急速にデータが増加しています。私が専門でやっている次世代シーケンサーは、ここ 5～6 年ぐらいにデータがたくさん出はじめ、本格的に集められています。

データベースの変遷をまとめた図があるので、借りてきました（図 6）。NCBI で、先ほどお見せしたようにいろいろな種類のデータベースが集められているのですが、もともと NCBI で全部メンテナンスされていたかというと、そうではなく、あちこちから NCBI に集めたという経緯になっています。

例えば、Nucleotide は、ここでは GenBank と書かれていますが、最初は 1972 年、ロスアラモス国立研究所で、物理学のデータの一つとして DNA の並び順がストックされるようになり、その後、1990 年少し前ぐらいから NCBI に集められました。

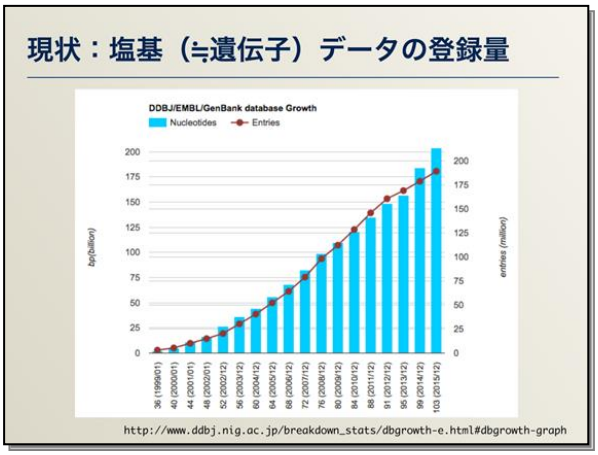
オレンジのバーは文献です。この後、PubMed が出てきますが、1960 年代からずっと集められています。疾患データベースの OMIM はジョンズ・ホプキンス大学で四十数年ぐらい、McKusick 博士が、「病気はこのような感じです」とカタログしていたのを、最近はインターネット時代なので全部デジタル化して、NCBI などから検索できるようにして、この中に一応入っていたのですが、最近は予算が切れて、またジョンズ・ホプキンス大学に戻りました。

このような感じであちこちで集めていたものを NCBI に集め、最近は無料でパブリックが自由に閲覧・検索できるような状況になっています。

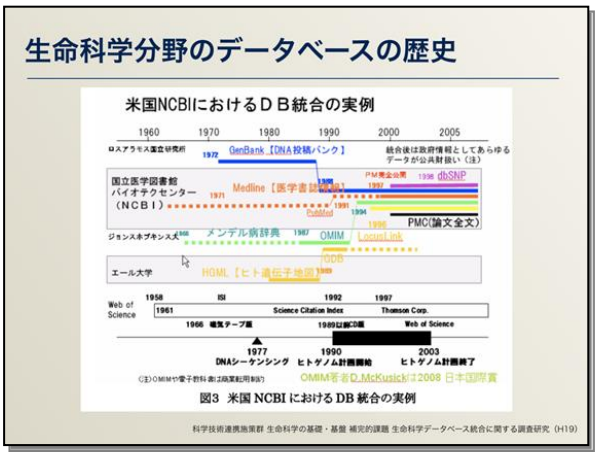
アメリカの NCBI、ヨーロッパの欧州バイオインフォマティクス研究所（EMBL-EBI）、日本の DDBJ は、国際塩基配列データベース（INSDC）というコンソーシアムのようなものを組んでいます。例えば、NCBI に入ったデータも大体は DDBJ から検索でき、逆もしかりで、DDBJ に集まったものが NCBI から検索できるという形になっています。

ちなみに、NCBI は、上位の組織がアメリカ国立衛生研究所（NIH）で、NIH の下にアメリカ国立医学図書館（NLM）という図書館の部門があり、この下に NCBI というデータベースをメンテナンスしている組織が入っているという構造になっています（図 7）。

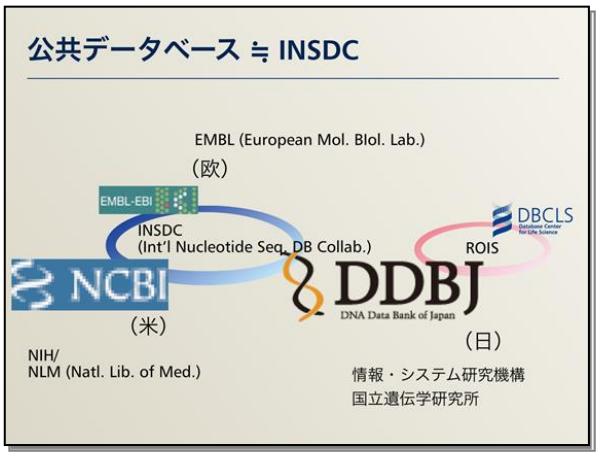
私は DBCLS にいるのですが、この 3 局で集めたデ



（図 5）



（図 6）



（図 7）

ータについては、特に DDBJ の場合は予算の兼ね合いもあって、集めるのにリソースを割いていました。私たちはデータをリサイクルさせるのが目的なので、DDBJ が集めたデータを検索できるようにします。これを同じ情報・システム研究機構の下位組織同士でやっているということです。

NCBI のデータベースはかなりたくさんあり、いろいろ集まっていると感じたかもしれませんが、データベースは NCBI だけではありません。「Nucleic Acids Research」という雑誌があり、これは年に一度、1 月に「Database Issue」というデータベース特集号、7 月に「Web Server Issue」というウェブツールの特集号を組んでいます。

図 8 は 2016 年のもので、23 回目の特集号です。一つのデータベースで一つのペーパー（論文）ではありませんが、ここには 178 のペーパーが入っています。そのうち 62 が新しいもの、95 がアップデートと書いてあります。これが毎年出ているので、ここに載らないようなデータベース、「Bioinformatics」「BMC Bioinformatics」など、いろいろな雑誌に載っているデータベースも合わせると、年間 1,000 ぐらいできているのではないかと言う人もいます。NCBI など大手のものもあれば、各研究者がつくるような小さいものもあるのですが、どんどんデータベースが生まれています。もう少し言うと、お金が切れると、マシンが壊れるとともに死んでいくということがあったりするのです。

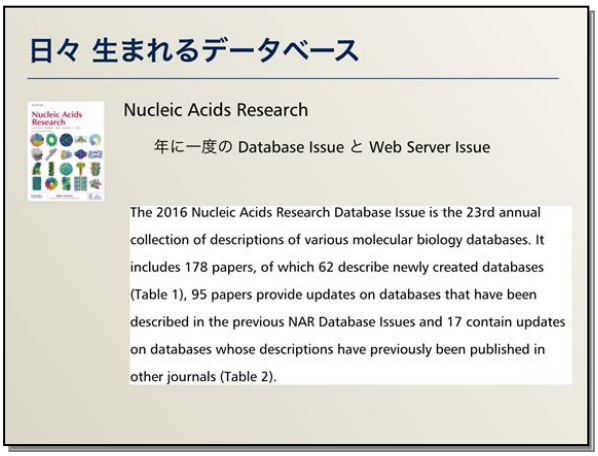
主要な生命科学データベース

特に注目すべき、私がまだ実験をしていた学生のころによく使っていたデータベース、ウェブツールが二つあります。一つは PubMed という文献のデータベース、もう一つは BLAST という類似した遺伝子を検索するツールです。

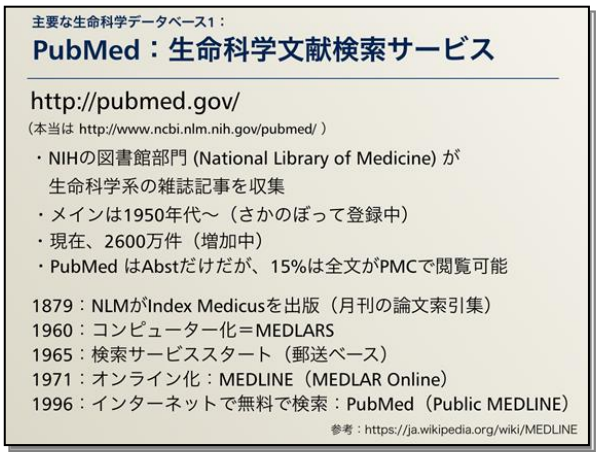
まず、PubMed について述べます。図 9 は NCBI の PubMed で、実際の検索結果の画面です。がん、糖尿病などのキーワードを入れると、関連する文献のリストが出てきます。詳細画面には、文献のタイトル、著者、掲載雑誌、発表日などの情報、Abstract（要約）が書かれています。バナーをクリックすると PDF を取れます。

ちなみに PubMed Central（PMC）という論文の全文を収載したものもあります（図 10）。PubMed にはデータが約 2,600 万件あるのですが、そのうちの約 15%

（図 9）



（図 8）



（図 10）

は PMC という別のデータベースにも収載されているのです。そこでは、論文の Abstract だけではなく、イントロ、メソッド、ディスカッションまで全部読めるようなものがあります。

物理で言うと、例えば Google Scholar、Citeseer などが昔やっていたように思いますが、生命科学分野では PubMed があって、今は全部読めるようにだんだんなってきています。

その歴史を見てみると、もともとは 1879 年に NCBI の上位組織 NLM が、「Index Medicus」という月刊の論文索引集を出しました。それが 1960 年代に MEDLARS としてコンピューター化され、その検索サービスが 5 年後ぐらいにスタートしました。それが MEDLARS Online としてオンライン化されたのは 1971 年です。電話をかけて取得したり、郵送でお願いしたりするのが、この時期のやり方でした。

私が研究を始める少し前の 1996 年に PubMed ができました。MEDLINE がパブリックになったので PubMed という名前が付いたと聞いています。インターネットに接続さえすれば、誰でも MEDLINE に入っている文献を検索して無料で見られるようになったという流れで、文献のデータが公開されてきました。

次はもう一つの BLAST についてです（図 11）。私は学生時代にウナギの血圧調節などに関係する遺伝子を探していました。当時は DNA シーケンサーを使うと、うまくいくと DNA の A・C・G・T の並びが 1,000 文字ぐらい取れました。それを BLAST に投げると、あなたの投げたものはここからここまであって、赤いものが似ているもの、緑がほんのり似ているもの、そしてだんだん似なくなってくるのですが、この辺がどのくらい似ているかというチャートが出てきます。具体的に A・C・G・T のどの辺りがどのぐらいマッチしたのかという詳細が出てくるというものになっています。

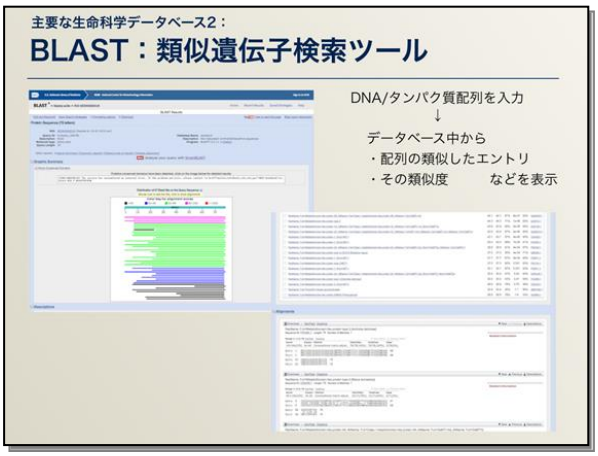
文献が MEDLINE に登録されると、PubMed でそれを検索できるようになります（図 12）。同じように、遺伝子の塩基配列が GenBank、EMBL、DDBJ などの DNA のデータベースに登録されると、BLAST で検索をかけられるようになります。BLAST は DNA の塩基配列情報がオープンでないと検索できないのです。結局、BLAST も PubMed も全てオープンデータの上に成り立っているツールだということです。

このようなものがないとわれわれは研究ができないと、学部 4 年のときにも言われていました。私が研究を始めたときには既にこのようなものがあったので、空気のようにオープンなデータを使っていたのだということを今あらためて感じています。

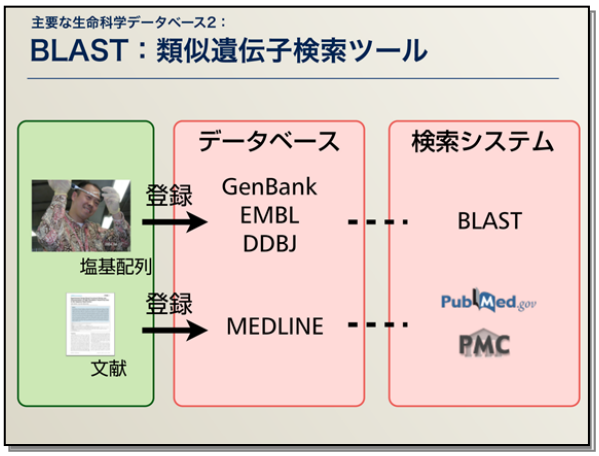
公共データベースにデータが集まる理由

それでは、なぜこのような公共データベースにデータが集まるのでしょうか。

理由の一つには、論文の投稿規定で要求されているということがあります。図 13 は「Nature」の投稿規定です。タンパク質配列、DNA 配列、遺伝子発現な



（図 11）



（図 12）

どのデータの種類、その横に GenBank、DDBJ、EMBL などのデータベースの名前が書いてあります。このデータが出てくる論文があれば、このデータベースに登録しなさいということです。登録してアクセッション番号を取らなければ、論文を掲載してくれないのです。

研究者は論文を出して何ぼで、成果は論文で測られます。生命科学分野では特に、データベースにきちんと登録しなければ論文が出せず、成果にならないのです。半分、強制、仕方なくというところもあるのですが、投稿規定で要求されているので、自然とデータが集まります。

もう一つの理由がファンディングエージェンシー（研究資金の配分機関）からの要求です（図 14）。例えば、NHI Data Sharing Policy には、1 年で 50 万ドル以上のグラントであれば、データをどのように公開するのかプランを考えなければいけないということが明記されています。

日本でも科研費の公募要領にも、「バイオサイエンスデータベースセンターへの協力」が掲載されています。バイオサイエンスデータベースセンターというのは科学技術振興機構（JST）の組織で、もともと私たちがしていたデータのリサイクル事業を今引き継いでもらっています。「論文発表等で公開された成果に関わる生データの複製物、又は構築した公開用データベースの複製物について、同センターへの提供に御協力をお願いします」と書いてあります。私たちも、データを出せ出せと言う圧力団体のような面があるので、このようにデータをオープンにするように公募要領に書かれるようになったのはそのような活動の成果です。

オープン化のインセンティブ

自分なりに、オープン化のインセンティブとは何かを考えてみました。研究者の視点から言うと、やはり究極の目的である論文が掲載されることでしょう。自分で論文を書くようになって分かったのですが、自分のデータが使ってもらえたり、論文が引用されたりするとうれしいのです。これは世代の差があるかもしれませんが、特にピペットや薬品を使うような生命科学の実験をする人は、もったいない精神にたけていて、少ないお金でデータを出し、それを自分だけで骨と皮しか残らないぐらいにしゃぶり尽くしたがる人が多いのです。

昔は、自分の物は自分の物、人の物は自分の物というジャイアニズムのように、データの囲い込みをする人たちが多かったです。しかし今は、オープンにした方がプレゼンスが上がります。例えば、論文をオープンアクセスのジャーナルに出すと、学会で「あなたの論文を読みました」と言われたり、読んだ人からコンタクトが来て、「私たちはこんなにいいデータを持っているから一緒にしませんか」と言われたりします。そうやって世の中を動かせるようになるので、研究者としても、実はオープンにした方がプレゼンスが上がるのではないかと思っています。そうなると、実際に

（図 13）

（図 14）

もらえる研究費も増えて、次の成果につながり、ハッピーなサイクルが回るのではないでしょうか。

これはお金のどろどろした話でしたが、別の観点で、付加価値の付与というインセンティブがあると考えています。私が本業で扱っている次世代シーケンサーのデータは大きいので、自分の手元のマシンで解析できない、そのスキルがないという人が結構いるのです。しかし、DDBJ にデータを登録すると、DDBJ の方で解析のパイプライン、ウェブツールを組んであるので、クリックするだけでそれにすぐに流し込めます。自分のところで解析せずに、国立遺伝学研究所のリソースを使って解析できるようになるのです。

また、自分が出したデータと、他のデータベースの他の人が出したデータが一緒になることで、自分のデータの付加価値が上がるということもあります。

別の観点では、オープンにすることで研究の再現性が上がるということがあります。図 15 は実は STAP 細胞の次世代シーケンサーのデータです。きちんと登録されているので、自分でデータを取ってきて、自分のコンピューターで追試できるのです。このデータについての論文がスピード感を持って取り下げられたのは、このようにデータが公開されていて研究の再現性が保たれていたからという理由もあります。

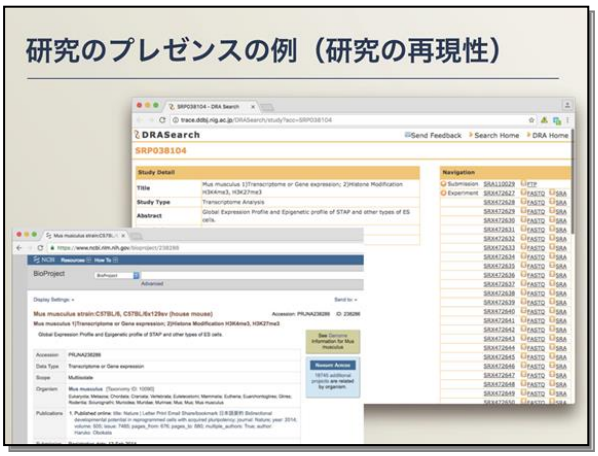
データのオープン化に求められること

データをオープン化するときは、公共データベースに登録する、自分でデータベースを作成するという手段もありますが、例えば「Scientific Data」「GigaScience」などのデータジャーナルにサブミットする、または機関リポジトリを利用するという方法もあると思っています。

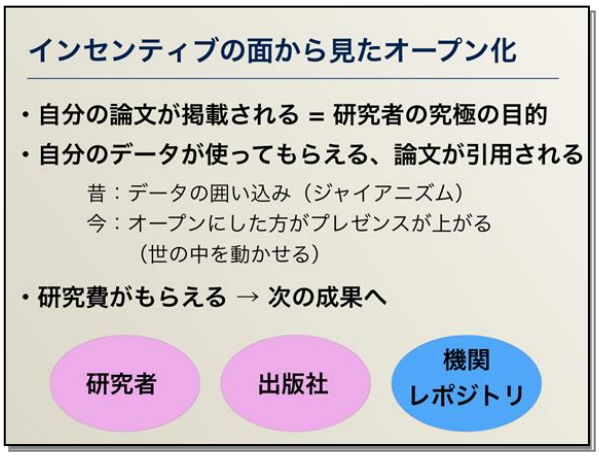
そのときに言いたいことが幾つかあります。まず、データを参照する仕組みです。遺伝子であれば、アクセッション番号や登録 ID などがありますが、そういうものできちんとポイントできるようにしてほしいのです。論文に番号を書いておけば、一意にこのデータだと分かるようにしてほしいと思っています。

また、実際にデータベースを維持しようとすると費用が掛かります。ミレニアム・プロジェクトなども、多額のお金が掛けられたのですが、お金が切れるとともにだんだん廃れていくので、私たちがレスキューしています。このような費用の問題は確かにあります。

そして、永続性です。例えば、researchmap の科学研究費助成事業データベースのリンクをクリックすると Not Found になることがあります。なぜかというと、科学研究費助成事業データベースの URL が変わったからです。名誉のために言っておくと、最近のものはきちんとリンク先のページが出ます。ただ、昔のものは Not Found になってしまっています。URL も一意のものとして機能するのですが、実際にはメンテナンスしている方で URL が変わってしまうとたどれなくなってしまうことがあるので、そのようなことも含めて永続性を強調したいと思います。私がインセンティブではないかと考えたことを図 16 にまとめました。



（図 15）

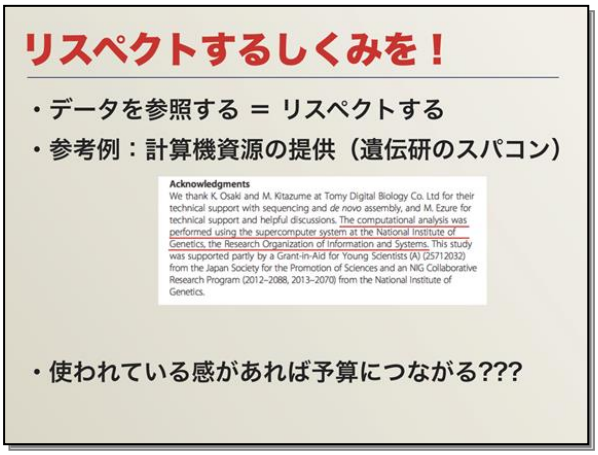


（図 16）

しかし、いろいろな立場があります。研究者の立場からすると、これは確かにプラスです。出版社も、例えば Scientific Data を維持している Nature などは、掲載したデータ論文を引用してもらえるのでプラスです。しかし、機関リポジトリにとって、これがインセンティブになっているかというと疑問もあるので、この後で議論できたらと思っています。

強調したいのは、データが水や空気のようにオープンになってしまっているということです。お金を掛けてデータを出しているので、それをリスペクトする仕組みがあると、インセンティブの一つになるのではないでしょうか（図 17）。

例えば、国立遺伝学研究所のスパコンを使った成果に関しては、「Acknowledgements」に「使いましたよ」ということが書かれています。このように、使われている感があれば予算につながるかもしれません。

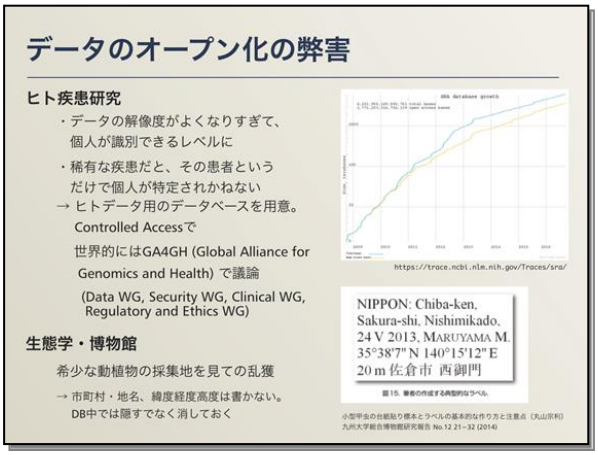


（図 17）

データのオープン化の弊害としては、最近、ヒトのデータがよく出てくるのですが、解像度が良過ぎて個人が特定できるのではないかという問題点があります。（図 18）

また、生態学や博物館では、データにラベルが付きます。詳細な地名、緯度・経度まで書いてあるので、貴重なランなどの植物は、それを見た花屋さんが採りに行くのです。それで絶滅がさらに加速されてしまうので、これらの情報は今全て消しているという状況になっています。

そして、オープン化はしていきたいのですが、実際には、図 19 のようにスライドガラスが山と詰まれて、倉庫に鍵が掛けられているので、カタツケていけたらと思っています。



（図 18）

（図 19）

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

日本古写真画像データのオープン化と大学図書館の役割

下田　研一
（長崎大学附属図書館）

講演要旨

長崎大学附属図書館は、1988 年以来、主に欧米に流出した幕末・明治期の日本古写真を収集し、1998 年からは、日本古写真画像データベースのインターネット公開に取り組んできた。今では、日本最大規模のコレクションとデータベースを有する。本発表では、日本古写真の資料特性、データベースの構築手法と公開後の反響、コレクションに対する研究者・学芸員・図書館員のアプローチの違い、博物館等の類縁機関との連携を紹介し、研究データのオープン化における大学図書館の役割を考える一助としたい。

下田　研一

1987年3月に長崎大学附属図書館に採用。受入・目録・参考調査等の図書館業務を経験し、2014年7月より学術情報サービス班長。この間に、学術情報センターと接続した図書館システムの導入、電子ジャーナルの導入、機関リポジトリの立ち上げのほか、日本古写真コレクションの構築・整理・電子化・各種公開事業の実施にも携わり現在に至る。

長崎大学では、幕末から明治にかけて日本国内で撮影された同時代の写真を「日本古写真」と呼んでいます。長崎大学の「幕末・明治期日本古写真コレクション」は、主に 1860 年代から 1890 年代にかけて、日本国内で撮影された約 7,700 点の写真から成ります。

当時はガラス板を使った感光原板でした。薬剤を塗って、それが乾かないうちに撮影しなければいけないものを「湿板」、その後、乾かして保存できる感光板が出てきたので、それを「乾板」と呼んでいます。さらに、感光したネガを卵の白身を使った紙に焼き付けました。このようにしてできた日本の鶏卵紙写真が現在世界各地に残っています。

長崎大学では、1988 年、当時の文部省の大型コレクションの予算で収集を始め、今年で 28 年になります。この間に欧米から入手した写真がコレクションのほとんどを占めています。例外的には、長崎出身で 1862 年に日本最初期の職業写真家となった上野彦馬のアルバムがありますが、このようなものはごく一部です。

1.写真術の渡来・伝播と日本古写真の資料特性

写真術は開港の地であった長崎・横浜・函館から入ってきました。長崎は主にヨーロッパから、横浜はアメリカから、函館はロシアから写真術を受け入れました。ここでは、長崎からの写真の国内導入と日本古写真の資料特性について考えてみようと思います。その

後、横浜からの写真の海外流出と日本古写真の資料特性についてお話しします。

1-1.長崎からの写真の国内導入と日本古写真の資料特性

まず、長崎からの写真の国内導入と日本古写真の資料特性について。出島は、鎖国時代、西洋に開かれた唯一の窓でした。写真術についても例外ではありません。1843 年に、最初の銀板写真機がオランダ船から出島を通じて長崎に持ち込まれました。このとき持ち込まれた写真機は、買い手が付かず、オランダ側に返されることになったのですが、これが写真機の持ち込まれた最初だと言われています。

その後、長崎では、1855～59 年の間、海軍伝習があり、西洋の科学技術の組織的な導入が行われました。ここで写真術がその一つとして教えられています。それに続く医学伝習でも、オランダ人医学教師が、日本人医学生に請われて、写真術を研究し教えました。これらは欧米列強の極東進出および日本の開国と同時期だったので、このようなところから日本古写真の資料特性が生まれてくることになります。

日本が開国すると、欧米から「イメージハンター」と呼ばれる人たちがやってきました。「東洋の神秘の国」日本の姿を欧米に伝えるために写真を撮りに来たのです。これは、その後の日本の近代化の諸相を写真が活写することにもなりました。日本古写真は近世日本への憧憬と日本近代化の記録として成立したのです。

長崎からの写真の導入について、もう少し詳しくお話しします。オランダ人医学教師ポンペが、1857 年に長崎奉行所西役所内で、幕府から長崎に派遣された松本良順とその弟子たちに、最初の医学講義を行います。これが医学伝習です。ポンペは医学だけを教えたのではなく、その基礎となる物理学や化学も一人で教えました。化学の応用技術として写真術も教えています。ポンペは写真を自ら撮ったことがなかったので、実例を学生に示すことについては大変苦労しました。

それを可能にしたのがピエール・ロシエです。彼は、ロンドンのネグレッティ＆ザンプラ商会から、ステレオ写真集制作のため、極東に派遣されたスイスの写真師です。ステレオ写真というのは、視差のある二つの画像を両目で見ることによって、立体に見える写真のことです。彼は 1859 年に長崎に少しだけ上陸し、すぐに江戸・横浜に向かいます。翌年再来し、ポンペの学生たちに写真術の実際を伝授しました。日本の写真は湿板写真により実用化されたのです。

ロシエが 1860 年に撮ったステレオ写真「Battle between Japanese soldiers」は、長崎の写真で最も古いものの 1 枚です。中央奥に松本良順が写っています。医学伝習所が長崎奉行所から近くの 2 階建ての長屋に移った頃で、その長屋の前で撮られたと考えられます。このとき、松本良順とその弟子たちはロシエが使っている写真機の方に興味津々だったと思われます。

「Costume of Japanese on a rainy day」という写真は、奥の山の形や海沿いの様子などを見ると、実は出島の荷揚げ場であることが分かります。

少し後の、長崎港の南側からの鳥瞰を写した写真は、江戸の海岸線をよく残しています。出島があります。

それより後の 1866 年に撮影された、反対側からの写真もあります。それを拡大してみると、ロシエの出島の荷揚げ場の写真に入っていた斜面が見えてきます。このような関連付けができることがあります。

ここで日本古写真の資料特性を考えてみようと思います。まずは、化学反応としての写真感光です。写真感光は化学反応であり、分子・原子レベルでの変化です。こうしてできたネガはガラス板に定着しますが、これを拡大することなくそのままの大きさで紙焼きするので、高精細な画像が得られます。当時の原板の標準的な大きさは A4 判ほどもありました。

次に、撮影主題と写真の意義のズレです。化学反応なので、写真師の意図しないものの写り込みがあります。また、外国人写真師と日本人の間で撮られた写真なので、写真師が撮ったときの意図と、今になって持つ写真の意義にズレがあります。最初からズレが潜んでいるのですが、それが 100 年、150 年という時間の

経過の中で、古写真の多層性と多義性を生むのです。

そして、豊富な内蔵情報と貧弱な出自情報です。これは古写真の一番の問題です。画像は化学反応によるものなので、1 枚の写真の中に非常に豊富な内蔵情報があります。それをさらに拡大して情報を発掘することができます。一方、個々の写真の同定には、多数を参照・比較することが必要になります。個々の写真は、「誰が、いつ、どこで、何を」撮ったものだという情報の欠落している場合が一般的です。

1-2.横浜からの写真の海外流出と日本古写真の資料特性

明治になると、横浜を中心に日本写真の商品化が進展します。いわゆる「横浜写真」と呼ばれるもので、工房で手彩色の写真が大量生産されます。さらに、日本の伝統技術である蒔絵や螺鈿（らでん）で装飾されたアルバムが大量に生産されます。日本を訪れた外国人旅行者にそれらを販売したので、個々の写真には英語でキャプションが付けられました。このようなものが日本土産アルバムとして欧米に流出しました。その中身としては、東京・横浜その他の日本の主要都市や名所・旧跡などの風景写真、日本の風俗を伝える人物写真、このようなものが含まれています。

日本古写真の資料特性が別の面から出てきます。それは、同一撮影地や同一主題の写真が多数存在するということです。あるポイントからの定点観測の記録と言えるものもあります。

もう一つは、図書館員として興味深い特性で、同一原板で、彩色の有無や状態、キャプションの内容や形式の異なる写真が多数存在することです。これは、著作権が未確立だったことと、原板が写真館の間で売買されたり、交換されたり、複製されたりしたことによります。同一原板の写真が複数の写真館のアルバムに存在したり、同一アルバム中に撮影者や撮影時期の異なる写真が混在したりします。個々の日本古写真の同定には、来歴をきちんと解きほぐす必要があるということです。

2.日本古写真画像データベースの構築

長崎大学では、これまで 3 種の日本古写真画像データベースを構築しています。

2-1.幕末・明治期日本古写真データベース

最初の「幕末・明治期日本古写真データベース」は 1998 年公開で 5,416 点を収録していました。当時は、インターネットが爆発的に普及して間もなかったので、画像データベースをインターネット上に公開することは研究者にとっても魅力的でした。このデータベースの開発は工学部の研究室が主導しました。日本語版のほか、インターネットということで海外からのアクセスを意識して、英語版も備えていました。さらに、多様な検索方式が導入され、クリックだけでも操作可能でした。研究者向けではなく、むしろ中学生でも操作可能なデータベースというコンセプトで開発が進められました。一時期は海外からのアクセスが国内からのアクセスを上回る現象がありました。研究者が魅力を感じたものだったため、科学研究費補助金（研究成果公開促進費）を使って構築しました。

研究者と図書館職員の分担は、研究者のチームが古写真の解説文やキーワードを作成し、図書館職員がそれ以前の古写真の整理・目録作成、その他の事務処理を担当しました。

図 1 が初代データベースのトップページです。最初に「文部省科学研究費補助金により作成されました」と明記されています。

（図 1）

図 2 が検索結果の詳細表示画面です。下の方に研究者が作成した解説文があります。上の目録データは主に図書館職員が作りました。左側の四つの虫眼鏡が 4 種類の検索方式を示しています。

図 3 は、2004 年 3 月、公開から 6 年位経って、そろそろ更新を考えなくてはいけなくなった頃の国別アクセス状況です。アメリカ合衆国、日本、その他の国々が 3 分の 1 ずつ位を占めています。

2-2.幕末・明治期日本古写真超高精細画像データベース

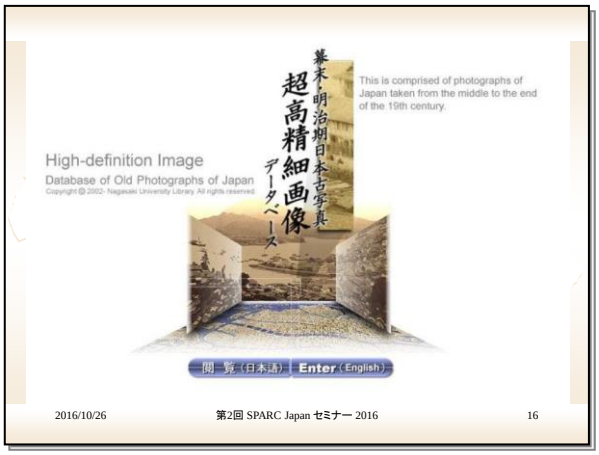
次に構築したのが「幕末・明治期日本古写真超高精細画像データベース」です。2003 年公開で 501 点を収録していました。画像をコンピューター画面上で 5～10 倍に拡大しても鮮明に見ることができます。当時の古写真コレクションは全 5,416 点でしたが、それから 501 点（長崎が 201 点、その他が 300 点）の特に画質の優れた画像を選んでつくったデータベースです。

このときも、画像の拡大には先進的な技術を要したので、研究者の研究課題として、同じように科学研究費補助金（研究成果公開促進費）を使って構築しています。また、解説文を詳しくしましたので、その写真が撮られた各地の学芸員を中心に、詳細な解説文の執筆を依頼しました。ずっと Windows のみに対応していて Mac に対応していなかったので、2011 年に画像拡大方式を更新し、端末を選ばないデータベースの閲覧を可能にしました。

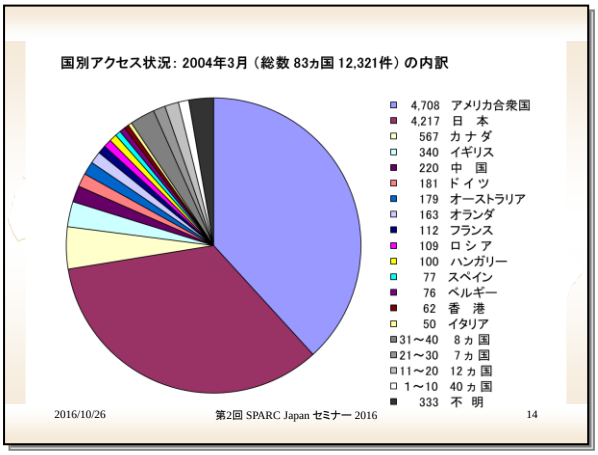
図 4 が最初の高精細画像データベースです。図 5 は 2011 年以降のトップページです。

現在、同じ拡大方式を使って、主だったアルバムの高精細画像データベースづくりを進めています。

（図 2）



（図 4）



（図 3）

（図 5）

図 6 はボードインコレクションというもので、ポンペの後任のオランダ人医学教師ボードインのコレクションです。日本に古写真ブームを起こしました。

2-3.幕末・明治期日本古写真メタデータ・データベース

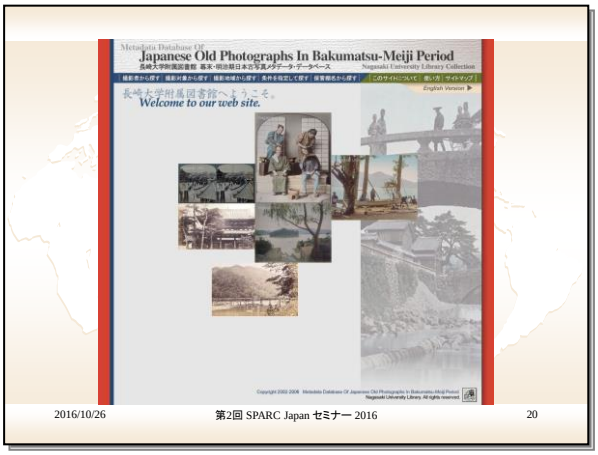
3 番目は「幕末・明治期日本古写真メタデータ・データベース」です。これは初代データベースの後継です。目立つ点は、同一標題を古写真の画像に与えることにより、同一主題の写真の通覧と比較を可能にしたことです。また、高精細画像データベースに収録されている写真については、該当のレコードへのリンクを貼っています。さらに、メタデータを Dublin Core に準拠させて、機関リポジトリに登録し、リポジトリ側からリンクを貼っています。同じように科学研究費補助金（研究成果公開促進費）で構築しました。

図 7 がトップページです。図 8 のような詳細表示画面で「同一タイトルの写真」というボタンを押すと同じタイトルの写真が出ます。「高精細画像を見る」ボタンから高精細画像データベースへ飛ぶと、詳細な解説と拡大可能な画像が出るようになっています（図 9）。また、リポジトリからは、「本文ファイル」というボタンを押すと、データベースの該当画像に飛べるようにしています（図 10）。

図 11 の円グラフは平成 27 年度の古写真等の申請者種別利用申請件数です。1 位が放送関係、2 位に出版印刷関係が来ています。これが最近の傾向です。下はデータベースのアクセス数です。2015 年アクセス数は 2,757,086、訪問者数は、アクセス元の端末数だと思いますが、117,767 です。私はオープンサイエンスと古写真を組み合わせて考えたことはありませんでしたが、利用の形態からいってもそのような印象でした。

（図 6）

（図 8）



（図 7）

（図 9）

3.日本古写真のオープン化と大学図書館の役割

最後に、古写真データベースの構築と大学図書館の役割についてお話して、まとめようと思います。

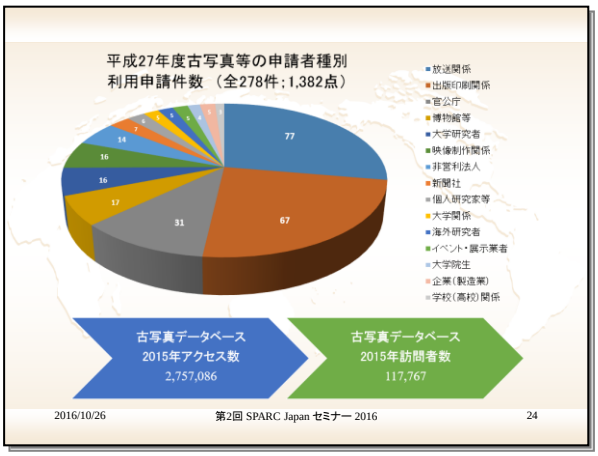
3-1.研究者・学芸員・図書館員のアプローチの違い

データベースの構築から見えてきた点は、研究者・学芸員・図書館員の古写真に対するアプローチに違いがあるということです。

研究者の場合、もちろん専門分野からアプローチします。そして、先駆的・先進的プロジェクトへの参画を歓迎します。さらに、学術論文、新聞・雑誌記事、図書の執筆につながるような仕事であれば、なおさら歓迎します。

学芸員は、図書館員から見ると、現物主義でモノに非常なこだわりがあります。データベースを構築する場合は、古写真から画像だけを切り取ってデータベースに収録しますが、学芸員は写真の表も裏も厚みさえも大事だと捉えます。そして、画像の電子化より、資料そのものの修復を優先させる場合があります。解説については、ギャラリートークなどで一般市民や子どもたちを相手にしているので、非常に分かりやすいと思います。

（図 10）



（図 11）

これに対して図書館員は、もちろん情報に定位したアプローチをしますが、研究者のように深く新しい情報を発見するというような目標を立てるのではなく、古写真にアクセスするための基本的な情報をそろえたいと考えます。図書館員の専門性としては、目録やメタデータの作成、資料がどのように利用されているかについての調査・研究があると思います。

学芸員は館の外に出かけて野外調査をすることが許されていますが、図書館員にはそれが許されていません。結局、図書館員のフィールドは、メタデータ・索引・リンクの作成になります。その目的は、資料と利用者（研究者・一般市民）を結び付けることです。

こうして、大学図書館の役割には、データベース構築の主体として構築後もそれを維持・管理すること、また、研究者に業績発表の場を用意してあげることがあると思います。古写真の場合は、展示会・講演会・シンポジウム・図書の出版を企画・実施しました。

さらに、プロジェクトを永続的なものにするためには、図書館長のリーダーシップも問題になると思います。法人化された後、大学は中期目標や中期計画に基づいて教育研究・社会貢献を進めることが求められるようになりました。長崎大学の場合、そのような目標や計画の中に古写真に関係する項目が入っています。古写真については「日本古写真の世界拠点を形成する」という文言が入っていますし、図書館については「地域と世界に開かれた知の拠点とした情報発信を行う」ことが掲げられています。

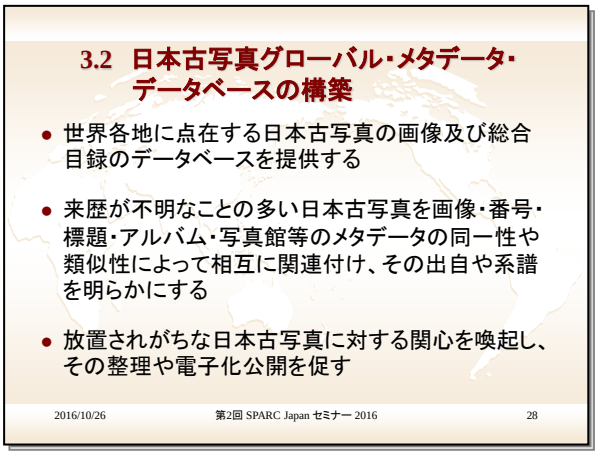
3-2.日本古写真グローバル・メタデータ・データベースの構築

これらを受けて、今取り組んでいるのは、グローバル・メタデータ・データベースの構築です。図 12 の三つが目標です。図 13 のような作業図式で進めようと思っています。パートナーは、フランス国立ギメ東洋美術館を考えています。ここは海外の美術館ですが、現在、世界最大の日本古写真コレクションがあります。ここと組んで、古写真の世界拠点を形成しようとしています。

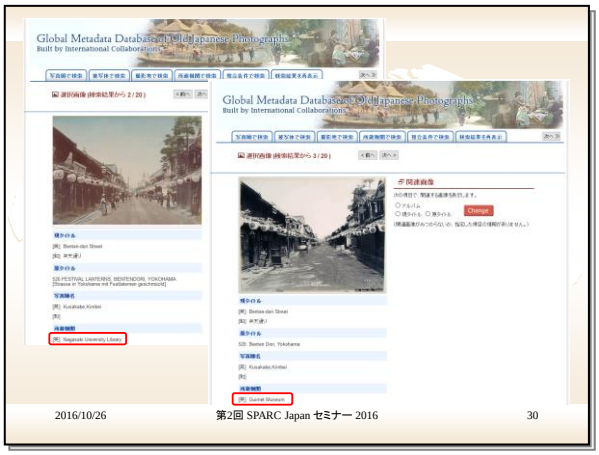
実際に取り組みはじめており、例えば、図 14 は横浜弁天通りの写真です。左が長崎大学、右がギメ東洋美術館に所蔵されています。少し時期が違いますし、彩色の状態も違います。

図 15 は日下部金兵衛という写真師の写真のリストです。番号の後にタイトルが続いているのですが、641、648、649、665、702、703 と、番号が飛び飛びになっています。このように欠番になっているところを、いろいろな機関に呼びかけて埋めていきたい、埋められたら楽しいな、と思っています。これを見るとよく分かるのですが、番号はただ振ってあるだけではなく、きちんと写真が分類されていて、上の方は全部東京です。このリストを完全なものにすることが今後の野望です。

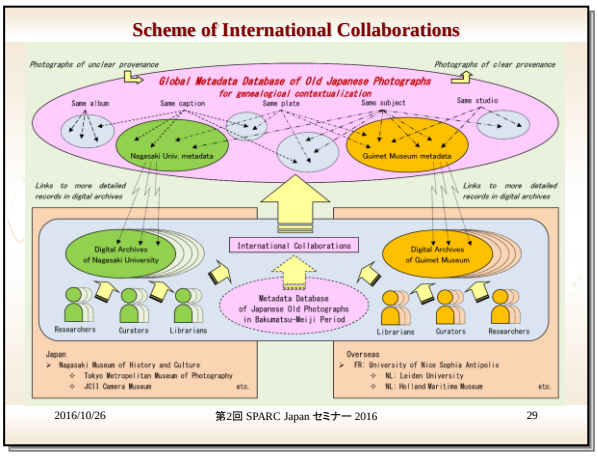
今日お話したデータベースについては、検索エンジンで「古写真」を検索すると、「幕末・明治期日本古写真メタデータ・データベース」が一番上に出てくると思います。そこに参考文献も置いています。他のデータベースを見たいという方がいらっしゃれば、「長崎大学電子化コレクション」を見ていただければと思います。



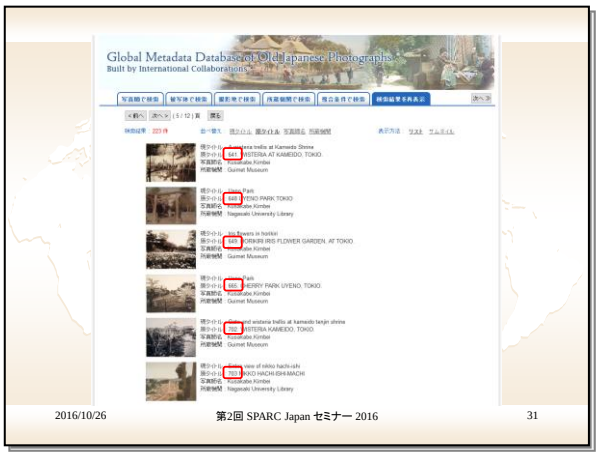
（図 12）



（図 14）



（図 13）



（図 15）

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

超高層大気観測データの メタデータ作成実験経過報告

南山　泰之
（国立極地研究所）

講演要旨

本報告では、研究者と図書館員における協働の一事例として、超高層大気分野における観測データのメタデータ作成実験の経過報告を行う。本実験は、大学図書館員の持つメタデータ運用スキルが、特定分野にフォーカスしたメタデータの作成に応用できるか、の検証を通して、①メタデータ作成支援による研究者の負担軽減、②図書館員による流通面での知見提供、といった具体的なインセンティブ付与、及び③図書館員による研究データマネジメントの今後の可能性を探るものである。幅広い関係者からのコメントをいただくことで、今後の両者の協働のあり方を考えるための一助としたい。

南山　泰之

2005年より国立極地研究所情報図書室に勤務。2007年～2008年、第49次日本南極地域観測隊に参加。その後、東京大学駒場図書館（2011年～2014年）を経て現職に戻る。2014年8月より機関リポジトリ推進委員会協力員。2014年第4回SPARC Japanセミナー企画ワーキンググループ（WG）メンバー、2016年SPARC Japanセミナー企画WGメンバー。

私は 2005 年に国立極地研究所に配属になり、2007 年から 2008 年にかけ、南極に観測隊として参加していました。帰ってきてしばらく極地研究所にいたのですが、その後、東大の駒場図書館に少しお世話になり、2014 年から極地研究所に戻ってきて、また仕事をしています。今は機関リポジトリ推進委員会をメインに活動しています。他に、図書館系の雑誌の編集や SPARC Japan の企画ワーキングを仰せつかっています。

機関リポジトリ推進委員会は、オープンアクセスリポジトリ推進協会と 2016 年 7 月から名前が変わりつつあるようですが、大学図書館、国立大学、公立大学、私立大学が連携した連携・協力推進会議が親組織です。大学図書館コンソーシアム連合や「これからの学術情報システム構築検討委員会」も連携・協力推進会議を親組織としています。私はオープンアクセスリポジトリ推進協会の中で、メタデータや、研究データに図書館がどう関わるか、を検討するタスクフォースに参加しています。

研究データ管理に関わる背景

国際的動向を踏まえたオープンサイエンスに関する検討会から、報告書「我が国におけるオープンサイエンス推進のあり方について」が公表されてもう 1 年以上たちます（図 1）。ここから「オープンサイエンス」という単語がメジャーになり、図書館は何をしようかという話が始まったと言ってもいいと思います。

今回の主題であるデータの話は、「公的研究資金による研究成果」の部分で少し触れられています。図書館の役割については、「研究成果等の収集、オープンアクセスの推進、共有されるデータの保存・管理を行う基本機能」を、持つべきと言われたのか、持つとありがたいと言われたのかは定かではありませんが、そのように書かれていました。

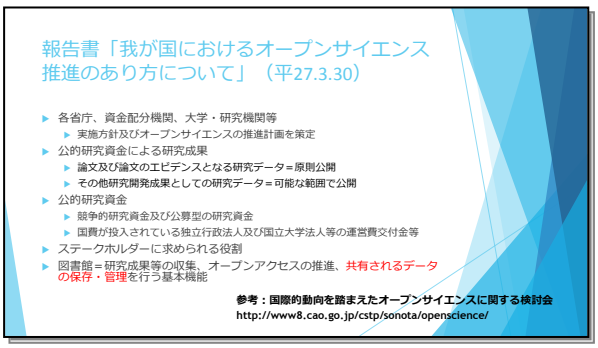
これを受けて、学術情報委員会で審議のまとめが出されました。そこから大学図書館の役割を幾つか抜粋して図 2 に書いてあります。論文のオープン化、研究データのオープン化、研究成果の散逸等の防止、人材育成です。今回、私の発表内容に関係するところしか赤字にしていませんが、研究データのオープン化に関しては、論文のエビデンスとしての研究データの公開に機関リポジトリを活用するようにというようなことが書かれています。

人材育成については、技術職員、URA および大学図書館職員を中心にデータ管理体制を構築すること、あるいは、機関リポジトリの構築を進めてきた経験等から研究成果の利活用促進を担うことが役割として挙げられています。

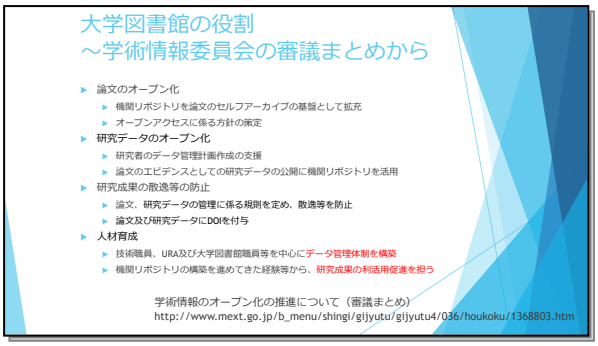
大学図書館はこれらの潮流を受けて、どのような動きをしているのかというと、2016 年 6 月に国立大学図書館協会の総会があり、そこで「国立大学図書館協会ビジョン 2020」が策定されました。その中で重点領域とされたのが、重点領域 1「知の共有：＜蔵書＞を超えた知識や情報の共有」、重点領域 2「知の創出：新たな知を紡ぐ＜場＞の提供」、重点領域 3「新しい人材：知の共有・創出のための＜人材＞の構築」の三つです。今回のデータの話に絡むのは、特に重点領域 1 と重点領域 3 です。

重点領域 1 では、学習教材やデータといった教育研究成果を対象として、知の共有のための方策を検討し、実現することが求められています（図 3）。

重点領域 3 には、これまで培ってきた学術資料に関する専門的知識やメタデータ運用スキルに加え、新たな知識やスキルを習得することによって、国立大学図書館に期待される新たな機能を実現するということが書かれています。



（図 1）

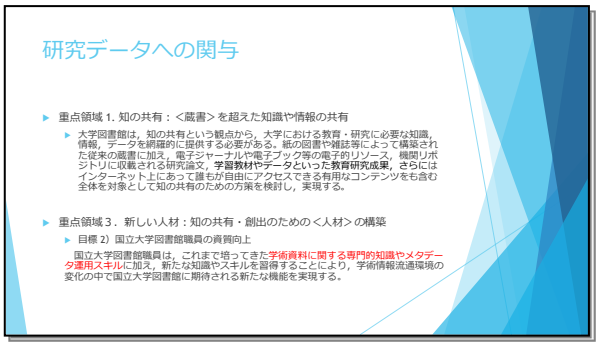


（図 2）

図書館ができること

ここまでの内容から、「図書館ができること」として、図書館の外に対して何をアピールできるか、私なりにまとめました（図 4）。

一つが、ネットワークの活用です。図書館は横のネットワークが非常に強いのです。委員会をつくって、タスクごとにその都度、最新のノウハウを共有し、検討してきたという経緯があります。図書館員はジョブローテーションがあり、3～5 年ごとに異動して違う



（図 3）

仕事を始めますが、そのような人事異動にも、委員会活動での知の共有があるので対応できるようになっています。

もう一つはメタデータ運用です。図書館員はメタデータが得意な人が多いです。日本はリポジトリ大国とよく紹介されますが、9 月末現在のデータを見ると、（構築中のものも含めれば）インスタンスを全部で約 744 持っていて、公に出ているデータとしては世界一リポジトリを持っている、すなわち、それを運用する図書館員は、一般的にメタデータに強いと言えると思います。

（日本のリポジトリ業務における）メタデータ運用で使われるスキーマは、junii2、Learning Object Metadata、LIDO などいろいろありますが、標準的な図書館のリポジトリで使われているのは junii2 です。LIDO は博物館関係のスキーマです。余談ですが、最近のアップデートでは今回お話しする SPASE というスキーマを、なぜか JAIRO Cloud に搭載してくださったので、非常にお話ししやすくて助かります。何が言いたいかと言えば、図書館にはこのように複数のスキーマを横断的に扱う人材がたくさんいます。（スキーマを横断的に扱うことが）業務の一環となっているので、メタデータ運用スキルは高いと言えるように思います。昔ながらの目録もスキーマの一つと言ってもいいですね。

実験の概要

ここまでは前振りで、次にわれわれが行った、図書館員による IUGONET メタデータ作成（超高層大気観測データのメタデータ作成）の実験についてお話しします。概要と問題意識、詳細、ここまでの検証についてお話しします。

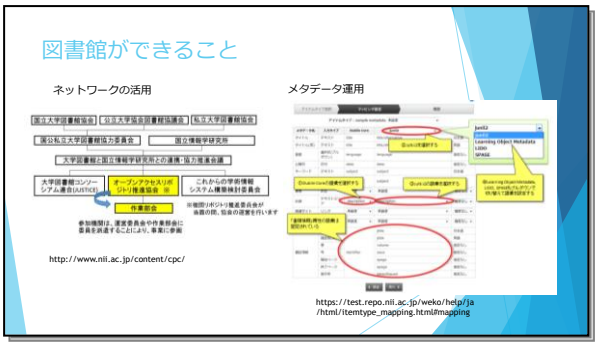
一つ目は概要です。今回の実験を通じて探りたかったことの一つは、研究者へのインセンティブの可能性です。研究者、特にメタデータ作成業務従事者の負担軽減、図書館が持つ流通面での知見の提供ができるかをこの取り組みを通じて考えたかったのです。

もう一つは、図書館による新規サービス展開の可能性です。図書館が研究支援の機能を持ってほしい、海外の機関と比べてこのようなことをやってほしいという声はいろいろなところで聞くのですが、実際にそれにどのように応えていくのかを探りたい、研究者との協同による直接的な研究支援を考えたいと思っていました。従って、オープン化というより、メタデータを通じて、研究データのありかが分かるようにして、オープン化に寄与しようという取り組みになります。

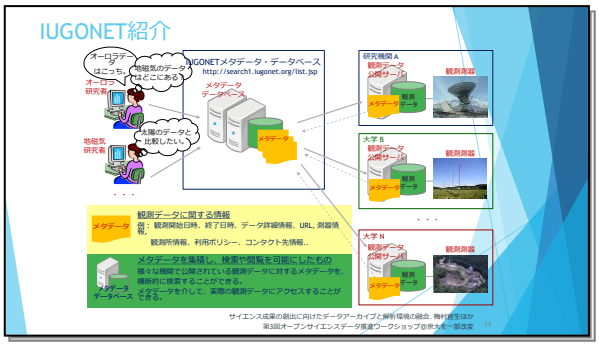
IUGONET 紹介

IUGONET は正式名称が Inter-university Upper atmosphere Global Observation NETwork です（図 5）。私が所属している国立極地研究所、それと東北大学、名古屋大学、京都大学、九州大学の 5 機関が連携して、観測データからメタデータを抽出し、それをネットワーク上で広く共有するシステムを構築するプロジェクトです。

IUGONET は 6 年以上進んでおり、やはりやっていくうちに課題が挙がってきます。2015 年の SPARC Japan セミナーで、極地研究所の田中良昌先生が現状を発表されました。図 6 はそのときのスライドです。



（図 4）



（図 5）

問題の一つは、データベースが個人レベルでのメンテナンスにどうしても依存してしまうので、研究者のモチベーションが尽きたら終わってしまう面があることです。

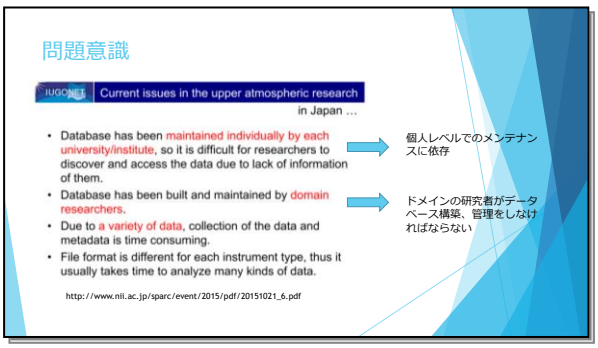
もう一つは、ドメインの研究者がデータベース構築、管理をしなければいけないことです。研究者は専門分野であるドメインの研究をしたいのですが、データベースの構築はドメインとはやや離れます。研究者は素人ではないので、ご自身で何とかしてしまうのですが、負担が大きいのは確かです。

そこで、個人レベルのメンテナンスに依存するのであれば、図書館のネットワークを利用できるのではないか。ドメインの研究者が構築・管理をしなければいけないのであれば、（サーバーの運用管理の話は少し置いておいて）少なくともメタデータ運用に関しては図書館側にこれまでの蓄積があるので、スキルをうまく共有できるのではないかと考えました。これが今回の実験につながっています。

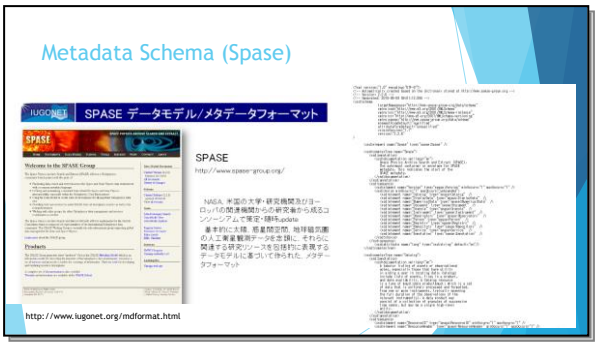
IUGONET で使っているメタデータスキームは SPASE というものです（図 7）。なぜか JAIRO Cloud に搭載されています。SPASE は NASA やアメリカの研究機関から成るコンソーシアムで作成しているメタデータスキーマで、太陽・惑星間空間・地球地磁気圏の人工衛星観測データを念頭に置いたメタデータフォーマットです。ご存じない方がほとんどだと思うのですが、私も半年前まではあまり分かっていなかったので、ここではこのようなものだと見ていただければと思います。

2016 年 6 月からパイロットを開始しました（図 8）。図書職員とのコラボレーションとして行ってみる内容を決め、その難易度も設定して行っています。現在は 3 番まで進んでいますが、重要な点として、私個人ができてもあまり意味がなく、水平展開をにらんで、あくまでも図書館のベースの知識の中でどこまでできるのかを探るのが今回の取り組みのメインです。図書館コミュニティの中で共有されていると思われる知見で、できる範囲を見極めようとしています。

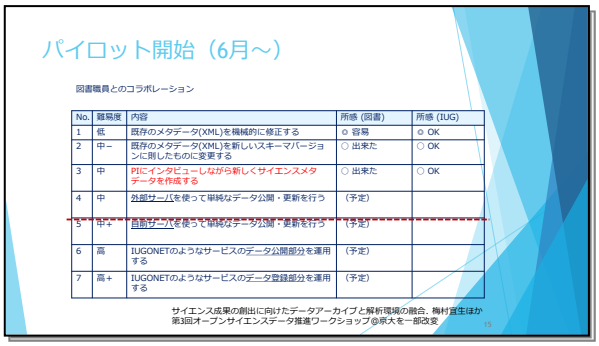
赤い線は業務境界線で、今は 4 番と 5 番の間に引いています。5 番以下は、国内の図書館の中で行っている例、自前サーバーを使って公開したりする例もなくはないのですが、図書館員の業務負荷が非常に大きくなります。また、機関リポジトリとの競合の問題もありますし、自前サーバーではなくクラウドにしたらどうかという話がすぐに出てくるので、共通化するのはこのあたりまでが一番いいのではないか、と IUGONET の担当者と話しています。



（図 6）



（図 7）

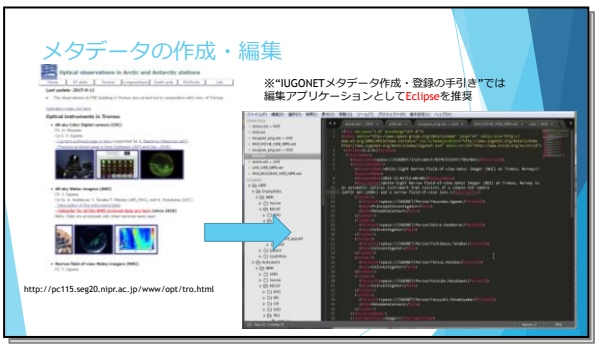


（図 8）

実際の作業はどのようなことをやっているかということで、メタデータの作成・編集です（図 9）。今、メインに行っている作業ですが、（前提として）ウェブ上に既に研究者が情報を載せています。「このようなことをやっています」というようなページをつくって、そこにデータを置いているのですが、それをスキーマに落とし込む作業をしています。IUGONET のメタデータ作成・登録の手引きは、Eclipse というソフトの使用を推奨していますが、私は自分で使いやすいテキストエディタで XML を直接編集してメタデータをつくっています。情報源をメタデータに起こすという作業は、通常の図書館の作業（目録や、リポジトリのメタデータ登録）に近いので、そのイメージで見ていただければと思います。マニュアルがあるほか、他のデータベースにも（同じデータに対する）簡単なメタデータが登録されているので、そのようなものも参照しながらやっています。

Validator という、XML Schema の構造にきちんと合っているかをチェックするソフトがあるので、そちらに掛けて問題がなければ、「このメタデータで本当によろしいですか」と担当の PI にメールで確認します（図 10）。当然 Validator を通ったといっても、あまり薄々のデータでは使い勝手が良くないというか、最終的にきちんと確認してもらわなければいけないこともあるので、それを確認していただきます。それで問題なければ、データベースへのメタデータの登録に進みます。

実際の登録は幾つかしています。Keogram data で検索すると、検索結果として私が元を作成したメタデータが出てきます。対象のメタデータの中には、私の名前もメタデータ・コンタクト・パーソンとして入れていただいています。

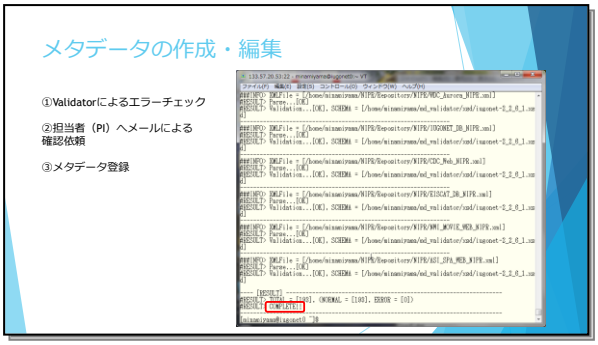


（図 9）

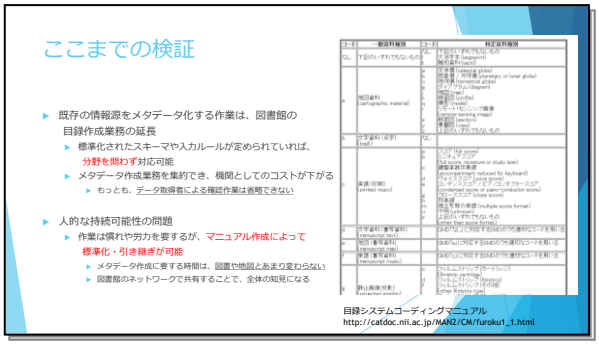
ここまでの検証

メタデータをつくる、機械的に修正するなどの作業を幾つか 10 月までにやってきました。ここまでの検証として分かったことの一つは、既存の情報源をメタデータ化する作業は、図書館の目録作成業務の延長だということです（図 11）。あえて「リポジトリ」と書かずに「目録」と書いたのは、リポジトリをやっていなくても、昔からそのような知識を持っている方、昔からそのようにやっている方であれば大体対応できるだろうと思っているからです。

標準化されたスキーマや入力ルールが、きちんと研究者との協同の段階で定められていれば、分野を問わず対応可能と言っていいと思います。分野を問わずと書いたのは、私はもともと法律系の人間で、そもそも理系ですらないですが、（IUGONET のメタデータにも恐らく対応できているため）大体何とかなるという



（図 10）



（図 11）

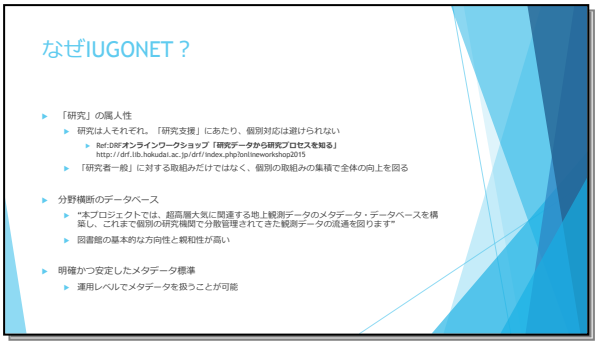
話をしたかったのです。右側に目録システムで取り扱ってきたコンテンツの表をあえて持ってきましたのも同じ意味ですが、地図・楽譜・静止画像など書籍ではないものであっても、最低限のレベルのメタデータは何とか作成できるということが言いたかったからです。

（図書館が研究データのメタデータ作成に関与できる、という前提からは）研究者が実際にやっている、あるいは研究者が非常勤の方を雇って、その方にやってもらっているようなメタデータ作成業務を、図書館の基幹業務として集約することができることになるので、機関としてはコストが下がるというメリットが挙げられると思います。もっとも、データ取得者による確認作業は省略できません。

もう一つは、人的な持続可能性の問題についてです。作業を実際にやってみたところ、慣れや労力は要しますが、マニュアル作成によって標準化・引き継ぎは可能です。作業に時間がかかり過ぎたらあまり意味がないと言われるのですが、普通の目録、リポジトリのメタデータ登録の作業とほぼ変わりませんでした。慣れれば 30 分ぐらいで一つできます。この知見をこれから図書館ネットワークで共有しようと思っているので、これは図書館全体の知見として共有されます。すぐにとは確約できませんが、誰でもできる体制になってくると思われます。

研究支援についての展望

IUGONET の研究集会が 1 週間前にあり、そこで質問として、なぜ IUGONET に協力したのか、個別のプロジェクトに対して図書館員がサービスする意義、



（図 12）

スタンスは何かと聞かれたので、研究支援について私が考えることを図 12 に書き出してみました。この点、皆さまからもぜひご意見を頂きたいと思っています。

当然、プロジェクトに関わるに当たっては、その中の人とやりたいかどうか、楽しく仕事ができるかどうかが初めに来るのですが、それは置いておきます。

一つ目は研究の属人性です。昨年度、DRF でオンラインワークショップ「研究データから研究プロセスを知る」が開催され、私もファシリテーターとして関わりました。自分の担当でもそうでしたが、報告書を見ると、やはり研究者の研究のやり方はばらばらでした。ばらばらなのが研究ですから、それに文句を付けるつもりで書いたのではなく、研究支援業務を一般化するのはかなり難しいのではないかと感じました。

一般化できないのですから個別対応は避けられないというスタンスに立つと、（研究者一般に対する取り組みのやり方を今まで図書館は考えてきた面があるのですが）どこの分野に取り組むとか、この先生とタイアップするというような個別の取り組みを集積させることで、全体の向上を図るような取り組みに変わっていけばいいのではないか、と思って取り組みを進めています。

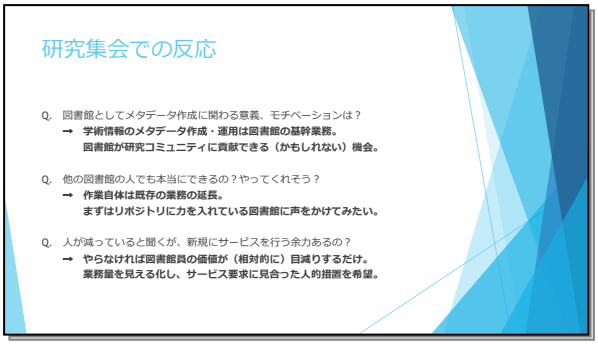
二つ目は分野横断のデータベースということです。IUGONET が分野横断のデータベースであったことが、やりやすさの一つでした。超高層物理という specific な分野ですが、その中でもさらに細かくデータの分野が分かれています。それを横断させるようなデータベースをつくっているということで、図書館は最終的な集合体として全部を大きく横断させるのが目的ですから、基本的な方向性として親和性が高く、図書館の業務としては触りやすいと感じていました。三つ目は明確かつ安定したメタデータ標準です。研究でメタデータを扱うときは、「このようなメタデータがあれば便利だ」という考えが研究を進めていく段階でどんどん出てくると思いますが、アイディアに一つ一つ図書館員の方で付き合っていくというのは、（運用担当として）なかなかやりづらい面があります。しかし、明確

に「これをフォーマットにしましょう」「当面これでやりましょう」というように定めていただくと、運用のレベルでメタデータを扱うことが可能になります。従って、図書館員にとってやりやすいです。

研究集会での質問と私の回答を図 13 にまとめました。「図書館としてメタデータ作成に関わる意義・モチベーションは何か」については、学術情報のメタデータ作成・運用はそもそも図書館の基盤業務であり、この仕事をきちんとしなければ、図書館は（オープンサイエンスの潮流の中で）最終的に何をやるのかという話になりかねないと思っています。もう一つ、前向きな回答としては、図書館が研究コミュニティに貢献できるかもしれない機会だからぜひ進めていきたいとお答えしました。

「他の図書館の人でも本当にできるか、やってくれそうか」については、正直、人によるという回答をしたと思います。ただ、各図書館のトップの方がどのように考えるかはともかく、作業自体は既存の業務の延長なので、やってできないことはありません。まずは、リポジトリに力を入れている図書館や、プロジェクトに関連する大学の図書館、具体的には IUGONET は京都大学、東北大学の方が多いので、そのあたりにぜひ声を掛けてみたいと思っています。

「人が減っていると聞くが、新規にサービスを行う余力があるのか」という質問については、初めの質問とも関連しますが、やらなければ図書館員は紙だけ扱うのかということになって、図書館員の価値が相対的に目減りするだけ、と考えています。業務量が多いから手を出さないという方向ではなく、業務量を見える化して、このサービスが本当に必要なら、（業務量に応じて）人員を増やしてほしいという話に持っていくのが筋だと考えます、という話をしました。



（図 13）

今後の展開

IUGONET で今後やりたいことの一つ目は、運用レベルでの支援（の続き）です。スキーマのバージョンアップは今後も起こり得るので、それに対応したアップデートを行ったり、そのときについでにメタデータクリーニング作業をしたり、あるいは新たにメタデータをリッチにしたりするような支援ができればと思っています。

二つ目は、水平展開です。これは特に重要です。複数人による安定的なメタデータ作成サービスでなければ、私が初めに提案したメリット（図書館のネットワーク活用）が薄れてしまうので、プロジェクト参加機関の図書館との連携を今後やっていきたいと思っています。

三つ目は、図書館側の知見の提供です。これは図書館側としては重要で、「協同」ですから、ただお手伝いするだけではなく、私の方からも何かしら提供しなければ、図書館側としてはあまり面白みがないのです。一番やりやすそうなのは機関リポジトリとの連携です。メタデータをハーベストして、図書館側のデータベース（IRDB）にデータを流して、CiNii などいろいろなところで検索できるようにしてみたりすることのほか、ライセンスに関する情報提供、識別子の話、あるいはデータベースをどのようなクラスターで検索させるか、検索しやすいためのファセット構造を考えるなどの話は、こちらから提案していければと思っています。

何にせよ、研究者の方々が、一番研究が進むやり方、あるいはやりやすいやり方を検討していければと考えています。

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

研究データマネジメントと日本の大学

青木　学聡
（京都大学情報環境機構）

講演要旨

オープンサイエンスへの貢献を含め、大学、研究機関において、研究データを適切に管理することは、新たな知の創生、研究成果の社会への還元、公正な研究活動の維持等、多くの理由からその整備が強く求められている。研究データの入手、保管、公開のルール、すなわち研究データマネジメントの確立は、欧米においては近年、必須の事項とされたが、日本ではまだ初期段階にあり、各大学は手探り状態にある。 本講演では、欧米における研究データマネジメントの位置づけ、大学での取組みを紹介し、日本の大学における研究データマネジメントの今後の展開を議論する。

青木　学聡

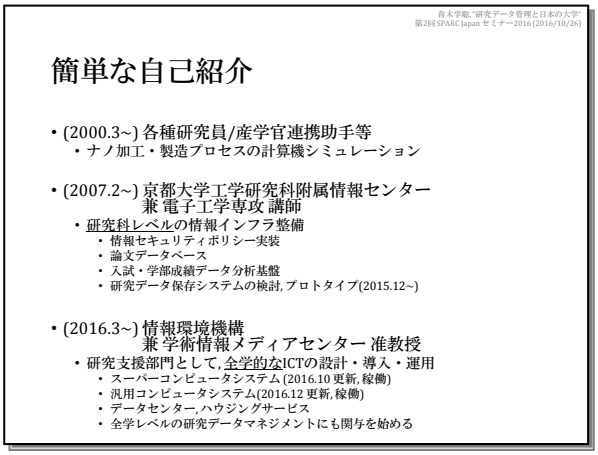
博士（工学）。ナノスケール加工・計測技術のシミュレーションを中心テーマに、2000年より各種研究プロジェクト研究員等として活動。2007年2月より工学研究科講師。教育研究活動と並行し、同附属情報センターにて研究科内の情報セキュリティ、情報インフラ整備、データ分析業務を担当。2016年3月より現職。 大学全体にわたる研究者、研究プロジェクト支援のためのICTシステムの計画、設計、運用に携わる。

1.研究データマネジメントへの関与

私は 2000 年にドクターを取得し、ナノ加工・製造プロセスの計算機シミュレーションを専門に仕事をしていました（図 1）。

2007 年から京都大学工学研究科附属情報センターという非常にマイナーな組織に属して、一方で健啖として電子工学専攻の講師をしていました。そこでは、研究科レベルの情報インフラ整備を行っていました。特に、必要だけれどインセンティブが起きにくい情報セキュリティポリシーの実装をしました。また、京都大学は論文リポジトリを持っていますが、工学研究科でも論文データベースを持っています。ともすれば形骸化しそうなものを、実体を持ったものにするために、それをどうやって研究評価に生かすかも検討しました。入試や学部成績データ分析の相談も受けました。本セミナーのテーマに関係する話では、2015 年 12 月から、研究データ保存システムの検討、プロトタイプの作製にたずさわりました。

このような仕事はあまり外に出ない仕事で、ついつい研究がおろそかになってしまって、どうしようかと



（図 1）

考えていたときに、全学機構である情報環境機構へ異動させていただくことになりました。現在は研究支援部門として、全学的な ICT の設計・導入・運用を行っています。例えばスーパーコンピューターシステムの運用支援、あるいは汎用コンピューターといい、京都大学は幾つか仮想サーバーサービスを提供しているので、それの基盤となるシステムの運用支援をしています。場合によっては、先生たちが持っているサーバー機器を空調の効いたデータセンターでお預かりすることもしています。

最近、全学レベルでの研究データマネジメントをどうすればいいのかという話が出てきて、こういうものにも関与を始めている状況です。

2.研究データ管理の多義性

私は研究データ管理に関わりはじめて半年です。半年間で人からいろいろと話を聞き、このように整理するのがいいのではないかと思った話をします。

研究データ管理には、三つの意味合いがあるのではないかと思います。「研究公正とコンプライアンスのため」「オープン（データ）サイエンス促進のため」「学術領域の発展と社会貢献のため」です。これは最初から順番に、研究者にとってみればつまらないものから面白いもの、といえるかと思います。

2-1.学術領域の発展と社会貢献のため

まずは、一番アトラクティブですが、あまり手を出せない、「学術領域の発展と社会貢献のため」についてからお話しします。

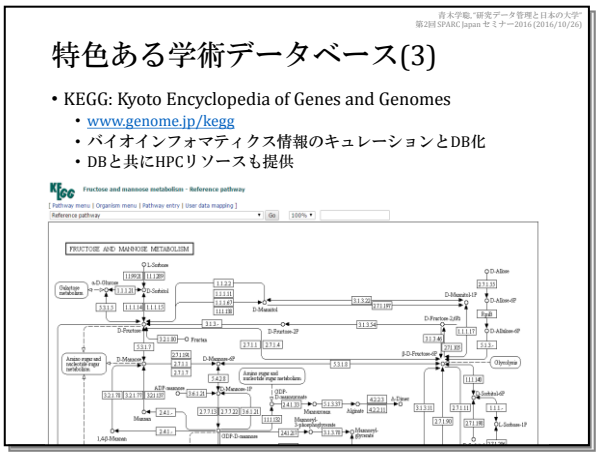
京都大学はやはり大きい大学で、しかも各部局の独立性が高いです。いろいろな小さい研究ユニットやセンターを持っていて、そもそもデータを集めて公開することが存在意義だというセンターもたくさんあるので、非常にたくさんの学術データベースがあります。

例えば、京都大学が保有している貴重資料画像を高解像度でお見せする図書館のデータベースがありますし（図 2）、Kyoto Encyclopedia of Genes and Genomes（KEGG）というバイオインフォマティクスのデータベースなどもあります（図 3）。このあたりのデータのアクティビティや位置付けについては下田先生や仲里先生に伺うのが一番いいと思いますが、このようなデータベースを各研究者が独自に運営し、世界的に有名になっているという状況です。

ただし、こういった研究・学術データは、主にコミュニティレベルで、部局やセンター、研究ユニットのような小さいユニットで、多大な労力と時間をかけて醸成されています。先ほどの KEGG は 20 年かけて育ててこられたという実績を伺っています。また、自発的な取り組みなので、研究者とその研究コミュニティが使いやすいように、独自に改良が重ねられてきました。

そういう意味で使いやすいデータベースになってい

（図 2）



（図 3）

ることは確かです。その一方で、持続的に発展させる、維持させることが大きな課題になっているということが、私が学術領域の発展と社会貢献の視点から見たこれらのデータベースの特徴です。

ただし、大学側は、一体どれぐらい学術データベースがあるのか、全容を全く把握できていない状況です。今後、偉い人が「大学としてのコンテンツ発信力を評価するぞ、もっと頑張れ」という新たな評価軸を設ければ、恐らく皆さんはせっせと調べるのでしょうけれども、まだそういうことは出てきません。

また、現行の運営側は、自分たちが頑張って運営してきたもので、お金がないから何とかサポートしてほしいという本音と、自分たちが好き勝手やるのだから、大学の上に余計な手出しをしてほしくないという相反する思いがあります。このあたりについては今、皆さんが自発的に、しかも先頭を走っているデータベースについて、研究データ管理をどうしなさいということを提言できる状態ではないというのが私の考えです。

2-2.研究公正とコンプライアンスのため

次は、研究公正とコンプライアンスのためのデータ管理についてです。研究現場でどんなインパクトがあったかという話をします。恐らくグッドプラクティスではなくバッドプラクティスになるのではないかと思います。

2014 年 8 月、文部科学省が「研究活動における不正行為への対応等に関するガイドライン」という文書を出し、各国公立大学・私立大学は対応しなくてはいけなくなりました。2015 年 2 月に京都大学は、「京都大学における公正な研究活動の推進等に関する規程」を出しました。これは公開されている内容なので、皆さんも読むことが可能です。ここには、教職員・部局（研究科や学部）・全学の役割、責任の所在、研究データ一定期間保存義務が明記されました。

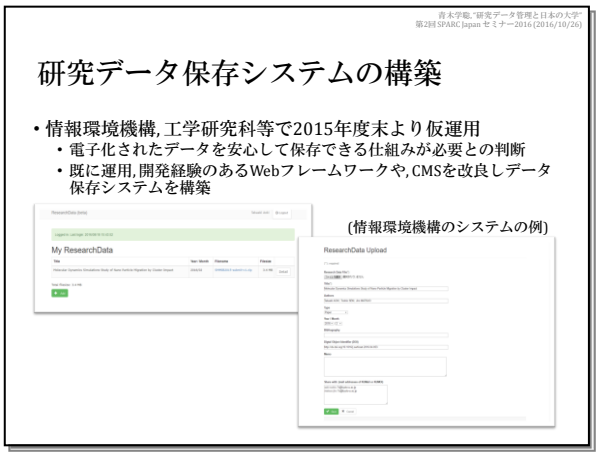
続いて 2015 年 7 月に、京都大学は、年度内に研究データの管理方法、保存方法を部局ごとに制定するように求めました。現在、2016 年度は、部局ごとに決定した管理方針ないしデータ保存計画を基に行動しているはずです。

私は部局ごとの制定が求められたときに工学研究科にいて、先生たちに研究データを 10 年間保存するように言っても、絶対にこれが一律に守られるわけがないということがあったので、慌てて研究データ保存システムを立ち上げることとなりました。工学研究科や情報環境機構は ICT のインフラを整備する底力がありました（図 4)。そのシステムを使って、研究科ないし情報環境機構という組織でデータを預かり、バックアップを取る活動を始めています。これはまだ試行段階です。

ファイルを zip ファイルにしてアップロードし、論文のテーマ、著者、発表日、発表メディア、書けるのであれば DOI を書きます。研究成果の正当性を保証する研究データがどこに保存されているかをトラックできるようにするための非常に簡単なシステムです。

それで始めたのですが、やはりやっつけでつくったので、いささか運用・安定性に不安があります。一方で、何もデータをきちんと管理するということは、研究データに限りません。事務の文書でも何にしても、データをきちんと保存したいという要求はあるはずなので、次に更新する計算機システムの中に、大容量の光ディスクを使ったアーカイブシステムを導入することにしました（図 5)。

これと連携させて、この上で、エンタープライズ・コンテンツ・マネジメントシステム、文書管理システ



（図 4）

ムを導入し、ユーザーは研究データとメタデータを文書管理システム上に登録する。ダークアーカイブをしなければいけないときは、コンテンツを光ディスクにコピーして、そちら側は誰も触れない状態でデータの正当性を保証するという仕組みを今検討して立ち上げようとしています。

ただし、研究データ管理の観点からすると、これはあくまでも組織あるいは研究者を保護する保険であり、オープンデータではありません。これはある種のバッドプラクティスの一つではないかと思います。とにかくデータを長期保存しなくてはいけないということで、2 年前にいきなり上から降ってきて、慌てていろいろなルールを決めて、すごく駆け足でやってしまったので、ルール（policy）が先行して、現実的な実施手順があまり検討されないままずっと走ってしまっているという現状があります。ルール（policy）を実施手順（procedure）と実施（deployment）が後追いしているのです。

私のいる情報環境機構は全学に対してシステムを提供することをミッションとしているのですが、この仕組みを使うと、各部局が定めた研究データ保存のプロシージャーポリシーに一致するかどうか、というレビューはまだ受けていない状況です。それはこれから擦り合わせが必要です。いびつな構造で、システムとポリシーと実施手順がかみ合っていない状況で、とにかく走らせなければいけない、という状態にあるということです。

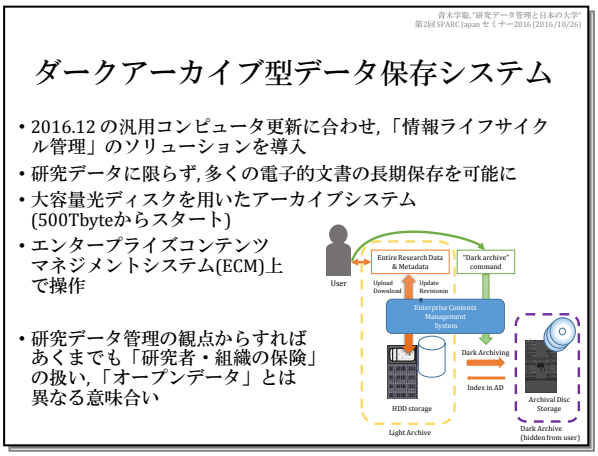
また、困ったことに、ほとんどのデータが死蔵されるものです。ここに収められたデータは何か研究不正が疑われたときに、ここに研究データがあると出すためのものだからです。だから、実はそれほど力を入れて構築したくないシステムで、できるだけ安く、それでなお皆さんに信頼されるシステムにする難しさも抱えています。研究公正とコンプライアンスだけのために IT システムを構築をしなさいと言われると、非常にやりにくい ICT システムになってしまうという事例になってしまっているのではないかと考えます。

2-3.オープン（データ）サイエンス促進のため

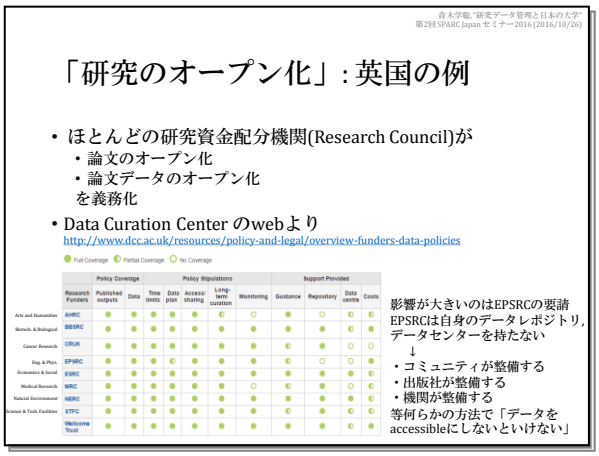
オープン（データ）サイエンスとは、今まで一部の研究者が自発的に実施してきたオープンデータ的な活動から広がり、研究者がデータを集めて、保存して、公開することを当たり前の作業にしていく運動だと私は認識しています。これに対して世界はどうなっているのか、あるいは日本国内はどうするべきかという話をします。

2-3-1.英国の事例

先日、私は大学のプロジェクトの資金を頂き、英国を訪問し、英国での研究データ管理がどうなっているのかという話を伺う機会がありました。英国ではほとんどの研究資金配分機関（リサーチカウンシル）が論文をオープンにすること、さらに踏み込んで、論文に載っているデータ、研究の成果として公開していいデ



（図 5）



（図 6）

ータは、オープンにして電子的にアクセスできるようにすることを既に義務化しています。具体的に星取表を英国の Digital Curation Centre が提示しています（図 6）。この機関の資金の場合は、このデータをオープンにしなければいけないということが書いてあります。これは非常に有名な絵なので、いろいろなところでご覧になった方がいらっしゃると思います。

また、英国に限らず、アメリカ、ヨーロッパの例は、科学技術振興機構（JST）が 2015 年 5 月に研究データの共有ポリシーに関するレポートを書いているので、それを読んでいただくと一番分かりやすいのではないかと思っています。

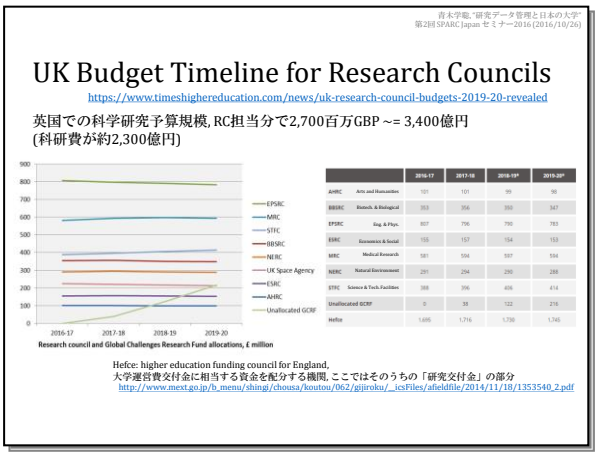
英国で一番インパクトが大きかったのは恐らく、英国工学・物理科学研究会議（EPSRC）から、工学と理学の分野はデータオープンアクセスにするように要請があったことでしょう。EPSRC はデータをオープンにするように言っていますが、データを置く場所は準備しておらず、自分たちで準備するか、コミュニティが準備しているリポジトリを使うようにと言っています。だから結局、工学・理学系の先生たちが研究をして、論文を出版したときに、そのデータにアクセスできるようにするためには、自分たちが所属している研究コミュニティがリポジトリを整備するか、出版社のデータジャーナルに投稿するか、自分自身でやるしかないのです。

英国の研究資金配分機関の予算規模は、約 2,700 万英国ポンド（約 3,400 億円）です（図 7）。科研費の規模が約 2,300 億円なので、それの 1.5 倍に相当する研究に対してオープン化を求めたということです。それ以外に、日本で言う厚生労働省系なども同様に、データをオープンにしなさいという宣言を次々出していて、ほとんど全てのファンドがオープンデータにするという方向に向かっているという現状です。

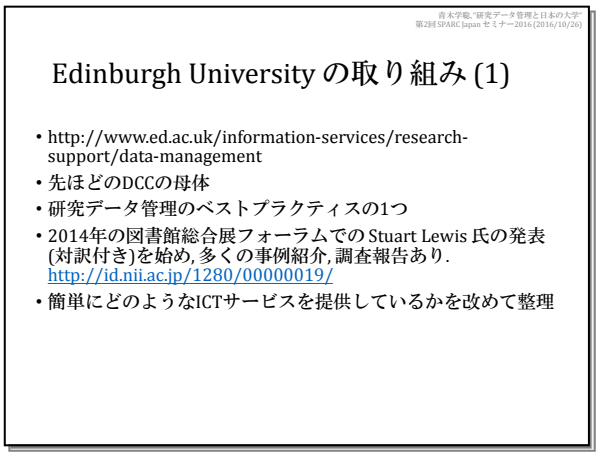
それに対して、英国の各大学はどうしているのでしょうか。一番先進的な取り組みをしているといわれているエジンバラ大学は、研究データマネジメントについて集中的に研究している Digital Curation Centre を生み出した母体になっている大学であり、研究データ管理を具体的に実践している大学の一つです（図 8）。

非常に有名な大学なので、何回も調査が行われ、調査レポートも出ています。一つ参考になるのは 2014 年の図書館総合展フォーラムで、ここのヘッドであった Stuart Lewis 氏が発表された内容です。英語と日本語の対訳のスライドが出ています。これを見ていただくと、歴史と今やろうとしていることが簡単にまとめられていて非常に参考になると思います。

私はインフラを提供する立場なので、ここで一体どんな ICT サービスを提供しているかに注目してお話しします。エジンバラ大学は、簡単なブローシャー「A guide to the Research Data Service」を出版しています（図 9）。研究を始める前、実際に研究を行っている間、研究が終わって成果を公開したいときという研究の各ステージ、そしてこれらの研究活動サイクルでどのようにトレーニングとサポートをするか、こうい



（図 7）



（図 8）

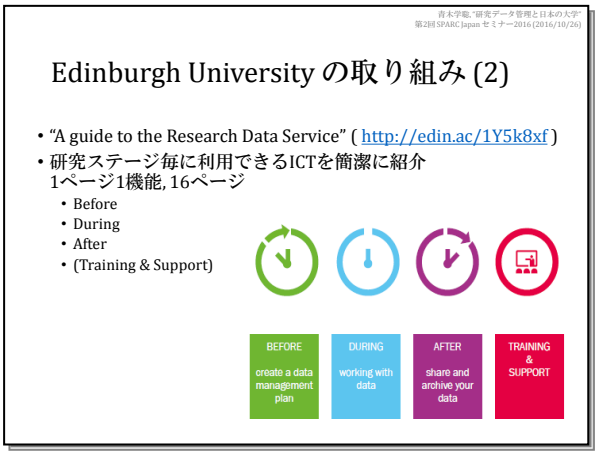
う四つの視点でエジンバラ大学では何ができるか、何を提供しているかをまとめています。1 ページ 1 機能で 16 ページ、ぱらぱらと読んで、なるほどと思える内容です。

最近、英国ないし米国では、研究申請調書を書く際に、こういうデータを集めて、どこで管理して、最後にどのように公開するかというデータ管理プランを 2 ページほど書くように、と言われます。エジンバラ大学では、出すファンドに合った Q&A がウィザード形式で表示され、それに答えていけばデータ管理プランを仕上げることができるツールを提供しています（図 10）。

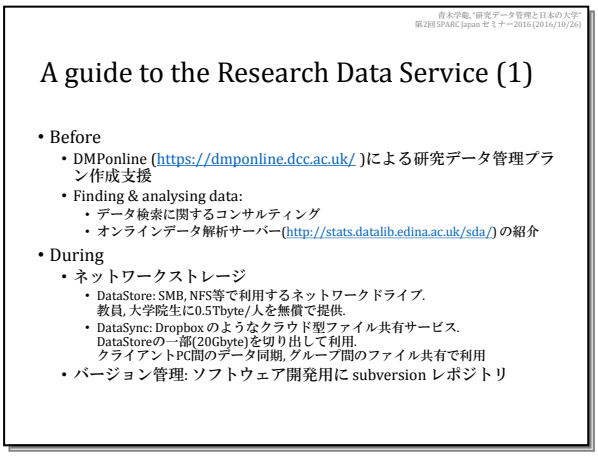
エジンバラ大学の経済・社会学系では、オンラインデータをその場で解析する仕組みも提供しています。このあたりはそれほどインパクトはないのですが、一番研究者にとって魅力的に移るのは、研究をしている最中にデータをどう扱うか、どう保存するかということです。そこでエジンバラ大学では、教員・大学院生に対して、0.5Tbyte、500Gbyte ぐらいのネットワークファイルストレージを使っていいという話をしています。当然、全員が使っているわけではないですが、最大これぐらいまでは自由に使っていいということ、だと推測しますが、少なくとも、自分たちの普段の生活で出てくるデータは全て大学が提供するネットワークドライブ上で扱ってよいというぐらいのキャパシティになります。また、共同研究者とデータがシェアできるように、Dropbox のようなクラウド型のファイル共有サービスを提供しています。

研究が終わった場合、論文を出版するということがありますが、PURE というシステムを持っていて、これは研究者総覧と機関リポジトリを合わせたような仕組みです（図 11）。内容はかなりいろいろなものが入っています。大学にどのような人がいるのか、どのようなプロジェクトが実施されたのか、研究成果（主にオープンアクセスの論文）、その他の研究活動（受賞等）などを、利用者の選択でアップロードして公開できるシステムです。これは恐らく日本の大学でだいぶ整備が進んでいるものだと思います。

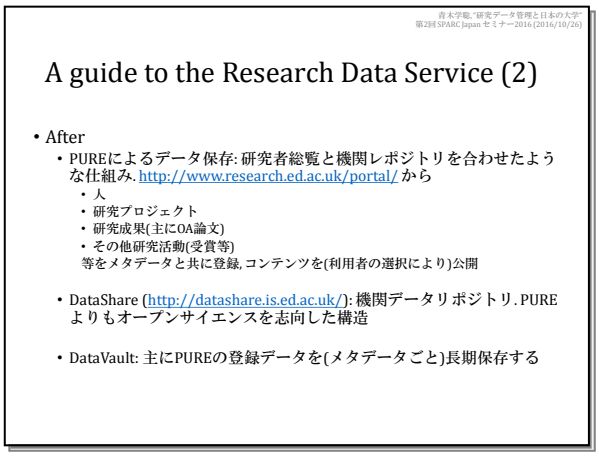
これとは別に、もう少しディテールにこだわった研究のデータセットを公開する機関データリポジトリも持っています。



（図 9）



（図 10）



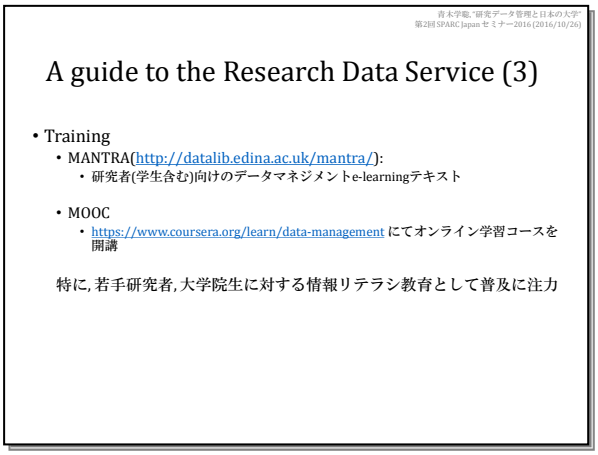
（図 11）

そしてトレーニングです（図 12)。向こうの方は、お年を召した研究者にいろいろ教えても駄目で、若い研究者、大学院生に対してしっかりと、「これからの時代はこうやってデータを管理して自分たちの研究をより良くするのだ」と教え込むことが重要だということを強調していました。今、研究データ管理は恐らくオープンサイエンスを促進したいということがバックグラウンドにあって、ファンドの要請に従って、データ管理プランを作成して実施しています。それに堪えられるだけの十分な研究用のインフラの提供（大容量のストレージ、機関リポジトリ）は、機関の体力勝負になるのではないかと思います。

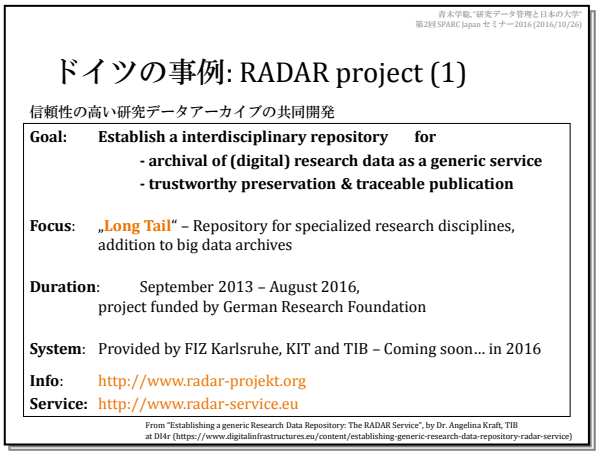
全員がやらなくてはいけないとなった場合は、非常にロングテールなものをサポートしなければいけません（図 13)。よく頑張っている、非常に頻繁にアップデートするデータリポジトリだけでなく、その後ろの方にある、何に使われるか分からないけれども残しておきたいデータまできちんとカバーしなくてはいけません。そうすると、可能な限りメンテナンスフリーで、汎用的な仕組みをしっかり入れなくてはいけません。ということは、できるだけコストを下げるという視点がどうしても必要になってきます。こういうものは果たしてあるのか、ないのかということを、これから各機関で整備しなさいと言われたら、どうしても考えなくてはいけないことになってしまいます。

2-3-2. ドイツの事例

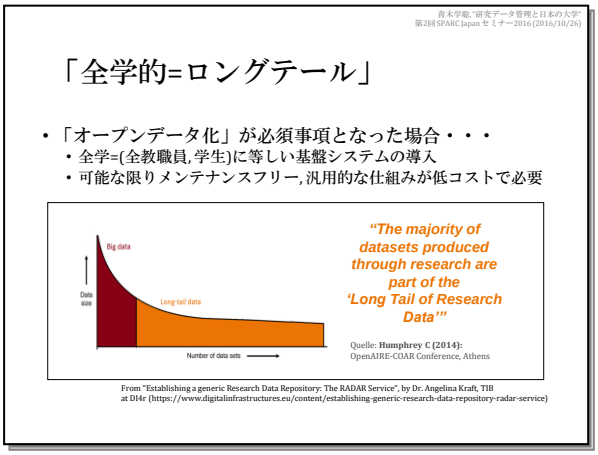
それに対応する形として、ドイツでは、RADAR project という、研究データアーカイブのシステムを共同開発して、それを各機関へ提供するというプロジェクトがあります（図 14)。ステージとしては、研究を始めていて、共同作業が終わり、研究が終わったとき



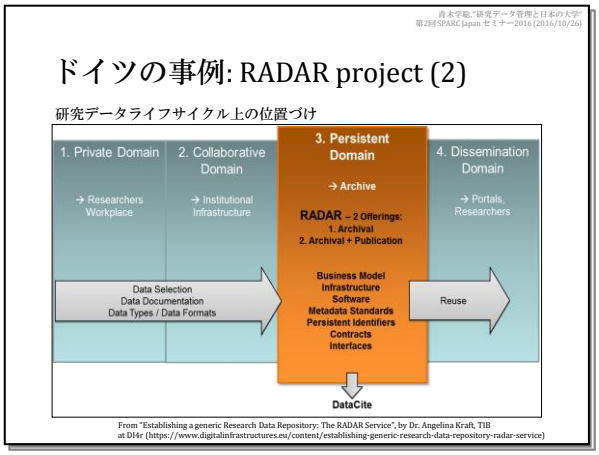
（図 12）



（図 14）



（図 13）



（図 15）

に研究の内容物を一気にアーカイブします（図 15）。目的は、一つはアーカイブするだけ、もう一つは研究者の選択によりその中の一部をパブリッシュすることです。

ビジネスモデルとしては、ベーシックサービスは何も言わずに黙って保管するだけです（図 16）。その中で、これとこれは公開したいと言われたときは DOI を付けて公開できるようにしようとしています。

各機関あるいは研究グループに提供するときの値段も決めています。2016 年 8 月にできる予定とされていますが現時点ではまだ公開には至っていない模様です。もうすぐ出てくると期待しています。

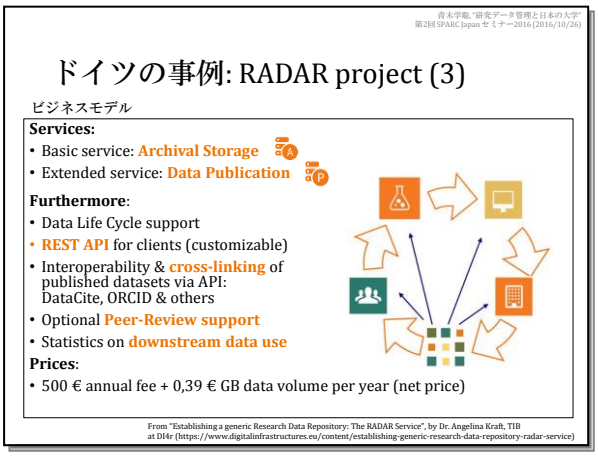
3.京都大学は何をするか

それでは、京都大学では何をしようということです。当初は、どこから手を付けたらいいのか分からない状況でうちひしがれていたのですが、図 17 のようなことをまとめています。本当にできるかどうかは分かりませんが、エジンバラ大学プラスアルファのものを理想高く目指すのであれば、これぐらいのものを準備すると完璧だ、という内容。まず、研究者・大学院生が自由に利用できる大容量のクラウドストレージは絶対に要るでしょう。そしてこれを、フラットな環境でできることです。アクティブな研究活動を行い、学内および学外でシームレスにデータ共有をし、きちんと協働できる環境は研究機関としては絶対に持つか、あるいはどこかで整備しなければいけないと思います。

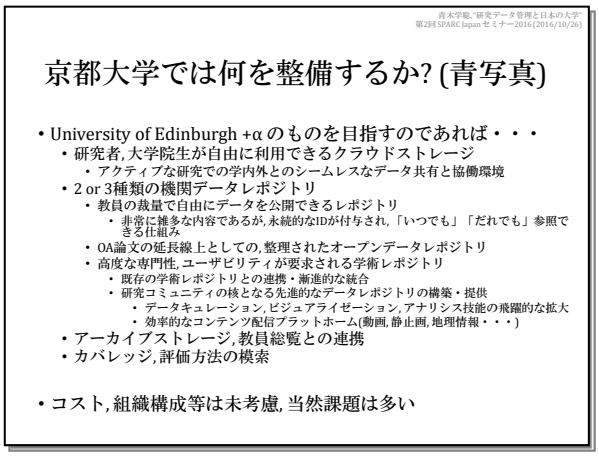
二つ目は、2～3 種類の性格の異なる機関データリポジトリです。データをアーカイブする場所、ためておく場所として必要になります。1 種類目は、教員の裁量で自由にデータを公開できるリポジトリです。先生たちが何かの機会で、このデータは面白いと思ったとき、あるいは、どこかの学会で発表して簡単なペーパーにしたので見てほしいというときに、それに永続的な ID、DOI かまたはそれに類するものを自発的に振って公開できる仕組みです。そこにはいわゆるキュレーションなどの作業はありません。つまり研究者が自発的に公開するドキュメントに対して永続的な ID を付けられる仕組みを持ったリポジトリです。

2 種類目は、今、各日本の大学はオープンアクセス論文を対象としてリポジトリを整備していますが、その延長線上として整理されたオープンデータリポジトリ、あるいはある種のキュレーションやある程度のフィルターを通ったリポジトリです。

3 種類目は、高度な専門性、ユーザビリティが要求される学術リポジトリです。これは、今業界の最先端を走っていて充実したリポジトリに迫る何かということです。現在アドバンストな、あるいはソフィスティケートされた学術リポジトリと連携・漸進的な統合ができるような基盤、さらに何か新しい学問領域を生み出すときには、新しいデータベース、新しい考え方、新しいメタデータを生み出せるような基盤になります。きちんとしたデータキュレーション、ないしはこのデータはこのように見るものだというデータビジュアラ



（図 16）



（図 17）

イゼーション、このように使うべきだというアナリシス技能を飛躍的に拡大させることができないことには、高度な専門性やユーザビリティにはなかなかたどり着けません。

ただし、これだけオープンデータが出てきたということは、今まで 20 年などすごい時間をかけていたものを、少しは早めることができる可能性があります。今まで 20 年かかっていたものが 10 年になる、もしかしたら 3 年ぐらいで立ち上がる基盤になるのであれば、ここを頑張る意味はあるのではないかと思っています。

研究公正のためのデータや、人には見せたくないけれども永年保存したいデータとうまく連携するような仕組みはやはり必要だと思います。

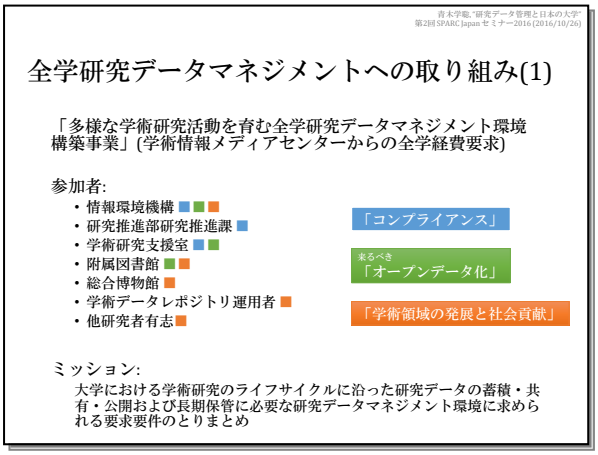
いろいろなデータベース、リポジトリが出現、整備された際には、結局、うまくいったのか、いかなかったのかということは絶対に言われると思います。パフォーマンスをどう評価するかは考えておかなければいけませんが、全くアイデアがない状況です。幾つかの研究でオープンアクセスは経済的価値を生んだというレポートは出ていますが、各機関がオープンアクセスにしてどれくらいの価値があったかということについては、何とも言い難い、評価のしようがない部分があるかもしれません。ここは非常に難しいところだと思います。

当然ながら、全部やるとなるとコストはうなぎ上りですし、幾つか選んで実施するか、NII が今準備しているようなフレームワークにうまく乗る、そういうところでお互いに情報連携を、パイロット的な事業に組み合わせて、どのように大学の中でうまくローンチさせるかを考えていきたいと思っています。

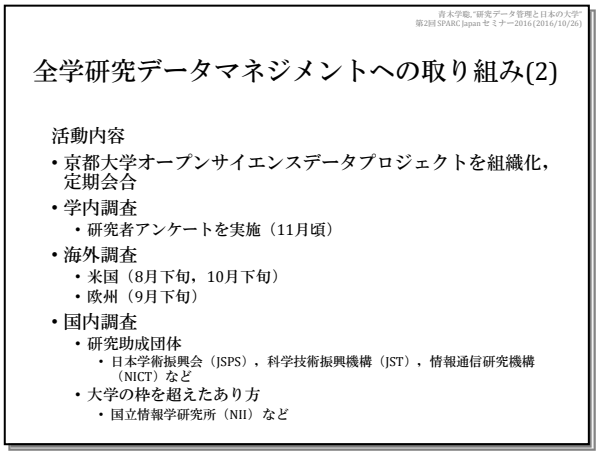
4.全学研究データマネジメントへの取り組み

こういうことを考える機会として、京都大学では今年度、どうやって研究データ管理の環境を構築するかというタスクフォースのプロジェクトを立ち上げました（図 18）。私が入っている情報環境機構の他、コンプライアン関係の研究推進部、そして学術研究支援室というファンディングのサポートを行う部署も入っています。実際にリポジトリで精力的に活動されている附属図書館、博物館、能勢先生をはじめとする、実際に学術データのリポジトリの運用をしていらっしゃる方、研究者有志、これらの方が集まって、大学における学術研究のライフサイクルに沿った研究データの蓄積・共有・公開、長期保管に必要な研究データマネジメント環境に求められる要求要件を取りまとめて、次年度以降に京大はどうしたらいいのかという提言ができれば、ということで、このような活動をしています。

具体的には、いろいろなミーティングを行い、関係各所からヒアリングをし、研究調査を行いました。私もエジンバラ大学をはじめ、幾つかの大学を回りました（図 19）。



（図 18）



（図 19）

5.まとめに代えて

私からは、京都大学における研究データ管理の現状を、「コンプライアンス」「オープンデータ化」「学術領域の発展と社会貢献」に分類し、それぞれに一体何が求められているかという話をさせていただきました。

ただし、この「コンプライアンス」と「学術領域の発展と社会貢献」は、いずれはオープンデータという大きな概念として、ポリシー・手順・システムなどが集約されていくと考えて、準備を進めていくことが重要です。

オープンデータを義務化することに関しては、恐らく大学あるいは機関全体の ICT インフラの底上げに直結する内容になるので、各組織が十分なサポート体制を、義務化のタイミングを見計らいながら構築していくことが大切だと考えています。

最後に宣伝です（図 20）。大学 ICT 推進協議会年次大会で、研究データマネジメントのセッションを持ちます。2016 年 12 月 14～16 日で、本セッションは 16 日 13 時半～15 時を予定しています。エジンバラ大学図書館の Dominic Tate さんもお呼びし、私たちの活動内容を報告させていただきます。

（図 20）

第 2 回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

研究データ利活用に関する国内活動及び国際動向について

武田　英明

（研究データ利活用協議会/国立情報学研究所）

講演要旨

研究データに関する利活用に関する関心が近年、国内外で高まっている。本講演では、オープンサイエンスの流れの理解とその上での研究データ利活用の枠組みについて概要を説明する。その上で、国内及び国際的な動向について概観する。国内では DOI の RA(登録機関)であるジャパンリンクセンターが 2014 年に行ったデータ DOI 実験プロジェクトを契機に分野横断的なつながりができ、それが研究データ利活用協議会の発足につながった。国際的には RDA(Research Data Alliance)が 4 年前より活動を始めており、funder、研究機関、出版社等を巻き込んで、横断的なつながりを形成している。その活動を一部紹介する。

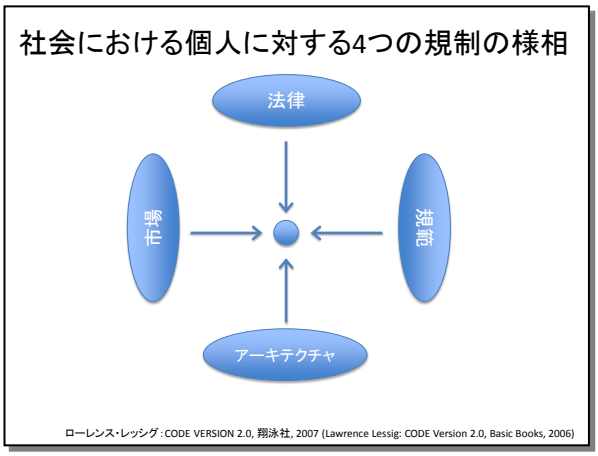
武田　英明

http://www.nii.ac.jp/faculty/informatics/takeda_hideaki/

現在、オープンサイエンスということが盛んにいわれています。しかしなぜ今、オープンサイエンスなのでしょうか。

なぜ今、オープンサイエンスなのか

一つだけ理解してほしいことがあります。それは、サイエンスをオープンにしたかったからオープンになったわけではないということです。それを理解するための一つの鍵はローレンス・レッシグの批判です。彼は、われわれ個人は四つの方向から規制を受けていると主張しています。それは市場（お金）、法律、規範、アーキテクチャです（図 1）。レッシグは法律家ですから、規制とは法律であると言いそうですが、そうではなく、市場（お金）からも受けているし、規範からも受けているとしています。最後のアーキテクチャが一番分かりづらいところです。これは社会の仕組みそ



（図 1）

のものから規制を受けているということです。かつてインターネットがなかった時代は、われわれは例えば鉄道網などの社会の規範から規制を受けていたけれど、現代はインターネットの仕組みで規制を受けているということが彼の主張です。

インターネットの世界、特にウェブの世界においてはオープンでフラットな仕組みが導入されました。それがわれわれのアーキテクチャになったということです。そのオープンやフラットということが、市場・法律・規範にも逆に影響を与えているのが現在なのです。

それと同じことが今、サイエンスにおいても起きているということがポイントです（図 2）。われわれはそのさなかにいるのでぴんとこないかもしれませんが、例えば研究者の規範として、成果はみんなで共有すべきといわれるけれど、インターネットがなかった昔には、みんなで共有するには出版するしかありませんでした。だから、論文出版が良かったのです。でも今は、インターネットだったらオープンで、もっとたくさんの人に見てもらえます。だからオープンデータにするのです。

われわれの規範は既に変わってしまっています。それはアーキテクチャが変わったからです。もちろん市場も変わりました。そうなったために、既存の出版社はそのオープン性を自分たちの商売へ入れ込んで、article processing charge（APC）を取るようになったのです。法律についても、各助成団体や国が制約を設けるようになりました。

オープンサイエンスの系譜

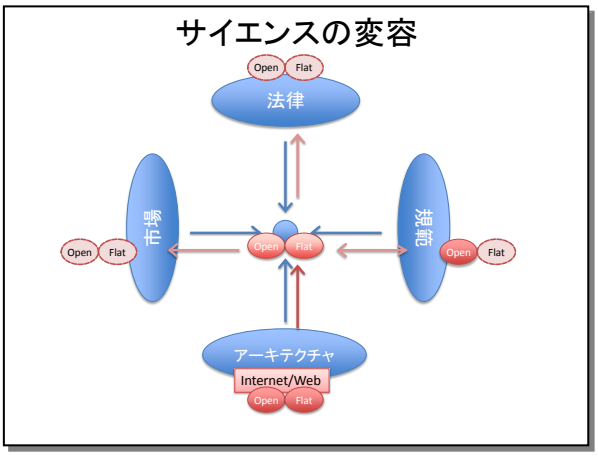
ネットワーク、コンピューターのアーキテクチャがどんどん変わりつつあり、それが直接・間接にわれわれに影響を与えているのが今のオープンサイエンスの世界です。図 3 の一番下が、ソフトウエア、コンピューター、ネットワークの世界で何が起きているかです。その世界から生まれてきたのがクリエイティブ・コモンズ、Web2.0 などです。

仲里先生の報告では、生命科学では 1960 年代から MEDLINE があったという話が出ました。でも、PubMed になったのは 1990 年代です。これは、もともと自分たちのコミュニティでデータを共有したいという考えがあったのですが、インターネットが発展してやり方が変わってきたという例です。

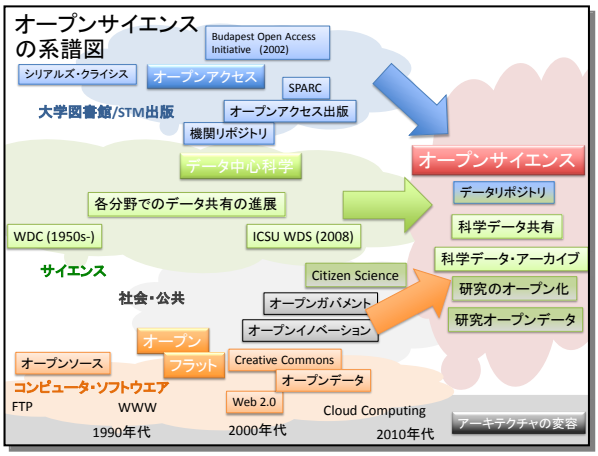
出版物も今までは紙で出版するのが良かったのですが、インターネットの発達によってウェブで出版することで、オープンとのつながりができてきたというのが一番上の青色です。

真ん中の部分は、むしろ社会・公共が変わって、オープンガバメント、オープンイノベーションが入ってきているということです。

オープンサイエンスは、この四つの絵から影響を受けて今があるということが理解を難しくしています。例えば、データリポジトリは図書館系とサイエンス系の中間辺りにあります。研究のオープン化、研究オープンデータは、どちらかというと政府・公共のオープン系と研究のオープン系の中間ぐらいにあります。そ



（図 2）



（図 3）

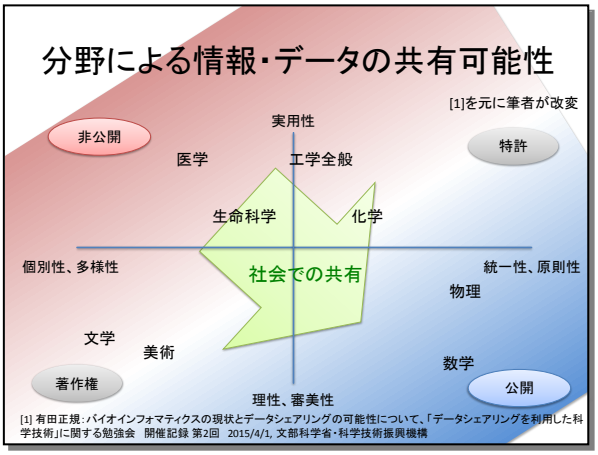
ういう状況で、まだら模様がオープンサイエンスの今だということです。

オープンサイエンスと研究データ共有

研究コミュニケーションの世界では、昔はファクスでやりとりをしたり、「あなたの論文を下さい」とはがきを書いたりすることがありました。それがオンライン化すると、雑誌購読料が高騰し、大学図書館に雑誌を購読するお金が足りなくなって購読雑誌が減少するというシリアルズクライシスが起きました。その対策として、SPARC やオープンアクセスが出てきました。

科学では、生命科学や天文学など、各分野でベストなデータ共有の方法を探してきました。ここで少し注意が必要なのは、分野ごとにデータ共有の特性が違っており、共有の度合いやデータ量、分散か集中なのかも違うということです。

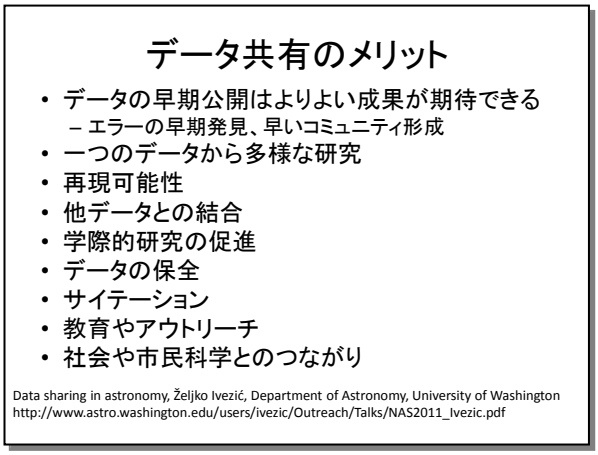
それを無理やり図にすると図 4 のようになります。二つのラインはかなり主観的に分けています。個別性・多様性を探求する学問なのか、統一性・原則性を探求する学問なのか、実用性を探求する学問なのか、理性・審美性を探求する学問なのかをマップしようとする試みです。例えば、理性・審美性を下に置いて、実用性を上に置いたときに各学問がどこにあるか。数学や物理は、より統一性や理性を探求するようなものなので、右下でしょうか。逆に工学は比較的上の方です。文学・美術は個別性・多様性を尊びつつ、ある種の理性・審美性を求めるので左下としています。



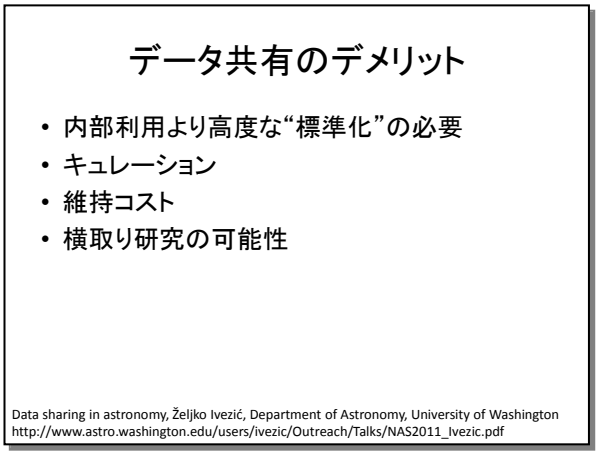
（図 4）

このとき、右下が公開、左上は非公開、その中間の右上と左下は、右上が特許で左下が著作権という世界が大ざっぱな理解で見えてくると思います。共有できるかどうかという観点では、右下に行くと比較的この問題はやりやすく、左上に行くと慎重にならざるを得ません。自分の学問分野によって立ち位置が変わります。この図で、自分はここにいるからこうなのではないかと理解していただけるといいと思います。そういう問題がある上で、各分野ではそれぞれの特性に合わせたことをやってきました。

図 5 は、天文学者が挙げたデータ共有のメリットです。天文学は最初のころ、300 年ぐらい前はデータを隠していたのです。ガリレオなども隠していました。下手をすると死ぬまでデータを出さない。それを天文学は早くに克服して、このようなメリットがあると言っています。



（図 5）



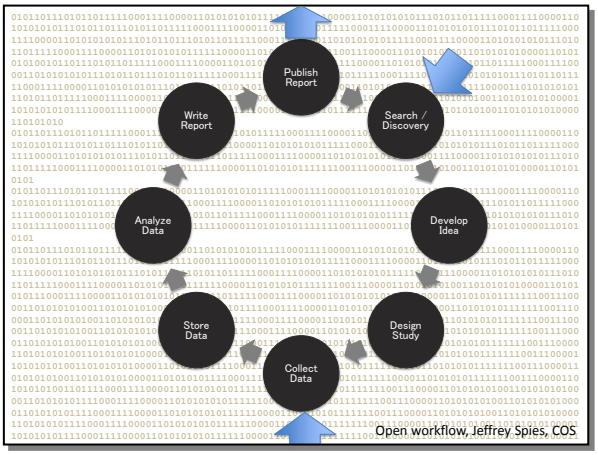
（図 6）

デメリットももちろん挙げています（図 6）。やはり人に見せるためにはコストが掛かります。キュレーションも必要です。維持するにもコストが掛かります。横取り研究の可能性もあります。ここまでが各学問分野での研究データ共有、メリット・デメリットの話でした。

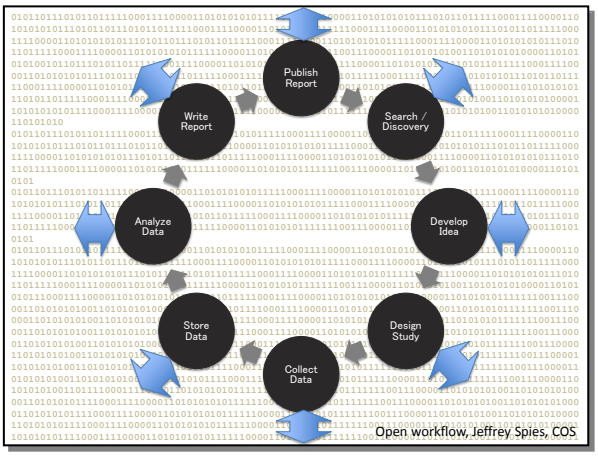
研究データ共有の枠組み

では一体、オープンサイエンスなりオープンデータは何をすることにならざるを得ないのかということです。デジタル化以前の研究者は論文を書いて、データも自分でつくっていましたが、だんだんデータをもらうようになり、論文もデジタルになり、データも論文もほとんど区別がつかなくなったのが現状です。

図 7 は、アメリカの Center for Open Science の Jeffrey Spies 氏がつくった Open workflow の図です。先ほどは



（図 7）



（図 8）

in と out しか描いていませんでしたが、もっと細かく描かれた、最初に探して、アイデアをつくり、実際に研究を計画し、関連するデータを集めて、分析し、レポートを書いてパブリッシュするというストーリーが回るのが研究のライフサイクルだというものです。誰が描いても大体このような図になると思います。

先ほどの私の図では、図 7 の青い矢印のところにしか入り口と出口はありませんでしたが、彼らは全体がオープンになるということを知っています（図 8）。研究プロセス自体がオープン化する。あるいは、最初は共有して、プライバシーやセキュリティが入っているものは永遠に公開されないかもしれませんが、そうでないものは時間が経つと公開されるという意味で、プロセス全体が公開されるという前提がわれわれの研究の未来像だと思います。これは今すぐにできるものではありませんが、5 年、10 年経ってみるとこれが当たり前の研究スタイルになっているだろうということは私も同意します。

ただ、研究プロセス全体というよりも、データ共有が当面の課題だと思います。データ共有のポイントはデータのライフサイクルです。研究者は研究途上のデータ、研究発表に使ったデータ、保存用データと切り分けますが、問題は担当が違うことです。研究中であれば研究者で、研究が終わるとそこから先は研究機関に任されます。データの作成から保存まで、研究データのライフサイクルを通してどうサポートするかということが課題です。

FAIR 原則はもともと FORCE11 がつくったものです（図 9）。FAIR というのは、Findable（見つけられる）、Accessible（アクセスできる）、Interoperable（相互運用可能）、Re-usable（再利用できる）の頭文字を取ったもので、研究データがどうあるべきかという原則です。今この FAIR 原則がコンセンサスになりつつあります。研究オープンデータがどうあるべきかという方向はかなり見えてきているように感じます。

Research Data Alliance（RDA）

ただ、それを実際にどう実施するかが問題です。今、大きく違う世界の人が関わってオープンサイエンスというコンセプトができているので、ステークホルダーが非常に多いのです。研究データ共有に関する国際活動として、Research Data Alliance（RDA）が 2013 年から、最初は 5 年という形で始まりました（図 10）。今まで研究に関するコンソーシアムは、研究者や研究機関が集まり、せいぜい拡張しても政府関係者、ファンディング、ファウンダーでしたが、RDA には研究者、研究機関、出版社、政府関係に加えて、社会の IT ベンダーや企業など、非常に多様なステークホルダーが入っていることが特徴です。

RDA は年に 2 回プレナリーミーティングというものを開いていて、先月デンバーで行われました。今度は 4 月にバルセロナで行われます。

（図 9）

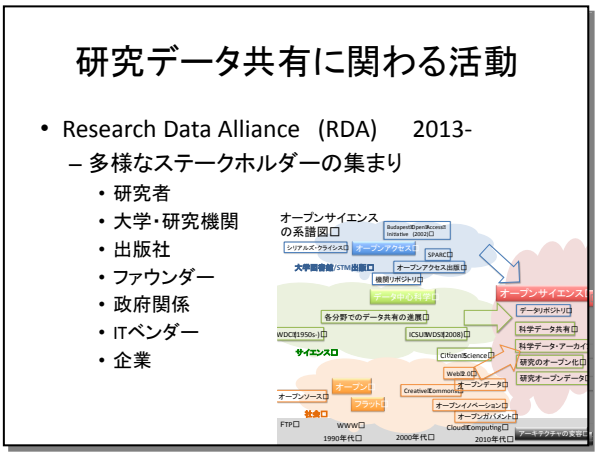
RDA は、研究データ共有と交換の障害を減らすためのインフラストラクチャーとコミュニティ活動の発展と世界的なデータドリブンイノベーションの加速に焦点を当てた国際的な活動です。皆が共有できるインフラをつくりたい、それを支える人の活動をつくりたい、プラス、データドリブンイノベーション、社会インパクトを与えることなど、かなりスコープを広く取っています。

RDA は、研究者とイノベーターが技術・ディシプリン・国境を越えてオープンにデータを共有するというビジョンを達成するために、データのオープンな共有を可能とする社会的・技術的架け橋をつくります。面白いことは、単に技術で応えるのではなく、社会的な仕組みでも応えるというところです。

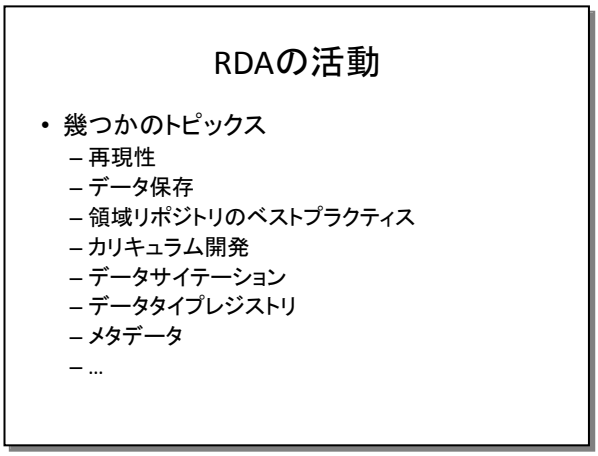
ここはトップダウンの組織ではなく提案ベースで、人が集まってインタレストグループとワーキンググループをつくり、リコメンデーションなどのアウトプットを出します。ワーキンググループでは 18 カ月で何かアウトプットを出すというプロセスだけが決まっていて、あとは自分たちで提案し、人が集まれば承認されるという制度です。

RDA で取り上げられるトピックスとしては、図 11 にあるものが挙げられますが、もっとたくさんあります。今日出てきたような話は大抵どこかのグループのテーマになっています。ポリシーをどうするかも取り上げられます。

RDA 自体は、組織はほとんどないも同然で、かな



（図 10）



（図 11）

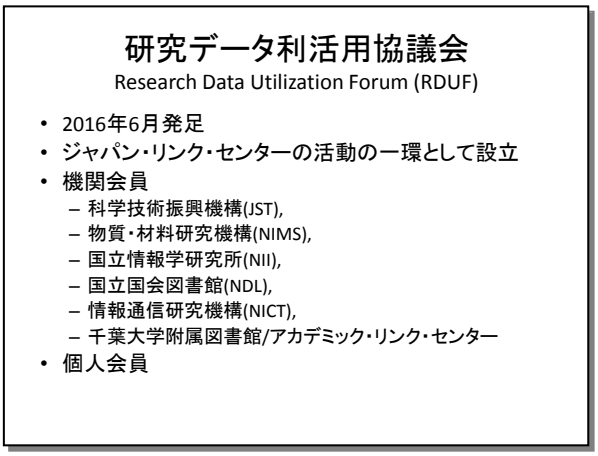
り小さな事務局があるだけです。ワーキンググループ・インタレストグループが勝手に集まって、勝手にオンラインで議論して、プレナリーミーティングのときにサマリーを報告する仕組みになっています。

ヨーロッパは RDA Europe、アメリカは RDA US があり、資金をもらって活動しています。それにオーストラリアを加えた 3 局が発足に尽力したと聞いています。でも、今影響力があるのはヨーロッパとアメリカの 2 局になっています。2016 年 3 月に、アジアで初の RDA のプレナリーミーティングが、本日の会場と同じビルの一橋講堂で、科学技術振興機構（JST）主催で開かれました。そこで、日本はどうしているのかとたくさん聞かれました。

研究データ利活用協議会

2016 年 6 月、研究データ利活用協議会が発足しました（図 12）。前身はジャパンリンクセンターの研究データへの DOI 登録実験プロジェクトというものです。DOI は文献に付けることが今までの多くの習慣でした。しかし、DOI の仕組みは文献に限らず使えるということは古くから気付かれていて、ヨーロッパでは 10 年前からデータサイテーションの活動があります。それが 5 年前から、DataCite という組織として活動しています。

その中で、研究データに永続的な識別子を付けるべきだという議論があり、日本はジャパンリンクセンターが主体となって、どうやってデータを見つけるべきかを一緒に考えようという実験プロジェクトを行いました。このときに初めて、データを扱う研究機関が一堂に会しました。図に挙げている以外にも、理化学研究所や海洋研究開発機構なども入っていただいて、一緒になって、データに DOI を付けることの意味は何か、どういう単位で付けるといいのか、どういう手順にしたらいいかを考えました。



（図 12）

それは 1 年で終わりましたが、初めて分野を超えて、まさにディシプリンを超えて、実務者レベルで顔を合わせて研究データについて議論することができた、非常に良い機会でした。それを母体にもう少し活動を継続したい、実験プロジェクトは終わったので別の名前を付けて広がりを持たせようということで「研究データ利活用協議会」と名乗ったのです。ジャパンリンクセンターのアウトリーチという位置付けで今は活動しています。

会員には機関会員と個人会員の二つがあります。機関会員は図に挙げている六つの機関です。上の四つはジャパンリンクセンターの共同運営機関です。それに情報通信研究機構（NICT）と千葉大学のアカデミック・リンク・センターが入っています。

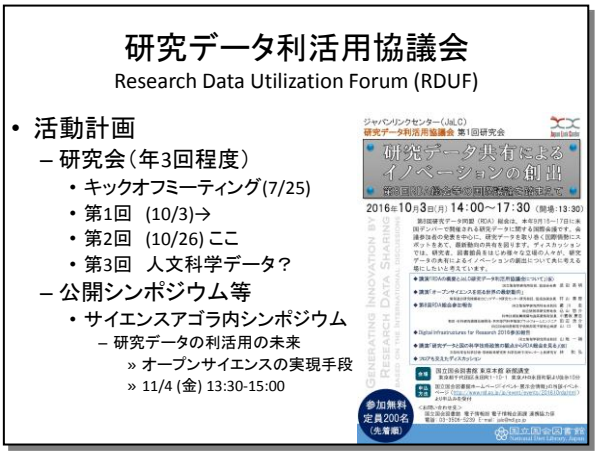
目的は、研究データに関する多様なセクター、特に実務者を集めることです。あまりお偉い方が顔を合わせる場所ではなく、実際にデータを取り扱う人たちが集まる場をつくりたいのです。集まって、研究データの共有と公開に関する課題を共有します。分野を超えると全く問題が違う、でも実は同じ問題もあるかもしれないということを共有したいのです。それを挙げるだけでも十分に価値があります。他には、うちではこういう技術を使っているということを共有できるとうれしいです。

その上で、研究データの共有と公開について、技術的・社会的な問題に関する議論を行いたいのです。自分たちの問題を解決するだけではなく、もっと広がりを持たせるためにどうしたらいいかということです。そして、RDA のような海外の関連組織とうまく情報共有やコラボレーションを図ることをミッションとし

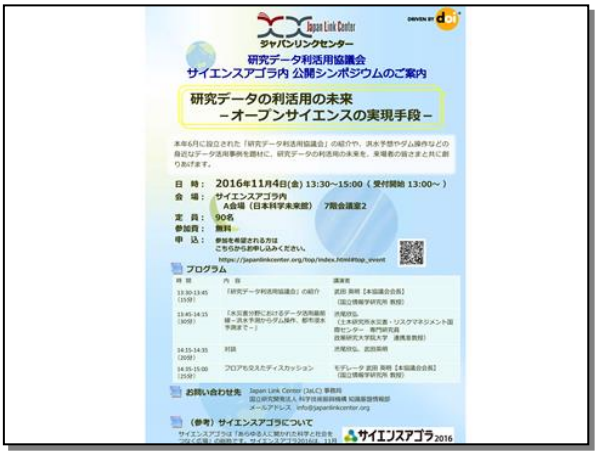
ています。

活動計画としては、研究会を年 3 回程度開きたいと考えています（図 13)。キックオフミーティングを 2016 年 7 月に開き、第 1 回の研究会は 10 月 3 日に国立国会図書館で「研究データ共有によるイノベーションの創出」という題目で行いました。ちなみに、研究会は、毎回担当が代われば興味が変わってくるので、機関会員の持ち回り担当制で行おうと思っていて、この回は国立国会図書館に担当になって企画していただきました。第 2 回は今回のセミナーと合同開催です。第 3 回はまだ決まっていませんが、人文科学データについてできればよいと考えています。もうじき公開できると思います。

11 月 4 日に、サイエンスアゴラ内で 1 時間半の一般向けシンポジウムを行います（図 14)。これは本当に一般向けで、研究データそのものに興味を持ってもらうことが目的です。今年は水の話題が多かったので水の専門家をお呼びして、実社会と研究データの話をしたいと思っています。



（図 13）



（図 14）

まとめ

オープンサイエンスは、ウェブの発展とともに変わりつつあります。オープンサイエンスの重要なステップとしての研究データ共有が、今のわれわれの焦点です。データ公開の原則（FAIR）があります。そして、横断的な対話が始まっているということも重要です。研究者だけで閉じるような議論では今や駄目ですし、大学や研究機関を飛び越えて、社会ともつながっているのが現在ではないかと考えています。

第2回 SPARC Japan セミナー2016 「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

ディスカッション 「インセンティブとデータマネジメントの今後のあり方」

蔵川　圭　（国立情報学研究所）
仲里　猛留　（情報・システム研究機構ライフサイエンス統合データベースセンター）
下田　研一　（長崎大学附属図書館）
南山　泰之　（国立極地研究所）
青木　学聡　（京都大学情報環境機構）
武田　英明　（研究データ利活用協議会/国立情報学研究所）

●**蔵川**　先生方がいろいろなテーマで、研究データ共有について情報を提供してくださいました。仲里先生からは、バイオインフォマティクスの連綿と連なる歴史的なデータベースの共有、またそれにつながる研究データについて、下田様からは、研究データ共有やオープンサイエンスと言われる前からずっと続けられている古写真の活動について、南山様からは、図書館の新しい試みとしての研究データの登録について、青木先生からは、京都大学の研究データ共有に関する今後の取り組み、それにまつわる世界の活動について、武田先生からは、特に世界における研究データのトレンド、または国内の活動として、Research Data Alliance（RDA）あるいは研究データ利活用協議会についての今後の方向性について発表を頂きました。

今回のパネルディスカッションは、それらの情報をベースにしながら、われわれ日本における研究データのオープン化というものを、図書館員と研究者の協同という観点から今後どのように推進していくことができるかを考えてみることが趣旨です。

研究データそのものをシェアする、共有するということは、広い観点からすると恐らく全員が得をするだろう、幸せになるだろう、イノベーションも生まれて、研究者だけでなく、人類全体が得をするだろうという総意の下で動きたい、動こうということで号令をかけて進もうとしているわけです。

個々の点からすると、単純に総意でもって研究データのオープン化に動いてしまったときには、不幸にも泣いてしまうとか、「それは私の利益とはつながらないのでやめます」ということがあったりするでしょう。そういうときに特に最初に問題になるのは、研究データそのものは研究者がつくっているので、研究者や科学コミュニティに対するインセンティブはどういうものが存在するのか、インセンティブがそもそもあるのか、あるとすると何か、何が不足で、今こういう状態になっているのかを考えなければいけないということです。

よく言われることとしては、新たな知見や価値が生み出せるということが一つのコミュニティにとってのインセンティブですし、または、今はありませんが、研究データを出した者にオープン化の成果に見合った処遇を与えるということもインセンティブでしょう。このようにインセンティブがなければ、最初に生み出されるデータ共有の点というものが存在しないということになるので、実はこれは非常に大事だということが確認できます。

それを受けて、では世界の中で研究データマネジメントというのは、どういうものが必要で、今どのくらいのことができていて、実際に何かに困っているのであればこれからどういうものをつくればいいかという話に展開すると思います。

このパネルディスカッションでは、主に二つのテーマを設けて、インセンティブには一体どういうものがあるのか、なぜそれが議論の点として必要なのかということを踏まえて議論をしていただきたいと思います。また、インセンティブを踏まえて、研究データマネジメントとしてどういうものが必要で、これからどういうものができてくるのか。そして、今回は SPARC Japan セミナーで、参加者の皆さまの中には図書館員の方も多いと思うので、図書館員としてそういう研究データマネジメントをするときに、研究者にどういうことが必要とされ、また自分たちはどういうことができるのかを明らかにしていきたいと思います。

まず、インセンティブですが、今回の企画は World Data Center for Geomagnetism, Kyoto（地磁気世界資料解析センター京都）で活動されている能勢先生が主導されました。研究データを実際につくっている側として、研究者にとって研究データの作成とは研究生活の中でどういう位置付けなのかということと、あまりデータをつくっていない人からするとよく分からないので、研究者のインセンティブとは何なのかということも含めてご紹介いただきたいと思います。能勢先生はパネリストではないですが、そこから始めたいと思います。

●能勢　私は京都大学理学研究科の地磁気世界資料解析センターという、地磁気のデータ、主に地球の磁場のデータをデータセンターとして扱っている部局にいます。大学の機関なので、研究と教育も大きなファクターを占めるのですが、それに加えて地磁気のデータを管理する、それからサービスをすることが一つの業務になっています。

そういう立場で、ご質問のありました研究データの作成についてですが、もちろんデータを作成したり、データを加工して他の研究者へ提供したりすることは喜びではあります。ただ、本業の教育・研究以上に時間が取られるので、使った時間、使ったコストに対して何らかの処遇を今後考えていけないのではないか、それはわれわれだけではなくデータセンターを管理されている方は必ず持っておられる意識だと思います。

かつ、先ほど仲里先生からもお話がありましたように、持続的にデータセンターを発展させていく、維持していくというのもやはり問題です。組織を維持していくためには人や費用が必要なので、何らかビジネスモデルと言うと少し大げさになりますが、そのようなものが考えられないのかというのが今回、このセミナーを企画させていただいた理由の一つです。

私が最初に趣旨説明で、馬の折り紙の写真をお見せしましたが、これが今、私が考えているインセンティブです（図 1）。ご講演を伺っていて感じたのは、局所的な成果に見合った処遇・研究費が得られるという即物的なものだけではなく、仲里先生から、データを公開することで、データが使われて世界を動かせるというような言葉がありましたが、まさにそのようなことが理想だと思います。そういう見方もあるのではないか、もしそれでコミュニティとして動いていけるのであれば幸せではないかと感じました。

インセンティブとは一体何かというご質問もありました。研究者がなぜ論文を書くかというと、もちろん出てきた結果をきちんと世に問うという理由もあると思いますが、自分の業績になるから書くわけです。それと同様な理由で、データセンターを維持していく、データサービスを恒常的に維持していくために使った時間、コストに対して、何らかのインセンティブを考えていく必要があるのではないでしょうか。

（図 1）

●**蔵川**　そもそも研究データ作成は研究者にとってやりたくないことなのか、やりたいけれども人に任せたくないことなのか、いろいろな位置付けがありますよね。研究論文を人に書かせているようでは研究者ではないので自分で書きます、でも後進のペーパーをチェックしながら書いていますなど、いろいろあると思います。そういう見方からするとどうなのですか。

●**能勢**　人によって感じ方は大きく違うと思うのですが、例えば生物や医学など、自分で実際に手を動かして実験をされて取られたデータは、もちろんその方にしか分かりませんし、その方が自分のデータとして管理する、処理するのは普通のことだと思います。データセンターだと、それ以外の他人が取ったデータの管理もあるので、そういうものの管理はその方に問い合わせたりして、時間が取られるのではないかと思います。

●**蔵川**　今のは地磁気の例だと思いますので、そこからするとかなりルーチンワーク化されていて、むしろデータを取ることが研究の中では当たり前であって、あえて自分から新しいデータを取ろうという意思が働かないような状態のルーチンワークさえあり、それをこなしていくという義務の中で今やっているという感覚ですか。

●**能勢**　もちろん自分自身の研究をするためのデータもフィールドに出て取るのですが、それ以外にも世界各地の地磁気観測所で取られたデータ、もしくは地磁気に限らず、われわれの分野であれば、他の観測所で、ある国の機関が取ったデータを集めてきてデータとして整形してサービスするので、その意味では自分が直接取ったデータではないけれども、データセンターとしてデータサービスを続けていく上で必要なデータの管理が出てくるという状態です。

●**蔵川**　そこにはインセンティブがちょっと足りないということですね。今は地磁気の例でしたが、ご発表者の中で、データをもうつくっているという方は仲里先生が挙げられると思います。そもそも研究データというものの研究する上での位置付け、価値、これはぜひ自分でやらなければいけないのか、人にできればやらせたいのか。また、やりたいという場合は、インセンティブがあるからそうなるのだと思います。その辺はご自分の分野の中ではどんな感じですか。

●**仲里**　私は最近はデータを取っているわけではないのですが、学生時代はデータを取っていました。今の議論を聞いていて思ったのですが、研究活動自体がデータを取ることなのです。測定しないと新しい知見が得られません。測定するということはもうデータが出てしまうのです。研究者だったら、測定したデータを煮たり焼いたり切ったり貼ったりというプロセスを経て、論文まで仕立てていきます。

私は今、日本 DNA データバンク（DDBJ）という生命科学系のデータを集めるところの隣にいて見ているのですが、「さあ登録するぞ」と言って登録しようとすると、少しパワーが要るのです。測定したそばからデータがデータベースに入って、簡単な操作をするだけで、「登録されました」となれば楽なのですが、さあ最後に登録ですとなると、今まで自分のハードディスクに死蔵されていた、もしかすると前のパソコンに入れっ放しだったかもしれないものまでひっくり返してきて、それをまとめて登録しなければいけないので、そこが研究者としてマイナスで、面倒くさいと思われているところではないかと思っています。

●**武田**　一研究者として見た場合、私も社会のネットワークデータのようなものを自分でつくることがあるのですが、先ほど偉そうにプレゼンしておきながら、自分のデータがどこに行ったか分からなくなっています。もう一回あのデータを使って別の研究者と共同研究をしようとなったときに、「ごめん、そのデータがない。前の共同研究者の手元にあるかもしれない」な

どと言って、やっと手に入ったりすることがあります。

結局、自分も研究データが大事だと言いつつ、自分の研究データをきちんと管理できないのです。理由はほとんど同じで、研究をやっているときには自分にとって大事なデータなので一生懸命やるのですが、終わった瞬間に、論文を書いた瞬間に、そのデータをどのハードディスクに置いたか、ほとんど記憶から飛んでしまうということはよくある話で、これは研究者あるあるではないかと思います。

●**蔵川**　青木先生は学内で研究データを出せ、出せと言っている立場、システムをつくっていたという立場ですが、そういう反応はありましたか。

●**青木**　まだ試行段階なので、あまり研究データを出せと宣伝ができていないのです。それぞれの研究者に対して、「あなたのところはデータをどうやって保存しているのですか」とインタビューして回ってつるし上げている状態ではないという状況です（笑）。

個人的なお話をすると、私はシミュレーションが専門なので、20 テラぐらいハードディスクを持っていて、そんなものを保存しろと言われてもできようがないということが一つあります。

今は、コンピューターサイエンス系では研究データをどう保存するかはまだかなり模索状態ですが、バイオの実験系では、きちんと実験計画を立ててそれに従ってデータを取りなさいということを事細かに最初に教育されるという話を聞いています。それに従ってデータの蓄積が進むということであれば、研究のプロセスが自動的にデジタル化され、ラボラトリー・研究室の中できちんとアイデンティファイできる形で整理されているというルールが構築されるのであれば、それはデータマネジメントの第一歩になるのではないかと思っています。

●**蔵川**　ルール化されればみんなやりますよということでしょうか。RDA のインターナショナル・データ・ウィークが 2016 年 9 月にありました。RDA には何回か出席したのですが、そこに科学技術データ委員会（CODATA）も併設されていて、データのインセンティブについて語り合うセッションがありました。データ共有はこれからどんどん拡大していく方向に向かいたいのですが、現状ではデータ共有のインセンティブがありません。

「あなたはこれをしなければ失職します」と言われたらやるかもしれませんが、そうでない限り、実質自由ですと言われた瞬間に、インセンティブがなければ研究者は誰も動きません。では、どんなインセンティブが大事かを皆さんで議論しましょう、もしそれを研究データを出す側として受け入れられるのであれば、それを仕組みとしてつくってくださいというようなことが議論されていたのです。

能勢先生にしても、これ以上何のインセンティブもなければ私は何もしませんという意見も含めて企画されたと思うのですが（笑）、どんなインセンティブがあれば「私は動きます」となるのでしょうか。

●**能勢**　難しいですね。何か物事を動かすためにはインセンティブは非常に重要で、特に内的・外的とよくいわれますが、研究者自身が、データを公開することにより、自分のデータが学術の発展や人類の知に貢献するという意識を持ち、内的なインセンティブによって公開してくことが理想的な状況ではあると思います。

データベースの維持には非常に手間が掛かります。CODATA でも、データベースを維持していくビジネスモデルについての議論があり、インセンティブはそれにも関わってくると思うのです。インセンティブは外的なものもある程度必要かなと思います。その際にどういう仕組みが可能なのかということが、現在自分でも考えているテーマです。

自分のデータを出すことによってたくさんの方に使われて、学術の発展、流れをリードしていけるのだという内的インセンティブもあると思いますが、投入したコストに対して見返りがあるようなインセンティブ

をつくっていけないかと考えています。

●**蔵川**　非常に奥ゆかしく表現されるので分かりにくい気がしなくもないのですが、仲里先生、インセンティブについてどうですか。

●**仲里**　先ほども言いましたが、私は DDBJ というデータを集めるところの隣にいて、人のデータが集まってくるのを見ているわけです。DDBJ はデータを集めることが仕事で、私たちはそれをリサイクルしてユーザーに使ってもらうことが仕事です。そのために、自己紹介でお見せしたような検索エンジンをつくっているのですが、検索エンジンをつくることは、研究者的には別にプラスにならないのです。それに対して何か論文を書くと初めて研究者として評価されますが、データベースをつくりました、ツールをつくりましただけでは、今までは研究者として評価されてきませんでした。

でも、今はデータジャーナルなどが出てきて、そういうところに、つくったデータベースそのものについて、「こんなものをつくりました」ということを記述・報告できるようになり、それがちゃんと他の論文と一緒にリファーできるようになってきていて、前よりは良くなったのかなという気はしています。

また、データをデータベースに入れることは、自分のデータに客観的に見てどのような価値があるのか評価してもらえる土俵に乗るということでもあります。具体的に言うと、私が手伝ったデータベースでも、「当然このデータはオープンだろう」と考えていたら、いきなり横から知財部が出てきて、「それはすごく重要なデータですからオープンにしないでください」と頼まれることも若干はあるのです。

オープンにするかクローズにするかは別として、データをデータベースにためるということは、評価してもらえる土俵に乗るということなので、「あなたのデータが知財として売り出せるかどうか分かるもしれないので、入れてみませんか」ということはインセンティブになるのではないかと個人的には感じています。

●**武田**　研究者がデータをつくるインセンティブにどのようなものがあるかというメニューは既に分かっていると思います。それは各分野でウエートが違うと思いますが、データサイテーション、データのオーナーシップ、マンダトリーなどです。逆に言うと、メニューにないような、今までにないインセンティブが降って湧くとは思えません。

掛けるコスト・労力とのバランスが取れればインセンティブになるし、取れなければインセンティブにならないわけです。コストを下げることは研究者単独ではできないので、大学・研究機関、インフラ側で頑張らなければいけないと思います。研究者のインセンティブは、それほど頭を悩ませて、今さらひねくり出すようなものではないのです。データサイテーションの仕組みやデータジャーナルというものも一応できたので、最低限の仕組みはできたと思っています。

そこで問題になるのは、今までのアーティクルと同じようにやればうまくいくのかというと、うまくいかないであろうということです。論文は読めば良いか悪いかが分かります。ところが、研究データになるとそれが簡単ではありません。そのデータが本当に価値がある良いデータなのか、クオリティが高いデータなのかは、データそのものを見ても分からないのです。それがむしろ、研究データに関わるかなり本質的な問題点ではないかと思います。

たとえサイテーションの仕組みができても、データジャーナルができても、それはある種の評価システムではあるのですが、本質的に中身を見られない、中身を評価できないというところがやはり論文と違います。それをどうしようという答えは私にもありません。

FAIR 原則で言うと、Re-usable（再利用できる）の項目で、「データは由来を付けるべき」とされていますが、まだ良い解決法はないのではないかと感じます。むしろ他の分野で、データのクオリティに関して何か活動があればご紹介いただければと思います。

●**仲里**　私は次世代シーケンサーのデータを扱っていますが、小さくても数ギガで、大きいとテラぐらいになってしまいます。前はダウンロードするのに一晩かかるくらいで、そういうデータを輸送するのに何がいいか測定した人がいて、ハードディスクを宅急便で送り付けるのがいいという結論が出たぐらいのデータでした。

今、検索エンジンをつくっていると、「じゃあ、どれを使えばいいの」と言われるのです。私たちはそれに対して二つの取り組みをしています。一つは、データクオリティをチェックするプログラムがあるのでそれにかけることです。そうするとリユースするときにあらかじめクオリティチェックがかかっているので、それを見てこれは使えるなと思って、落とします。もう一つは、データを登録することと論文を出すことは独立なので、論文がきちんと出ているデータはこれですというマークを付けることです。そうすると、論文が出るぐらいクオリティがいいようなデータだろうとユーザーは分かる、そのような仕組みを入れています。

●**南山**　今、データジャーナルなどの品質管理の問題が少し出たと思うのですが、Nature の「Scientific Data」などでは、品質管理というほどかは分かりませんが、一つ一つバリデーションはどうやった、メソッドはどうやったという項目がきちんと書かれています。これが書いてあるのだから、きちんとしたデータだろうというような、形式面から品質を担保するような取り組みは行われていると思います。

●**蔵川**　品質という話が出ました。青木先生、死蔵されたデータという話もありましたが、死蔵されたデータの品質についてコメントはありますか。

●**青木**　実験系の場合はほとんどが失敗データですよね。何らかのはずみで一度掘り出して、何かを見つけたというときは追試して、もう一度確認を取るわけです。死蔵されたデータというのは、研究者自身しか気が付かないこと、あるいは測定条件などを残していますが、そこから書き漏らした何かが入っているようなデータだと思うのです。

少し話がずれてきましたが、研究データをオープンにすることには、何か心理的な抵抗があるのです。例えば、ストレージの容量が足りないのにこれをずっと残しておいていいのか、データを取ったけれどどんな名前を付けようか、（仮）のような名前のままデータリポジトリに上げるわけにはいかないといった心理的な抵抗と物理的制約によって、自分がやっている実験そのものも上手にマネジメントできない状況があります。それで、何かの弾みで再利用しようとしてもできないというところがあるのかなと思います。

そういうところをうまくサポートできるようなパーソナルな研究データマネジメントをして、そのうちの一部をアーカイブとしてパブリッシュできるというようなサイクルが描けたら、研究者自身にとっては非常に有用なのではないかと思います。

その場合は物理的制約はできる限り外したいのです。アップロードは 1 日 1 回までとか、あなたが使えるファイル容量はこれぐらいとか、自分のハードディスクのバックアップも取れないような量しかもらえないという制約はできる限り外すことが、インフラを提供する、気持ち良くサービスを使ってもらう側のやるべき内容ではないかと考えています。

●**蔵川**　結構いい感じで、研究者にとっての研究データマネジメントについてスライドしていただきました。仲里先生は分野としてもかなりされていますよね。どんなことが今問題になっているのか、青木先生が言われたような話は既に終わっているのか、これからもやりたいのか聞かせてください。

●**仲里**　まさに同じ問題に直面しているのではないかと思います。次世代シーケンサーのデータはとても大きいので、例えばアメリカの NCBI（国立生物工学情報センター）などでは集めているのですが、一度、予

算が付けられないからやめるというアナウンスが出たことがありました。その後やはりやることになったのでよかったのですが。やはりデータを持って運用する側は、ディスクを食って、いくらディスクがあっても足りないという状況に直面していて、そうすると要らないデータは消そうかとか、非可逆圧縮で、クオリティが悪いところは切って圧縮するといった話がこの界隈では出てきます。

●**蔵川**　今度は図書館にとっての研究データマネジメントの話を下田様に振りたいと思います。長い間古写真を展示されていて、研究データという観点からしても何かマネジメントの必要性はありますか。

●**下田**　古写真について言えば、研究データと言えるレベルのデータは扱っておらず、あくまでもメタデータの一要素としてのデータを扱ってきました。だからこそ研究者の間で共有できました。研究者はもともと、これぐらいの古写真に関する解説は研究成果にならないと思っていたと思います。むしろ、そういうものをたくさん集めて、それを自分が通覧することで、そこから一歩踏み込んで自分の研究成果を出していくというスタイルだったのではないかと思いますので、あまり研究データそのものを図書館で扱ってきたということはありませんでした。

●**蔵川**　南山様に聞きたいのですが、既に研究データについて、新しい IUGONET という組織でいろいろやろうとしている、その経験から言って研究データ管理はどんな感じですか。

●**南山**　研究データ管理そのものは、研究プロセスを通じて研究データがどのように生成されていくかという、全体のもっと大きな話だと思うのですが、今回はその中でメタデータに関するところを取り出して研究データ管理にコミットしはじめたというスタンス、位置付けで考えています。

図書館員が研究データ管理全体にどこまで手を出せるかというのはこれからの話だと思うのですが、研究データ管理にはいろいろなプロセスがあります。そもそも管理のトレーニングをしなければいけない、あるいは実際にメタデータをつくるときに検証まで手伝えるのか、メタデータをつくっても実データの保存は誰がやるのか、そのようなところがあると思います。

こちらからはやれることをとにかく提示していく形ですが、（協同に対する）私の感覚としては、研究者の方々が研究データ管理をどこまでやりたいか、その中で図書館員に何を求めてくれるかです。分野によっては、これは業績として認められるし、きちんと若手の育成に使っているのでそんな手数は不要ですというところもあるかもしれませんし、そうではないところもあるかもしれません。まだ全くやっていないのでぜひ力を借りたいと言ってくれるところもあればうれしいというスタンスです。これで回答になったでしょうか。

●**蔵川**　また私が長々質問すると申し訳ないのですが、図書館から見たときの研究環境をつくるということと、研究者から見たときの研究環境をつくるということは、実は違うことが多いのです。研究者の人が研究環境をつくっているというときに、何をやっているのか聞くと、「学生にどれだけ論文を読ませるか、学生にどれだけ実験をさせるかという研究室の環境をつくっている」と答えます。例えば、本の購入をしたり、学習の場を与えたりするような、図書館でやっているようなことは、「あれはインフラであって、研究環境ではないですよね」とまで言うのです。

一方で図書館は、「私たちは根本的な研究環境をつくっています」とよく言われるのですが、そういうギャップは感じましたか。あるとしたらどんな感じですか。

●**南山**　おっしゃるとおりのギャップがずっとあると感じています。私は今は研究所勤務なので、まだ研究

者の方々と話す機会が多く、近しい立場にあるとは思うのですが、前に東京大学にいたときや、一般の大学図書館では、図書館は図書館として独立していて、「図書館の」仕事をやっている人たちというスタンスがあり、研究者は使えるところだけ使ってくださるという感じで、そもそも研究としての枠組みとだいぶずれてしまっていると感じています。

（図書館からの研究支援が）研究者一般に対する取り組みとしてある以上は、研究室の環境を整えることには全く影響しないというか、関われないような立場だと思うのですが、今後関わっていくのであれば、そもそも研究のプロセスを知った上で図書館の人がそこに何ができるかという視点で取り組んでいかなければいけないのだろうことが、今回の IUGONET の動機の一つです。

（機関リポジトリで実データを扱うこととの関係について、）インフラを別に図書館の方で持ちたいと思って IUGONET との協同をやっているわけではなく、IUGONET でできている研究環境の中で図書館員が役に立つスキルとは何か、ということをまず検証したいのです。その上で、今まで既存でやってきたものにプラスアルファでできるものがあれば提供したいというスタンスで考えるべきだと思っています。

●**蔵川**　下田様は長い間、研究とは無縁の世界で活動してきたと言われていますが、そういうギャップのようなものはありますか。

●**下田**　古写真以外ではあまり経験していないので、ギャップというのはよく分かりません。長崎大学で古写真の研究に関わる先生方は複数おられたのですが、それぞれが写真研究のプロということではなく、いろいろな別の研究分野を持っておられたので、本当の研究フィールドの深いところで古写真に関わるというよりは、むしろ自分の専門分野から見たときに古写真をどんなふうに利用できるかということでした。

研究者との間で流れているデータは、研究上それほど深いものではなかったと思います。むしろ、新聞記事になったり、街に出かけていって古写真展をしたり、テレビに出たり、本を出したりするのに使われていました。図書館が研究者に期待されたことは、書いた新聞記事の原稿をひととおりチェックすること、本を出版するときの多少編集的な仕事などであって、サイエンスだとか、そういうものではなかったように思います。むしろ一般の人たちに公開するというところだったので、一緒にできてきたのではないかと思います。

●**蔵川**　ちょうど研究者と図書館員のコラボレーションという観点からいろいろなストーリーをご紹介いただいたのですが、南山様にまた聞きたいことは、IUGONET を通して、研究者のニーズが何だと理解して、それにどれだけ応えたのか、応えられたのかということです。本質的なディスカッションに向かっていきたいと思います。研究データマネジメントではハードディスクがふんだんにあればいいなど、情報システムには関係あるけれど、図書館員との絡みでは難しい点があったので、その辺の接点という意味で。

●**南山**　研究者が困っていることに対して、図書館はどのように貢献できるかという視点から私が今考えている答えの一つが、スライドでもご紹介したネットワークの活用とメタデータ運用です。

研究者のニーズに対してどれだけ応えられたかというと、私としては、作業が減って役には立つのだろうなというスタンスで関わってきているのですが、実際にやった人、IUGONET 側の人に聞くのが面白いのではないかと思います。梅村（宜生）さん（名古屋大 ISEE）、いらっしゃいますか。僕は役に立ちましたか？（笑）

●**梅村**　非常に役に立っております。というのは、研究の現場では死蔵化されつつあるデータ、つまり、まだメタデータが付与されていないデータがまだまだたくさんあります。では、研究者サイドではどういう状

況かというと、「そんなところ構っていられないよ」という、手を付けられない状況なのです。

そんな中で南山さんから、「ちょっとやらせてください」とトライアルのアプローチを頂いて、われわれの期待以上に成果を出していただきました。われわれの期待以上に図書館側ができそうだという考えを持っています。そういう意味では非常に役に立っています。

今回はトライアルで体制を組ませていただきましたが、実は内々で、オフィシャルパートナーとしてやったらどうなのかという意見もちらちら出ています。現場と図書のサイドでそういう体制が組めればいいなと思います。

まだまだ現場では死蔵されつつあるデータがあるのですが、まさに蔵書管理ですから、図書館の人はそういったものの整理整頓が得意だと思うのです。「現場はこういうデータを持っていますよね。図書に任せてください。メタデータをつくりますよ」というリーダーシップをとってぜひやってもらえればと思っています。

●**南山**　どうもありがとうございます。大変うれしいです。協同の話はまた別として、今のでお答えになったでしょうか。

●**蔵川**　役に立ったということで、トンネルの遠くの方に光が見えている感じがしてはいるのですが、まだコラボレーションしていない研究者からすると、「本当に図書館員にできるのですか、私の仕事の何ができるのですか」という意見を言う人が多いのです。仲里先生はまだコラボレーションされていないと思うので、図書館員に何を任せればいいのかという意見はありますか。

●**仲里**　個人的な意見というよりは全体的に聞いて思ったことですが、同じ話を何度もしていますけれど、DDBJ では生命科学のデータ、例えば遺伝子配列や次世代シーケンサーのデータを集めています。「集めています」と言いますが、それは研究者が Excel なりのフィールドを埋めてサブミットするのです。すると、DDBJ にアノテーターという人がいて、それを眺めて、きちんと書けているのかどうかチェックします。例えば、これは絶対にオスのデータなのにメスになっているとすると、はねます。

そういうプロセスは、南山さんがおっしゃったメタデータをきちんと付けるというプロセスとまさに同じではないかと思っています。職業は違いますが、結局やっていることは同じなので、生命科学のアノテーターが回している仕事を図書館の人が同じように回しているということを今日強く感じました。

●**南山**　今聞いていて、発表で話し損ねたことを 1 点思い出しました。実際に研究者の方々でデータベースを運用されていても、自分の専門分野以外の、超高層物理の中でも幾つか細かく分野があって、例えば地磁気のデータ、宇宙系のデータなどがあるのですが、当然、研究者間でも（細分化された）自分の分野でなければ分からない。どのようにそれ（アノテーション）を対応するのか聞くと、「担当の人（PI）に聞いてやっている」と話されています。「担当の人に聞いてメタデータをつくるのであれば、それを図書館の人がやっても同じ仕事ができると思います」という殺し文句で今回のお話を進めてきたのですが、そういう関わり方も図書館員としてはありなのではないかと思います。図書館員は「人と人をつなぐのが専門性である」と表現されている方もいらっしゃるので。

●**蔵川**　青木先生は今度全学システムをつくろうと言われていますが、図書館員とのコラボという点で、何か意見はありますか。

●**青木**　全学のプロジェクトで意見交換を行うまで、私は恥ずかしながら、図書館の方々の、オープンアクセスやオープンデータなどデジタルデータに対する力の入れようを実は全く知りませんでした。ここ数カ月

いろいろ話をしていて、図書館の方が情報を扱うことに関する専門性を有していて、その成果の一つとして各機関リポジトリが存在するのだという認識に至っています。ですから、情報を上手にアノテートして整理してもらうということについて、今後図書館の方とコラボレートする機会は存分にあると思っています。

ただ、図書館側のヒューマンリソースが限られている状況で、山のようにある研究データをどう切ったり貼ったりするかは非常に難しい問題です。南山さんが今回、一例としてデータキュレーションを採用されたということですが、いろいろなところにデータキュレーションは必要なので、データキュレーションのスキルを持った人を、図書館の専門職員、学生、研究者への教育といったものとセットにして、リテラシーの底上げができるような体制ができればそれがベストではないかと考えています。

●**蔵川**　武田先生、研究データ利活用協議会の中で、図書館とのコラボレーションについて何かありますか。

●**武田**　研究データ利活用協議会はどちらかというと、データを直接扱う人をまずターゲットにするものなので、それほど図書館を強く意識はしていません。しかし今回、キックオフの時点で千葉大学のアカデミック・リンク・センターに入っていただきました。アカデミック・リンク・センターは自分自身がデータをつくるというよりは、教員などからもらっているデータをキュレーションするような立場にあるので、そういう意味ではそういうメンバーがいないわけではありませんが、DOI を付けるというスタンスで言うと、あまりそこに重きを置いていなかったところはあります。

でも、今日の議論を聞いていて、そこはやはり重要だなと思いました。結局、見えないデータを見える化するためには、豊富なメタデータを付けなければいけません。単に形式的なメタデータだけではなく、中身まで入ったメタデータを付けなければ、データの中身のクオリティが保証できないからです。それはデータの価値を高めますが、それをやることは研究者にはものすごく大変です。そこでキュレーションはある程度別のセクターが担ってくれたら、先ほどの研究者側のコストを下げることにつながります。

図書館、キュレーションセンターのようなところ、アノテーション専門のセクターが、実は同じ役割をしていることを理解して、役割分担が明示的にできると、研究者も、図書館も、キュレーションセンター的なところも、お互いにハッピーになれるのではないかと今の議論を聞いていて思いました。お互いに役割を認識し合うことができたら、もう少し前向きに進めそうな気がしてきました。

●**蔵川**　そろそろ時間ですので、会場を含めて質問をしていただける方、ご意見がある方はいらっしゃいますか。

●**フロア**　WDS-IPO（World Data System-International Programme Office）の渡邉です。私自身も研究データを取って論文を書いています。私や能勢さんが当てはまるようなデータセンター型研究者は、データアーカイブを大事にしていて、データアーカイブで自分の論文が書けるし、データベースもつくり、データ解析システムまでつくってしまいます。そうするとそれが共同研究になり、一般公開もしようという話になります。IUGONET などはまさにそれです。データベースがないと自分の研究が進まないというタイプの研究をしている人たちは、データの公開に関して比較的問題が少ないです。もちろんデータセンターをどう維持するか、金の問題、人の問題はありますが、それこそ WDS が取り組むべき問題で、これははっきりしています。

研究者が論文を書いて、そのデータを公開するのはある程度義務でしょう。裏付けデータを何も公開しないで、「書きました」ではおかしいわけです。やはりそれは何らかの形でアーカイブするというか、いつも見られるように公開しておくことが研究者の義務であり、モラルの問題です。学会やアカデミーなどの議論

によって、それは進められるでしょう。

一番問題なのは研究データです。先ほど失敗データとありましたが、時系列データ、例えば太陽面爆発のデータなどはいつ起きるか分からないので延々とデータを取っています。論文を書くときは、何かあったところのデータだけを使います。他の延々と取ったデータも使い道があるかもしれませんが、あまり興味のない人が多いので、放ってあるデータも結構あるのです。

ところが、何もない、延々と取ったデータが大事だという人もいるのです。自分が考えていることとは違う使い方をされる可能性もあります。私は電波天文学で、太陽から来る風をリモートセンシングで見ていたのですが、先駆けて公開しました。「そんなデータを公開していいのか」と言われ、「それで誰かに論文を書かれたらどうするのか」とさえ言われました。でも、われわれは、公開することによって共同研究が広がるのではないか、自分たちが考えづらいような使い方をする人もいるのではないかと考えて公開を始めたのです。

そこで大事なことは、自分が研究に使わなかったデータでも、それを自分は持っている、こういうデータがどこにあるという所在情報だけでも公開してもらうことです。そうすれば新しい共同研究ができるかもしれません。データを取った人に断りもせず、名前も入れずに研究論文を出す人などいないのです。それが研究者としての一つのインセンティブにもなるのではないでしょうか。

あまり最初から、エラーも含めて全部保存しなさいというのは無理ですが、とにかくデータの所在情報だけでも公開する、それが手始めではないでしょうか。そういう印象を持ちました。図書館あたりにもそういうところからまずやってもらうと、研究者としてもインセンティブになると思います。

●**武田**　インセンティブはいろいろあり得ますが、今は人的コスト、実際のコストも含めて、掛かるコストをいかに下げるかということにかかっているのではないかと思います。

今、渡邉先生がおっしゃったことも、取っているデータが、研究者がわざわざ触らなくても、そのままダークアーカイブなどに取り込まれるのであれば研究者はやると思います。黙っていてもできてしまうのだから、イエスかノーかと聞けば、きっとイエスと言います。そこで「これはぜひ隠しておきたい」と言う人は、今時は多分いないでしょう。

ただ、「あなたがコピーして、どこそこに置いて、名前を付けて登録してください」と言われると、優先度が低いからやらないのです。小さなインセンティブ、もしかすると 1 万人に 1 人の研究者が利用するかもしれないという程度なら、普通の人はそんなことをしないのです。でも、それをボタン一つでできるようにして、コストが下がればやるでしょう。

インセンティブの種類よりもむしろ、それを実現するためのコストをいかに安くできるかに尽きるのではないかというのが僕の思うところです。

●**蔵川**　最後に南山さん、いいインセンティブの話がありましたが、その中に図書館はどうやって入っていきますか。

●**南山**　今の武田先生のお話を受けると、公開コストをいかに下げるかということにまず図書館がコミットしていくのが良いように思います。いろいろな研究者に聞いても、（図書館業界の）大御所の方に聞いても、メタデータの仕事はもともと図書館が専属でやってきた、それは専門性であると言うので、それならばメタデータの仕事を図書館で引き受けて、図書館に全部まとめることで機関としてのコストを下げることで、貢献できればと考えます。

（実際の運用では）ボタン一つで図書館の人にデータが飛び、図書館の人はそのメタデータを必要な分だけつくる。その中から公開したいデータは公開する、というような理想的なワークフローがつくれるといいのではないかと思います。この辺はインフラがあって

こその話ですが、インフラがあれば図書館としては絡みやすいですし、絡むべきだと思っています。

●**蔵川**　ありがとうございました。時間になりましたので、登壇者の方々へ拍手をもってこのパネルを終わりにしたいと思います。