SPARC Japan

ISSN 2188-8248

国際学術情報流通基盤整備事業
（SPARC Japan）
年報

平成28（2016）年度

国立情報学研究所

目次

巻頭言 1
１　概要 2
１．１ 第５期の活動概要 2
１．１．１ 第５期の基本方針 2
１．１．２ 第５期事業計画 2
１．２ 平成２８年度活動 3
１．２．１ SPARC Japan セミナー 3
１．２．２ 海外動向調査 3
１．２．３ arXiv.org コンソーシアム事務局 4
１．２．４ SCOAP³ 支援 4
１．２．５ CLOCKSS 支援 5
１．２．６ 高エネルギー物理学分野の情報サービスに係る国際連携協定への対応 5
１．２．７ 平成２７年度 SPARC Japan 年報の発行 5
１．２．８ 論文公表実態調査 5
２　委員会等開催記録 6
２．１ 国際学術情報流通基盤整備事業運営委員会 6
２．２ SPARC Japan セミナー企画ワーキンググループ 6
３　委員名簿 6
３．１ 国際学術情報流通基盤整備事業運営委員会 6
３．２ SPARC Japan セミナー企画ワーキンググループ 7
４　SPARC Japan セミナーの記録 8
５　総合年表 9
６　刊行物一覧 20
６．１ 国際学術情報流通基盤整備事業(SPARC Japan)年報 20
６．２ SPARC Japan ニュースレター 20
６．３ SPARC Japan セミナー資料 20
７　資料　ニュースレター再掲 23

巻頭言

平成 15 年から始まった本事業も、平成 28 年度で第 5 期の一年目を迎えました。この一年の活動を年報としてまとめました。SPARC Japan セミナーの内容をまとめたニュースレターもすべて再録しております。

第 5 期は第 4 期（平成 25～27 年度）の活動を継承し、オープンアクセス（OA）を推進し、学術情報流通の更なる発展に取り組むことを基本方針としています。第 4 期から第 5 期に引き継がれる課題としては(1) 国際的な OA イニシアティブとの協調、(2) 学術情報流通にかかわるアドボカシー活動、(3) オープンサイエンスへの活動スコープの拡大、(4) オープンアクセスに関する基礎的情報の把握があります。

そして第 5 期の初年次から、国際的な OA イニシアティブとの協調を活発に行うことになりました。特筆すべきは、CERN（欧州原子核研究機構）の SCOAP[3] の責任者である Dr. Salvatore Mele を招聘し、日本の高エネルギー物理学分野の研究者を対象としたシンポジウム開催を支援したこと、そして arXiv.org について、Member Advisory Board に日本コンソーシアム代表として京都大学引原隆士図書館機構長に出席いただき、グランドデザインや財政面の施策に対し、日本コンソーシアムの意見を発言いただいたことです。国際的イニシアティブをとりまく外的要因の変化や財政面等の課題にわが国も迅速に対応するために、国立情報学研究所が中心となって意見をとりまとめ、窓口を担当していることの意義はさらに高まっています。

また、アドボカシー活動として、「オープンサイエンス時代の文献とデータの流通」を年間テーマとした、SPARC Japan セミナーを 3 回開催しました。第 1 回セミナーでは、文献のグリーン OA とゴールド OA と日本としての対応について講演・ディスカッションが行われました。第 2 回セミナーでは、研究データについて、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行いました。第 3 回セミナーでは第 1 回・第 2 回のまとめも兼ねて、「オープンサイエンス再考」とし、オープンサイエンスの実現にむけ理念から一歩踏み込んだ、より実践的な活動に関する情報提供を行いました。

第 5 期も「大学図書館と国立情報学研究所との連携・協力推進会議」の下、大学図書館コンソーシアム連合（JUSTICE）やオープンアクセスリポジトリ推進協会（JPCOAR）等との協調を一層強化し、学術情報流通の発展に向けて参加意識を強める方向でアドボカシー活動を継続的に行っていく所存です。

SPARC Japan は学術情報流通の更なる発展のため様々な活動を行います。引き続き、皆様の厚いご支援をお願い申し上げます。

平成 29 年 4 月 1 日

国際学術情報流通基盤整備事業委員長

安達　淳

１ 概要

１．１ 第５期の活動概要

１．１．１　第５期の基本方針

第５期においても、第４期の活動を継承し、国内外の OA イニシアティブや関係組織と連携し、オープンアクセス等を推進し、学術情報流通の更なる発展に取り組むことを基本方針とする。

特に米国 SPARC と連携し、日本のオープンアクセス活動を国際的に発信する。オープンアクセス等の推進にあたっては、まずその課題を把握することに努めると共に、「大学図書館と国立情報学研究所との連携・協力推進会議」の下の機関リポジトリ推進委員会および大学図書館コンソーシアム連合等との協調を一層強化し、学術情報流通の発展に向けて参加意識を強める方向でアドボカシー活動を継続的に行っていく。

１．１．２　第５期事業計画

SPARC Japan　第５期の事業は次の 4 つを柱として計画することが、平成 27 年度第 3 回国際学術情報流通基盤整備事業運営委員会で決定した。

⑴国際的な OA イニシアティブとの協調

国際イニシアティブに参画し、日本の窓口としての役割を果たすとともに、その活動・成果のアピールに努める。これらも含めて、国際的な動向を注視し、必要な対応を行う。

⑵学術情報流通にかかわるアドボカシー活動

「大学図書館と国立情報学研究所との連携・協力推進会議」等の組織と連携しつつ、オープンアクセスやオープンサイエンス、学協会出版の国際流通に係るアドボカシー活動を継続して実施する。

⑶オープンサイエンスへの活動スコープの拡大

研究成果のオープンアクセス、イノベーションの基盤となる可能性を秘めたオープンデータ、加えて高等教育の基本的構成要素の再考を迫るオープンエデュケーションなどへの関心の高まりにあわせて、理工学分野だけではなく、人文科学・社会科学分野の動向等に関して適時の情報提供を実現する。また、大学図書館におけるオープンサイエンスの取組み、研究データの管理等への関与について、戦略的な検討を行う。

⑷オープンアクセスに関する基礎的情報の把握

第４期に引き続き、オープンアクセスに関する基礎的情報を把握するために実態調査等を行う。各大学・研究機関の研究戦略を考える上で、データを集め分析するために、図書館が一定の役割を果たすことも検討する。

１．２ 平成２８年度活動

１．１の事業計画のもと、平成 28 年度は次のプロジェクトを実施した。

１．２．１　SPARC Japan セミナー

アドボカシー活動として、SPARC Japan セミナーを 3 回実施した。セミナーワーキンググループ（以下、「WG」という。）を立ち上げ、WG メンバー全員で年度を通した全体テーマ、セミナー各回のテーマの割り振りを検討し、各回に担当者を置いて企画・実施した。セミナー終了後、SPARC Japan ニュースレター（以下、「NL」という。）を発行・ウェブ配信した。

第 1 回 平成 28 年 9 月 9 日	オープンアクセスへの道 http://www.nii.ac.jp/sparc/event/2016/20160909.html	NL 第 30 号 （2016 年 11 月）
第 2 回 平成 28 年 10 月 26 日	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 http://www.nii.ac.jp/sparc/event/2016/20161026.html	NL 第 31 号 （2017 年 1 月）
第 3 回 平成 29 年 2 月 14 日	科学的知識創成の新たな標準基盤へ向けて： オープンサイエンス再考 http://www.nii.ac.jp/sparc/event/2016/20170214.html	NL 第 32 号 （2017 年 4 月）

１．２．２　海外動向調査

下記の国際会議等に参加し、情報収集を行った。

・COAR (Confederation of Open Access Repository) Annual meeting 2016
（4 月 11-13 日 Wien、Austria）に NII 山地准教授を派遣した。

・ORCID (Open Researcher and Contributor ID) Board Meeting
（5 月 18-19 日 Toronto、Canada）に NII 武田教授を派遣した。

・CRIS2016 (The 13th International Conference on Current Research Information Systems)
（6 月 8-11 日 St Andrews、Scotland）に機関リポジトリ推進委員会協力員の大学図書館員 1 名を派遣した。

・OR2016 (The 11th Annual International Conference on Open Repositories)
（6 月 13-16 日 Dublin、Ireland）に機関リポジトリ推進委員会協力員の大学図書館員 1 名と NII 山地准教授を派遣した。

・RDA (Research Data Alliance) 8th Plenary Meeting
（9 月 15-17 日 Denver、USA）に NII 武田教授と NII 蔵川特任准教授を派遣した。

・PIDapalooza (PID：Persistent ID)
（11 月 9-10 日 Reykjavík、Iceland）に NII 武田教授を派遣した。

・IDF Strategic Meeting
（1 月 17-19 日 Barcelona、Spain）に NII 武田教授を派遣した。

・Open Access に係る海外機関の調査
（1 月 25-27 日 Universitätsbibliothek Regensburg、Bayerische Staatsbibliothek、Max Planck Digital Library）に大学図書館コンソーシアム連合（JUSTICE）作業部会委員の大学図書館員 2 名を派遣した。

１．２．３　arXiv.org コンソーシアム事務局

arXiv.org は物理学のプレプリントサーバで、コーネル大学図書館が運用している。2014 年 12 月に 100 万論文を突破し、新規登録数は年間 9 万件、ダウンロード数は年間約 8,100 万件である。利用件数上位の機関による財政支援があり、2013～2017 年の「arXiv 会員制プログラム」には 24 カ国 183 機関が参加している。

日本においては NII が各大学の意思確認を取りまとめて支援してきた。利用回数順位 300 位までの大学に会員申請の意向調査を行った結果、2017 年 3 月末現在の会員数は 14 機関である。

2014 年 4 月にコーネル大学から、日本の会員館でコンソーシアムとして参加することについて打診があり、会員に確認の後、コンソーシアム契約に切り替えを行った。会費はコンソーシアム価格のため 10%減となった。

さらに 2015 年度には、本コンソーシアム名を日本研究図書館コンソーシアム（英文名：Consortium of Japanese Research Libraries: Coordinated by National Institute of Informatics (NII)、英文略称： NII Japan Consortia）とした。また、引原隆士 京都大学図書館機構長が本コンソーシアムの代表に就任し、2016 年から arXiv.org の Member Advisory Board（MAB）に参加し、9 月 12-13 日開催の会議に出席した。

１．２．４　SCOAP³ 支援

2014 年から開始した SCOAP³ について 2015 年も参加意向および連絡先を確認し、日本の大学図書館からの拠出金を、日本のナショナル・コンタクト・ポイントである NII がとりまとめて支払った。なお、日本からは 2016 年 3 月末現在 34 機関が参加している。

SCOAP³ により OA 化される論文数は着実に増えている。拠出額を算定する根拠とした 2011 年には当該 10 誌で 3,552 論文出版されたが、2014 年中に 4,280 論文が、2015 年には 4,477 論文が発行された。2014-15 年の平均 APC(Article Processing Charge)は 1,100Euro であり、これは Gold OA 誌の APC の水準から見ても低くなっている。

SCOAP³ の OA 論文を収載した SCOAP³ リポジトリとその API が既に公開されているが、リポジトリは DOI 付与、CC BY ライセンス表示、XML 形式で公開されており、テキストマイニング、データマイニングが可能である。なお、2016 年 3 月には OA 論文が 1 万件に達したことが報じられた。

本年度は、SCOAP³ 担当責任者である Salbatore Mele 氏（CERN）を招へいし、SCOAP³ に関する講演・意見交換会（12 月 7 日 NII、12 月 8 日京都）を開催した。また、SCOAP³ 評議会日本代表委員である NII 安達教授が Executive Meeting（10 月 14-15 日、2017 年 3 月 20 日）及び Governing Council Meeting（2017 年 3 月 23-24 日）に出席した。

１．２．５　CLOCKSS 支援

CLOCKSS（Controlled Lots of Copies Keep Stuff Safe）は、全世界の研究者のためにデジタル資源（Web ベースの学術文献等）の長期保存を実現することを目的とし、アーカイブとそれを運営するコミュニティを構築して、コンテンツが出版社から提供されなくなった場合にアーカイブ上のコンテンツを広く利用できるようにするなどの取り組みを行っている。

日本においては 2013 年から NII が各大学の意向確認および年会費支払いのとりまとめを行っている。なお、日本からは 2016 年 6 月末現在で 99 機関が参加している

１．２．６　高エネルギー物理学分野の情報サービスに係る国際連携協定への対応

高エネルギー物理学分野に関する国際連携協定を欧州原子核研究機構（CERN）、高エネルギー加速器研究機構（KEK）及び国立情報学研究所（NII）の 3 機関が締結した。CERN が提供する高エネルギー物理学分野の文献情報サービスである INSPIRE データベースを中心にデータキュレーションの実務に関する共同事業を行うために、平成 28 年 3 月から約 1 年間、大学図書館員 1 名を CERN に派遣するとともに、双方の関係者によるテレビ会議を複数回行った。

１．２．７　平成２７年度 SPARC Japan 年報の発行

平成 27 年度の活動状況をまとめ、平成 28 年 10 月に発行した。また、平成 26 年度 SPARC Japan 年報について、平成 29 年 1 月に英語版を発行した。

１．２．８　論文公表実態調査

平成 27 年度に、大学図書館コンソーシアム連合（JUSTICE）と連携し、JUSTICE の下に論文公表実態調査チームを設け、我が国における論文公表と APC（Article Processing Charge）の実態調査を開始した。平成 28 年度は引き続きフォローアップを行った。

２ 委員会等開催記録

２．１ 国際学術情報流通基盤整備事業運営委員会

開催日	議題
第１回 平成 28 年 9 月 6 日	1. 前回議事要旨（案）について 2. 平成 28 年度 SPARC Japan セミナー進捗状況について【報告】 3. 論文公表実態調査チームの活動状況について【報告】 4. 国際連携の状況について【報告】 5. Open Access に係る海外機関の調査について 6. その他
第２回 平成 29 年 3 月 14 日	1. 前回議事要旨（案）について 2. 平成 28 年度 SPARC Japan 事業報告【報告】 3. 論文公表実態調査チームの活動状況について【報告】 4. オープンアクセスに係る海外機関の調査について【報告】 5. 国際連携の状況について【報告】 6. 平成 29 年度 SPARC Japan 活動計画について【審議】 7. 平成 29 年度 SPARC Japan セミナー企画ワーキンググループの設置について【審議】 8. その他

２．２ SPARC Japan セミナー企画ワーキンググループ

開催日	議題
キックオフミーティング 平成 28 年 6 月 9 日	1. 年間テーマ確定 2. スケジュール・分担検討 3. その他

３ 委員名簿

３．１ 国際学術情報流通基盤整備事業運営委員会

氏名	所属・役職	備考
逸村　裕	筑波大学 図書館情報メディア系 教授	１号委員 （研究教育職員）
今井　浩	東京大学大学院 情報理工学系研究科 教授	１号委員 （研究教育職員）
森　重文	京都大学高等研究院 院長・特別教授	１号委員 （研究教育職員）
土屋　俊	独立行政法人 大学改革支援・学位授与機構 教授	１号委員 （研究教育職員）
倉田　敬子	慶應義塾大学 文学部 教授	１号委員 （研究教育職員）
小林　富雄	高エネルギー加速器研究機構 教授 国際連携推進室 室長	１号委員 （研究教育職員）

深貝　保則	横浜国立大学大学院国際社会科学研究院 教授	１号委員 （研究教育職員）
尾城　孝一	東京大学附属図書館 事務部長	２号委員 （大学図書館関係者）
荘司　雅之	早稲田大学図書館 事務部長	２号委員 （大学図書館関係者）
林　和弘	科学技術・学術政策研究所 科学技術予測センター 上席研究官	３号委員 （学会の関係者）
安達　淳	国立情報学研究所 副所長	１号委員 （研究教育職員） 委員長
酒井　清彦	国立情報学研究所 学術基盤推進部 次長	２号委員 （大学図書館関係者）

３．２ SPARC Japan セミナー企画ワーキンググループ

氏名	所属・役職
梶原　茂寿	北海道大学附属図書館 学術システム課 システム管理担当係長 （国立情報学研究所推薦）
小野　亘	東京学芸大学 教育研究支援部 学術情報課長 （これからの学術情報システム構築検討委員会推薦）
林　和弘	科学技術・学術政策研究所 科学技術予測センター 上席研究官
能勢　正仁	京都大学大学院理学研究科附属地磁気世界資料解析センター 助教
西脇　亜由子	明治大学 学術・社会連携部 図書館総務事務室 （大学図書館コンソーシアム連合運営委員会推薦）
坊農　秀雅	情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター（DBCLS） 特任准教授
南山　泰之	国立極地研究所　情報図書室 主任 （機関リポジトリ推進委員会推薦）
蔵川　圭	国立情報学研究所　学術コンテンツ課 特任准教授

4 SPARC Japan セミナーの記録

平成 28 年度 SPARC Japan セミナー実施記録

回	実施日	テーマ	企画 WG（所属） ◎主査、五十音順	講師（所属） 登壇順	形態	参加人数
1	平成 28 年 9 月 9 日（金） 13:00～16:40 （NII 12 階会議室）	オープンアクセスへの道	梶原　茂寿　（北海道大学附属図書館） 蔵川　圭　（国立情報学研究所）◎ 西脇　亜由子（明治大学） 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）	○土屋　俊　（大学改革支援・学位授与機構） ○尾城　孝一　（東京大学附属図書館） ○安達　淳　（国立情報学研究所） ○荘司　雅之　（早稲田大学図書館） ○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） ○〔モデレーター〕山本　和雄（琉球大学附属図書館）	オープン	119
2	平成 28 年 10 月 26 日（水） 13:00～17:15 （NII 12 階会議室）	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	蔵川　圭　（国立情報学研究所） 能勢　正仁　（京都大学大学院理学研究科）◎ 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） 南山　泰之　（国立極地研究所）	○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） ○〔モデレーター〕蔵川　圭（国立情報学研究所）	オープン	112
3	平成 29 年 2 月 14 日（火） 10:30～17:20 （NII 12 階会議室）	科学的知識創成の新たな標準基盤へ向けて： オープンサイエンス再考	小野　亘　（東京学芸大学） 梶原　茂寿　（北海道大学附属図書館）◎ 能勢　正仁　（京都大学大学院理学研究科） 林　和弘　（科学技術・学術政策研究所） 南山　泰之　（国立極地研究所）	○Ron　Dekker（European Commission （DG Research & Innovation）） ○北本　朝展　（国立情報学研究所） ○小賀坂　康志（科学技術振興機構） ○小野寺　千栄（物質・材料研究機構） ○小野　雅史　（東京大学地球観測データ統融合連携研究機構） ○大向　一輝　（国立情報学研究所） ○梶原　茂寿　（北海道大学附属図書館） ○田村　峻一　（滋賀医科大学附属図書館） ○尾城　孝一　（東京大学附属図書館） ○〔モデレーター〕林　和弘（科学技術・学術政策研究所）	オープン	118
合計						349
平均						116

5 総合年表

年度	評議会　運営委員会	主催イベント	その他のイベント
平成 15 （2003）	06/25　第 1 回評議会		
	07/14　事業参画提案の募集開始	07/02　学協会向け事業説明会（於：日本教育会館）	
	08/01　第 1 回運営委員会		
	09/11　第 2 回運営委員会	08/19　事業説明会（於：東北大学　東北大学附属図書館との共催）	
	09/17　第 2 回評議会（事業参画提案決定） 09/17　記者発表		
	10/08　作業グループ合同会議		
			11/05　第 5 回図書館総合展フォーラム「SPARC／JAPAN：日本の国際学術コミュニケーションの変革」開催（於：東京国際フォーラム　国立大学図書館協議会・私立大学図書館協会主催）
			11/20　国立大学図書館協議会電子ジャーナルタスクフォース（生物系、物理系、医学系の購読交渉）
		01/21-29 Project Euclid 説明会（於：学術総合センター、東北大学、京都大学、名古屋大学）	
		02/23　SPARC/JAPAN 懇談会：参加学会への成果報告、新雑誌創刊構想説明（於：学術総合センター）	
		03/11　SPARC/JAPAN セミナー：生物系学協会誌をめぐる学術情報流通体制の将来　-UniBio Press のめざすもの-（於：東京大学附属図書館）	
	03/22　第 3 回運営委員会		
	03/23　第 3 回評議会		
平成 16 （2004）	05/28　第 1 回運営委員会		
	06/02　第 1 回評議会		
	06/07　参画提案募集開始		
			07/01　国立大学図書館協会総会ワークショップ：「国際学術情報流通基盤整備事業の活動」（於：大阪大学コンベンションセンター）
		07/07　学協会向け事業説明会（於：学術総合センター）	
	09/15　第 2 回運営委員会		

	09/22 第 2 回評議会（事業参画提案選定） 10/14 作業グループ合同会議 03/07 第 3 回運営委員会 03/10 第 3 回評議会	09/27 Project Euclid 懇談会（Project Euclid への参画に関する技術的打ち合わせ、DPubS についての説明） 10/15 シンポジウム：学会出版と学術コミュニケーション活動の変革～SPARC/JAPAN を事例として～（於：広島大学中央図書館　広島大学図書館、国立情報学研究所、国立大学図書館協会国際学術コミュニケーション委員会共催） 10/19 緊急シンポジュウム「どうする日本の学術誌！」（於：早稲田大学総合学術情報センター　（社）高分子学会、（社）電子情報通信学会、東北数学雑誌編集委員会、（社）日本機械学会、（社）日本金属学会、（社）日本動物学会、（社）日本分析化学会、日本哺乳動物卵子学会、日本哺乳類学会、国立情報学研究所共催） 11/05 OUP 懇談会「Open Access の現状について」 11/25 第 6 回図書館総合展フォーラム「学術コミュニケーションの最先端：オープン・アクセスとセルフアーカイブ」（於：パシフィコ横浜） 01/27 ワークショップ「電子ジャーナルのビジネスモデル構築と学術出版をめぐる動向」（於：日本教育会館） 03/24 シンポジウム「SPARC の現状と課題：学術雑誌・機関レポジトリ・オープン・アクセス」（於：早稲田大学）	10/19-20 Project Euclid DPubS Conference に参加（於：コーネル大学）
平成 17 （2005）	06/06 第 1 回運営委員会 06/08 第 1 回評議会	05/19 SPARC/JAPAN 連続セミナー第 1 回「Nature の歴史、今、未来を語る－Nature の編集方針」 06/29 SPARC/JAPAN 連続セミナー第 2 回「電子投稿査読システムとは何か－今、日本で使えるシステム」JST「J-STAGE 投稿審査システム」 07/09-10 電子ジャーナル利用の現在と未来に関するクローズド・ワークショップ（於：経団連ゲストハウス、静岡） 07/15 SPARC/JAPAN 連続セミナー第 3 回「オープン・アクセスの理念と実践－研究者・図書館・学術誌」	06/21-22 JISC International Solutions for the Dissemination of Research に出席、討議（ロンドン） 07/07-08 エルゼビア・ライブラリ・コネクト・セミナー2005「ユーザーを理解する（Understanding Users）」（於：京都・東京、エルゼビア・ジャパン主催、NII 後援）

		07/20 UniBio Press の挑戦 － 学会の新しいビジネスモデル（於：茨城大学　茨城大学図書館主催）	09/15 山口大学図書館セミナー2005「日本の電子ジャーナルの現況」学術コミュニケーションの今日：SPARC/JAPAN の挑戦（於：山口大学学術情報機構図書館主催）
		09/22 SPARC/JAPAN 連続セミナー第 4 回「電子ジャーナルをどう作成し、どう公開するかー学協会、企業の試み」	09/16 京都大学学術情報・電子ジャーナルシンポジウム「「大学における学術情報資源の整備–電子ジャーナル時代の学術コミュニケーションの変革–」（於：京都大学　京都大学附属図書館と NII の共催）
	10/13 第 2 回運営委員会 10/26 第 2 回評議会（事業参画提案選定）	10/06 SPARC/JAPAN 連続セミナー第 5 回「主体である研究者は何をすべきかー電子ジャーナル時代を迎えて」（於：つくば国際会議場社団法人日本動物学会第 76 回大会関連シンポジウムとの共催）	
		11/24 SPARC/JAPAN 連続セミナー臨時回「Journal of Bioscience and Bioengineering WEB 投稿審査システム」説明会・デモンストレーション	
		11/30 SPARC/JAPAN 連続セミナー第 6 回「第 7 回図書館総合展フォーラム COUNTER プロジェクト：オンライン利用統計の国際標準について」（於：パシフィコ横浜）	
		12/01 COUNTER プロジェクトに関するクローズド・ワークショップ	12/09 長崎大学附属図書館連続講演会第二回講演会「学術情報発信の新しい動向」：SPARC/JAPAN の活動と課題（於：長崎大学附属図書館主催）
		12/12 SPARC/JAPAN 連続セミナー第 7 回「日本の学術誌における英文校閲を考える」	
		01/31 SPARC/JAPAN 連続セミナー第 8 回「学術情報流通をめぐる最近の動向と技術標準：Google Scholar、 CrossRef、 OAI-PMH、 etc.」	
	02/15 第 3 回運営委員会 02/24 第 3 回評議会	02/10 SPARC/JAPAN 連続セミナー第 9 回「SPARC/JAPAN 選定誌によるラップアップセッション」	
平成 18（2006）		06/30 SPARC Japan 連続セミナー2006 第 1 回「海外商業出版社から見た日本の学術コミュニティ」	03 米国研究図書館協会（ARL）と MOU を締結
		07/26 SPARC Japan 連続セミナー2006 第 2 回「e-Journal の販促とライセンシング：海外の状況と海外市場における日本ジャ	07/03-04 エルゼビア・ライブラリ・コネクト・セミナー2006 「From “ Search ” to “ Find ” ～ 必要な情報を見つけやすい環境づくり ～」（於：東京・大阪、エルゼビア・ジャパン主

		ーナルの展望」	催、NII 後援）
	09/08　第 1 回運営委員会	09/05　Sally Morris 氏講演会「Introducing ALPSP」 09/29　SPARC Japan 連続セミナー2006 第 3 回「Web 投稿審査システムの検証：ビフォーアフター」 11/02　SPARC Japan 連続セミナー2006 第 4 回「大学図書館から学術出版社への要望：COUNTER を例にして」 11/20　第 8 回図書館総合展フォーラム「TRANSFERー出版社間のジャーナル移行に伴う問題点とその解決に向けて」（於：パシフィコ横浜） 12/14　SPARC Japan 連続セミナー2006 第 5 回「著作権：学会の権利、著者の権利、機関リポジトリへの対応」 12/18-19　「デジタル巨人の肩の上に立つ」機関リポジトリ、e-サイエンス、および学術コミュニケーションの将来に関する国際シンポジウム（於：都市センターホール）	
	01/30　第 2 回運営委員会	01/30　SPARC Japan 連続セミナー2006 第 6 回「e-Journal の販売とライセンシング(2)- 販売のプロに学ぶ成功の秘訣」 03/05　SPARC Japan 連続セミナー2006 第 7 回「計量書誌学からジャーナル・論文のパフォーマンスを測る」	
平成 19 （2007）			05/15　UniBio Press セミナー「生物系ジャーナル の挑戦－より明確に、より広く、その情報を伝えるために」（於：学術総合センター　UniBio Press 主催） 05/17　UniBio Press セミナー「生物系ジャーナル の挑戦－より明確に、より広く、その情報を伝えるために」（於：京都大学附属図書館　UniBio Press 主催）
	06/12　パートナー誌合同会議 07/19　第 1 回運営委員会	07/17　SPARC Japan 連続セミナー2007 第 1 回「計量書誌学からジャーナル・論文のパフォーマンスを測る-2-」	08/05-11 41th IUPAC　（International Union of Pure and Applied Chemistry）化学会議出展（トリノ） 08/20-22　234th ACS 秋季大会出展（ボストン）
		10/02　SPARC Japan 連続セミナー2007 第 2 回「Web 投稿審査シス	

	12/14　パートナー誌と大学図書館の合同会議「SPARC Japan パートナー誌のコンソーシア購入に向けて」 02/29　第 2 回運営委員会	テムの検証パート 3 稿より良いシステムを目指して―」 11/02　SPARC Japan 連続セミナー2007 第 3 回「メタデータ Publishing の現在―電子ジャーナル主体の製作・出版に必要なもの」 11/09　第 9 回図書館総合展プレゼンテーション「日本の英文トップ電子ジャーナルの挑戦―図書館総合展プレゼンテーパートナー誌からの提案―」（於：パシフィコ横浜） 01/17　SPARC Japan-ALPSP 特別セミナー（第 4 回 SPARC Japan 連続セミナー2007）「学術出版と学会 Journal Publishing and Scholarly Societies」 01/18　ALPSP トレーニングコース「Introduction to Journal Publishing」	11/07-09 第 9 回図書館総合展出展（於：パシフィコ横浜）
平成 20（2008）		04/22　SPARC Japan セミナー2008 第 1 回「研究成果発表の手段としての学術誌の将来」 06/24　SPARC Japan セミナー2008 第 2 回「学術出版と XML 対応-日本の課題」 07/10　SPARC Japan セミナー2008 第 3 回「韓国コンソーシアム事情 - 海外展開を目指して -」 09/02-03 RIMS 研究集会（第 4 回 SPARC Japan セミナー2008）「紀要の電子化と周辺の話題」（於：京都大学数理解析研究所 京都大学数理解析研究所主催） 10/14　SPARC Japan セミナー2008 （Open Access Day 特別セミナー）「日本における最適なオープン・アクセスとは何か？」	06/15-17 SLA（Special Libraries Association 米国専門図書館協会）年次総会出展（シアトル） 06/26　第 55 回国立大学図書館協会総会出展（於：東北大学） 07/13-15 中国化学会学術年会出展（於：天津） 08/17-19 236th ACS National Meeting & Exposition 出展（於：フィラデルフィア） 09/11-12 私立大学図書館協会総会出展（於：國學院大學） 09/16-20 2nd EuCheMS Chemistry Congress 出展（於：トリノ） 09/25-26 KESLI（Korean Electronic Site License Initiative）電子情報 EXPO での発表、出展（於：大田） 10/12-15 15th North American ISSX (International Society for the Study of Xenobiotics)) Meeting での広報（於：サンディエゴ） 10/27-30 ISAP2008（International Symposium on Antennas and

			Propagation）出展（於：台湾） 11/13-14 INFOPRO2008 プロダクトレビュー参加・出展（於：日本科学未来館）
		11/17-18 SPARC Digital Repositories Meeting 2008（於：ボルチモア SPARC、SPARC Europe、 SPARC Japan 共同主催） 11/25　SPARC Japan セミナー2008 第 6 回「IF を越えて-さらなる研究評価の在り方を考える」 11/27　SPARC Japan セミナー2008 第 7 回（第 10 回図書館総合展・学術情報オープンサミット 2008 フォーラム）「Open Access Update」 12/16　SPARC Japan セミナー2008 第 8 回「日本で使える電子ジャーナルプラットフォーム」	12/17-20 EUC2008（International Conference On Embedded and Ubiquitous Computing）出展（於：上海）
	12/24　第 1 回運営委員会	01/22-26 Project Euclid と数学系ジャーナルの打ち合せ（於：国立情報学研究所、京都大学、東京工業大学） 02/13　SPARC Japan セミナー2008 第 9 回「SPARC Japan 選定誌がやってきたこと」	
	03/10　第 2 回運営委員会 03/27　パートナー誌合同会議 03/27　第 3 回運営委員会		03/16-20 APS March Meeting 2009（米国物理学会年会）出展（於：ピッツバーグ）
平成 21 （2009）	10/05　第 1 回運営委員会	06/25　SPARC Japan セミナー2009 第 1 回「研究者は発信する－多様な情報手段を用い、社会への拡がりを求めて」 08/04　SPARC Japan セミナー2009 第 2 回「非営利出版のサステイナビリティとは－OUP に学ぶ」 09/08-09 RIMS 研究集会（第 3 回 SPARC Japan セミナー2009）「数学におけるデジタルライブラリー構築へ向けて－研究分野間の協調のもとに」 09/17　日本動物学会大会（第 4 回 SPARC Japan セミナー2009）「ZS プロジェクトについて」 10/20　Open Access Week （第 5 回 SPARC Japan セミナー2009）「オープンアクセスのビジネスモデルと研究者の実際」 11/11　第 6 回 SPARC Japan セミナー2009 （第 11 回図書館総合展学術情報オープンサミット 2009 フォーラム）	11/25　第 9 回アジア太平洋生物化学工学会議（APBioChEC 2009）

	03/23　第 2 回運営委員会	「NIH Public Access Policy とは何か」 12/11　第 7 回 SPARC Japan セミナー2009「人文系学術誌の現状ー機関リポジトリ、著作権、電子ジャーナル」 02/02　第 8 回 SPARC Japan セミナー2009「Marketing to Libraries Worldwide」 02/03　ALPSP トレーニングコース「Effective Journals Marketing」	に SPARC Japan の化学系パートナー誌が出展 12/03-04 DRFIC 2009　デジタルリポジトリ連合国際会議 2009（於：東京工業大学　DRF（デジタルリポジトリ連合）と NII の共催）
平成 22（2010）		06/23　第 1 回 SPARC Japan セミナー2010「学会の仕事とその経営を知る」 07/06　第 2 回 SPARC Japan セミナー2010「ジャーナル出版―海外学会の現状」 08/24　第 3 回 SPARC Japan セミナー2010「図書館の仕事を知る -学術雑誌の購読と利用-」 09/16　第 4 回 SPARC Japan セミナー2010（RIMS 研究集会）「数学におけるデジタルライブラリー構築へ向けて」 09/24　第 5 回 SPARC Japan セミナー2010（社団法人 日本動物学会 第 81 回大会）「日本の学術情報流通　10 年後を見据えて」 10/20　第 6 回 SPARC Japan セミナー2010 Open Access Week「日本発オープンアクセス」 11/08- 09 SPARC Digital Repositories Meeting (デジタルリポジトリ	08/19　International Congress of Mathematicians (国際数学者会議) に出展 08/22- 26 American Chemical Society (ACS) 2010 年 秋季大会に出展 08/29-09/02 3rd EuCheMS Chemistry Congress（第 3 回ヨーロッパ化学会議）に出展

	03/16　第１回運営委員会	会議）（於：ボルチモア SPARC、SPARC Europe、 SPARC Japan 共催） 12/10　シンポジウム 「大学からの研究成果オープンアクセス化方針を考える」（於：東京大学　国立大学図書館協会と NII の共催） 01/14　第 7 回 SPARC Japan セミナー2010 「著者 ID の動向」 02/03　第 8 回 SPARC Japan セミナー2010 「世界における”日本の論文／日本の学術誌”のインパクト」 03/08　TIB（ドイツ技術情報図書館）/ ZB MED（ドイツ医学中央図書館）/ NII（国立情報学研究所）MoU 締結記念 講演会 「ドイツと日本における学術情報流通基盤の未来」（於：学術総合センター　東京ドイツ文化センターとの共催）	
平成 23 （2011）	10/06　第１回運営委員会	10/28　第１回 SPARC Japan セミナー2011 Open Access Week 「OA 出版の現況と戦略（ジャーナル出版の側から）」 12/06　第 2 回 SPARC Japan セミナー2011 「今時の文献管理ツール」ワークショップ 01/31　第 3 回 SPARC Japan セミナー2011「学術情報流通の新たな展開 ― 研究者・学会とオープンアクセス ―」 02/10　第 4 回 SPARC Japan セミナー2011「学術情報流通の未来を切り開く ― 電子ジャーナルの危機とオープンアクセス ―」 02/29　第 5 回 SPARC Japan セミナー2011「OA メガジャーナルの興隆」	08/28-09/01 American Chemical Society (ACS) Fall 2011 National Meeting & Exposition（第 242 回米国化学会秋季大会）に出展 (於：デンバー) 09/04-09 14th Asian Chemical Congress 2011 (14 ACC)（第 14 回アジア化学会議）に出展(於：バンコク) 10/26　2011 Open Access Korea(OAK) Conference での発表 (於：ソウル)

	03/27　第 2 回運営委員会	03/26　第 6 回 SPARC Japan セミナー2011「数学出版に関するワークショップ」（於：東京理科大学 Project Euclid 主催、日本数学会共催ワークショップ）	
平成 24 （2012）	12/10　第 1 回運営委員会 03/26　第 2 回運営委員会	05/25　第 1 回 SPARC Japan セミナー2012「学術評価を考える」 06/19　第 2 回 SPARC Japan セミナー2012「ジャーナルの発展をもとめて～プラットフォーム移築を中心に～」 07/25　第 3 回 SPARC Japan セミナー2012「平成 25 年度 科学研究費補助金（研究成果公開促進費）改革」 08/23　第 4 回 SPARC Japan セミナー2012「研究助成機関が刊行するオープンアクセス誌」 10/26　第 5 回 SPARC Japan セミナー2012「Open Access Week ー 日本におけるオープンアクセス、この 10 年これからの 10 年」 12/04　第 6 回 SPARC Japan セミナー2012「オープンアクセスによって図書館業務はどう変わるのか～図書館のためのオープンアクセス講座～」 02/19　第 7 回 SPARC Japan セミナー2012「図書館によるオープンアクセス財政支援」	07/02-07 European Congress of Mathematics (ECM) に出展（於：クラクフ、ポーランド） 08/19-21 American Chemical Society (ACS) Fall 2012 National Meeting & Exposition（第 244 回米国化学会秋季大会）に出展 （於：フィラデルフィア） 08/26-30 4th EuCheMS Chemistry Congress（第 4 回ヨーロッパ化学会議） に出展（於：プラハ） 12/26-27 RIMS 共同研究　研究会にて基調講演（於：京都大学）
平成 25 (2013)		06/07　第 1 回 SPARC Japan セミナー2013「SPARC と SPARC Japan のこれから」 08/23　第 2 回 SPARC Japan セミナー2013「人社系オープンアクセスの現在」 10/25　第 3 回 SPARC Japan セミナー2013「オープンアクセス時代の研究成果のインパクトを再定義する：再利用と Altmetrics の現在」 12/19　第 4 回 SPARC Japan セミナー2013「今日の問題を解く、学術情報の受信と発信－Think Globally, Act Locally」 02/07　第 5 回 SPARC Japan セミナー2013「アジアを吹き抜けるオープンアクセスの風―過去、現在、未来」	08/06　第 1 回　OA ジャーナルへの投稿に関する調査ワーキンググループ開催 10/02　第 2 回　OA ジャーナルへの投稿に関する調査ワーキンググループ開催 12/04　SCOAP[3] と MOU を締結 01/27　RIMS 共同研究　研究会にて基調講演（於：京都大学） 03/02　COAPI Meeting へ参加 （於：カンザスシティー）

	03/24　第 1 回運営委員会		03/03-04 SPARC2014 Open Access Meeting 本会議への参加 (於：カンザスシティー) 03/13　第 3 回　OA ジャーナルへの投稿に関する調査ワーキンググループ開催
平成 26 (2014)	01/15　第 1 回運営委員会 03/19　第 2 回運営委員会	08/04　第 1 回 SPARC Japan セミナー2014「大学/研究機関はどのようにオープンアクセス費用と向き合うべきか―APC をめぐる国内外の動向から考える」 09/26　第 2 回 SPARC Japan セミナー2014「大学における OA ポリシー：日本版 OA ポリシーのモデル構築に向けて」 10/21　第 3 回 SPARC Japan セミナー2014「「オープン世代」の Science」 03/09　第 4 回 SPARC Japan セミナー2014「グリーンコンテンツの拡大のために我々はなにをすべきか？」	05/21-23 COAR(Confederation of Open Access Repository) 2014 Annual meeting への参加　(於：アテネ) 06/09-13 OR2014 (The 9th Annual International Conference on Open Repositories)への参加　(於：ヘルシンキ)
平成 27 (2015)	08/04　第 1 回運営委員会 11/30　第 2 回運営委員会 03/24　第 3 回運営委員会	09/30　第 1 回 SPARC Japan セミナー2015「学術情報のあり方 - 人社系の研究評価を中心に - 」 10/21　第 2 回 SPARC Japan セミナー2015「科学的研究プロセスと研究環境の新たなパラダイムに向けて - e-サイエンス，研究データ共有，そして研究データ基盤 -」 01/19　第 3 回 SPARC Japan セミナー2015「研究者向けソーシャルメディアサービスの可能性」 03/09　第 4 回 SPARC Japan セミナー2015「研究振興の文脈における大学図書館の機能」	04/15-16 COAR(Confederation of Open Access Repository) 2015 Annual meeting への参加　(於：ポルト) 05/18-20 ORCID-CASRAI Joint Outreach Conference & Codefest 及び ORCID Board Meeting への参加　(於：バルセロナ) 06/08-11 OR2015 (The 10th Annual International Conference on Open Repositories)への参加　(於：インディアナポリス) 11/03-06 ORCID Outreach Meeting & Codefest, November 2015 及び ORCID Board Meeting への参加(於：サンフランシスコ) 02/02-03 ORCID Board Meeting への参加　(於：ロンドン) 03/07-08 SPARC Meeting on Openness in Research & Education への参加　(於：サンアントニオ)

平成 28 (2016)	09/06　第 1 回運営委員会 03/14　第 2 回運営委員会	09/09　第 1 回 SPARC Japan セミナー2016「オープンアクセスへの道」 10/26　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」 02/14　第 3 回 SPARC Japan セミナー2016「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」	04/11-13 COAR (Confederation of Open Access Repository) Annual Meeting 2016 への参加　(於：ウィーン) 05/18-19 ORCID Board Meeting への参加　(於：トロント) 06/08-11 CRIS2016 (The 13th International Conference on Current Research Information Systems)への参加 (於：セントアンドリュース) 06/13-16 OR2016 (The 11th Annual International Conference on Open Repositories)への参加　(於：ダブリン) 09/12-13 arXiv.org Member Advisory Board Meeting への参加 (於：イタカ) 09/15-17 RDA (Research Data Alliance) 8th Plenary Meeting への参加　(於：デンバー) 10/14-15 SCOAP³ Executive Meeting への参加(於：ジュネーブ) 11/09-10 PIDapalooza (PID：Persistent ID) への参加 (於：レイキャビック) 12/07-09 SCOAP³ に関する講演・意見交換会への協力 (於：NII、京都) 01/17-19 IDF Strategic Meeting への参加　(於：バルセロナ) 01/25-27 Open Access に係る海外機関の調査 (於：レーゲンスブルク，ミュンヘン) 03/20 SCOAP³ Executive Meeting への参加(於：ジュネーブ) 03/23-24 SCOAP³ Governing Council Meeting への参加 (於：ジュネーブ)

６ 刊行物一覧

６．１ 国際学術情報流通基盤整備事業(SPARC Japan)年報

〔日本語〕

・国際学術情報流通基盤整備事業（SPARC Japan）年報 平成 27（2015）年度
http://www.nii.ac.jp/sparc/publications/annual/pdf/sparc_annual_2015.pdf

〔英語〕

・SPARC Japan (International Scholarly Communication Initiative) Annual Report FY2014
http://www.nii.ac.jp/sparc/en/publications/pdf/sparc_annual_2014-E.pdf

６．２ SPARC Japan ニュースレター

〔日本語〕

・SPARC Japan NewsLetter 第 30 号（2016 年 11 月）
http://www.nii.ac.jp/sparc/publications/newsletter/PDF/sj-NewsLetter-30.pdf

・SPARC Japan NewsLetter 第 31 号（2017 年 1 月）
http://www.nii.ac.jp/sparc/publications/newsletter/PDF/sj-NewsLetter-31.pdf

・SPARC Japan NewsLetter 第 32 号（2017 年 4 月）
http://www.nii.ac.jp/sparc/publications/newsletter/PDF/sj-NewsLetter-32.pdf

〔英語〕

・SPARC Japan NewsLetter No. 26, Nov. 2015
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter26E.pdf
・SPARC Japan NewsLetter No. 27, Nov. 2015
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter27E.pdf
・SPARC Japan NewsLetter No. 28, Feb. 2016
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter28E.pdf
・SPARC Japan NewsLetter No. 29, Mar. 2016
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter29E.pdf
・SPARC Japan NewsLetter No. 30, Nov. 2016
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter30E.pdf
・SPARC Japan NewsLetter No. 31, Jan. 2017
http://www.nii.ac.jp/sparc/en/publications/pdf/sj-NewsLetter31E.pdf

６．３ SPARC Japan セミナー資料

【第 1 回 SPARC Japan セミナー 2016】（平成 28 年 9 月 9 日）

「オープンアクセスへの道」

http://www.nii.ac.jp/sparc/event/2016/20160909.html

「趣旨説明」 蔵川　圭 （国立情報学研究所）

「オープンアクセスのあり方，グリーン OA とゴールド OA」

土屋　俊 （大学改革支援・学位授与機構）

「学術雑誌のキャッシュフロー転換の可能性を探る～JUSTICE/SPARC Japan 合同調査チームによる調査結果の概要～」
尾城　孝一　（大学図書館コンソーシアム連合(JUSTICE)/東京大学附属図書館）
「SCOAP³ による学術誌のオープンアクセスへの転換」 安達　淳　（国立情報学研究所）
「大学図書館におけるオープンアクセスの取組み」 荘司　雅之　（早稲田大学図書館）
「生命科学分野における研究者の投稿先雑誌選択趣向と OA への意味づけ」
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）

【第 2 回 SPARC Japan セミナー 2016】（平成 28 年 10 月 26 日）
「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」
http://www.nii.ac.jp/sparc/event/2016/20161026.html
「趣旨説明」 能勢　正仁　（京都大学大学院理学研究科）
「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）
「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）
「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）
「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）
「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）

【第 3 回 SPARC Japan セミナー 2016】（平成 29 年 2 月 14 日）
「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」
http://www.nii.ac.jp/sparc/event/2016/20170214.html
「開会挨拶/概要説明」 梶原 茂寿　（北海道大学附属図書館）
“Open Science in a European Perspective”
Ron Dekker　（European Commission（DG Research & Innovation））
「ディープラーニングとオープンサイエンス～研究の爆速化が引き起こす摩擦なき情報流通へのシフト～」
北本 朝展（国立情報学研究所）
「JST におけるオープンサイエンスへの対応（DMP 導入試行をはじめとして」
小賀坂 康志　（科学技術振興機構）
「材料科学分野におけるデータ利用のライセンスの考え方 - 図書館からの視点」
小野寺 千栄　（物質・材料研究機構）

「研究データ共有の理想と現実，そして実践可能性～地球環境分野の研究基盤に関する意識調査から～」
小野 雅史 （東京大学地球観測データ統融合連携研究機構）
「研究データを用いたサービスの調査・企画」 大向 一輝 （国立情報学研究所）
「A 班：機関内研究情報統合システム（IRIIS）の企画構想」
田村 峻一 （滋賀医科大学附属図書館）
「B 班：オープンデータ利活用総合基盤システム「ODEN」の企画」
梶原 茂寿 （北海道大学附属図書館）
「RDM トレーニングツールの紹介」 尾城 孝一 （東京大学附属図書館）

SPARC Japan NewsLetter
ISSN 1883-826X
NO.30 2016 年 11 月

SPARC Japan ニュースレターは、本号から、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 の活動をご紹介することとしました。

CONTENTS

■ SPARC Japan 活動報告
SPARC Japan 運営委員会

■ SPARC Japan セミナー報告
概要
参加者から
企画後記
ドキュメント
（講演・パネルディスカッション）

■ SPARC Japan 活動報告

SPARC Japan 運営委員会

CC BY

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/

■ SPARC Japan セミナー報告

CC BY

第 1 回 SPARC Japan セミナー 2016

「オープンアクセスへの道」

2016 年 9 月 9 日（金）　国立情報学研究所 12F 会議室　参加者：119 名

本セミナーでは、ゴールド OA とグリーン OA の最近の動向を踏まえつつ，オープンアクセス化のあり方と今後の日本の取り得るべき戦略について，オープンアクセスに関わってきた実務家や研究者をお迎えして議論しました。

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/2016/20160909.html）

概要

学術雑誌の購読価格高騰を原因としたシリアルズクライシスへの対応として，大学図書館はコンソーシアムを組んで大手海外出版者を相手にビッグディールという包括的な契約方式を実現させ，同時に，機関リポジトリによる雑誌論文のオープンアクセス化とポリシー策定による組織内のルール化を行ってきた。

現在のオープンアクセス化の方法は，購読料支払いを前提に出版者サイトからアクセスコントロールされた論文公開に対し，機関リポジトリに出版された論文のコピーや著者最終稿を掲載して無料でアクセスさせるグリーン OA(open access)と呼ばれる方法と，APC(article processing charge)という論文加工料を著者が出版者に支払って，出版者が出版者サイトから論文を無料でアクセスさせるゴールド OA と呼ばれる方法が主流となっている。

2012 年に英国政府がゴールド OA を主体としたフィンチレポートを実現する意向を表明して以降，欧州ではオープンアクセスのあり方をめぐってグリーン OA とゴールド OA を 2 極とする議論が沸き起こっている。

英国内では，JISC は購読料と APC が混在したハイブリッド誌のゴールド OA 化を促し，機関の総支払額を抑えるオフセットシステムを導入したパイロット契約を実施している。オランダは自国の研究者が出版したシュプリンガーの 2000 誌の雑誌論文をゴールド OA 化することを決定し，エルゼビアとも同様の OA 化を約束した。欧州委員会は，“オープンサイエンスポリシープラットフォーム”を推進する中で，ゴールド OA 化への道を探ることを含めている。素粒子物理学分野では，SCOAP3 という名のもと，分野の主流な雑誌を対象にゴールド OA 化するビジネスモデルを開発し，実施している。

一方で，セルフアーカイブを推奨し続けるスティーブン・ハーナッドは，見境なくゴールド OA 化を進めようとしている現状に警鐘を鳴らし，COAR はユネスコと共同で同様の声明を発表している。また，様々なステークホルダーは，急進的なゴールド OA 化への懸念をブログ上に発信している。

このような状況の中で，我が国はどう振る舞うべきか。主要な学術雑誌の大半を欧州や米国のプラットフォームに依存している我が国は，欧州や米国の OA 化への流れを避けて通ることを許されない。我が国にとっての，機関リポジトリの存在意義を見出してきたのと同様に，ゴールド OA 化へのあり方の議論とその反動としてのグリーン OA への回顧が必要である。

本セミナーでは，オープンアクセスに関心のある実務家や研究者が一堂に会し，オープンアクセス化のあり方と今後の日本の取り得るべき戦略を議論する。

参加者から

（大学/図書館関係）

・論文が増加する中でゴールド OA が進むと、APC の支払い額が増える一方なのではないかと思っていたのですが、JSPS へ持ち込まれたような案が実現すれば、一定の質のある論文量に歯止めがきいてよいと思いました。また図書館員が国際的なつながりをもつことの大切さにハッとさせられました。

・「図書館不要」「グリーンはもうムリ」など、ズバリとした話がつづき、今後何ができるのか考えなければいけないと思いました。

・今後、SPARC Japan にて対応方策が議論されるかと思いますが、パネルディスカッションで出てきた APC 実態把握については出版社に聞けばよいのでは、という部分が気になりました。当日は、出版社に対する応報としての面白さはあったのですが、これから SPARC Japan として APC 把握を具体的に進めていく際に、交渉相手である出版社に根拠数字を頼るのは、作戦としてよろしくないと思います。難しいとしても大学として独自に把握すべきであり、それは図書館員的な書誌方面からの把握でなく会計方面からの把握のほうが合理的だろうと思います。

（企業/その他）

・Redirection、flipping、APC の量的把握の困難さ等、論点が明確になりました。

・SNS が研究の中でどう活用されているのかが分かり、今後どう関わっていくべきかを考えるきっかけとなりました。

企画後記

☺ 長年にわたって日本の OA 活動を実質的に牽引したキーパーソン三傑を主軸に据え、いまこのタイミングでグリーン OA かゴールド OA かを問うことの必要性を具現化したセミナー企画。オープンサイエンスという地に足のつかないキーワードが叫ばれている傍らで雑誌論文を取り巻く契約交渉は着実に進められており、実務では待った無しの状況に違いない。大学図書館は、これまで取り組んで来た機関リポジトリをどのように生かすのか、ゴールド OA という枠組みにどのように対処すべきか、それぞれの担当者の頭の中に考慮すべき案件として喚起できたと思う。

蔵川　圭

（国立情報学研究所）

☺ 昨年度に引き続き、SPARC Japan セミナーの企画 WG を担当させて頂いております。「グリーンオープンアクセスとゴールドオープンアクセスの関係性やあり方について」今一度再確認することができ、企画する側としてもとても興味深いテーマでした。多くの方にご来場また同時配信での視聴を頂き、有意義なセミナーとなったのではないでしょうか。

梶原　茂寿

（北海道大学附属図書館）

☺ 自らオープンアクセス(OA)を推進してきた一研究者として、「生命科学分野における研究者の投稿先雑誌選択趣向と OA への意味づけ」と題して話しさせていただいた。投稿先雑誌の実例を紹介させていただいたのが、図書館業界の人には普段こういった話を聞かないようで、参考になったようである。今後はさらに踏み込んでオープンデータを推進すべく、活動していきたい。

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 SPARC Japan セミナーでは、今後のオープンアクセスの方向性、可能性をあらためて考えるよい機会となりましたが、企画 WG を通じて、グリーンとゴールドは対抗概念としてではなく相互補完的な概念として捉えた方が、議論が進む気がします。理想と現実という表現も不十分でしょうが、オープンアクセスの実現、継続、発展を考えていく上では、図書館側の立場として、安達先生から図書館員への叱咤激励のメッセージをしっかり受け止めなくてはと感じました。

西脇　亜由子

（明治大学 学術・社会連携部 図書館総務事務室）

SPARC Japan 事務局より　第 5 期（平成 28~30 年度）においても、第 4 期（平成 25~27 年度）の活動を継承し、国内外の OA イニシアティブや関係組織と連携し、オープンアクセス等を推進し、学術情報流通の更なる発展に取り組むことを基本方針とします。「大学図書館と国立情報学研究所との連携・協力推進会議」の下の機関リポジトリ推進委員会および大学図書館コンソーシアム連合等との協調を一層強化し、学術情報流通の発展に向けて参加意識を強める方向でアドボカシー活動を継続的に行っていきます。

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

趣旨説明

蔵川 圭
（国立情報学研究所）

蔵川 圭
http://researchmap.jp/kurakawa/

第 1 回 SPARC Japan セミナー2016 のテーマは「オープンアクセスへの道」です。これまで、日本の大学図書館は学術雑誌の価格高騰を原因としたシリアルズクライシスへの対応として、コンソーシアムを組んで大手海外出版社を相手にビッグディールという包括的な契約方式を実現させ、同時に、機関リポジトリによる雑誌論文のオープンアクセス（OA）化と、ポリシー策定による組織内のルール化を行ってきました。

オープンアクセスをめぐる動向

現在の OA 化の主流な方法は二つです。一つは、購読料支払いを前提に出版社サイトからアクセスコントロールされている論文の公開に対し、機関リポジトリに出版されている論文のコピーや著者最終稿を掲載して無料でアクセスさせるグリーン OA と呼ばれる方法、もう一つは、article processing charge（APC）という論文加工料を著者が出版社に支払い、出版社が出版社サイトから論文を無料でアクセスさせるゴールド OA と呼ばれる方法です。

2012 年に英国政府がゴールド OA を主体とした Finch レポートを実現する意向を表明して以降、欧州ではオープンアクセスの在り方をめぐって、グリーン OA とゴールド OA を二極とする議論が沸き起こっています。英国内では、情報システム合同委員会（JISC）が、購読料と APC が混在したハイブリッド誌のゴールド OA 化を促して、機関の総支払い額を抑えるオフセットシステムを導入したパイロット契約を実施しています。

オランダは自国の研究者が出版したシュプリンガーの 2,000 誌の雑誌論文をゴールド OA 化することを決定し、エルゼビアとも同様の OA 化を約束しています。欧州委員会はオープンサイエンスポリシープラットフォームを推進する中で、ゴールド OA 化への道を探っています。

高エネルギー物理学分野では、SCOAP[3] という名の下、分野の主流な雑誌を対象にゴールド OA 化するビジネスモデルを開発し、実施しています。

一方で、セルフアーカイブを推奨し続けるスティーブン・ハーナッドは、このように見境なくゴールド OA 化を進めようとしている現状に警鐘を鳴らし、

COAR はユネスコと共同で同様の声明を発表しています。また、さまざまなステークホルダーは、急進的なゴールド OA 化への懸念をブログ上に発信しています。

日本の取るべき戦略

こうした状況の中で、わが国はどう振る舞うべきなのでしょうか。主要な学術雑誌の大半を欧州や米国のプラットフォームに依存しているわが国は、欧州や米国の OA 化への流れを避けて通ることはできません。これまでわが国にとっての機関リポジトリの存在意義を見いだしてきたのと同様に、ゴールド OA 化の在り方の議論とその反動としてのグリーン OA への回顧が必要です。

本日のセミナーでは、過去 10 年以上にわたって、次世代学術コンテンツ基盤共同構築事業開始前から日本のオープンアクセスを牽引されてきた講師陣、オープンアクセス誌に投稿する現役の研究者をお招きして、OA 化の在り方と今後の日本が取るべき戦略を議論します。

なお、本セミナーは、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特任准教授の坊農秀雅先生、明治大学学術・社会連携部図書館総務事務室の西脇亜由子様、北海道大学図書館の梶原茂寿様と私の 4 名で企画いたしました。どうぞよろしくお願い申し上げます。

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

オープンアクセスのあり方、グリーン OA とゴールド OA

土屋　俊

（大学改革支援・学位授与機構）

CC BY

講演要旨

ブダペスト宣言以来、学術的研究成果へのアクセスのオープン化への手法にはグリーンな道とゴールドな道という 2 つの方途があることが共通認識となった。当初、たんなる自費出版モデルとも見えるゴールドな道の妥当性は認められず、既存の商業出版に寄生し究極的にはそれを崩壊させる(はずの)グリーンな道を機関リポジトリの利用によって実現することが真剣に試みられた。しかし十年余を経て、後者の道の非効率性は検証され、同時に前者の道の沿う事業の持続可能性が理解されるようになった。したがって、そもそもオープンアクセスの実現にまともな関心をもつならば、ゴールドな道の到達点の姿を論ずるべき段階に至っている。学術情報流通の全体像の変貌、その中での出版者、図書館等の仲介的情報提供者、研究をいまや実施する基盤となった国家的助成システム、そして「研究」そのものの社会的位置づけという観点から、あり得べきそのゴールドな道の到達点の姿を描く。

土屋　俊

1982年から千葉大学で哲学や認知科学を専門に教鞭を執り、1998年に千葉大学図書館長を併任し、以来、図書館コミュニティに参加する。その後、図書館や図書館員と共にコンソーシアによるEライセンス価格交渉、著作権者や出版社との著作権交渉、機関リポジトリやオープンアクセスのプロモーションなどの活動を続けた。2011年には千葉大学から大学評価・学位授与機構(現在の名称は、大学改革支援・学位授与機構)に異動し、同機構の各種事業を支援するとともに、高等教育とその質保証に関して調査・研究している。

研究成果へのオープンなアクセスを推進することは良いことである、だから実現しなければいけないということは当然で、誰も否定しないでしょう。従って、今回それについての議論はしません。しかし、昨今の「オープンサイエンス」という言葉の使い方の空虚さ、「オープンサイエンス」というものは基本的に無意味であるということを申し上げたいと思います。

まず、「オープンサイエンス」の中で唯一意味のある要素はオープンアクセスであろうということで、オープンアクセスが今までどのようになっていたかについてお話しします。最終的には、ゴールドをやるしかない、グリーンはやめようという議論をしたいと思います。

基本的に、日本の各機関ではゴールドではなくグリーンの方が柔らかいので受け入れられやすいのですが、オープンアクセスへの実現には非効率であることは既に実証済みです。10 年たってもまだグリーンと言っていること自体が非常識だと感じるようになりました。

いわゆる「オープンサイエンス」の空虚さ

オープンサイエンスのブームは、六つの要素から成っていると思われます（図 1）。すなわち、オープンソース、オープンアクセス、オープンデータ、オープンな再現可能な研究、オープン教育資源および MOOCs、市民科学です。

最初によく使われるようになった「オープン」が付く単語は、オープンソースです。これは 1980 年代から定着したものだと思います。

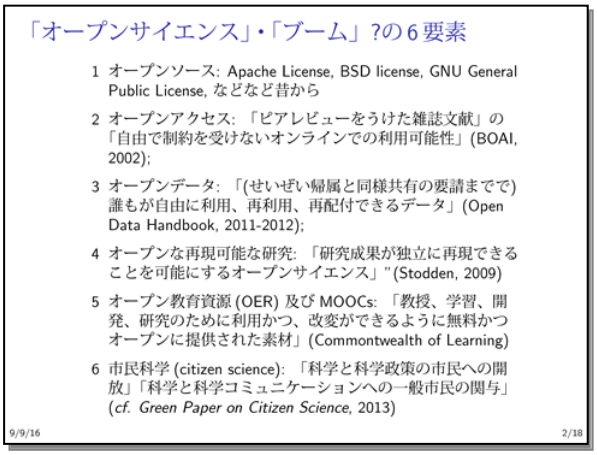
その次にオープンアクセスが出てきて、そこから先は指数関数的に「オープン」の付いた単語が増えていきました。研究不正が増えると、データは公開されていて検証可能な方がいいと考えられるようになり、2010 年前後からオープンデータ、オープンな再現可能な研究の話がたくさん出てきました。

さらに、オープン教育資源（open educational resources）という話も出てきました。アメリカにおいて教科書高騰、高等教育学費高騰が問題になったときに、画期的な解決法として、無料の高等教育を提供する MOOCs が現れました。2011～2012 年ごろの熱気は冷めましたが、まだ続いています。

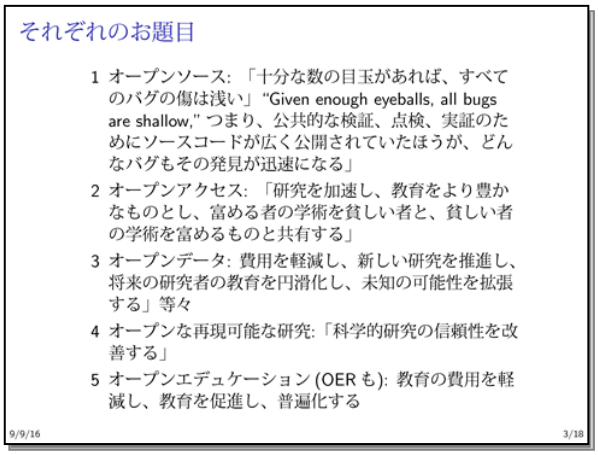
市民科学（citizen science）は、サイエンスコミュニケーションが問題になった 1990 年代あたりから出てきました。科学と科学政策の市民への開放、科学と科学コミュニケーションへの一般市民の関与を citizen science と定義しようというものです。そして、オープンサイエンスという議論の中で、EU の Green Paper on Citizen Science のようなものが出てきました。

図 2 はそれぞれのお題目です。これらはそれぞれ、図 1 に記載した引用文献から引いた言葉です。オープンアクセスは「研究を加速し、教育をより豊かなものとし、富める者の学術を貧しい者と、貧しい者の学術を富める者と共有する」ことが可能になるといわれています。オープンデータは「費用を軽減し、新しい研究を推進し、将来の研究者の教育を円滑化し、未知の可能性を拡張する」と書かれています。このように素晴らしい話しか書いてありません。

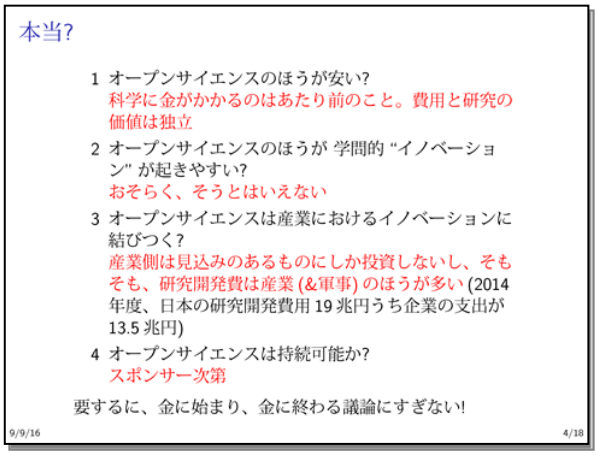
しかし、それらは本当なのでしょうか。オープンデータとオープンエデュケーションのプロモーターたちは、安くなるということを強調しています（図 3）。もちろんオープンなのだから利用者にとっては安くなります。多くの場合が無料になるので、これ以上安いものはないでしょう。しかし、社会全体のコストとし



（図 1）



（図 2）



（図 3）

て本当に安くなるのかどうかは、また別の問題だと考えます。

私の意見としては、科学に金が掛かるのは当たり前なので、学術について「安いからいいのだ」と言うのはおかしいのです。学者たるもの、価値あるものならば、いくら金が掛かってもいいではないかと本来言うべきだと思います。すなわち、費用と研究の価値は、本来は独立であると考えざるを得ず、両者の間に緊張関係が生ずるのは、単に予算が有限であるからにすぎません。それを前提にして考えろとみんな言うのですが、科学の問題を考えるときには、安いからいいのだという話ではないはずです。

オープンサイエンスの方が学問的なイノベーションが起きやすいのかというと、恐らくそんなことはないでしょう。オープンサイエンスでないものが何千年も続いた結果、現在の 20 世紀後半の科学が生まれたのですから、オープンサイエンスでなければ科学が発展しないということはないということは既に実証されています。

オープンサイエンスは産業におけるイノベーションに結び付くのか、これが一番冗談に近い話だと思います。産業側は見込みのあるものにしか投資しないはずです。しかも、2014 年度の日本の研究開発費用は 18 兆 9,000 億円で、そのうち企業の支出が 13 兆 5,000 億円なので、公的な支出はほんの 5 兆円にすぎないのです。そんな 3 分の 1、4 分の 1 のものをいじって一体何になるのかということなので、イノベーションを欲する企業は勝手にやるのです。

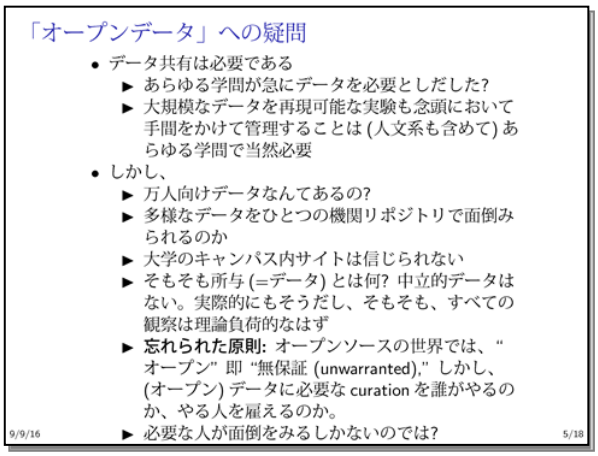
そもそもオープンサイエンスが持続可能かというと、これは要するに利用者に負担を求めないということですから、基本的には社会全体で負担することになるので、そのときのスポンサーの発想次第です。ここで言うスポンサーとは納税者です。納税者というのは普通、有権者と同じ意味になります。実際には有権者のサブセットですが、昨今の日本やイギリスの投票結果などを見て、有権者の気まぐれ度合いを考えれば、有権者を信用してはいけないということがよくお分かりになると思います。

「タダだ」と言ってうまくいく仕組みはどこかにうそがあり、タダであることを保証しようとすれば社会が負担することにならなければいけないのですが、本当に社会に依存していて持続可能であり得るだろうかと考えます。やはり Brexit はなかなかいい例だったと思います。

さらに、オープンデータへの疑問として、データ共有が必要であるというのは当たり前ですが、なぜオープンにしなければいけないのかは私にはよく分かりません（図 4）。使いたい人が共有すればいいだけで、そのために良いデータをつくればいいのでしょう。いろいろな人が使えるデータをつくることはほとんど不可能だと思います。これは単なる偏見と無知ですが、どうやってできるのか、本当に見当がつきません。

最近、気が付いたこととして、オープンソースの時代を思い出していただくとすぐ分かるのですが、オープンであるということは、基本的には品質保証がないということを意味します。ですから、間違いなくこのデータに基づいて何かができる、信用できる、科学研究ができるようなデータをつくるということは、オープンであることと、少なくとも歴史的なコンテクストの中では語義的には矛盾していると思われます。

従って、「オープンデータ」という言葉は自己矛盾を感じないでもありません。もちろんオープンデータではなく、データを使って研究する必要がある研究コミュニティが、昨今のオープンデータブームを使って



（図 4）

予算を取ってくることについては全く反対しません。どんどん頑張ってやってください。ただし、持続可能性をどうやって担保するかについては気を付けていただきたいのです。でも、そうやって一時期ブームになって廃れた研究はたくさんあるので、駄目になったら捨てていけばいいとも考えられます。それでも人類は今まで生き延びてきたのです。

そう考えてみると、まず分かりやすいものとして、オープンエデュケーションはあまり意味のないことだとすぐお分かりになると思います（図 5）。オープンエデュケーションでは、MOOCs が出てから既に 5 年たちますが、アメリカの教育費は依然として上昇傾向にあります。ですから、オープンエデュケーションによって教育費が安くなるということは、基本的にはありません。もっとひどいことに、MOOCs で世界中の人がアメリカの一流の教育機関で行われている授業に基づくコース提供を享受できるようになった結果として、より多くの人がアメリカへ行くようになってしまいました。要するに、世界中から才能を吸い上げる一番いい仕組みになってしまったということです。これは本当に怖いことです。

オープンな再現可能な研究というのは、しょせん研究不正を防げるという後ろ向きの話で、大きな声で言うのは実際には恥ずかしいことです。

オープンデータも、データ共有を頑張ろうということであれば、昔からいろいろなコミュニティはそれなりに努力しています。さらに、データを共有するだけでなく、データをつくる仕組み自体を共有しなければやっていけなくなってしまった高エネルギー物理学のような分野では、研究そのものがシェアされている状態です。その中で、研究成果をシェアすることに意味を見いだすのは難しいでしょう。

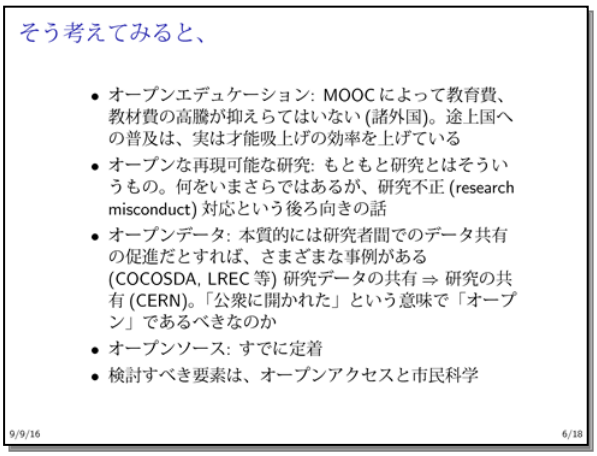
そして、「全ての人に開かれた」という意味でのオープンであることの意味は果たしてあるのでしょうか。オープンソースは既に定着していますが、これは唯一の成功例なので、放っておけばいいのです。

従って、冒頭で挙げた六つの要素のうち、オープンアクセスと市民科学だけがオープンサイエンスの中で意味のある要素として残ると考えられます。

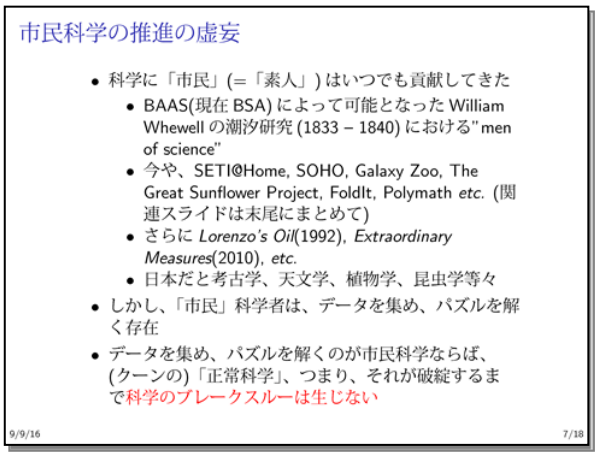
市民科学とオープンアクセス

しかしながら、市民科学の推進は虚妄のようなものです（図 6）。しょせん市民科学で市民に期待されている役割はデータを集めてパズルを解くことなのです。

SETI@home という地球外生命体を探索するプロジェクトは市民を巻き込み、市民に一生懸命空を見てもらいました（図 7）。The Great Sunflower Project は、世界中のどの花にどの虫が来たかという情報を集めるものです（図 8）。Foldit は、蛋白質の構造に基づいたパズルを解いてもらうことによって、蛋白質の 3 次元的な分子構造の予測を試みるコンピューターゲームです（図 9）。Galaxy Zoo は、宇宙にたくさんある銀河系を分類するために、いろいろな分類案を出して適合するかどうかを見ていくプロジェクトです（図 10）。



（図 5）



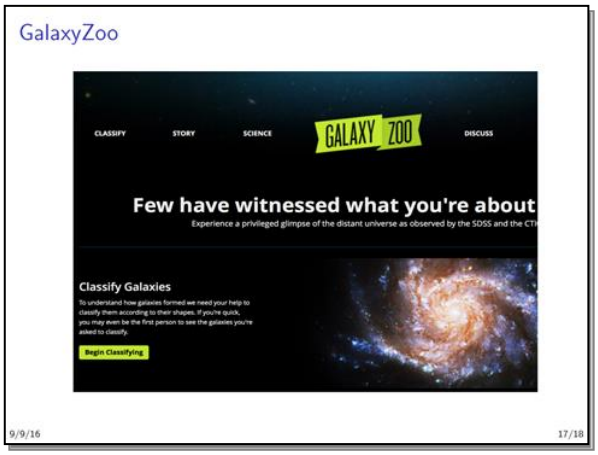
（図 6）

さらに市民科学として有名なのは、「Lorenzo's Oil」という有名な映画があったように、自分の息子が病気になってしまい、図書館で文献を一生懸命調べて、その難病を治療する薬をつくったという現実にあった話です（図 11）。

こうしたものがいい話としてたくさん出てきますが、SETI@home、Galaxy Zoo などは、よく見ると単に市民を搾取しているだけです。もちろん搾取の一番優れた形態は、言うまでもなく本人が搾取されていると気付かないことなので、参加している市民たち自身は極めて自己満足していますが、客観的に見ると単なる搾取なのです。科学史のトーマス・クーンは「データを集めてパズルを解くということが正常科学であり、これがずっと続くと最後に新しい科学が出るが、データを集めてパズルを解くことをやっている限りブレークスルーは生じない。これは既に科学史が示したことだ」と言っています。科学の歴史は時々大きく変わることがあると考えると、こんなことをしていても学術的なブレークスルーが生じそうもないというぐらいのことは言えるのではないかと思います。

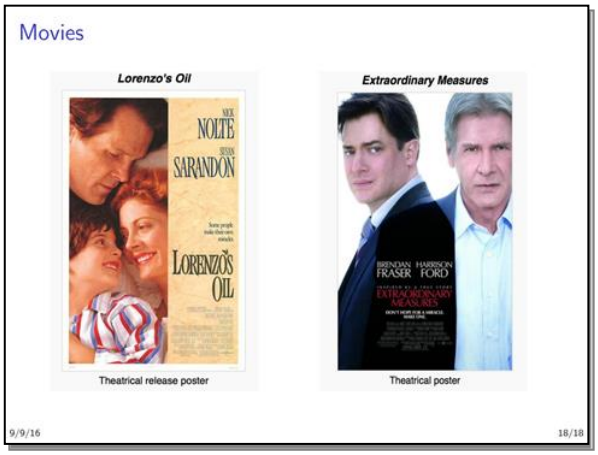
（図 9）

（図 7）



（図 10）

（図 8）



（図 11）

真に論ずべきオープンアクセスの方向性

このように、オープンサイエンスの六つの要素のうち五つは、ほとんど真に語る意味がありません。かつてグリーンな道を探求する人々が、正直に言うと私を含めて存在しました。最初から疑問はあったけれど、ひとまず乗ってみたのです（図 12）。

約 20 年前の 1994 年、Harnad と Okerson が Subversive Proposal を言いだしました。約 10 年前の 2002 年には、ブダペスト・オープンアクセス・イニシアチブが出てきました。それ以来の 10 年間もいろいろなことを行いました。そんなことをしても、ペイウォールの裏側・内側にある文献へのアクセスは進まなかったということを尾城さんなどが一生懸命調べてくださって、なかなかうまくいっていないというのが事実だろうと分かりました。ですから、非効率だということです。PubMed Central はかなりの出版社が入っているので成功したと言っていいかもしれませんが、それ以外のものはうまくいっていません。

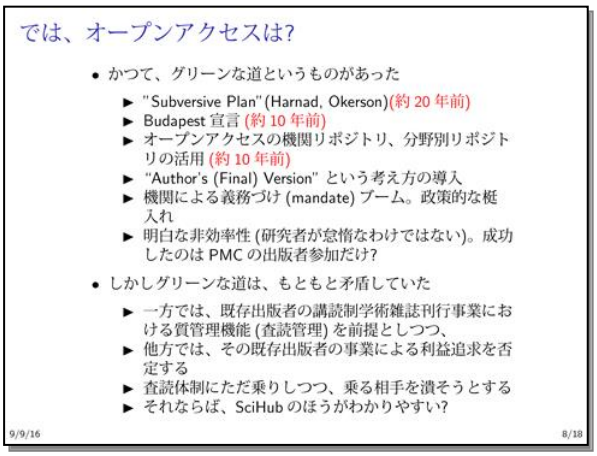
よく考えてみると、グリーンな道というのは、もともと矛盾点をうまくごまかしてきただけだと言えなくもないのです。グリーン路線の本質は、既存の出版社が行ったサブスクリプションベースの刊行事業における査読機能による品質保証は利用しつつ、それにタダ乗りするというものです。その結果として既存出版社がつぶれてしまった後の出版モデルについては、実は明確な展望を述べていません。

そもそも Subversive Proposal のときは、既存の強欲な出版社をつぶすことが目的だったので、それをつぶした後どうするかということは言いませんでした。本当に Harnad は優れた政治家だとしか言いようがありません。

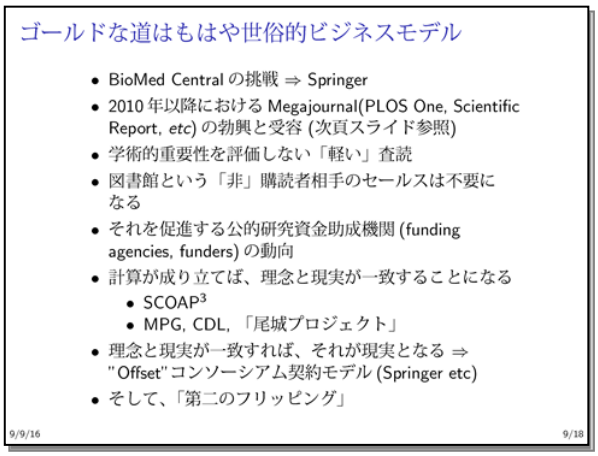
要するに、査読体制にタダ乗りしつつ、乗る相手をつぶそうとしていたので、Sci-Hub のような単純な著作権法違反の方がもっと分かりやすいものになっているはずだということです。

ところが、幸か不幸か、ゴールドな道が実施可能であるということがだんだん見えてきました（図 13）。典型的なのは、BioMed Central、PLOS ONE、Scientic Reports といったメガジャーナルの勃興です。最近は新聞などでも、日本のどこかの大学の偉い先生がこんな素晴らしい研究を発表したというときに、PLOS ONE や Scientic Reports を引用していることが非常に多いです。業界に通じている方は、Scientic Reports への掲載は素晴らしい業績を挙げたこととイコールにならないと思う方が多いのではないかという感じがしますが、学者の中では受け入れられているように思われます。

なぜかというと、日本から出る論文数はここのところ微減ぐらいの状態が続いているのに、なぜかマルチディシプリナリー分野の雑誌への掲載数がやたらと増えているからです。2009 年で 700 本ぐらいだったものが、2015 年では 3,500 本を超えています（図 14）。下の青と赤が PLOS ONE と Scientic Reports です。ですから、日本のリサーチコミュニティはこれを受け入



（図 12）



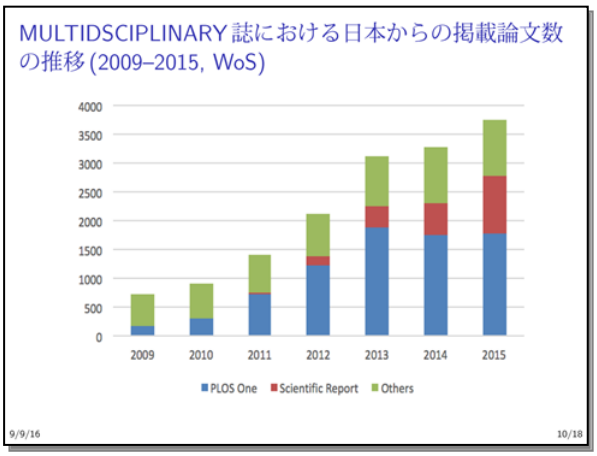
（図 13）

れている状態なのだと思われます。

彼らの言っていることは、学術的な重要性・価値は評価せず、科学的な基準に合っていれば載せることができるということです。これはオンラインだから可能なことです。プリントでページ数が限定されている場合には不可能なことができるのです。そうすると、単価×数量なので、その気になればいくらでも稼ぐことができます。もちろん、そのことによっていわゆるプレダトリージャーナルが出てきていることも事実だと思います。軽い査読を行うことで、きちんとした経済モデルが可能になることを確認できたということです。

おまけに、出版社にとっては図書館に売るという余計な仕事をする必要がなくなります。ですから、セールスチームは要りません。あったとしても、それは研究者に投稿してくださいと言えばいいのであって、その雑誌を読まない図書館員に雑誌を買ってくださいと言う必要はないという、とてもいい状態になりました。しかも、公的研究資金助成機関はそれに乗ろうとしています。

そして、うまく計算が合えば、理念と現実が一致します。SCOAP[3]がうまくいっているかどうかは安達先生からお話しいただけると思いますが、取りあえず1年はうまくいって、2年も大丈夫かと思います。マックス・プランク研究所や、カリフォルニア電子図書館（CDL）、尾城プロジェクトなど、ひょっとして実現可能なのではないかという計算結果がだんだん出てきています。



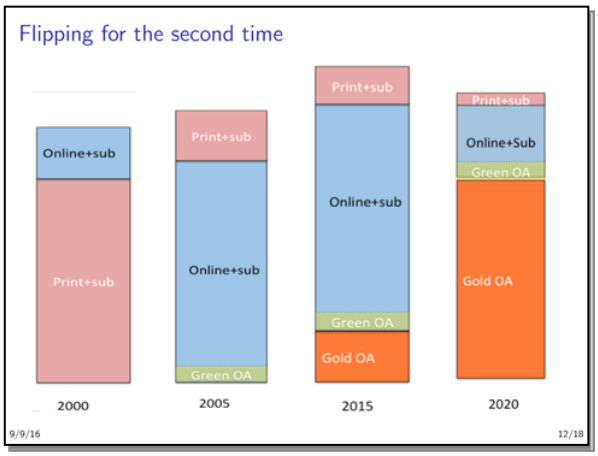
（図 14）

理念と現実が一致すればそれは現実となるので、一番大きいところではシュプリンガーと情報システム合同委員会（JISC）のオフセットモデルのようなものが存在するということになります。

かつて電子ジャーナルが登場したときに、フリッピング、すなわち、プリントに対する価格付けと電子に対する価格付けのどちらをベースにするか逆転させようということがありました（図 15）。2000 年だとプリントでサブスクリプションが多かったのですが、2005 年にはもうかなり逆転していました。

今はオンラインのサブスクリプションがほとんどを占めていて、ゴールドオープンアクセスは少数です。しかし、グリーンオープンアクセスよりはある意味では機能しているという状態が来ます。この後を予測することは不安ですが、ゴールドオープンアクセスがほとんどになって、グリーンオープンアクセスは伸びず、オンラインサブスクリプションの部分がこのぐらい残るということになるのではないかと思います。

今までは宗教団体が唱えていたようなオープンアクセスがいわば世俗化したので、あとは金の計算の問題です（図 16）。そうすると、各国のファンディングエージェンシーは計算をします。情報流通のための経費はごくわずかです。極端でアンフェアな比較をするならば、先ほど述べた 2014 年度の日本の研究開発費用約 19 兆円に対して、日本の図書館が払っているお金は 300 億円です。ですから、どう計算してもものすごく少ないということになるでしょう。



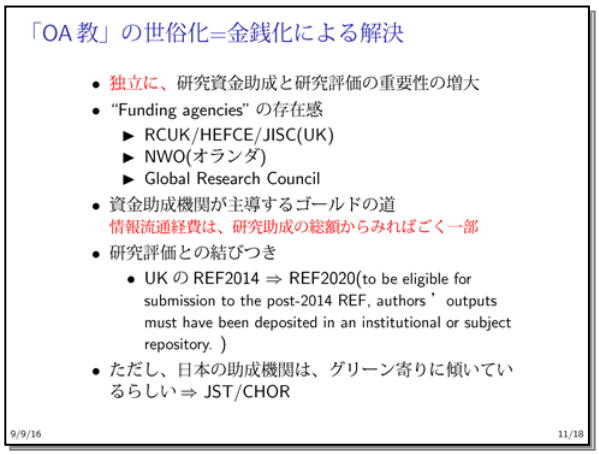
（図 15）

そのように各国の資金助成機関はゴールドの道を主導しています。にもかかわらず、日本の資金助成機関である科学技術振興機構（JST）はグリーン寄りに傾いているので、CHOR と組もうという話をしています。グリーンについてはそれしかやりようがない、だから図書館は要らないという話です。

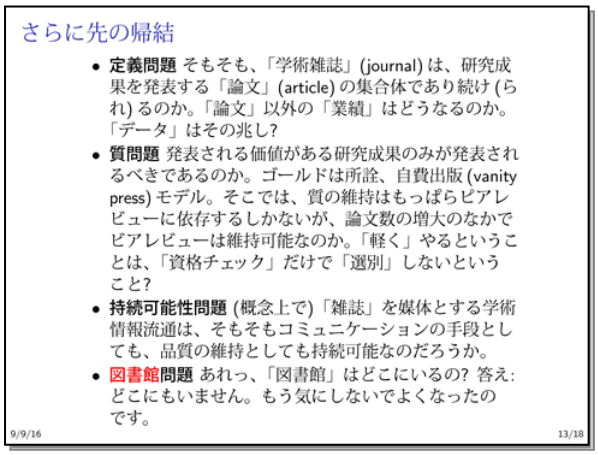
結論としては、まず、論文によって研究成果を公表することの意味付けがどう変わっていくかについて、真剣に考えなければいけません（図 17）。つまり、データというものの制作管理について何かの評価が必要だという動きは、ある種の兆しなのかもしれません。

そして、質の問題として、本当にピアレビューモデルはいつまでも持つのかという疑問は当然あり得るでしょう。また、そもそも論文やレビューを入れた雑誌が本当に学術情報流通の媒体として何か概念的な意義を持ち続けることができるのかは疑問です。

図書館についての問題は、あるように見えますが、ありません。なぜかというと、もはや学術情報流通は図書館には関係なくなったからです。図書館についてこの問題は完全に解消されたと言うべきだと考えられます。



（図 16）



（図 17）

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

学術雑誌のキャッシュフロー転換の可能性を探る

～JUSTICE/SPARC Japan 合同調査チームによる調査結果の概要～

尾城　孝一

（大学図書館コンソーシアム連合（JUSTICE）/東京大学附属図書館）

CC BY NC SA

講演要旨

これまで大学図書館コンソーシアム連合（JUSTICE）は、出版各社との間で購読費の交渉を行ってきたが、昨今のいわゆるゴールド OA の進展のなかで、APC を含めたトータルな経費に関する交渉を行う必要に迫られている。JUSTICE は、毎年会員館に対して契約状況調査を実施し、各館における購読費の支出状況を把握しているが、トータルな経費に関する交渉を行うには、購読費に加えて APC 支出の実態をつかむことが前提となる。一方、2015 年の Berlin 12 において、購読費を APC に振り替えることにより、OA モデルへの転換を世界的規模で実現しようという提案が行われた。こうした状況を踏まえ、JUSTICE と SPARC Japan の運営委員会は、2015 年に合同の調査チームを設置し、国内研究者による論文公表の実態と APC 支払い額を把握するための取組みを開始した。本発表においては、当調査チームの活動とこれまでのデータ集計結果の概要を報告し、今後のオープンアクセス化戦略についての議論に資する。

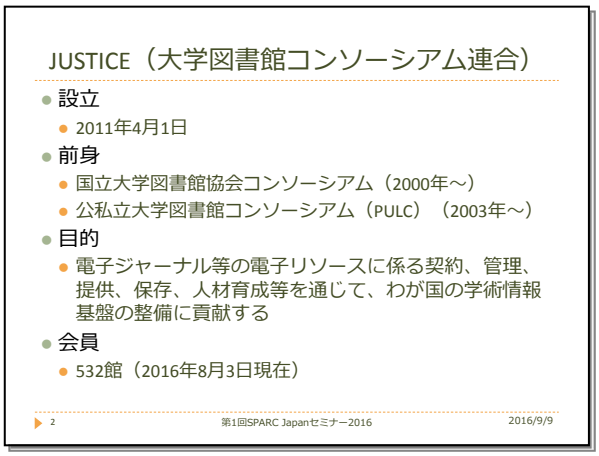
尾城　孝一

東京大学附属図書館事務部長。1983年1月、名古屋大学附属図書館に採用され図書館職員としてのキャリアを開始。その後、東京工業大学附属図書館、国立国会図書館、千葉大学附属図書館、国立情報学研究所、東京大学附属図書館、大学図書館コンソーシアム連合事務局を歴任。2015年4月より、現職。

今日は「学術雑誌のキャッシュフロー転換の可能性を探る」と題して、JUSTICE と SPARC Japan の運営委員会の下に置かれている調査チームの調査結果について報告させていただきます。最後のディスカッションの材料にしてもらえればと考えています。

JUSTICE の概要

JUSTICE は日本の大学図書館のコンソーシアムで、出版社との価格交渉を一本化して、大学における電子ジャーナルの普及に貢献してきました（図 1）。



（図 1）

出版社との交渉状況は図 2 のような具合になっています。

図 3 のグラフは、国公私立の大学図書館における購読雑誌数と支出額の大学ごとの平均値を示したものです。このデータを見ると、大学図書館全体としてはコンソーシアムの交渉で値上げを抑えつつ、かつ冊子体、紙のジャーナルの予算を電子ジャーナルへうまく振り向けながら経費をコントロールして、利用できるジャーナル数を増やしてきたことが分かると思います。

コンソーシアムを通じた電子ジャーナルの導入が進んだ結果、大学間の情報格差もかなり改善されてきました（図 4）。

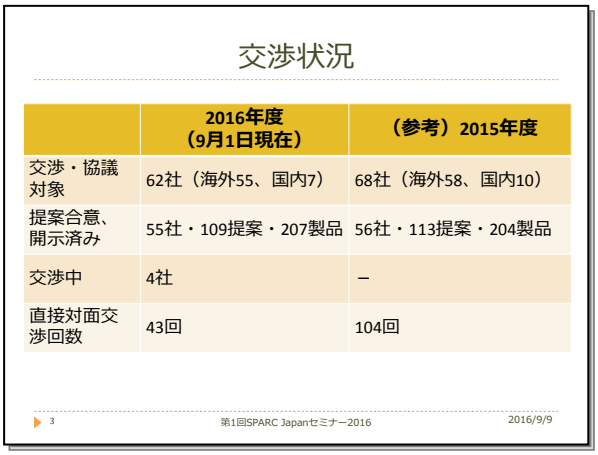
調査の背景

次に、JUSTICE が今回の調査を始めるに至った経緯や背景についてお話しします。

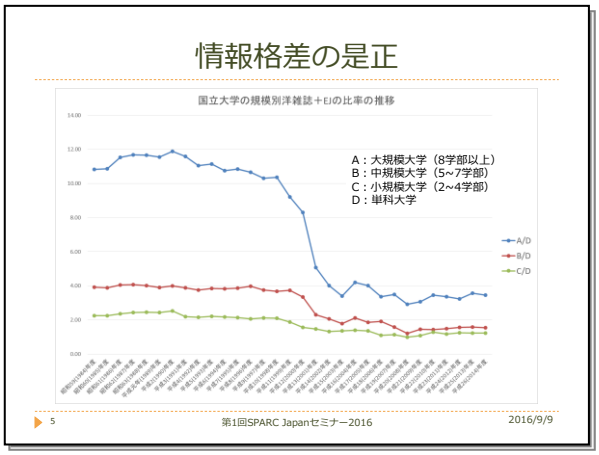
第一の背景として、ゴールドのオープンアクセスジャーナルや論文が増えてきて、それに対応する必要性に迫られていることがあります（図 5）。いろいろな調査のデータがありますが、オープンアクセスのジャーナルやオープンアクセスの論文はどんどん増えており、これからも増えると予測されます（図 6）。

図 7 はエルゼビアから頂いたグラフで、2014 年に購読モデルで世界で 210 万論文、オープンアクセスで 42 万論文が出版されています。

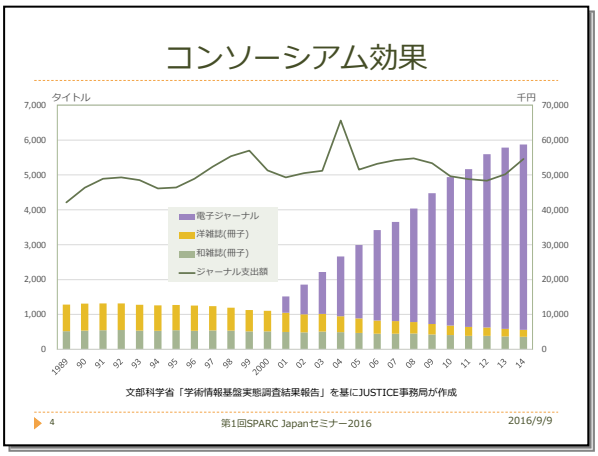
このようなゴールドオープンアクセスの状況を踏まえると、これまで JUSTICE は購読費（サブスクリプション）の交渉を出版社と行ってきましたが、これからはそれだけでは足りず、article processing charge（APC）といわれるオープンアクセスの出版費用を含めたトータルの費用についての交渉が求められます。そのためには、日本の大学が一体どのぐらい APC を



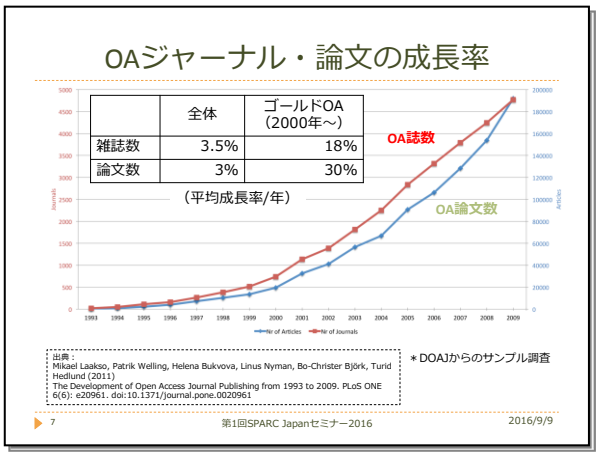
（図 2）



（図 4）



（図 3）



（図 5）

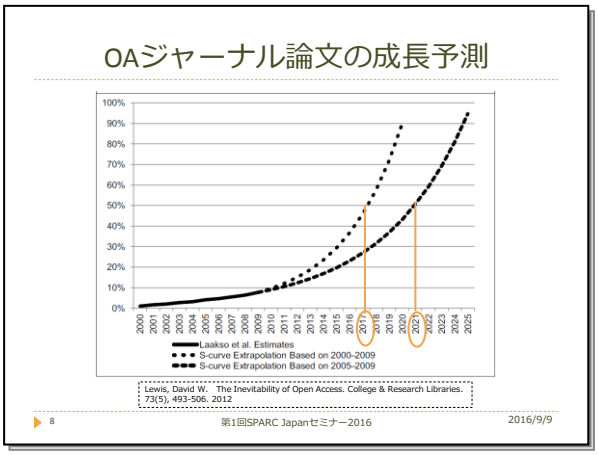
払っているのかというデータをつかんでおかなければ話になりません。

第 2 の背景として、購読モデルでの交渉の限界があります。大学図書館は JUSTICE の交渉を通じて、値上がりを抑えつつアクセスできるジャーナルの数を増やしてきましたが、それでももちろん値上がりは続いています。図 8 は自然科学系の電子ジャーナルの平均定価の推移を示すグラフで、この 5 年間で年平均 5.6%値上がりしています。人社系でも年平均 4.1%の値上がりとなっています（図 9）。

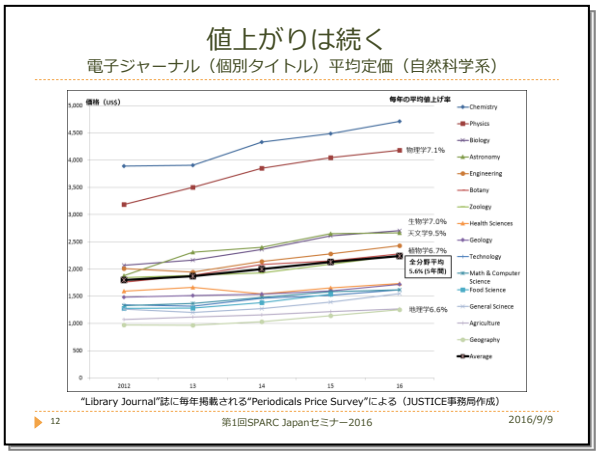
よくいわれる電子ジャーナルの高騰というのは言い過ぎですが、着実に値上がりを続けていることは間違いありません。もちろん大学図書館は定価で買っているわけではなく、コンソーシアムで契約をして交渉し、もう少し値上げを抑えた価格で実際には買っています。とはいうものの、購読費の交渉はあくまで対症療法であり、問題を根本的に解決するわけではありません。

そもそもなぜ値上がりが続くのか、その原因はいろいろ指摘されているのですが、根本的な原因は、どうも購読というシステムに内在しているのではないかと考えられています。購読モデルでは、出版社はアクセス権を販売しています。出版社は論文の著作権の譲渡を受けているので、アクセスを提供する権利を独占しています。ですから、そこで価格競争が生まれる余地は全くありません。

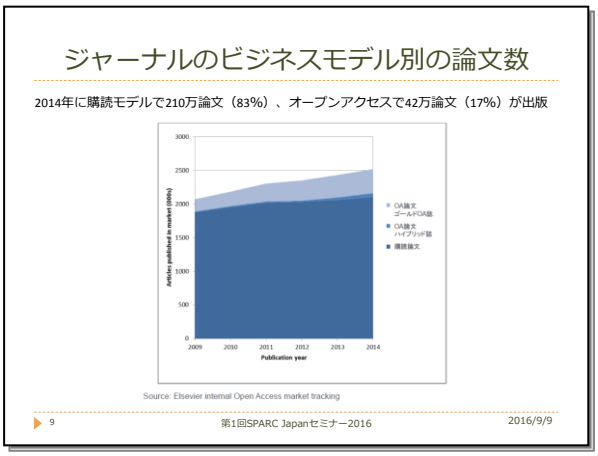
また、購読者から見た学術雑誌は代替財ではなく補完財となっています。例えば、ある分野の A というジャーナルと B というジャーナルは代替可能ではありません。お互いに補完し合う関係になってしまっています。読む側からすると両方の雑誌を購読しなければならず、需要に歯止めがかかりません。売る方は安心して値上げができます。



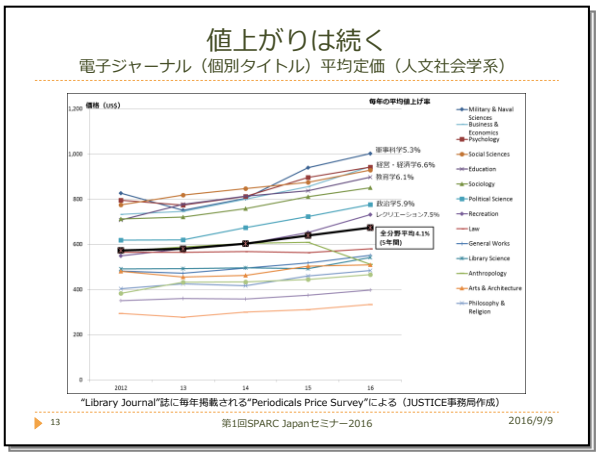
（図 6）



（図 8）



（図 7）

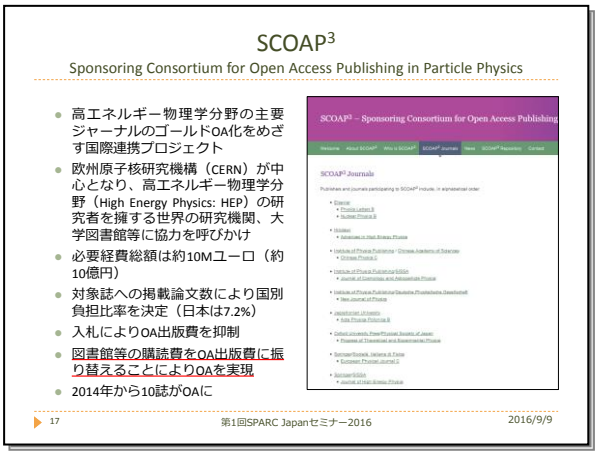


（図 9）

しばしば今の購読モデル、すなわち読者が払うモデルは機能不全を起こしていると指摘されています。購読というシステムの中で新しい電子ジャーナルの契約モデルをつくるという試みは、もう何年もコンソーシアムと出版社の間で協議を続けてきているのですが、結局のところ、ビッグディールと呼ばれているパッケージで買うか、個別タイトルごとに契約するかという選択肢しかありません。その中間の第 3 のモデルを出版社も図書館も模索していますが、出てきません。

これがオープンアクセスのモデル、すなわち著者が払うモデルになると、ガラッと変わる可能性があります。オープンアクセスのモデルでは、出版社が著者に売るのは出版サービスです。アクセス権の販売の独占ということはあり得ません。また、著者から見ると学術雑誌は補完財ではなく、代替財に変わります。より質が高く、より APC の安い雑誌に論文を投稿するという投稿先の選択ができます。そこに競争が生まれ、さらには価格を抑えるメカニズムが働く可能性が出てきます。かつ、出た論文は全部オープンにアクセスできるようになるので、オープンアクセスを一挙に進めることができます。

従って、購読モデルからオープンアクセスモデルへの転換の可能性については検討してみる価値があり、データに基づいてきちんと検討する必要があるということが二つ目の背景、問題意識になります。



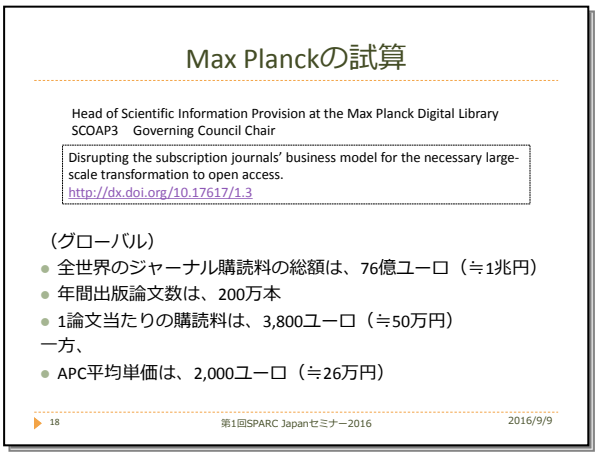
（図 10）

世界の動向

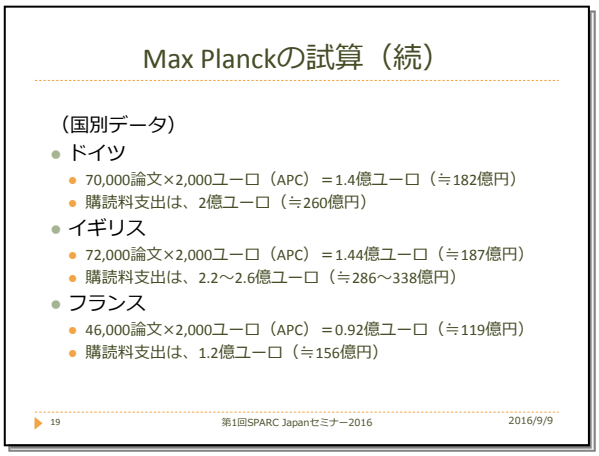
調査結果の報告に入る前に、少し世界の動向を眺めておきたいと思います。SCOAP³ は、それまでに図書館が払っていた購読費をオープンアクセスの APC に振り替えることによってオープンアクセスを実現しようとするプロジェクトです（図 10）。

この SCOAP³ にヒントを得て、ドイツのマックス・プランク研究所がある試算を行いました（図 11）。世界で年間 200 万本の論文が出ていて、1 論文当たりの購読費を計算すると 3,800 ユーロになります。一方、APC の平均単価は 2,000 ユーロだといわれているので、APC の方がはるかに安いのです。

マックス・プランク研究所は国別のデータも示しています（図 12）。例えばドイツを見ると、年間約 7 万本の論文を出しています。それを全部オープンアクセスにするための経費は 1.4 億ユーロです。一方、ドイツ全体で購読料を購読費として払っている額を足し合



（図 11）

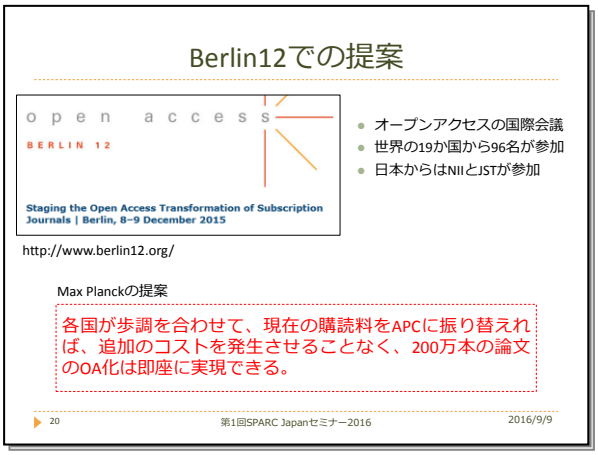


（図 12）

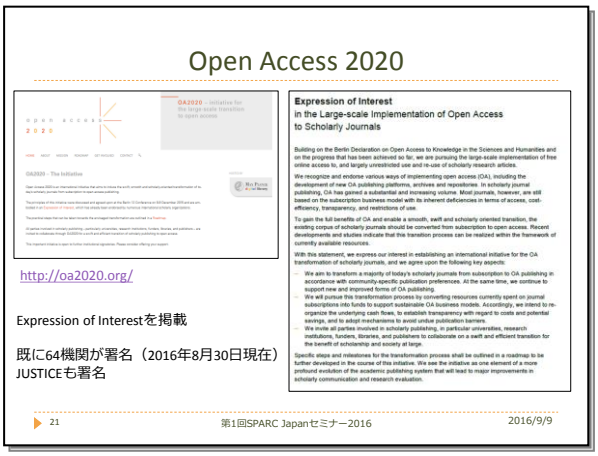
わせると 2 億ユーロになります。ですから、この 2 億ユーロを APC に振り替えれば、7 万本の論文はすぐオープンアクセスにできて、かつ、お釣りが来るという計算になります。イギリスやフランスも同じような状況になっています。

2015 年 12 月に、Berlin12 というオープンアクセスの国際会議が開かれました。そこで、こうした試算に基づき、マックス・プランク研究所が、各国が歩調を合わせて現在の購読料を APC に振り替えれば、追加のコストを発生させることなく、世界中の 200 万本の論文のオープンアクセス化は即座に実現できるという提案を行いました（図 13）。

その後、マックス・プランク研究所は Open Access 2020 というサイトを立ち上げ、そこに関心表明（Expression of Interest）という文書を掲載して、関心を持つ大学や機関から署名を募っています（図 14）。



（図 13）



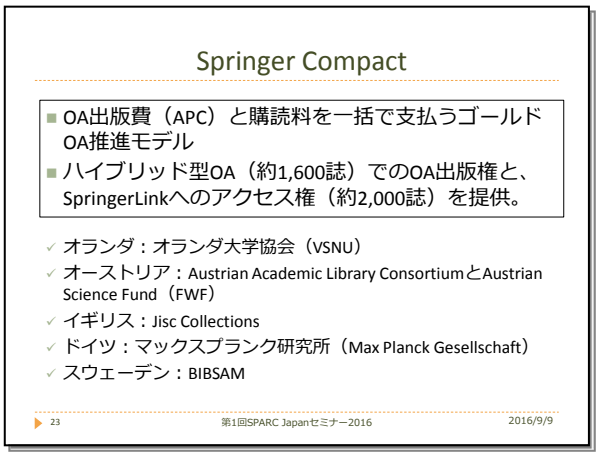
（図 14）

今、六十数機関が署名しており、JUSTICE も 8 月に署名しました。

もちろん、このマックス・プランク研究所の提案に対してはいろいろな問題が既に指摘されています。まず、1 論文 2,000 ユーロという設定は安過ぎるのではないかということです。また、SCOAP[3] は高エネルギー物理学分野の 10 のジャーナルに限定されたプロジェクトでしたが、それでも購読費を APC に振り替えるのにかなりの時間と労力がかかりました。それを本当にグローバルに展開することが本当にできるのかということもあります。

そして、出てくる論文の数が増え続けると、いずれは今の購読料の振替だけではそのコストは賄えなくなるということもいわれています。これはそうだろうと思います。これまで以上に特定出版社にロックインしてしまうのではないかという懸念も表明されています。

このような動きをにらみつつ、具体的なオープンアクセスベースの契約モデルを打ち出している出版社もあります。シュプリンガーは Springer Compact というゴールドオープンアクセスの推進モデルをつくって、ヨーロッパやイギリスなど各国と契約を結んだり、パイロットを始めたりしています（図 15）。



（図 15）

国内研究者による論文公表実態調査

やはり日本でも自分たちできちんとデータを持っていなければいけないと考え、JUSTICE と SPARC Japan が協力して調査チームをつくり、集計作業を始めました（図 16）。

調査の目的は、トータルなコスト（TCP）の交渉を行うための APC 支払い額の把握と、購読モデルからオープンアクセスモデルへの転換の可能性を探ることです。先述の二つの問題意識に対応した目的を設定して、ちょうど 1 年ぐらい前から調査を始めました。

どんな作業をしているかというと、トムソン・ロイター社のデータを基にして集計作業を進めています（図 17）。論文公表年が 2012～2014 年の 3 年分の論文データを入手し、それを対象とします。著者に日本の研究機関に所属する研究者が含まれている論文データだけを抽出しました。さらに、Web of Science にはドキュメントタイプが設定されているのですが、Article と Review のものに絞りました。マックス・プランク研究所の調査がこの二つのドキュメントタイプに絞っているので、それに合わせました。そして、corresponding author や第一著者などといわれている reprint author の国が日本である論文に絞り込みました。それで集計のための母集合をつくり、それに基づいて各種の集計作業を行っています。

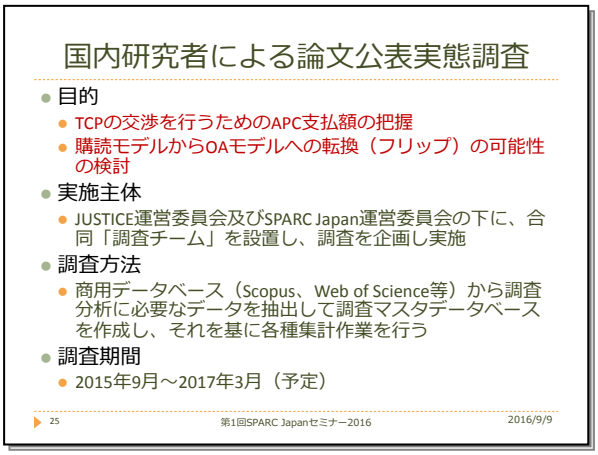
今回、オープンアクセスの論文を数えていますが、それはフルオープンアクセスジャーナルに掲載された論文に限られます。いわゆるハイブリッドジャーナルに載ったオープンアクセスの論文にはフラグが付いていないので、それは残念ながらカウントできていません。

最後に、APC の額については、出版社やジャーナルのサイトから APC 額の情報を集めてきて、当該年度の為替レートにより日本円に換算しています。

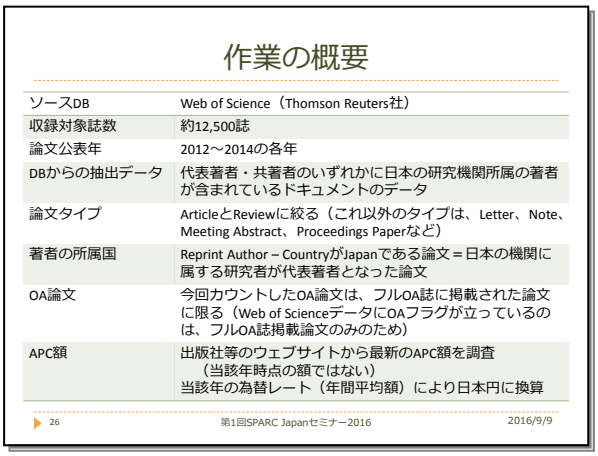
このような作業をして出てきた結果が図 18 のグラフです。全体の論文数は約 6 万 5,000 件となっています。そのうちフルオープンアクセス誌に掲載されたオープンアクセス論文の数は少しずつ伸びてきています。

APC 支払い額も、2012 年には約 6 億 8,000 万円だったのですが、2014 年には約 12 億 8,000 万円に増えているということで、伸びてきています。

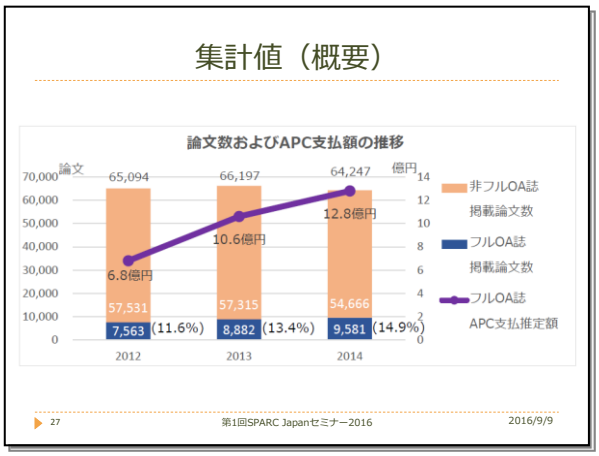
図 19 は分野別掲載論文数を集計したグラフです。医学、生物学、物理学、工学、化学といった論文がかなりの割合を占めています。これはオープンアクセス



（図 16）



（図 17）



（図 18）

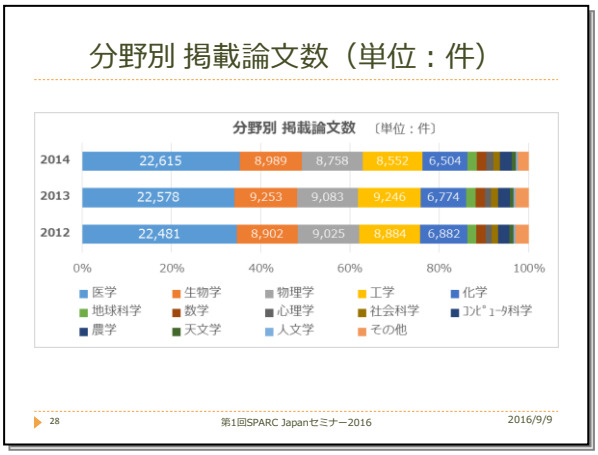
だけではなく、全論文の比率です。

フルオープンアクセス誌に掲載された分野別の論文数を集計すると、医学と生物学の割合が大きいことが分かります（図 20）。当然、APC 支払い推定額で集計しても、医学と生物学の割合が大きいことが分かります（図 21）。

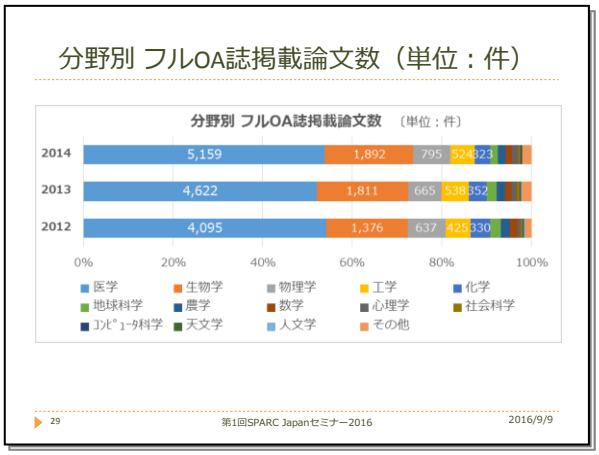
JUSTICE の会員大学は全部で 532 ありますが、そのうち 2014 年に Web of Science に掲載された論文を出していて、かつ購読費が判明している大学がちょうど 300 あります。その 300 大学分を抜き出して、大学ごとに発表論文全てをオープンアクセスにするための経費と購読費を比較したグラフが図 22 です。

このオープンアクセス化必要経費は、大学ごとの発表論文数に論文当たりの APC を 2,000 ユーロと想定して掛けて、2014 年の換算レート 141 円 85 銭を掛けて円価格を出しました。一方、購読費は、JUSTICE が毎年行っている契約状況調査からカレントの電子ジャーナルの購読費が分かるので、それを大学ごとに取ってきました。そうすると、JUSTICE の 300 大学全体で論文数が 4 万 8,967 件あり、それを全部オープンアクセスにするための経費が約 140 億円になります。一方、購読費を 300 大学足し合わせると、約 240 億円になります。ということで、JUSTICE の 300 大学全体では、購読費を APC に振り替えれば 100 億円のお釣りが来てしまうという計算になります。

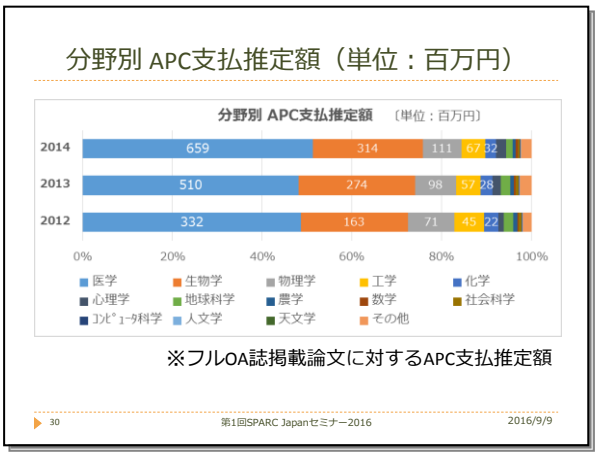
オレンジ色の折れ線グラフは、オープンアクセス化必要経費の多い大学順に左から右に並べて、オープンアクセス化するための経費と購読費を比較したものです。300 大学あるうち、39 大学だけがオープンアクセス化経費の方が高くなっています。このような結果が出ています。



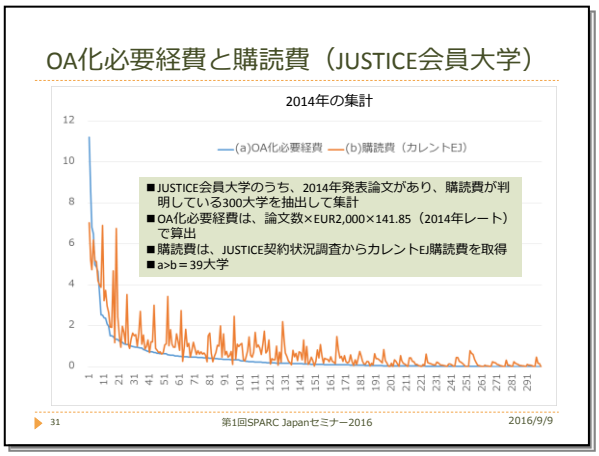
（図 19）



（図 20）



（図 21）



（図 22）

調査のまとめ

APC 支払い額の把握については、ハイブリッドジャーナルに掲載されたオープンアクセス論文の特定が簡単ではありません。簡単ではないといっても何とかしなければならず、出版社はデータを持っているはずなので、主な出版社 20 社にデータの提供を求めています。今のところ 17 社からはデータが出てきて、3 社からはまだ出てきていません。ちなみにその非協力的な 3 社はアメリカ化学会（ACS）、ケンブリッジ大学出版局（CUP）、Nature です。

また、今出している APC 額はあくまでその出版社のサイトに載っている定価を使っているので、実際に大学が払った額ではありません。APC も割引価格などいろいろあるようなので、実際の支払い額をつかむことは容易ではありません。論文を書いた著者と出版社の間に立って支払いを管理するエージェントのようなものがなければ把握できないと思います。それは図書館の役割なのか、あるいはひょっとすると、こういうサービスを提供する業者が出てくるかもしれません。いずれにしても、大学の APC 支払いを一元的に管理するサービスがなければ実際の支払い額はつかめないのではないかと思います。

購読からオープンアクセスへの転換の可能性については、マックス・プランク研究所などが主張しているとおり、日本全体で見れば恐らく可能だと推測されます。その一方で、論文発表の多い大規模大学を中心とした幾つかの大学では、購読費とオープンアクセス化必要経費にまだギャップがあり、オープンアクセス化必要経費を購読費で賄えません。そういうところは購読費以外の何らかの資金を投入しなければいけないのではないかというところまで分かりました。

今後のタスク

最後に、今後の作業について触れます。継続して調査したいことは、ハイブリッド誌掲載オープンアクセス論文の特定、APC 額データベースのジャーナルごとの整備です。毎年 APC の定価を調査しておけば、APC の値上がり状況などもつかむことができるでしょう。

また、調査対象範囲の拡大です。今は 2014 年までのデータがあるので、2015 年のデータも入手して、合わせて集計します。

そして、JUSTICE 会員館へのデータ提供です。大学ごとのデータを何らかの形で JUSTICE の会員館に提供して使ってもらおうと考えています。

国際連携や国際協調・協力も大事だと考えています。具体的には、2016 年 10 月に ICOLC がアムステルダムで開かれるので、そこに JUSTICE から 2 人派遣して、情報収集や情報提供、意見交換などを行いたいと思っています。また、Open Access 2020 の関心表明に JUSTICE も署名しましたが、それに関連する会議である Berlin13 が 2017 年 3 月にベルリンで開かれるので、これにも参加する予定です。そして、マックス・プランク研究所にぜひ関係者を派遣したいと思っています。さらに、カリフォルニア大学電子図書館（CDL）などが中心になって、Pay It Forward という調査を行っています。われわれと同じようなことを目的として、詳細なデータの収集と分析を行い、分厚い報告書を出しています。できればその関係者とも意見交換ができるような場をつくっていきたいと考えています。

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

SCOAP³ による 学術誌のオープンアクセスへの転換

安達 淳
（国立情報学研究所）

CC BY SA

講演要旨

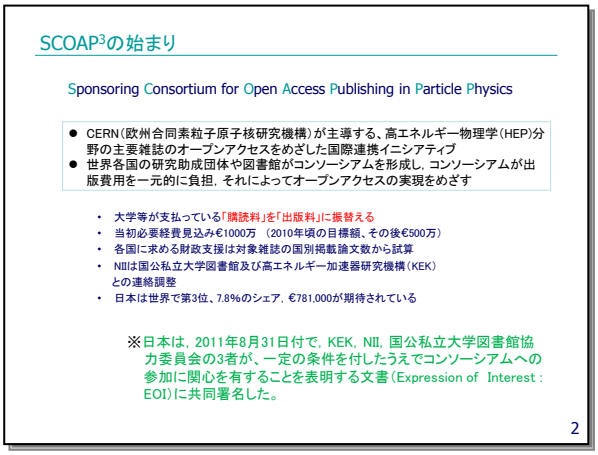
SCOAP³ は、高エネルギー物理学分野の主要学術雑誌をオープンアクセスに転換し、著者への経費的障壁なしに論文出版できるようにするためのコンソーシアムである。CERN（欧州原子核研究機構）が提唱し、44 カ国の 3,000 を越える大学図書館等を中心に構成する国際コンソーシアムとして、2014 年 1 月から 10 の雑誌をオープンアクセス化した。この活動のフェーズ 1 は 2017 年末までの 3 年間で、現在、フェーズ 2 への継続発展の方向性が定まってきたところである。本講演では SCOAP³ の経緯と現況について紹介するとともに、我が国の困難な状況に対する支援のお願いをいたします。

安達 淳

1981年東京大学大学院工学系研究科博士課程修了。工学博士。東京大学大型計算機センター助手、学術情報センター助教授、教授を経て現在国立情報学研究所教授。東京大学大学院情報理工学研究科教授を併任。SCOAP³ Executive Committeeの委員。データベースシステム、情報検索、テキストマイニング等の開発研究等に従事。電子情報通信学会、情報処理学会、データベース学会、IEEE、ACM 各会員。

SCOAP³ の始まり

SCOAP³ は Sponsoring Consortium for Open Access Publishing in Particle Physics の略です（図 1）。これは、欧州合同素粒子原子核研究機構（CERN）という研究機関が各国に呼び掛け、高エネルギー物理学分野の主要な定評のあるジャーナルをオープンアクセス化しようするプロジェクトです。そこで提案されたのが、大学が払っている当該雑誌の購読料をオープンアクセス出版料にリダイレクトする、すなわち「振り替える」というやり方です。この分野の当該雑誌をオープンアクセス化するのに必要な経費を当初 1,000 万ユーロ（約 14 億円）と見積もりました。各国に求める財政支援は対象雑誌の国別掲載論文数から試算するという提案も含まれており、日本は当時世界 3 位の 7.8%の



（図 1）

シェアで、78 万ユーロが期待されていました。

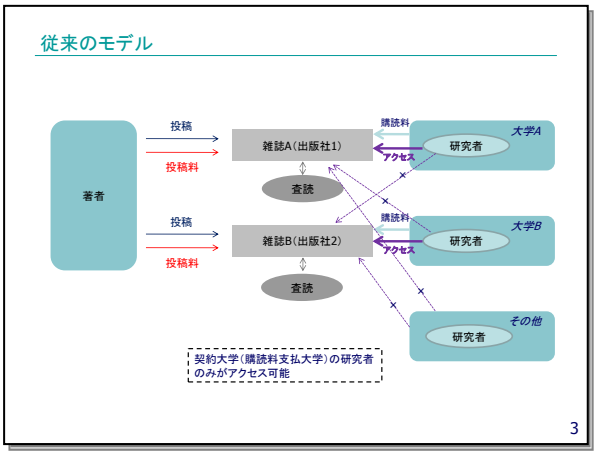
2011 年に、高エネルギー加速器研究機構（KEK）、国立情報学研究所（NII）、国公私立大学図書館協力委員会の 3 者で関心表明（EoI）にサインしました。

図 2 が従来の出版モデル、すなわち購読料のモデルですが、SCOAP³ は図書館などから拠出金を募り、一括して出版社に article processing charge（APC）を払い、オープンアクセスにするというモデルです（図 3）。幾つかの基本的な原則があり、まず、著者からは APC を取りません。この分野では研究者は APC を払わないで自由に論文を書ける世界をつくりたいということです。SCOAP³ が、1 論文当たりの APC を出版社の入札で決め、それに従って APC を払います。予定していた論文数よりも少なければ、その分の APC だけ払います。研究者がたくさん論文を書いたときには、APC の総支払額の上限を入札条件で決めていて、その上限までは APC を払うが、それ以上は無料で発行するという条件になっています。また、リポジトリにも CC-BY のライセンスでデポジットします。

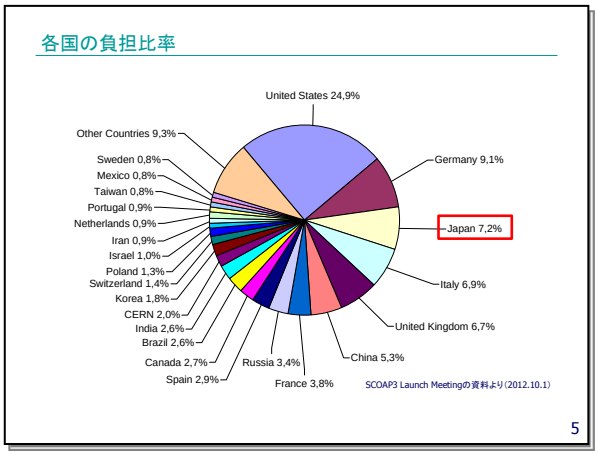
図 4 が 2011 年当時の各国の負担率です。アメリカが 24.9%、ドイツが 9.1%、日本は 3 番目で 7.2%です。高エネルギー物理学では日本は世界で 3 番目に活躍している国であり、それに沿った貢献が求められたわけです。

オープンアクセス出版開始までの交渉

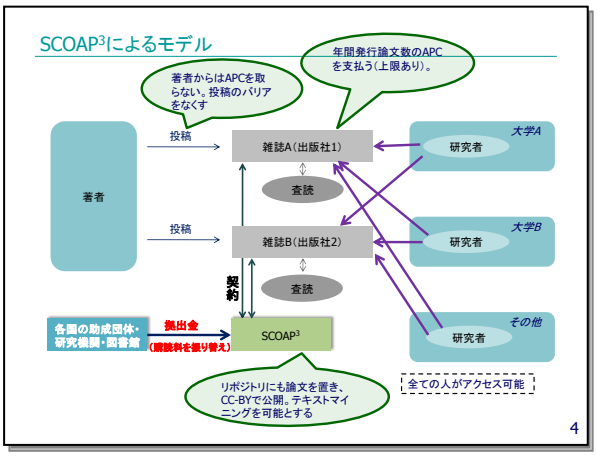
SCOAP³ が大きく動いたのが 2011 年です（図 5）。主としてヨーロッパでこの運動が進んだのですが、アメリカの SPARC US がアメリカの大学などに対して、これへの参加を促す大キャンペーンを始めて、うまくいきそうな機運が出てきました。日本はずっと国内のコンセンサスが取れなかったのですが、2011 年に図



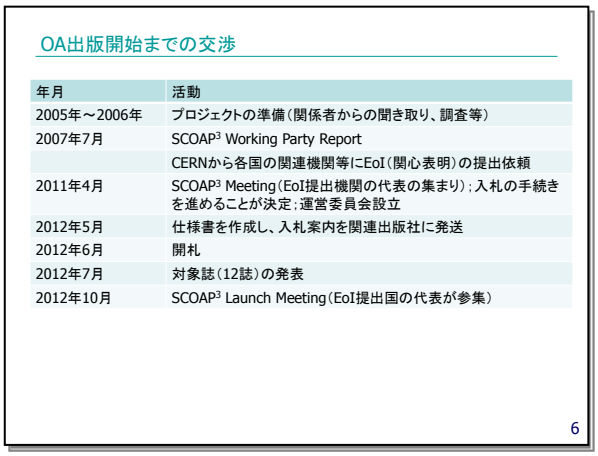
（図 2）



（図 4）



（図 3）



（図 5）

書館を中心に再検討し KEK とも合意して EoI を出して始まりました。

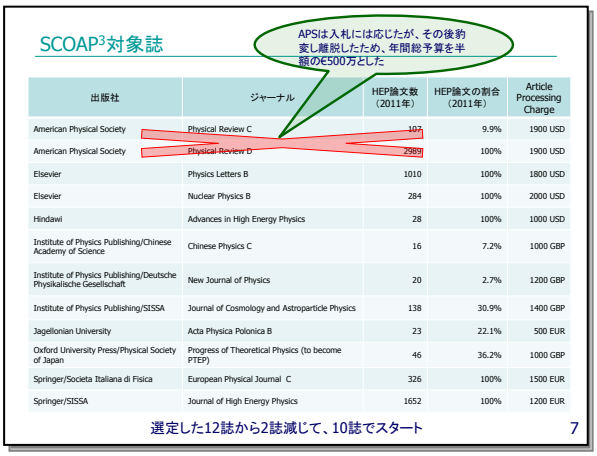
2012 年、SCOAP³ は仕様書を作り、入札案内を関連出版社に送り、開札して雑誌を決めました。この入札は、日本の国立大学の政府調達と大体同じようなもので、総合評価、つまり APC の額、インパクトファクター、その他の条件を点数で評価し、点数の良い雑誌から買うという入札を行いました。EoI を出す国が少しずつ増えて、必要とする 14 億円のうち 60%ぐらいまでは賄えそうだということになり、OA 化を始めると決定しました。

当時の日本では、図書館が今まで高エネルギー物理学の雑誌のために払っていたお金を CERN に払うように振り替えてもずっとはうまくいかないだろう、オープンアクセスで絶対に出てくるタダ乗りはどうするのかという議論がありました。また、基本的にはオープンアクセスは著者支払いモデルです。根岸正光先生（国立情報学研究所）が Journal Citation Index を駆使して、日本の大学がそれぞれ何本論文を書いているのかを調べたところ、東京大学が一番多くなっていました。ということは、東京大学が一番払わなければいけないモデルなのです。恐らく、これは既存の購読型の経済に極めて重大なインパクトを与えるだろうという予感がありました。また、他の分野の先生から、高エネルギー物理学以外の分野の雑誌の値段高騰に跳ね返るような悪影響を避けるようにと、きつく言われました。つまり、高エネルギー物理学の雑誌に論文を出すコストが安くなったとして、出版社が別の分野で減った分を取り戻すようなことになってはならないということです。

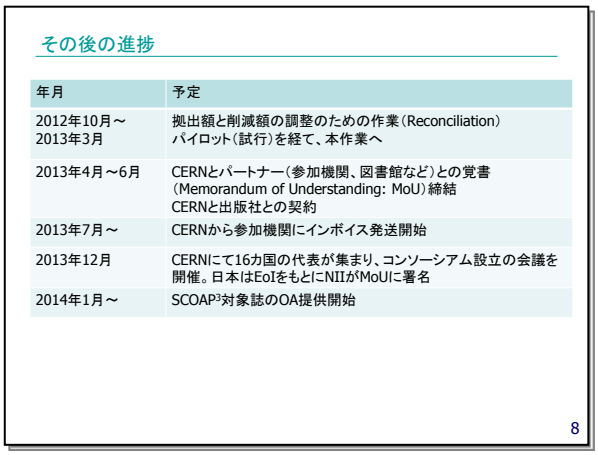
さらに、これは切歯扼腕（せっしやくわん）なのですが、SCOAP³ が始まる少し前に、京大がもともと発行し日本物理学会に移行した「Progress of Theoretical and Experimental Physics（PTEP）」という雑誌がオープンアクセスになったのです。それまでは日本物理学会の雑誌だということで、多くの物理の先生が買ってくれていました。その購読料がオープンアクセス化により消えてしまった後で SCOAP³ が始まったのです。PTEP の購読料が SCOAP³ への拠出金に含まれればよかったのですが、タイミングが悪く叶いませんでした。

結果として、図 6 の 12 誌が入札で残りました。しかし、残念なことに、アメリカ物理学会（APS）の「Physical Review C」と「Physical Review D」は、入札した後で態度を変えて脱落しました。総論文数の約半分を占めていたので、総額を 1,000 万ユーロの半分の 500 万ユーロにして、残りの 10 誌で SCOAP³ が始まりました。

まず行ったのが reconciliation という作業です。これにより各参加機関の拠出額が定まり、2014 年 1 月から、これら 10 誌のオープンアクセスが開始され、今はその 3 年目です。従って、今は次のフェーズ、次の 3 年間をどうするかという議論の真っただ中にいます（図 7）。reconciliation とは、図書館ないしコンソーシアムと出版社との間で拠出額と削減額の調整を行う一



（図 6）



（図 7）

連の作業です（図 8）。例えば、ビッグディールで契約しているエルゼビアの雑誌 2,000 誌のうち、高エネルギー物理学の雑誌 2 誌の占める割合が幾らかを計算するのです。出版社によってはどういうわけかこのタイミングに合わせたように雑誌の価格が安くなったケースもあります。定価表と契約書にある金額でビッグディールの契約を幾ら安くするかという計算を出版社も大学図書館もします。その値を双方でぶつけて合意が取れるまで、それを繰り返す作業をしました。

日本の大学は大体は問題なく終わりましたが、マックス・プランク研究所やカリフォルニア大学は出版社とかなり厳しい交渉を行ったようです。ビッグディールの時代になり、契約している個別雑誌価格がよく分からなくなりましたが、他国には用意周到に対応してきた国もあるようで、reconciliation を通じて削減額をなるべく大きくするような交渉を続けたようです。

その後、各国で MoU を結んで SCOAP³ が始まりました。日本は私どもの NII が National Contact Point になり MoU に署名しました（図 9）。また、Executive Committee に私が入り、いろいろ貴重な情報を得ることができています。

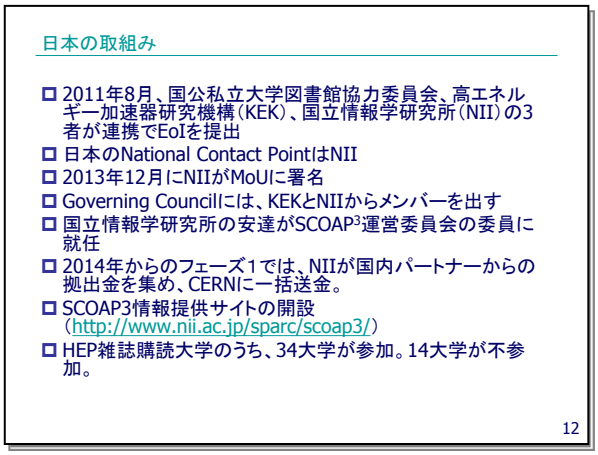
NII が毎年日本のパートナーの図書館に請求書を送りお金を集め、それを一括して CERN に送る形で進めています。先ほどの 10 の雑誌を購読している大学のうち、日本では 34 大学が参加して、14 大学が不参加です。この 14 大学にはぜひ参加していただきたいと思っております。

SCOAP³ の成果

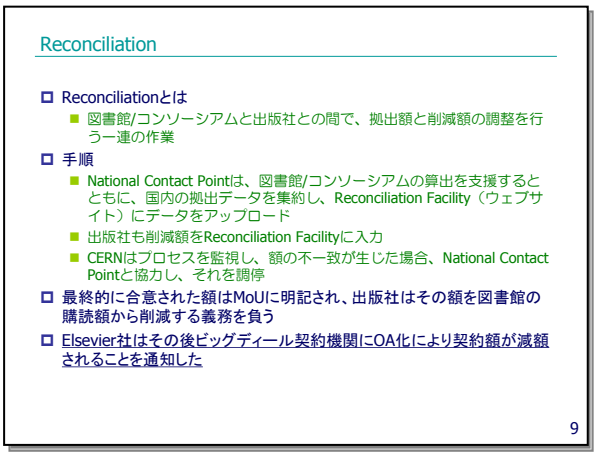
フェーズ 1 の最初の 3 年間の成果は、44 カ国から 3,000 を超える図書館などのパートナーが参加していることです。10 の雑誌をオープンアクセスにし、1 万 3,000 論文以上が出版されました。日本からは先ほどの 34 大学がパートナーです。

図 10 の紫色が参加国で、黄色は著者がいる国です。主な不参加国はインドとロシア、ブラジルです。旧共産圏の国際的な機関があり、まとめて東ヨーロッパ各国や北朝鮮、ベトナムなどが入っています。ほぼ主要な国は入っている状況です。

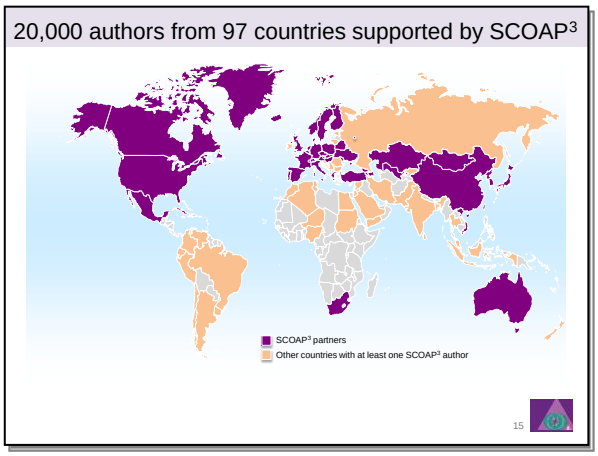
図 11 が論文の内訳で、68%が学会誌であることが一つの特徴です。エルゼビアの雑誌が二つ入っています。シュプリンガーおよび英国物理学会出版局（IOPP）もあります。オックスフォード大学出版局（OUP）の雑誌とは、先述の PTEP という日本物理学



（図 9）



（図 8）



（図 10）

会の雑誌です。ドイツ、中国、ポーランドの物理学会の学会誌も入っています。

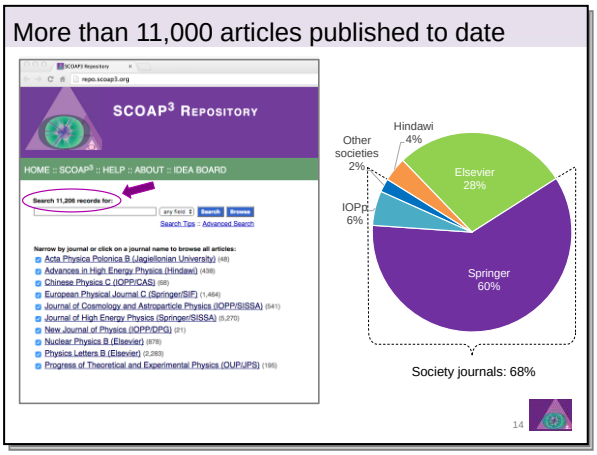
図 12 が実績です。IOPP3 誌、Jagiellonian Uni、Oxford Uni Press が学会誌です。入札で提示してきた APC です。実際には 2011 年に想定した論文数よりも 1,000 論文ぐらい多く、その上限数を超えた APC をキャップする仕組みによって、実質的な APC は 1,100 ユーロぐらいにまで下がっています。この実質的な APC を下げたことが一つの重要な成果になっています。

尾城さんのプレゼンにあった「大学は一体幾ら払えばいいのだ」という課題と、APC を抑えることは直結する話です。

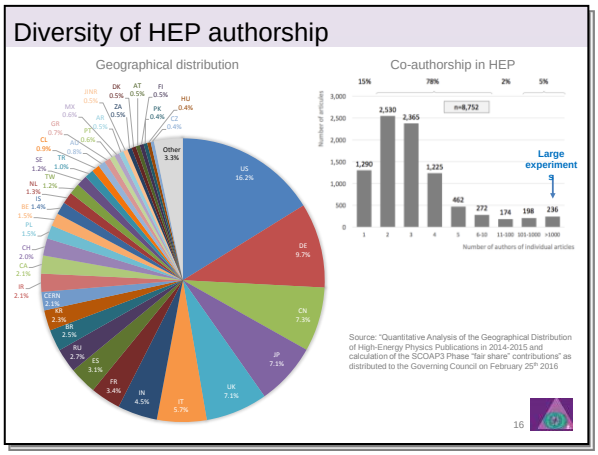
図 13 が著者の人数分布です。高エネルギー物理学では著者が 1,000 人以上いる論文が書かれることもあり、他の分野の研究者を驚かすのですが、それは 236 論文にすぎず、理論の人は単著の論文も多いことが分かります。

また、フェーズ 2 を考えるに当たって、この各国の比率が拠出金額になります。残念ながら APS の雑誌が入らなかったため、アメリカの比率は 16%まで落ちました。ドイツが 2 位で、中国が 3 位に躍進、そして、日本、イギリス、イタリアと続くという出版状況です。

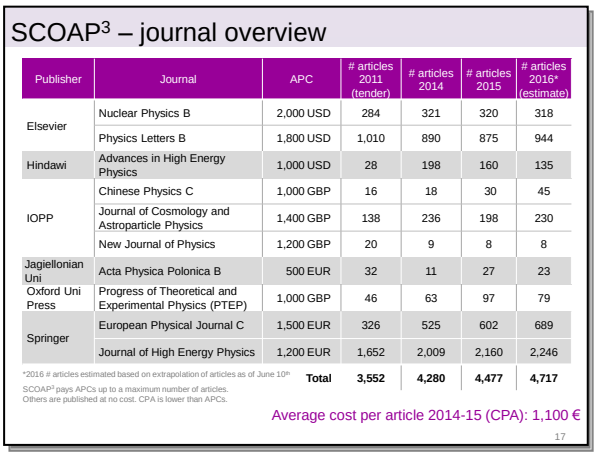
図 14 は成果の説明です。イギリスのウェルカム財団等のオープンアクセスの場合、コンプライアンス、つまり書誌の体裁や項目をきちんとそろえた電子データがリポジトリにきちんと納入されるかどうかが完璧に管理されていないようですが、SCOAP³ では非常に厳しくチェックし、そのような不備を回避できています。



（図 11）



（図 13）

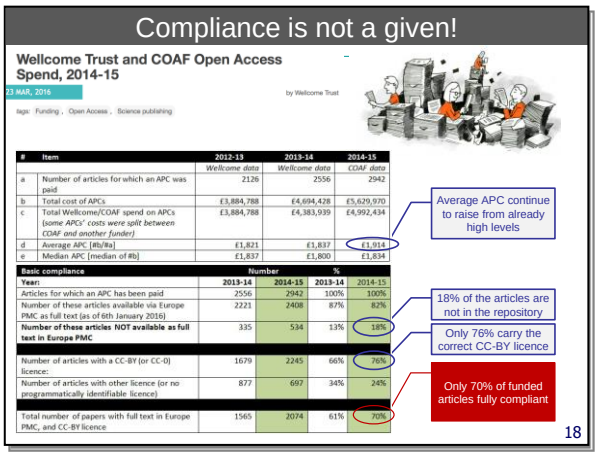


Publisher	Journal	APC	# articles 2011 (tender)	# articles 2014	# articles 2015	# articles 2016* (estimate)
Elsevier	Nuclear Physics B	2,000 USD	284	321	320	318
	Physics Letters B	1,800 USD	1,010	890	875	944
Hindawi	Advances in High Energy Physics	1,000 USD	28	198	160	135
IOPP	Chinese Physics C	1,000 GBP	16	18	30	45
	Journal of Cosmology and Astroparticle Physics	1,400 GBP	138	236	198	230
	New Journal of Physics	1,200 GBP	20	9	8	8
Jagiellonian Uni	Acta Physica Polonica B	500 EUR	32	11	27	23
Oxford Uni Press	Progress of Theoretical and Experimental Physics (PTEP)	1,000 GBP	46	63	97	79
Springer	European Physical Journal C	1,500 EUR	326	525	602	689
	Journal of High Energy Physics	1,200 EUR	1,652	2,009	2,160	2,246
		Total	3,552	4,280	4,477	4,717

*2016 # articles estimated based on extrapolation of articles as of June 10th

SCOAP³ pays APCs up to a maximum number of articles. Others are published at no cost. CPA is lower than APCs.

（図 12）



#	Item	2012-13 Wellcome data	2013-14 Wellcome data	2014-15 COAF data
a	Number of articles for which an APC was paid	2126	2556	2942
b	Total cost of APCs	£3,884,788	£4,694,428	£5,629,970
c	Total Wellcome/COAF spend on APCs (some APCs' costs were split between COAF and another funder)	£3,884,788	£4,383,939	£4,992,434
d	Average APC [#b/#a]	£1,821	£1,837	£1,914
e	Median APC [median of #b]	£1,800	£1,800	£1,834

Basic compliance	Number		%	
Year:	2013-14	2014-15	2013-14	2014-15
Articles for which an APC has been paid	2556	2942	100%	100%
Number of these articles available via Europe PMC as full text (as of 6th January 2016)	2221	2408	87%	82%
Number of these articles NOT available as full text in Europe PMC	335	534	13%	18%
Number of articles with a CC-BY (or CC-0) licence:	1679	2245	66%	76%
Number of articles with other licence (or no programmatically identifiable licence)	877	697	34%	24%
Total number of papers with full text in Europe PMC, and CC-BY licence	1565	2074	61%	70%

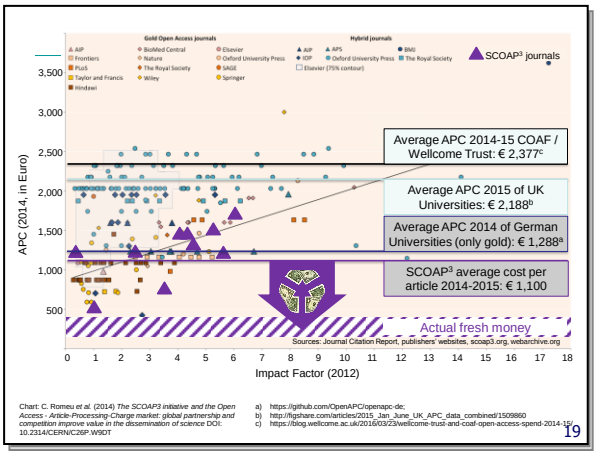
（図 14）

図 15 が APC の効果の説明です。ウェルカム財団やイギリスやドイツのオープンアクセスの平均 APC よりも、SCOAP³ は APC を下げました。結果として「もし SCOAP³ がなかったら図書館はもっと APC を払わなければいけなかった、SCOAP³ はその余裕分をつくったのだ」ということが一つの成果として主張されています。

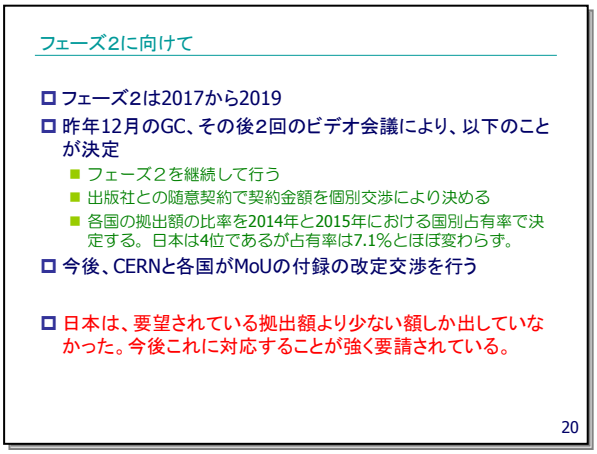
今後の方向性

さて、フェーズ 2 に向けての方針が決まりました。フェーズ 2 は全員賛成で継続するということになりました（図 16）。また、出版社とは入札ではなく、随意契約で金額を決めることになりました。論文のシェアでは、日本は、中国の躍進により、4 位に一つ下がったのですが、占有率は 7.1%とほぼ変わりません。

また、IOPP の「New Journal of Physics」と「Journal of Cosmology and Astroparticle Physics」という二つの雑誌が SCOAP³ から離脱することになりました。他の出版社が全部 APC を下げてきたのに、逆に高い価格を要求してきたので、交渉が決裂してしまいました。これに合わせて、各国の拠出額を調整し、MoU の改定を行います。日本は期待される額よりも少ない額しか出していなかったので、今回もっとたくさん出し、不足分を埋めることが強く求められています。



（図 15）



（図 16）

アメリカは SCOAP³ に関する限りは立派です。フェーズ 1 では 24%を占める額を彼らは多くの図書館から集めて全額払っています。そして、24%が 16%に減ったにもかかわらず、今までの額を維持する見通しだそうです。ドイツも全部出すと言っています。ドイツの大学は日本の大学と同じようにいろいろ苦しいのですが、日本の学術振興会に相当するドイツ研究振興協会（DFG）が差額を補填して全額出します。中国も全部出すと即答しました。日本は今、検討中です。イギリスは英国情報システム合同委員会（JISC）が全部出すと言っています。

お分かりのように、論文数上位の国は、従来の購読モデルでは少ない購読料で済んでいたので、論文出版数に応じたモデルに従うと、追加で払わなければならないことになります。ところが、あるところから下の国にとっては、今まで高い購読料を払っていたのに reconciliation をして金額を計算すると、SCOAP³ に拠出する金額は今までよりも少なくて済む、SCOAP³ は大変良いシステムだということになります。従来の購読料のモデルでは、小さい国からそれなりの購読料を取って、大きい国は相対的に安く論文を読んできた状況が分かります。これをどう解消するかがオープンアクセス化のマクロな問題です。

国際的には国単位で問題を線引きするというのが SCOAP³ の考え方で、国内問題は国内で解決しなければいけません。論文数の占有率は伝統のある大きな大学が高いのですが、先ほどと同じ国別の状況が国内にも存在し、著者支払いモデルでは、生産論文数比率に応じて負担することが最終的な到達点となります。わ

が国でこの問題をどう解決するかが SCOAP³ の呈示する最大の問題で、OA の根本的問題と言うことができます。

日本での議論について

3〜4 年ほど前に行った議論の「リダイレクションでうまくいくのか」ということを振り返ってみたいと思います。今まで購読料として払っていたお金を振り替えればいいというのは、あくまでも初期値です。プロジェクトをスタートするときに、新しい金をどこからか探してくるのは大変なので、購読料の振り替えであれば新規持ち出しは取りあえずしなくて済む。従って、まずそこから始めようという提案により、出発時の合意を取りやすくしたというわけです。問題を全て解決するための合意ではないのです。初期値が今まで払っていた金額よりも少ない国もあれば、多い国もある。多い国は余力ができてハッピーなのですが、今まで少なく払っていた国は、その差額をどこかから持ってこなければいけないことになります。わが国は今これに対処することが求められています。

フリーライダーはどうするか。これは研究機関の意識とアカウンタビリティの問題です。研究助成機関などが強い理念と方向性を示すことが重要だと考えられます。

基本的にオープンアクセスは著者支払いモデルで進めようという暗黙の合意があるように思います。OA を進める際にはこの点が極めて重要です。これを受け入れるのは、研究者の知識と見識の問題になります。

次に、研究助成機関からの効果的な誘導策が必要です。他の分野の雑誌にしわ寄せが行くのではないかということですが、これへの対応は簡単で、全ての分野をオープンアクセスにすればいいのです。「こんなこと無理だ」とお笑いになるかもしれませんが、昨年始まったマックス・プランク研究所の OA2020 はそれを言っているのです。3 年前は、高エネルギー物理学のような特殊な分野だからできる話でしょうと、皆さんが言っていました。しかし、状況は急速に変わっているように思います。

PTEP の問題については、国際学術誌としてさらに発展していただきたいと本当に思います。しかしまだ著者の多くが日本人です。

私と尾城さんで数年前につくった案を紹介します。日本学術振興会（JSPS）は約 2,000 億円以上の研究費を配分しているのですが、その 1%、20 億円でも基金にして、APC を補填するための助成金として何年間かプロジェクトを実現することができないかというものです。そのときに JUSTICE が間に入って、科研費でどれくらいのオープンアクセス論文を書いているかをきちんと把握し、そのデータに基づいて JSPS から APC 相当分のお金を払い戻してもらうという仕組みにして、とにかく日本から APC として払っているお金を適正な額までに抑えるようなメカニズムをつくらないと、オープンアクセスは破綻するでしょう。

まとめ

1 点目、国際的なオープンアクセスの潮流の中で、日本は科学先進国として協調していくことが重要です。G7 の科学技術大臣会合などではこのようなレベルでの方向性を言っています。日本が科学先進国であることは事実で、高エネルギー物理学分野では 3 位か 4 位を占めています。

2 点目、世界的には、この種の話は、図書館がリーダーシップを出す問題であると認識されています。日本では誰が主体なのかよく分かりません。

3 点目、OA2020 が夢物語のように提案されていますが、SCOAP³ がその trail blazer です。SCOAP³ で起こるさまざまな課題を実践的に解決していくことがオープンアクセスを深化させることに直結しています。そのような実践的プロセスなしに未来は開けません。

4 点目、今まで以上に研究者の協力と援助が必要です。

以上、オープンアクセスの実践的なプロジェクトの紹介でした。

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

大学図書館における オープンアクセスの取組み

荘司　雅之
（早稲田大学図書館）

CC BY NC SA

講演要旨

大学図書館におけるオープンアクセスの取り組みは、主に機関リポジトリにより展開されてきた。積極的に商業雑誌掲載論文を収集し「グリーン」を進めている図書館、あまり商業雑誌掲載論文収集に積極的になれなかった図書館などについて機関リポジトリを運営している立場から紹介する。

荘司　雅之

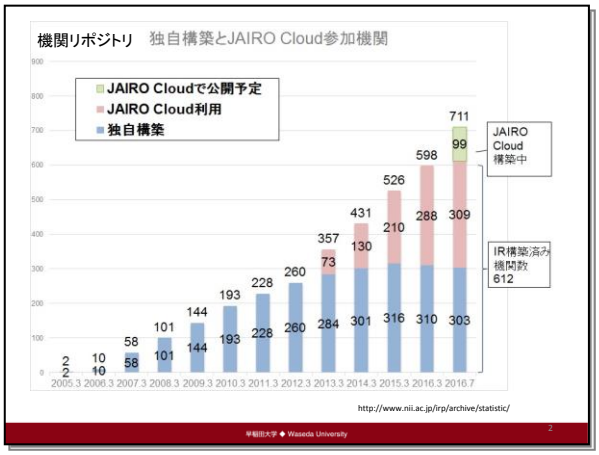
早稲田大学図書館事務部長。早稲田大学情報科学研究教育センター、情報システムセンター、図書館学術情報課、メディアネットワークセンター等のコンピュータ関連部門に勤務した後、図書館学術情報課長、理工学図書館担当課長、高田早苗記念研究図書館担当課長、図書館事務副部長兼総務課長を経て現職。

私からは、過去から今まで、大学図書館が取り組んできたオープンアクセスの取組みをご紹介します。主に機関リポジトリについての話になります。

今までの機関リポジトリ

図 1 は NII のサイトから頂いた資料で、機関リポジトリ独自構築数と JAIRO Cloud の参加機関のグラフです。2005 年は 2 機関だったものが、2016 年 7 月には構築中を含めると 711 です。恐らくこれは世界最多のリポジトリ設置数になるのではないかと思います。ちなみに、2005 年の 2 のうちの一つが早稲田大学です。

図 2 は機関リポジトリの現在のコンテンツの内訳です。今、グリーン、ゴールドと言っているのは journal article（学術雑誌論文）に分類されるものです。これは IRDB（Institutional Repositories DataBase：学術機関リポジトリデータベース）のコンテンツ分析のサイトから取ってきました。サイトには IRDB のコンテンツ増減のグラフが 2007 年から載っていて、journal article



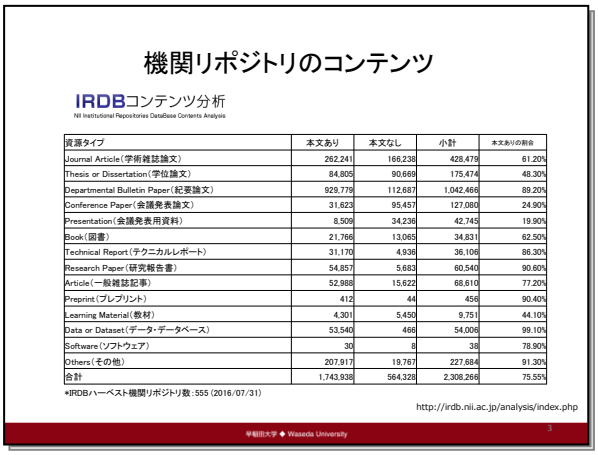
（図 1）

は約 15%あることになっています。

尾城さんの発表で、トムソン・ロイター社のデータを基にした集計作業の結果、2014 年に日本人が出版した論文 6 万 4,247 のうち 9,581 がフルオープンアクセスになっていました。リポジトリの場合は過去のものを登録することがあるので、出版年と登録年が必ずしも一致しないのですが、調べてみると、1 年間で 2 万 103 件増えています。9,581 の倍ぐらいの数が、リポジトリでグリーンオープンアクセスになっているということです。

2005 年にリポジトリをつくったときには、何だかよく分からないものでしたが、コンテンツ登録のために何をするかははっきりしていました（図 3)。まだデジタル化されていない学術資料は何らかの権利処理をしてデジタル化し、メタデータを作成して登録します。デジタル化されている学術資料は、権利処理を行ってメタデータを作成して登録します。登録作業は、当時はセルフアーカイブと図書館員が行う登録がありました。

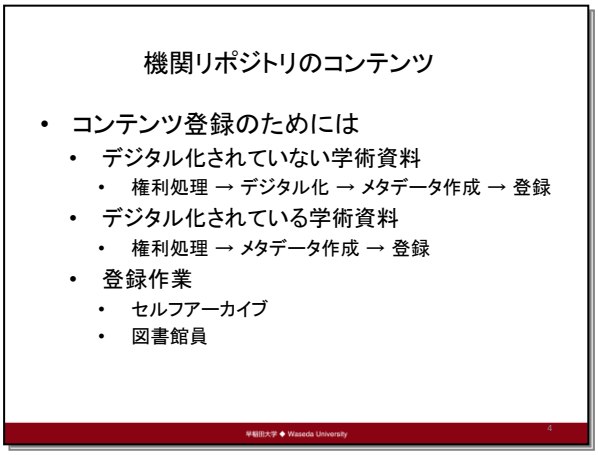
われわれは DSpace というマサチューセッツ工科大学（MIT）がつくったリポジトリサービスを導入しました。MIT の体制はコアサービスとプレミアムサービスがあり、コアサービスは MIT コミュニティに対して無料で提供されているもので、研究者がコンテンツ登録のための作業を全部自分でできるというものです。プレミアムサービスは代行サービスで、図書館員が代わりに権利処理、デジタル化、メタデータの作成をやってあげるというものです。当時はそのようになっていました。それを愚直に信じていたので、それぐらいだったらリポジトリはつくれるのではないかと思いました。MIT の体制では、図書館員が関わるのは有料サービスだったというのが 2003～2004 年ごろの話です。



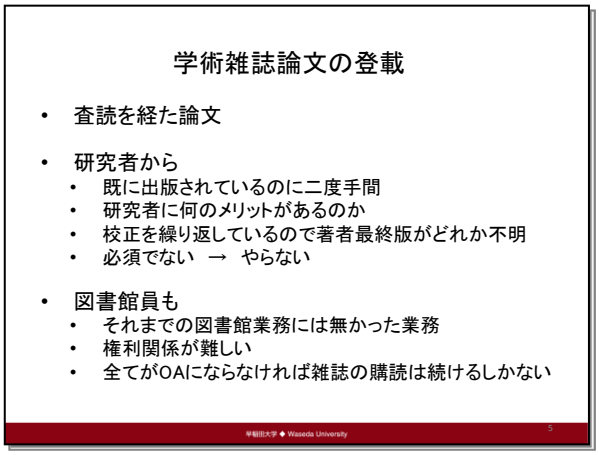
（図 2）

学術雑誌論文の登載

グリーンと呼ばれている学術雑誌論文の登載については、査読を経た論文は権威があるので、コンテンツとして非常に重要だろうと思いました（図 4)。当時の研究担当理事をはじめ、いろいろな先生にインタビューしてみたところ、既に出版されているので二度手間である、つまり、自分がいる研究者コミュニティでは論文はみんながもうシェアできている、それをなぜわざわざまたやらなければいけないのだという話が多



（図 3）



（図 4）

かったです。また、著者最終版を入れてくださいと言ったときにページの記載もないし、そのようなものを出してもきちんと引用してくれないのだったら意味がないということ、いろいろな校正のやりとりをしているから、自分の手元にあるどれが最終版かよく分からないという話もありました。一番多かったのは、やはり研究者に何のメリットがあるのかと問いただされることです。当時はきちんとした答えは返せていませんでした。また、やらなければいけないのか聞かれて、これはオプションだと答えると、必ずやらないということになります。

図書館員も、コアサービスだけであればシステムで何となく解決できるのですが、それまでの図書館業務になかった業務、例えば、権利処理、メタデータの作成、リポジトリシステム自体の立ち上げといった多くの業務が増えることになります。それでオープンアクセス誌ができたから、雑誌の購読料はタダになればいいのですが、そんなことはなく、例えば早稲田大学全体、日本全体の著者の論文がオープンアクセスになったとしても、中国やインドの著者の論文がオープンでなかったら、その雑誌の購読は続けなければいけません。ということは、人件費や何かの新しいバジェットを取ってくることが不可能で、全くメリットになっていませんでした。

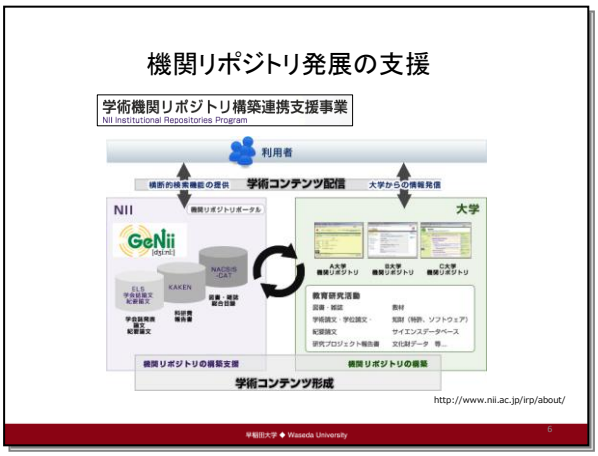
機関リポジトリ発展の支援

そこでどうにかできないかと考えていたのですが、NII で学術機関リポジトリ構築連携支援事業を平成 16 年ぐらいから平成 24 年まで行っていただきました（図 5）。学術機関リポジトリ構築支援事業では、委託事業というものがあり、リポジトリを立ち上げる事業、コンテンツを増やす事業、先進的な取組みの事業等々を支援していただきました。そして、コンテンツ登録が図書館員の業務にはなく、お金もなかった状況から、最初にご覧に入れたグラフのように、リポジトリの数が増えていきました。

難しかったのが権利関係ですが、それに関しては、例えば、日本の学会誌について、著者版を登載して良いのか悪いのか等をデータベース化したのが学協会著作権ポリシーデータベース（SCPJ データベース）です（図 6）。このようなデータベースは、世界的には SHERPA/RoMEO があり、雑誌名や出版社名を入れると、その雑誌が査読前・査読後の両方についてリポジトリ登載を許諾するもの（グリーン）か、査読後のみを許諾するもの（ブルー）か、全く許諾しないもの（ホワイト）かが分かるようになっています。

また、リポジトリは新しい仕事でノウハウがないので、図書館員同士でノウハウを共有して進めていくためのデジタルリポジトリ連合（DRF）という組織もできました。

学術機関リポジトリ構築連携支援事業は平成 24 年度までで終わり、その後を引き継いで、NII と国公私



（図 5）

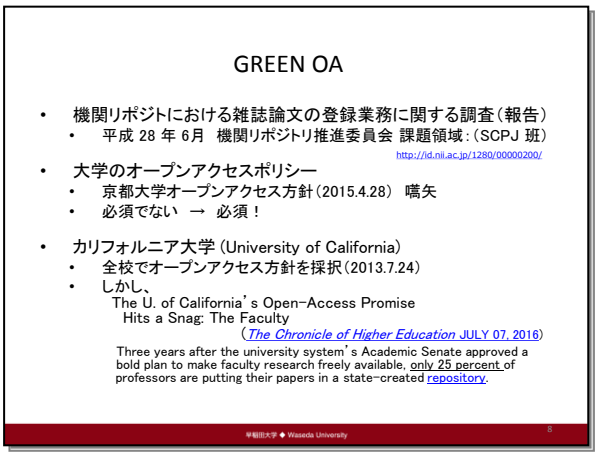
（図 6）

立大学図書館協力委員会の連携・協力推進会議の下に、機関リポジトリ推進委員会が発足しました。そこと、まだ少し残っていた DRF など、今まであったいろいろな機関を 2016 年 7 月にオープンアクセスリポジトリ推進協会という形に統合して、推進していこうとしています。

機関リポジトリ推進委員会から、「機関リポジトリにおける雑誌論文の登録業務に関する調査（報告）」が 2016 年 6 月に出ているので、こちらもご覧ください（図 7）。図書館員が積極的に論文を集めているところなど、いろいろなコメントを読むと面白いものがあります。

先ほど、必須でないことが障壁になったと述べました。大学自体のオープンアクセスポリシーは海外では結構あったのですが、日本では京都大学が 2015 年 4 月 28 日に「京都大学オープンアクセス方針」を策定して、その後も 7〜8 大学・機関が続いています。これで必須でなかったものが必須になったので素晴らしいと思っていたところ、カレントアウェアネスである記事を発見しました。カリフォルニア大学全校では 3 年前からオープンアクセス方針を採用しているが、3 年たってもたった 25%の研究者しか登録していないという「The Chronicle of Higher Education」の記事でした。これは最初オプトアウトを取っていたそうで、出版社が文句を言ってきて駄目になったのかと思ったのですが、記事を読むと、出版社が文句を付けて落としたのは 5%にすぎないということでした。やはり研究者の理解と協力がカリフォルニア大学でもあまり進んでなかったのではないかと思います。

この先はディスカッションでお話しします。



（図 7）

第 1 回 SPARC Japan セミナー2016

「オープンアクセスへの道」

生命科学分野における
研究者の投稿先雑誌選択趣向と OA への意味づけ

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）

講演要旨

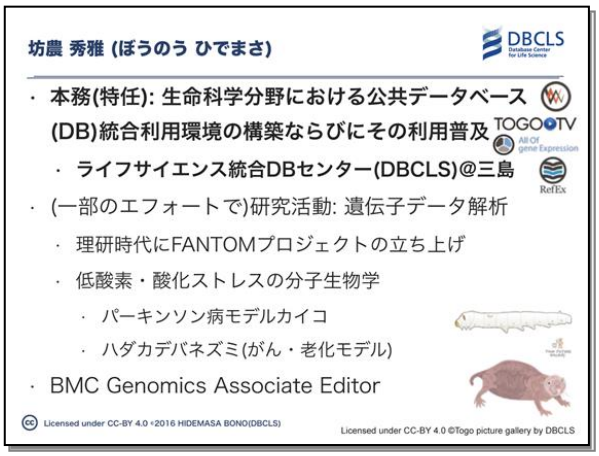
オープンアクセス（OA）誌に掲載された論文は、インターネット検索エンジンの発達に伴い、目に触れる機会が格段に増えることとなり、その結果 PLOS（Public Library of Science）や BMC（BioMedCentral）といったゴールド OA 誌に投稿される論文が増え、古くからある論文誌も OA オプションが選べるようになってきた。その一方、OA 紙で必要となる論文掲載料は、研究費から捻出する状況となっており、実験するのに必要不可欠な試薬代や消耗品代を圧迫し、研究活動を行う上で無視できない存在となっている。OA 紙に論文を載せて広く自分の研究を知ってもらいたいが、それにはお金がかかる、そういった状況となっているのが現状である。本講演では、生命科学研究者の周辺でおこってきた投稿先雑誌選択趣向の変化について実経験に基づいて紹介する。

坊農　秀雅

理化学研究所においてFANTOM（Functional annotation of mouse）プロジェクトの立ち上げに関わった後、埼玉医科大学ゲノム医学研究センターを経て、2007年7月より情報・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（DBCLS）にて統合データベースプロジェクト（統合DB）に従事。統合DBの広報・普及活動として統合TVや統合データベース講習会AJACSの立ち上げに関わり、現在は国立遺伝学研究所にて日本DNAデータバンク（DDBJ）と連携して大規模塩基配列データ利用環境の構築とそれを利用した生命科学研究を行っている。京都大学博士（理学）。

私は生命科学の研究者で、普段はライフサイエンス統合データベースセンター（DBCLS）にいます（図 1）。本務は生命科学分野における公共データベース統合利用環境の構築ならびにその利用普及で、利用普及のためのウェブツールを作成したり、講習会を行ったりしています。また、遺伝子配列データを使ったデータ解析もしています。このようなことをずっと 20 年ぐらい続けています。もう一つ、「BMC Genomics」のアソシエートエディターも務めています。

私は生命科学の分野にいますが、それが全てではなく、researchmap で私のエントリーを見ていただくと図 2 に書いてあるように、研究分野はゲノム科学、分



（図 1）

子生物学などなので、こんなことを言っている人もいると参考程度に受け止めてください。

生命科学研究者から見たオープンアクセス

私は生命科学の研究者と普段よく話をしたり、学会へ行ったりしますが、「オープンアクセス」という言葉は知っていても、ゴールド、グリーンという話は誰も知りません。そのような話になったことは一度もありません。生命科学の研究者がオープンアクセスと言うと、それはゴールドのことを指しています。

グリーンオープンアクセスへ出すということは、生命科学者から言ってみればインパクトファクターの付いていない雑誌に出すようなもので、よほど特殊な事情がない限り、まず出さないです。また、特に生命科学がそうなのかもしれませんが、ピアレビューを経たものでないと論文の業績として認めないという伝統があります。ピアレビューを経ずに、誰からでも見られるようにすることには強い抵抗があります。

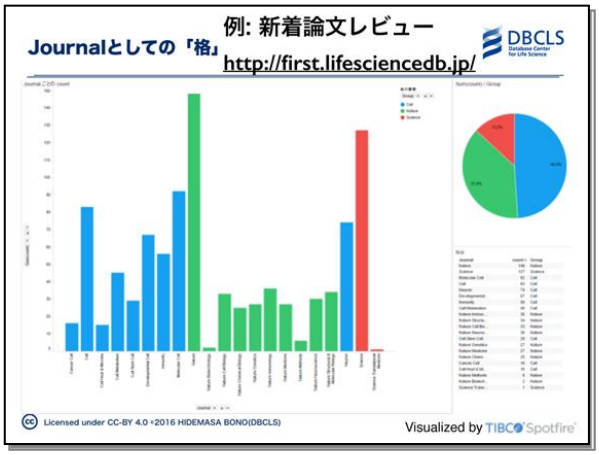
私が思う研究者から見たオープンアクセスの良い点は、誰にでも自分の論文を場所を問わず見てもらえることです。インターネットでキーワードを検索されれば自分の論文がヒットします。生命科学でもいろいろな分野があって、自分の専門誌の人たちが見て、その人たちだけで分かっていればいいという時代は終わっています。他の人に見てもらって広めていかなければ駄目で、専門誌に出しても見られないのが現実ではないかと私は思います。

ネット検索に引っ掛かることは重要です。別にPubMed に入っていればそれでいいではないかといわれますが、PubMed は抄録部分、アブストラクトだけです。一方、オープンアクセスに出していると全文があるので、検索ワードがどこかにヒットすれば、興味を持って論文を見てもらえるので、やはり全文がインターネットで検索されるのが良い点ではないかと思います。

悪い点は APC の問題です。今は APC を自分の研究費から出していることがほとんどです。もちろん大学によっては手当てがあるのかもしれませんが、基本的には自分が取ってきた少ない研究費の中から出すのが普通です。また、オープンアクセス誌は長くても 10 年ぐらいですから、ジャーナルとして格がないのです。格がある雑誌とは何かというと、インパクトファクターで測る場合もありますが、「いい雑誌」とみんなが言うものです。

DBCLS では、新着論文レビューを行っています。これは、一般に生命科学者がプレステージャスジャーナル、いいジャーナルと言っている約 20 誌に掲載された論文について、著者自身にレビューを書いてもらい、誰でも自由に閲覧・利用できるよう Web 上にていち早く無料で公開するサービスです（図 3)。既に 6 年間やっていて、レビューは 1000 ぐらいあります。この対象になっている雑誌は飯田啓介という編集長が決めたものですが、これが生命科学で言うところのいい雑誌だという意見もあります。

（図 2）



（図 3）

論文の探し方の変化

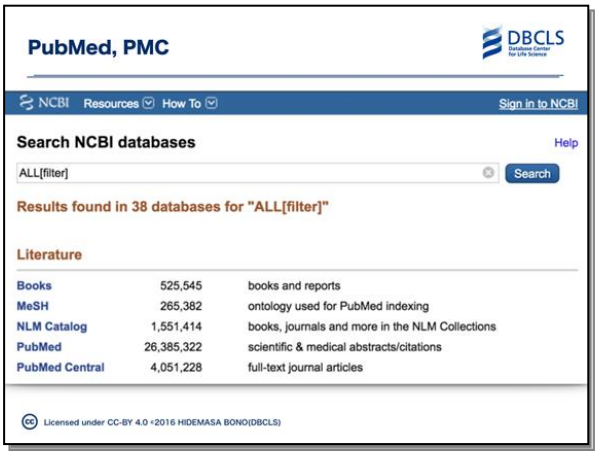
そのような良い雑誌とは昔からある雑誌がほとんどですが、論文と一口に言っても、昔からある雑誌、最近になって出てきた雑誌というように、論文の探し方は私の中でもだいぶ変化しました。私は研究を始めて20年ぐらいですが、研究を始めた当時はまだPubMedが試用段階でした。そのころは図書館に来ている雑誌を見に行っていて、そのうちに電子メールが普及して電子メールでタイトルだけは送られてくるようになりました。

今でも私は古い人間なので、電子メールで来るサービスを利用してどんな新しい雑誌が出ているか確認しますが、最近の若い研究者はこんなことをせずに、自分の研究に関係ありそうなものをPubMedで定点観測して、差分を見てそれで合ったものを見ている、また、さまざまなサービスを使って、注目している論文を引用した論文を見ているのではないかと思います。

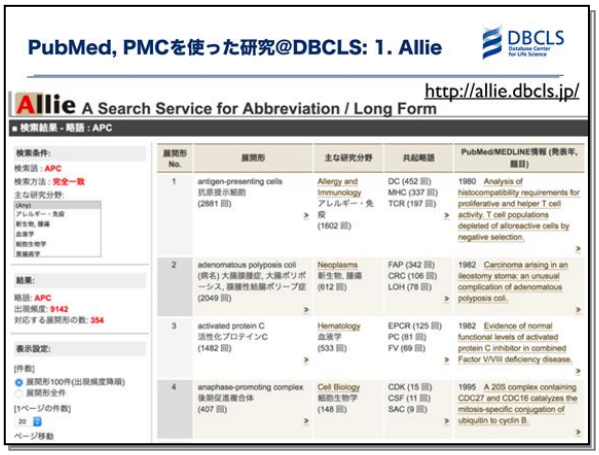
今、PubMedはエントリー数がものすごく多くなり、全文が入っているPubMed Centralも、PubMedの15%ぐらいの割合であるので、昔よりはだいぶ信頼できるリソースになったのではないかと思っています（図4）。

PubMedのデータは全て研究目的で使えるので、DBCLSはそれを使ってサービスを幾つかつくっています。一つは、PubMedの中に出てくる略語が何のことを意味しているのか検索できるAllieというサービスです（図5）。「APC」で検索すると、PubMedではantigen-presenting cells（抗原提示細胞）やadenomatous polyposis coli（大腸ポリポーシス）の略語としてよく使われていることが分かります。ちなみに、article processing chargeの略語としてのAPCは19番目でした。だいぶ下の方です。

もう一つは、Colilという、論文の引用情報に関する検索サービスです（図6）。ある論文が他の論文から引用されているとき、本文中ではどのような文脈で引用されているかについて、効率良く検索します。PubMed IDまたはキーワードで検索すると、論文がコンテクスト、イントロダクションなどどこに引用されたかが出てきます。これは全文が使えるようになっています。限定ですが、PubMed CentralにあるデータだけでこういうものをつくることもDBCLSで行っています。



（図4）



（図5）

（図6）

論文の投稿先、投稿順序

投稿先雑誌の選択趣向については、生命科学者はインパクトファクターの高いジャーナルからとなっています。先ほどから議論を聞いていて「あれ？」と思ったのですが、研究者はジャーナルを選ぶ自由を持っていると思っています。投稿先は、インパクトファクターの高さと、自分がその雑誌に出したいという思い入れ、自分の研究している分野でよく見られている雑誌かどうかで決まると考えています。

論文はピアレビューがあるので、落とされることがままあります。ほとんどの場合は 1 回では通りません。何回も挑戦しなければいけないので、何種類か論文誌のリストを持っていて、それで投稿していきます。インパクトファクターが高いもので PubMed に入っていないものはないのですが、PubMed に収録されていることは必須です。PubMed に登録されていないジャーナルがありますが、それは制約条件になると思います。皆さんは就職するときなどに論文リストを書きますが、PubMed で検索して真偽を確かめることが多いので、ここは必須です。

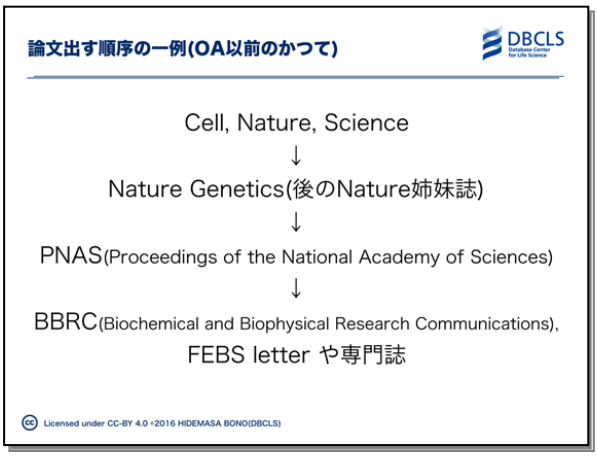
分野によっては所属学会誌に投稿します。高エネルギー物理学だとそうなると思うのですが、私がやっている分子生物学などの分野は所属学会誌が一応ありますが、皆さんあまり出したがらないです。こういう分野もあります。大体、分野の雑誌にはあまり出さないのではないかと思っています。

オープンアクセスがはやる前の論文を出す順序は、私が周りを見たところ図 7 だったと思います。Cell、Nature、Science に出して駄目だったら、だんだん下に行き、最後の受け皿は速報誌や専門誌になるのではないかと思います。

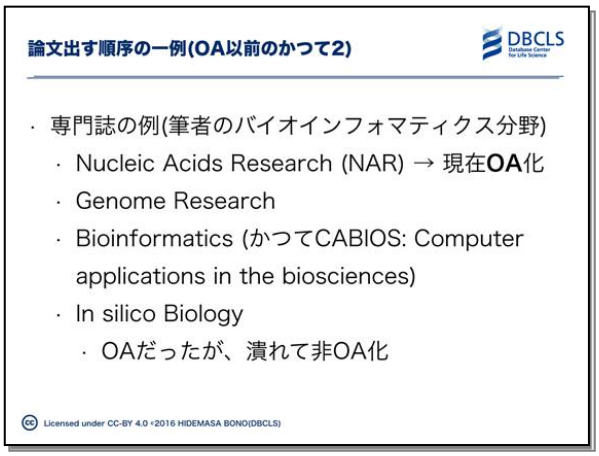
私のバイオインフォマティクスという分野でも、専門誌、ゲノムの研究に特化した雑誌などに出していました（図 8）。いろいろな道筋をたどっていて、Bioinformatics という雑誌は、昔は Computer applications in the biosciences（CABIOS）という変わった名前だったのですが、バイオインフォマティクスという言葉が 1998 年ぐらいからはやりだした途端に名前が変わって、いい雑誌になりました。In Silico Biology という雑誌は一度、オープンアクセスで誰でも全部読めるということを打ち出していたのですがつぶれて、どこかの雑誌社に買われた途端に非 OA 化になってしまったという非常に悲しい歴史があります。私もここに 2 本ぐらい論文があるのですが、今は自分も読めなくなっています。

生命科学には Nature ラブな人が多いのですが、そういう人の場合、Nature に出して通らなくても、いいレスポンスがあれば Nature 姉妹誌に出します（図 9）。太字になっているのはオープンアクセスの雑誌です。それでも駄目だったら、Nature Communications、Scientific Reports です。

私自身は Public Library of Science（PLOS）に出しています。マイクロアレイという、遺伝子の発現を測定する実験ツールを開発した人たちが、実験するときに、



（図 7）



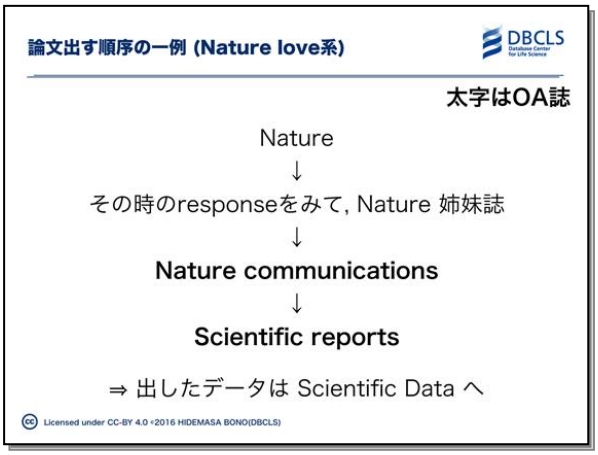
（図 8）

ゲノム全体から新しい遺伝子が発現した、ある遺伝子がすごく発現していたといったことを探し当てるのですが、その遺伝子に関して書いてある論文が読めないことがままありました。そこで、誰でも読めるものを始めようと言って PLOS を始めたのです。

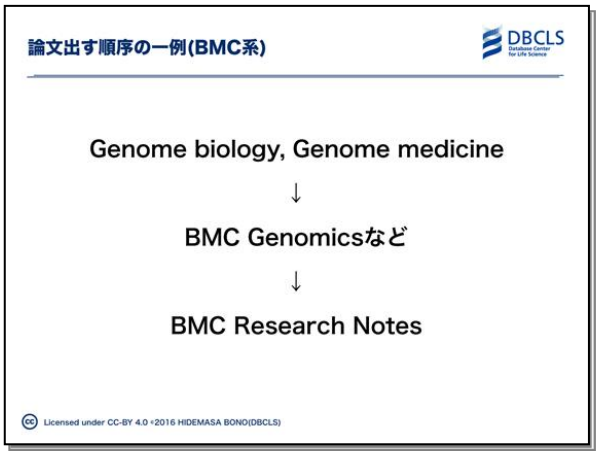
私は PLOS の方が好きで、実際に論文を出すときは図 10 の道筋です。PLOS が一番推している古い機関誌は PLOS Biology で、それ以外の PLOS Genetics や PLOS Medicine など、割と中間のものもたどって PLOS ONE に行きます。いきなり PLOS ONE という場合ももちろんありますが、やはりいい仕事をしたという場合は、皆さんはより上流からトライするようです。

私は BMC Genomics のアソシエートエディターをしているので BMC 系の場合も出しておきます（図 11）。BMC Genomics は多分中間層で、その前には Genome Biology、Genome Medicine といういい雑誌があります。BMC Genomics など中間層で駄目だったら、最終的には BMC Research Notes に出してくださいと言われていますが、これも今後変わるそうです。どうやって変わるか見ものです。

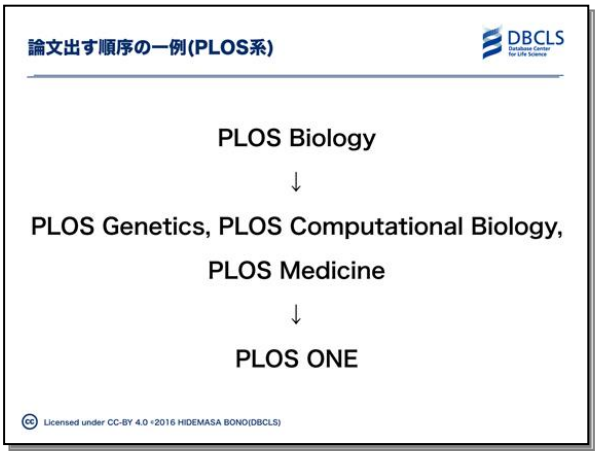
では、ジャーナルはどんなところに出したらいいのかとよく皆さんから質問を受けるので、それを解決するサービスが PubMed のテキストからできないかということでつくったのが Journal Sug-gestion using MedliNe（JSMN）です（図 12）。これは PubMed のアブストラクトとの類似度からどの雑誌に投稿するのがよいかサジェストするものです。今は動いていないのですが、かつて DBCLS が東大の中にあったとき、統合牧場という学生アルバイト組織があり、そこの技術開発部が JSMN をつくりました。その中で生まれてきたのが統合 TV です。



（図 9）



（図 11）



（図 10）

（図 12）

今はこのようなサービスを雑誌社や査読を引き受ける会社がつくっていて、Edanz Journal Se-lector、Journal Guide、Elsevier Journal Finder などがあります。そこに自分の論文のタイトルとアブストラクトを貼り付けると、どんな雑誌がいいかを出してくれますが、自分のところの雑誌ばかり出てなかなか面白いです。

まとめると、生命科学研究者はやはり Cell、Nature、Science に固執してしまっている感があります。それが無理なら姉妹誌、それが駄目なら PubMed に収録されるオープンアクセス誌にという傾向がかなりあると思います。

PubMed に収録されるオープンアクセス誌は、大学によらず全文が見てもらえるし、ネットでインデックスされるので検索したら出てくるようになっています。しかし、それを出すときは APC が掛かり、それは自分の研究費から出すことがほとんどなので、ますます減っている研究費からそれを捻出するのは非常に大変になっています。

グリーンオープンアクセスの余地は？

受け皿をつくってオープンアクセスにすることは、やった方がいいと思います。私もデータを使う人間ですから、そういうものを蓄えることで、今後のテキストマイニングに使えるような材料となればと思います。

PubMed 収録、DOI 付与など、サイテーションの手段を強化して、誰がどのようにアクセスしているか、いつも統計を取って見えるようにすれば、インセンティブになって、皆さんもっとやる気が起きるのではないかと思います。しかし、研究者からしてみれば、一回論文を出したからもういいではないかという気持ちが強いです。論文を一回出してしまうと、続きの研究をされる場合もありますが、忘れてしまっている研究者も多いです。それは、今の論文が、掲載前にピアレビューで評価するという昔のシステムを引きずっているからではないかと思っています。評価は後から皆さんがするものだという意識がもっと広まっていくと、変わっていくのではないでしょうか。

そういう意味で、オルトメトリクス（altmetrics）のツールなどは非常に有効なのではないかと思っていて、よく使うようにしていますが、それすらも面倒ということでやらない人も多いです。

そういうものは研究者サイドからやるのは無理で、ある程度トップダウンで強制にしてしまうしかないのではないかと思います。自発的な動きははっきり言って期待できません。アメリカ国立衛生研究所（NIH）のように、「研究費をもらっているからには出せ」と強制にしなければ無理ではないかと考えています。その辺はいろいろな意見があると思いますが、この後で議論できたらと思います。

2016 年 10 月 29 日に、NII のすぐ近くで「国立遺伝学研究所　公開講演会 2016　出張!!　遺伝研」が行われます。生命科学者の生態が都内で見られますので、ぜひよろしくお願いします。

第1回 SPARC Japan セミナー2016 「オープンアクセスへの道」

ディスカッション 「グリーンOAとゴールドOAと日本としての対応」

山本　和雄　（琉球大学附属図書館）
土屋　俊　（大学改革支援・学位授与機構）
尾城　孝一　（大学図書館コンソーシアム連合（JUSTICE）/東京大学附属図書館）
安達　淳　（国立情報学研究所）
荘司　雅之　（早稲田大学図書館）
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）

●**山本**　今回のセミナーのテーマは「オープンアクセスへの道」ということです。

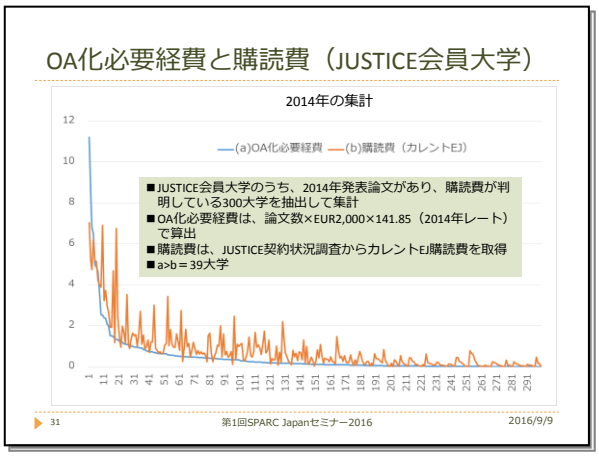
今回、5名の方々に、オープンアクセスに関して、さまざまな観点からご発表いただきました。最初は、それぞれの方から何か追加のコメント等がありましたら、お話しいただきたいと思います。土屋先生のいつもの図書館不要論がまた聞けて、ますます洗練された不要論になってきたので楽しいところなのですが。

今回、図書館の皆さんが大勢参加されているので、その辺が少し気になるところかと思いますが、もともと図書館は、最初からそんなに研究には関わってきていなかったのですよね。今さら不要と言われても、もともと何もしていなくて、電子ジャーナルの世界が始まって、ようやく先生方の研究に図書館が踏み込んできたのだと思います。それがオープンアクセス化ということで、リポジトリのグリーンオープンアクセスの仕事をしていく中で、図書館が先生方とコンタクトを取り、「先生方は実際どんな研究活動をしているのですか」というようなインタビューなどを行って、その中で、大学全体として、よりブランド力を高める、研究成果をきちんと大学の外に出していくという仕事を図書館員が担うようになってきたという流れではないかと思います。

その中で、早稲田の荘司さんなどから、幾つか具体的なお話を頂けたと思います。センセーショナルな不要論がありましたが、まだ少し図書館員の仕事もあるのではないかと見ています。

●**土屋**　安達先生は発表で、アメリカは立派だと言っていましたが、あれはむしろアメリカは図書館員が立派だということですよね。基本的に図書館がお金を払っているのです。けれども、日本の場合には、安達先生が物理学者の話を集めた部分もなきにしもあらずです。要するに図書館側は、「そんなことを言われたって、私たちはお金がありません」と言って終わってしまう部分もあるのです。でも、逆に言えば、その機能を持ち得ないとすれば、ますます図書館は要らない、むしろ負の存在にすらなってしまうのではないでしょうか。

尾城さんの発表にしても、図1のグラフで一番左側



（図1）

にあった大学など、オープンアクセス化必要経費と購読比の差が大きい大学では、誰がどうやってその差を埋めるのかということについてどうお考えなのでしょうか。私はむしろ、尾城さんと安達先生のお考えを伺いたいです。同時に、荘司さんにも、図書館の偉い人としてのお考えを伺いたいです。早稲田大学は左から何番目にいたのかは存じ上げませんが。

●**安達**　最後の方でお話しした案は、実際に日本学術振興会（JSPS）に尾城さんと二人で説明しに行ったものです。もう少し詳しく紹介させてください。尾城さんのプレゼンテーションにあったように、例えば坊農先生のような研究者が研究費でどんどん APC を払っていくと、日本国が外国の出版社に払う経費はどんどん増えていくでしょう。それは当然のことで、例えば研究費の 10～20%がそのための経費として消えていくのを日本の国がよしとするのであればそれでいいのですが、それが捕捉できないということが一番の問題です。捕捉するためのメカニズムとマンデートを組み合わせた案が最後にお話ししたものです。

まず、JSPS が「研究成果をオープンアクセスで出しなさい」とマンデートを課します。その際に「1 年後にその APC を JSPS が還付します、大学単位で金額の計算をやってください」という方策を出します。今は、どの雑誌に幾ら APC を払ったか捕捉できていませんから、学振から還付を受けようとすれば大学内のどこかで集中的に OA 論文をカウントしなければいけません。大学中の先生方がどこに幾ら払ったか、例えばシュプリンガーに APC を 20 万円払ったなどという情報を集め「わが大学全体で今年は APC に科研費から 5,000 万円払いました」と申告すれば、翌年、JSPS がそれを還付します。

仮に今、10%がオープンアクセスの論文だとすると、購読費の 10%というと 20 億～30 億円、科研費の 1%は 20～30 億円です。大体それぐらいのお金を初期のプロジェクトの経費として出して、もちろん 100%還付するかどうかはもっと現実的に考えるべきでしょうが、大学が APC を幾ら払っているのかを捕捉し、JSPS が補填します。これを研究者に直接返すのではなく、大学に返すというのがみそです。機関としてコミットするようにします。

こうして、シードマネーを与えることによって、「わが大学は一体幾ら APC に払っているか」「この先生は OA 誌のカスケードの一番下にばかり払っている、何とかならないか」という話が大学の中で起こってきて、もう少し効果的に大学の成果を外に出し、研究者の評価などの仕組みと相まって合理的になっていくという姿が考えられないかということです。

その事務をするとしたら、科研費による論文だけではなく、他の全ての発表論文の APC も集めなければ駄目です。そうすると、日本から外国の出版社に APC を幾ら払っているのか捕捉でき、次は SCOAP[3] と同じ考え方で、それをどう抑えるかという交渉ができます。そのデータがない限り、こういう話は全くできません。

このような案をつくって提案したのですが、反応が悪くて、日の目を見ないというのが現状です。ぜひ、皆さんもうまいビジネスモデルを考えて提案していただきたいと思います。尾城さんのプレゼンテーションのポイントは、成果発信などにお金を幾ら用意すればいいのかということです。出版社はどこも、発表論文数が増えるからお金がたくさん要るようになるのは仕方がないと言います。SCOAP[3] でもオープンアクセスにすると確実に発表論文数が増えていっています。その中で、経費をもっと抑える仕組みを取り込まないと、このシステムは破綻するだろうというのが私の予測です。

●**山本**　APC を幾ら払っているのか把握が難しいという点ですが、先ほど、アメリカの大学できちんと払った、図書館員が立派だったということだったのですが、日本の大学は図書館に購読費・資料費がごっそり来ていて、それで何もかも買っているというわけではありません。基本的には研究費から拠出していただい

て、先生が選ばれたものを図書館が発注代行しています。そんな時代が延々と続き、恐らく多くの大学ではいまだにそういうことをやっています。電子ジャーナルが始まったときも実はどの出版社で幾ら買っているのか全然分かっていませんでした。ようやくそこを一生懸命調べて、ビッグディールの交渉ができるようになってきました。そんな時代もあるので、今回、APC が俎上に上がってきたということは、APC に関してもきっちり把握していくための事務手続きなりをこれからしっかり組み立てていくことになる気がします。

今はもう各大学でかなり会計システムを導入されているので、APC の区分をしっかり立てて事務方が運用すれば集計できるようになると思いますが、まだ具体的に進まないという状況です。

●**土屋**　会場にいらっしゃる出版社の人に聞きたいのですが、著者からサブミッションがあったときに、大学に通知するというシステムをつくったら、完全に捕捉できるようになる気がします。もちろん、最初のサブミッションのところで、アクセプト・リジェクト関係なしに捕捉しておいて、「結果こうなりました」というのを大学に通知してくれると、大学は完全に捕捉できると思うのですが、そういう要望を大学からもらったことはありますか。そんなことをやると投稿しようとしている研究者の心を傷付けるからやりたくないという可能性もありますが。合理的だと思うのですが、なぜそういう方向が今あまり見えないのかについて、出版社の方にご説明いただけると一番いいと思います。

●**フロア１**　オックスフォード大学出版局の的場と申します。まず、そのようなリクエストが大学から今まであったかどうかというご質問ですが、一度もなかったです。全ての出版社が同じような対応を約束できないということも、今までちゃんと行われなかった理由の一つだと思います。例えば、オックスフォード大学出版局にある論文が投稿されて、それが仮にリジェクトされたとします。その論文が違う出版社に投稿された際に、その次の出版社が必ずしも大学にそれを伝えるということを標準化してやるとは限らないと思うので、その辺が最終的にフォローアップするのが難しい理由の一つではないかと思います。

●**尾城**　日本の大学からなかったというのは分かるのですが、海外、例えばイギリスの大学などからもそういう要望は出ていないのですか。

●**フロア１**　私が把握している限りでは、そういった話は聞いたことがないです。

●**安達**　SCOAP[3] に携わってきたものとして、ぜひ図書館の方に言いたいのは、今のような情報は外国の図書館とコンスタントに付き合っていないと引き出せないということです。国際図書館コンソーシアム連合（ICOLC）に行くなどの交流はそれでいいのですが、そこでメールですぐに質問できるようなネットワークをつくることが一番重要です。このような情報はウェブなどには絶対出てこないので、人とのネットワークをつくらないと集められません。

欧米には手厚いネットワークがあって、情報交換しながら OA の流れに対応しているので、日本も同様な努力をしないと日本だけが孤立してしまう懸念があります。そこをぜひ頑張っていただきたいと思います。つまり、ここで話しているようなことは、絶対彼らだって考えているのです。それで、何か方策を立てようとしているのだと思います。

●**尾城**　おっしゃるとおりだと思います。先方から情報を出してもらうためには、やはり自分たちも情報を出さないと向こうだって本気で付き合ってくれませんから、自分たちが出せるデータを持っていないといけないので、それも今回の調査を始めた背景の一つに当然あるわけです。

●**山本**　把握に関してはこれから行くのではないかと

思います。

先ほど、土屋先生から、オープンアクセス化必要経費と購読比の差が大きい大学はどうするのかという質問がありました。昔はもっとたくさん購読額があったような気がしたのですが。

●**尾城**　そのギャップをどう埋めるかという話は、ビジネスモデルをどうつくっていくかという話になるので、実際のところ、まだそこまでいっていないのが現状です。

また、今日のグラフでお示ししたのは、あくまでも、カレントの電子ジャーナルに払っている購読費を集計して出した金額です。もちろん、あの対象論文の中には既にフルオープンアクセスの論文も相当数含まれていて、特に左側の大規模な研究大学はフルオープンで出している論文の数も当然多い。そういうところは、別途、研究者が自分の研究費で APC を払っている分があるので、その金額を押さえて、そのプラスマイナスをきちんと考えた上で、何かモデルを考えていかなければいけないと思います。

どう考えても、今捕捉できていないハイブリッドジャーナルに載っている OA 論文が相当ありそうだというところまでは何となく分かっているのです。ハイブリッドジャーナルの APC は、フルオープンアクセスジャーナル誌の APC よりもかなり高い。ですから、論文を全てオープンアクセスに転換するための財源としては、カレントの電子ジャーナルに払っている経費だけではなく、既にオープンアクセスのための APC として払っている額を押さえた上で、それを足してグラフを描かなければいけないと考えています。まだ道半ばで、そこからさらに実際のビジネスモデルをつくっていくためには、相手方との交渉が必要となってくるので、それは時間がかかると思います。

●**山本**　SCOAP³ が立ち上がったときも、国立情報学研究所（NII）からかなりご尽力いただいて、参加すれば得になるという枠組みで最初ご提案いただいたと思うので、いまだに参加していない大学が 14 あるというのは何でだろうという気がしていたのですが、あれはフリーライドしてしまっているのですか。

●**安達**　先ほどのプレゼンテーションでは、その 14 大学のことをフリーライダーとは言いませんでした。なぜなら、Nature のような雑誌を考えてみたらよく分かるのですが、論文は投稿しないけれども、読むための雑誌というのがあるのです。ある雑誌に 1 本も論文を書かないけれど、購読している大学があるとします。先生が勉強するために雑誌を買っていた。その後、著者支払いモデルになり、OA で読めるようになった。このような状況の場合に、引き続き支払い続けなさいと言えるのか、OA になったのでやめるのはフリーライダーと言うのかどうかを躊躇したために使いませんでした。

●**山本**　あの枠組みが SCOAP³ でスタートしたことによって、オープンになっているのですよね。そのオープンになっているのは、お金を頑張って出してくれている機関があるという。

●**安達**　先ほどの尾城さんのプレゼンであったように、マクロに見ると、今購読で使っているお金が余るのです。しかし、ミクロなレベルで、そこにどのようにして移行していくのかという問題が最大の問題なのです。公的機関の助成が必要だというのは、最初の谷を乗り越えるためのシードマネーがないと乗り越えられないということを申し上げているのです。ドイツもイギリスもそれが分かっていて、投資をしているという現実がある中で、日本はずっと乗り越えられず、ずるずるいくというのが一番危険だと思っています。

●**山本**　また、入らない大学がある中で、お金を払う側にすればかなりリーズナブルなモデルになっているかと思いますが、出版社側からすると、あまりぼろもうけはできません。撤退した理由は分からないのです

が、この枠組みから抜け落ちていくジャーナルも出てきているということですね。

●**安達**　アメリカの図書館が素晴らしいというのは、SCOAP³に関して、わが大学からはこういう雑誌には論文は投稿しないのだけれども、オープンアクセスを進めるということで出しますという大学があるのです。アメリカも日本も本来は得してきた国だと思うのですが、きちんと100%出して、しかも、アメリカ物理学会（APS）の不参加で比率が下がったのだけれど、今まで出していた金額を出すと図書館が言っているということは本当に立派です。フリーライダーどころではありません。

ドイツもこの3年間で変わりました。3年前はドイツの大学は、どうやって配分するのかという議論に結論が出なかったために、ドイツ研究振興協会（DFG）が差額を補填することになりました。次の3年間もさらに補填するのですが、3年後には各大学が全てを負担するようにするということのようです。全ての大学がこの3年間で変わりました。それが世の中のスピードです。

●**山本**　言い換えると、乗り超えるための山があり、その乗り超えた先はもう新しいモデルになってしまっているので、昔のようにお金を出したどうこうではなくて、そこの枠組みを支えるためにお金を出すという発想に変わってくるということなのでしょうか。新しいモデルでは、以前とお金を払う根拠はだいぶ変わってくるということですか。

●**安達**　3年前、日本人はリダイレクションで本当にうまくいくのかという議論ばかりしていて、それはやらない理由を考えているようなものでした。今考えているモデルが未来永劫に続くかどうかは誰も言えないのです。問題が出てきたらそれを解決していくというようにしてしかこの問題は対処できません。もう一つのやり方として、欧米のやり方に倣って、日本が全部それを受け入れて後からついて行くのでうまくいくというのであれば、ほったらかしておけばいいという話ではないかと思います。お金があり、それで済んでいればそれが一番いいのです。

●**山本**　ありがとうございます。

先ほど土屋先生から出た、あのグラフの中で新しい世界に行けるのかということについて、早稲田大学さんはいかがでしょうか。

●**荘司**　先ほど山本さんもおっしゃったのですが、アメリカのように図書館長が全てのバジェットや人事権を握っている組織は、日本ではほとんどないと思います。ですから、国立大学風に言うと、部局ごとに持っているお金を預かって執行しているところが多いと思います。自分のお金をダブルディッピングで余計に払っているのだったら、これはけしからんとなりますが、アメリカでは、人のお財布のような感覚が少しあるのではないかと思います。日本では選書権が館長にもわれわれにもないので、そういうところだと、まとまった意見を言う動機がなかなか得にくい感じがしています。

モデルが変わるのであれば、お金の使い方などはきっと図書館しか知らないので、図書館が学内のいろいろなところに丁寧にお話ししていかなければいけないのではないかと思います。解決策をいきなり提示するというよりは、現状をまず把握して、本当に実現可能性があるところを探っていかなければいけないのではないかと思っています。

●**土屋**　JUSTICEの版元説明会か何かがありましたが、JUSTICEがExpression of Interest（EoI）に署名したことをどのぐらいの図書館の人が意味を理解しているのですか。つまり、今日の話がリアリティを持って受け止められているかということについて、何か感触はお持ちですか。

●**荘司**　恥ずかしながら、本学では休みにも入っているので、まだその話をきちんと学内の図書館員には周知していなくて、これからという段階になります。ただ、枠組みができても動かなければいけないというのが現場の考え方だと思うので、「何かやらなければいけないんですか」という質問は来ています。「当面は関心表明です」という話だと、そこで止まってしまいそうな気はしています。

●**安達**　先ほど SCOAP³ のことで言いましたが、日本で大きく動いたのは、アメリカの SPARC が SCOAP³ に対応するために大学図書館に手紙を出して動きだし、結構いけそうになったことがきっかけでした。

昨年の Berlin12 では、アメリカの人たちはばらばらで特に方向性があるようには見えなかったのですが、この 9 月にマックス・プランクでそれを取りまとめている Ralf Schimmer をアメリカに招待して、SPARC との間で密な話を持つというようなことが SPARC の委員会で計画されています。そのような流れの中でアメリカとしての方向性が出てくると、具体的に一歩進む可能性が出てきます。

このタイミングで JUSTICE が EoI を出したのは、大変時宜を得ていていいことだと思いました。みんなどういう方向でいけばいいのか分かっていませんが、そういう動きの中に絡まないと後からついて行くだけになってしまう気がします。

●**尾城**　JUSTICE で EoI に署名したのは 8 月で、まだ 1 カ月もたっていないので、JUSTICE の会員間でこの署名の意味がそんなに広まっているとはとても思えません。まさに今日のセミナーは、その意味をきちんと考えてもらうためのいい機会だと思います。

最後の方で触れましたが、大学ごとのデータをつくって、それを会員大学に配ることはできます。そういうことを通じて、今までの図書館は要するに外のものを買ってくることしか考えていませんでしたが、自分の大学からどれぐらい論文が出ていて、そのためのコスト負担はどうなっていくのかというところにも少し目を向けてもらうきっかけになるのではないかと思っています。

●**土屋**　そんなのんびりした感じですか（笑）。

●**尾城**　世界全体で今の購読費を APC に振り替えれば、200 万論文を一挙にオープンアクセスにできるという話をしましたが、ある日突然、全てオープンアクセスに変わるということは現実ではあり得ないと思うので、当然、その移行期間があります。やはり購読しなければ読めない論文が相当数、相当な期間残っていくわけで、そのときのビジネスモデルを考えていかなければいけません。そうなると、アクセスするために払うお金と、オープンに出すためのお金をどのような比率にしたらいいのかなどを、これから数字を出して、きちんと出版社と交渉していかなければならなくて、それには時間はかかると思います。

●**安達**　つい数日前、SPARC の委員会があったときに、慶應の倉田敬子先生が大変いいことをおっしゃって、それを自分なりに解釈しました。それは機関リポジトリとは少し性格が変わりますが、プレプリントサーバーというものがあって、日本の主要な大学図書館は、コーネル大学の arXiv.org にお金を送って、コンソーシアムをつくってサポートしています。オープンアクセス発祥のプレプリントサーバーが、世界の主要な大学によるサポートへと動いているという現実が数年前からあります。倉田先生は、グリーンといいましょうか、セルフアーカイブというのが適切かもしれませんが、出版社ではないサーバーがあることが出版社に対するカウンターバランスになるとおっしゃっています。土屋先生のお話にあるように、ゴールドオープンアクセスは普通にやればもうかる商売です。ですから、出版社はどんどんやるし、それに合わせて、坊農先生のプレゼンにあったように、APC で研究費をかすめ取っていくビジネスが成り立っているのです。

物理の研究者達は、コーネルのプレプリントサーバーarXiv に入れた論文とその後の査読をうまくつなぎ合わせて、インパクトファクターの付いた定評ある雑誌にどうやって自分の論文を持っていくかというプロセスを睨んで、SCOAP³ のような形で APC を抑えることをしようとしています。そのリポジトリは、出版社の独走を抑えるために重要という考え方で見直すことができるのではないかと思っています。

そういう意味で、今後、データのオープン化などが進んできたときに、論文以外の情報とうまくつながるということで、大学の機関リポジトリが最下層のオープンなコンテンツを提供するサイトになって、その上にサブジェクトリポジトリや一般の意味でのアーカイブデータベースが分野別にできていくという構造が一つ考えられると思います。

そのようにして、オープンアクセスの論文が多くなってきたときに、例えば先ほど紹介された PubMed Central などがそれなりの力となって存在感を示して、商業的な出版活動をなるべく適正にするという方向に持っていくのに使えるのではないかと思いました。

●**土屋**　例えば PubMed Central に入っているもののほとんどが出版社デポジットになっているはずなので、実際には出版社に対する抑制機能ではなくて、出版社は多分、グリーンと十分に共存する方法を見つけたのだと思います。それだったらあまり抑制作用にはならないのではないかという感じがします。

グリーンというのはそもそも矛盾したプロジェクトで、あれに反論してくれる人がいるととても楽しいのですが。そもそも矛盾していたものを Stevan Harnad のレトリックでうまく言いくるめて、われわれはだまされたふりをしていました。本当にだまされた人もいるかもしれませんが。

そのような状況だったと思うので、基本的にはパブリッシャーがパブリッシュしたものを何とかするということを前提にしたときには、パブリッシャーなんてつぶしようがないのではないかというのが当たり前の話のような気がするので、カウンターバランスにはならないと思うのですが。

●**安達**　先ほど私が申し上げたのは、解決策ではなくて、これから未来に向かって考えるときの一つのフレームワークです。ですから、土屋先生のように事後的に今から昔を振り返ってみれば、グリーンというのはそうだったかもしれません。けれども、最初はやはり雑誌の価格高騰に何らかの形で対抗しなければいけないというところから出てきました。PubMed Central は結果としては出版社からインジェクトするような形になったというのは、全体の期間を歴史的に見るとそうなのだけれども、その時々の問題を解決するための方策として位置付けて考えるべきなのではないでしょうか。

●**土屋**　それに関してあえて自己弁護すれば、リポジトリについて話しはじめたのは 2004～2005 年ぐらいなのですが、そのときには、これでは下がらないということを度々申し上げたはずです。機関リポジトリ自体の重要性は、Clifford Lynch のコミュニティサービスという概念で捉えるべきであって、ジャーナルの価格を安くするというだけなら、リポジトリの存在意義はないのだということは度々申し上げてきました。誰も聞いてくれませんでしたが。

●**山本**　少し話を膨らませていきたいと思うのですが、ゴールドとグリーンということで、最初にブダペストで考えていたのは、全分野、雑誌論文それぞれに二つのやり方があるということで、グリーンを一生懸命やってきたというところだと思います。その後いろいろ経験を積んで、ゴールドで出ている出版社最終版とリポジトリにある著者最終稿、あれは同じものではないということはだいぶ皆さん気が付いてきました。

また、分野によっても全く状況が違います。最近、文科省の議論では、日本学術会議の答申などでも、やはり分野によってだいぶ違うだろうということは出て

きたと思います。最初に土屋先生からこれからの課題として示された市民科学の話も分野によってだいぶ違う話ではないかと思います。

最初にブダペストで考えたグリーンやゴールドの枠組みから 10～20 年たって、いろいろな広がりが出てきていると思います。坊農先生からは、生命科学はかなりゴールドに特化して、そこがきちんと押さえられていればいい分野であるとお聞きしましたが、その中でも、グリーン対応するにはどうしたらいいのかと、最後の方でご配慮いただきました。かなり無理やり感があったのかもしれませんが、そのあたりで何かコメントいただけますか。

●**坊農**　私自身はオープンアクセスのものしかもう出さないようにしていて、それが研究を一緒にするときの条件になっているぐらいなのですが、そういう事情にならないものもあるわけです。ハイブリッドジャーナルの筆頭である Nature などがそうだと思うのですが、そういうところに出たものは誰でも見るからいいとして、特定の学会誌などはいろいろな事情からオープンアクセスにもならず、ずっと残されていると思います。

そういうところのものが見たいというのがどうしても出てくるので、何もしないで放っておくよりは、何らかの形でアーカイブされるという手段は大学図書館でやっていただけると、各大学が成果を把握するという意味でもいいと思います。タイトルだけやるよりは中身も一緒にやったらいいのではないかぐらいに私個人は思っているのですが。

●**山本**　一度、電子ジャーナルとしてネットワーク上に流れたものは技術的に捕捉していこうという CLOCKSS（Controlled Lots of Copies Keep Stuff Safe）のような動きも出てきて、図書館員が地道に汗を流すよりは、かなりスマートなやり方も今は出てきているという状態です。ただ、そうやって出版社レベル、流通レベルで捕捉するのと、また別に大学レベルでも捕捉して、大学の業績として取っておかなければいけないという部分もある気はしています。

●**土屋**　今日ここで「やはりグリーンでいくべきだ」と言う人が誰も出ないのであれば、もうこれは総意だとしていいと思うので、無理やりにでも理屈をつくって発言いただいて、それを一生懸命たたきつぶすという作業をやるべきではないかという気はするのですが、荘司さん、どうですか。

●**荘司**　ゴールド、グリーンといったときには、STM（科学・技術・医学）のインターナショナルな学術雑誌での議論になると思うのですが、日本の人文系学会は非常に小さくて、そういうところに入っていて、論文も自分のところで電子化できない先生は、実はリポジトリに登録してくれと積極的に言ってくるのです。登録すると CiNii で引っ掛かるからでしょう。

落穂拾いではないですが、学術成果はいろいろな形態があると思うのです。国立の旧七帝大などは理数系の先生が多いですが、私立などは逆に文科系の先生が多い大学が多いです。文科系の先生が多いところでは、グリーンというか、セルフアーカイブで研究成果をネット上に出すと、特に CiNii のおかげで幅広くアクセスされるという話も聞いているので、そこは分野によってすみ分けをしていくべきではないかと思っています。

●**土屋**　それは人文系弱小学会の出版機能は要らないということを意味するわけですよね。

●**荘司**　その機能が要るか、要らないかはちょっとコメントしませんが、仏教の中の狭い何とか宗派についての学会誌など、非常に流通量の少ない、大学図書館に 1 部ずつしか回らないような学会誌もあるので、そういうところは拾ってあげられるのではないかと思います。

●**山本**　5 人の皆さまから発表を少し掘り下げていただいたので、今度は会場の方から、ご質問のある方は挙手していただければと思います。

●**深貝**　横浜国立大学の深貝と申します。今日のお話は非常に大事なことで、ラジカルな土屋先生の問題提起に他の名だたるスピーカーたちが必死に答えていました。しかも、パブリッシャーの方々がいらっしゃるところで作戦会議を公にやっているような不思議な感じです。

尾城さんのお話は、SCOAP³は、基本的にヨーロッパで開発され、アメリカも頑張ったことで進んでいて、それを日本にも取り入れるという話だったと思います。国際的に乗ることを希望する人は乗れる仕組みに一応なっているのですよね。購読料を APC に振り替えて安定的なオープンアクセスモデルをつくることについて国内で理解を得るために、どのようなストラテジーを取る必要があるのでしょうか。日本学術会議は、オープンアクセスは学術情報の基盤を支えるために重要であると言っていますが、具体的なことは恐らく JUSTICE に投げられています。

図書館は自力で図書館の予算を調達できないので、組織内の研究者の理解をどれだけ得られるか、かつ大学の予算の中でジャーナルをどこまで支えられるかという問題になります。新しい仕組みに移ると安くなると錯覚して拠出額が減ってしまってはいけないので、逆に図書館の方が理解を相当積極的にしなければいけないのです。

乗り移るためには相当試練があるということで、見通しを尾城さん、安達先生に伺いたいです。

●**尾城**　安達先生のお話の最後でも出てきたと思うのですが、やはり研究者の皆さんの理解・協力を得なければ仕方がないとは思います。ただ、どうやったら理解してもらい協力を得られるのかが、なかなか難しい。そもそも、この SPARC Japan を始めたときから、図書館と研究者の連携ということはずっといわれてきましたが、この 10 年ぐらいでどれぐらい進んだのかというと、少し心もとないところがあって、なかなか進んでいない。どうやったら進められるのでしょうか。

●**安達**　私はオープンアクセスにどう対応するかを研究者個人に任せておくと負けると思います。科研費をたくさん取ってきて、学生に年に何十本も論文を書かせて、APC を払って業績リストを増やしていくという形でそのままやらせていくと、いろいろな不具合が出てくることが懸念されます。

強調したいのは、機関として方向性を持ってやらなければいけないということです。JSPS や JST（科学技術振興機構）のようなところが「オープンアクセスは良いことである」と言っても、強制力がなければ、言ってみただけになってしまいます。

先ほど言った私と尾城さんが提案したアイデアは、例えば今年科研費から APC に 20 万円払ったとすると、来年その 20 万円が戻ってくるという明確なインセンティブを与えて動かすというものです。他の手立てとしては、本当に強制して、オープンアクセスジャーナルで出さければいけない、それ以外は科研費の報告書に掲載できないとするなどです。その中で一番研究者を誘導しやすいものをつくっていくことが一つの解決法です。

尾城さんの試算では、マクロに見ると購読費より何十パーセントか安く済むのですが、ほっておくと、研究者が一生懸命成果を出し、論文をたくさん書いてしまうので、その分はすぐに消えてしまうという構図が考えられます。

●**土屋**　論文が増えるのはいいことではないですか。

●**安達**　そこのところをどう考えるかです。つまり、「イノベーションが大事だ、研究しなさい」と言って若い人を育てたら、みんな論文を書きます。そうすると、それに比例してどうしてもお金が掛かります。この状況をそういうものだと受け止め、きちんと APC

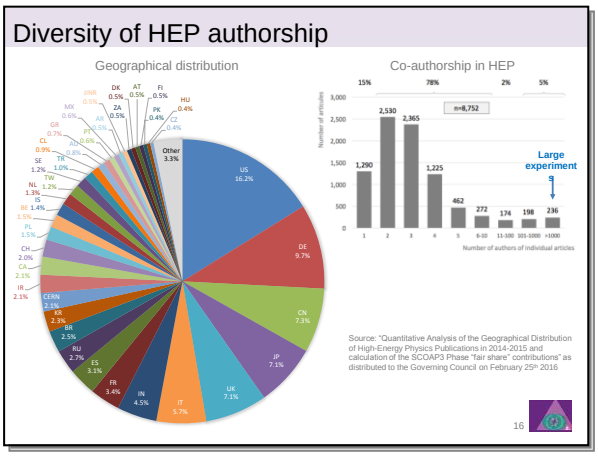
に使うお金を与えずに、「今までどおりでやれ」と言ったら、つまりは論文を出せないということになります。それでいいのでしょうか。

電子ジャーナルにしたからコスト・単価は少し安くなるだろうという気は何となくするのですが、それ以上に研究者が論文を書いているという現実はあると思います。

●**土屋**　今の深貝先生の質問に対する答えとしては、基本的にはうそをつくしかないと思います。尾城さんの発表にあった、JUSTICE の 300 大学全体で、購読費は 240 億円、オープンアクセス化必要経費は 140 億円という差額は、厳然たる事実であると言い張るしかありません。あれだけ丁寧に調べた結果なのですから、100 億円得するのだと。出版社は 100 億円損するのだから、何をするか楽しみですが、そのように考えるべきだと思います。

その前の段階について一つだけコメント申し上げたいのは、ヨーロッパの考え方を日本にとおっしゃったけれど、基本的には少なくとも科学論文の生産と流通・消費という中では、日本とヨーロッパは今のところ変わりありません。気持ちとしては、脱亜入欧という世界がまだ感覚的に残っているのは分からないではないのですが、もう差はないのです。高エネルギー物理学分野の論文生産力の順位を安達先生に見せていただきましたが、全般的にあんなものなのです（図 2）。

幸いにして、これらの国や日本は人口が増えていま



（図 2）

せん。アメリカだけは他の国から入ってくるので増えていますが。従って、研究者の人口も増えないはずなので、基本的には留学生の数などを入れなければ生産性はそれほど向上しません。だから、今のとんとんのところで手を打つのが一番賢いのです。これから論文が増えるとすれば、いわゆる途上国からの論文が増えることによって増えることが予想され、期待されます。彼らは科学技術立国と依然として言っています。ですから、一番困るのは途上国のはずです。

われわれが今のところで手を打てば、それ以降は定常状態を維持できる可能性は十分ありますが、途上国でこれから増えていくところは増えた分だけ全部自分で負担しなければいけません。ある意味でそれだけのお金を今まで払ってきていないので、それこそどこから出すのかが不思議だという感じです。

そういう意味では、非常に大雑把な言い方をすれば、日本は今のタイミングで、ヨーロッパとつるんで塗り替えるというのが一番分かりやすい戦力だと思います。議論としては、やはり 20 万円という話をするよりは、100 億円得するという話をし続ける方が、まだ政策的な反映はあり得るという感じを持ちます。

●**山本**　まあ、図書館的にも多分「これがお得ですよ」という言い方しかできないのだろうと思います。ふたを開けてみると、SCOAP³ のこれまでの状況を見ても、こういう世界に入っていけば、また論文が増えていきます。前回、1 社抜け落ちて、今度また雑誌が抜け落ちようとしていますが、やはり出版社側でもこの枠組みを活用していこうという理解がまだ普及していない気がします。

●**土屋**　出版社の方に伺いたいのは、SCOAP³ と Ralf Schimmer が考えている世界的乗り換えを本質的に同じものだと捉えているのか、違うものだと捉えているのかということです。巨大出版社の現役の社長あたりに伺いたいです。

●**山本**　大学側がこの枠組みに乗れば短期的に得をするのだ、すなわち出版社は短期的に損をするのだというそに出版社もだまされてしまっているのでしょうか。

●**フロア 2**　アンテナハウスの小林といいます。出版社の社長ですが、超弱小です。

安達先生にお聞きしたいのですが、SCOAP³というのは素晴らしい企画だと思いますが、強力な出版社にはあまりメリットがなくて、下位の出版社にメリットがあるような気がします。結局、先ほどの入札でも、アメリカの出版社は大きなところは抜けてしまいました。つまり、自分の力でお金を稼ぐことができる強力なブランドを持っている出版社にとってはSCOAP³に参加するメリットはなくて、そうでないところに参加するメリットがある気がするのですが、その辺はいかがでしょうか。

●**安達**　今のSCOAP³に残っている雑誌はヨーロッパ系が多いのですが、やはりアメリカとヨーロッパという肌合いの違いがあって、研究者はお互いに競争しながら論文を書いているのです。それは先ほどの坊農先生のどの雑誌に出すかというところに如実に出てきます。それが露骨に表れて、アメリカ人はやはりAPSの雑誌に書くことが多いということがあります。ヨーロッパ人はエルゼビアやシュプリンガーの雑誌に書くことが多いということも何となく現れてきました。しかし、APSを入れようという動きがずっと今も続いています。やはり学会ですから、会長や担当者が代わったりすると、協力的になったり、そうではなくなったりというのをこの数年間繰り返してきました。

ストレートなお答えになっていないのですが、商業出版社の方が、ゴールドオープンアクセスでどのようにビジネスをしていくかの対応がしゃっきりしています。学会は研究者コミュニティに寄り添いながら、出版ビジネスもやっていかなければいけないということでやはり心が乱れるようで、どのように今のオープンアクセスに向かうトレンドの中でやっていけばいいかは分野ごとにいろいろ違うようです。

弱小の学会誌をサポートするというのは、日本の高エネルギー物理学分野からの要望もあり、きちんとSCOAP³のEoIに書きました。つまり、地域の特性を考慮した出版活動を支援する、と。そのとおり、各国の物理の雑誌も入れるということになりました。ただ、APSも入れた12の雑誌で高エネルギー分野の75%をカバーしています。そういう意味で、メジャーな雑誌はそれだけのウエートを占めているということは現実です。

補足ですが、坊農先生はもう完璧に欧米の研究者と同じレベルで研究・出版活動をするというご紹介でしたが、私のようにエンジニアリングの分野では、日本語で書いたり、グレーな雑誌に評価がきちんとできていない研究途中のものを書いたりと、メディアを分けながら学術コミュニケーションをしています。そういうものについては、グリーンのリポジトリは大変重要な機能を果たしています。エンジニアリングの分野では日本語でのコミュニケーションが結構多いこともあって、グリーンの意義はそれなりにあります。択一的、つまりゴールドかグリーンかというのではなく、全部共存して進んでいくのだと思います。図書館がどこに力点を置くのかは、制約されたリソースの中で難しいことがあることは認めますが、全部やっていかなければいけないのが現実なのではないでしょうか。

●**土屋**　最後のものはグリーンではなくて、要するに弱者のための出版機能というものでしょう？　評価が定まった論文に対するオープンなアクセスを保証する、つまり評価が定まった論文を機関リポジトリ等に載せてオープンにするというのがグリーンロードだったと思うので、最後のものはグリーンというよりは、ただリポジトリです。

私自身の主張としてはリポジトリはそれをやればいいということに尽きるので、グリーンの話ではないような気がしました。リポジトリの機能としては重要だ

けれど、グリーンロードの話ではないと思うのですが。

●**安達**　ある活動を幾つかのアスペクトで切って、その言い方を変えているだけのように思います。例えば、一番簡単にセルフアーカイビングのような仕事をしようと思ったら、先生方が書いた論文を全部リポジトリに入れるという形でいく方がよくて、査読付きの定評のある論文誌のセットをどうつくるかは、それこそ研究者コミュニティの間で一番大きな問題です。そういうことにコミットしてああだこうだするより、全部入れることにした方が仕事は合理的にいくような気がします。

それで、「Web of Science に載るような論文だけを入れるリポジトリをつくりましょう」と言うとこれはまた大変になるので、そういう言い方はしない方がいいのではないでしょうか。

●**土屋**　日本の図書館の人は、そのようなことは言えないですよね。言った先生の中で、その先生は多分、3 人に 1 人とか、そんな感じなのですから。

●**深貝**　先ほどから土屋先生と安達先生とで「グリーン」という言葉が出てくると微妙に違ったことをおっしゃっているのですが、オープンサイエンスの最近の話が出てきた後で言われはじめているのは、データも出しなさいということです。そうすると、オープンデータということを考えると、差し当たっては機関リポジトリは使い勝手がいいという結論にたどり着いています。

従って、ジャーナルであれ、他の形であれ、論文として書き終わったものを載せることを考えると、グリーンかゴールドかは大事な二者択一の話になりますが、それとは違って、さまざまな学術データの入れ物としては可能性があるものとして今後のリポジトリを考える。これは第 2 回の SPARC セミナーで予定されていると思うので、そのときに掘り下げていただければいいのではないかと思います。

●**山本**　ありがとうございます。では、それは第 2 回に譲るとして、会場の皆さんで研究者の方が何人かいらっしゃっていると思うのですが、「自分のところではこんな感じだよ」というお話をどなたか頂けないでしょうか。

いらっしゃらないようですので、その枠は外して、どなたか自由に何かお聞きになりたいことがございましたら。

●**フロア 3**　松浦といいます。京都から来ました。コーネル大学の arXiv.org など、プレプリントサーバーの動きについては、例えば日本は京都大学が貢献していると思いますが、そのあたりで何かコメントがある先生がいらっしゃったらお願いします。

●**安達**　arXiv.org に対して、日本の 14 の大学が資金的サポートをしています。ですから、もっと日本のサポートを評価していただくように国際的にもアピールしていきたいし、日本の大学がそういうことできちんと活動しているという一つのいい例だと認識しています。

単に物理のみならず、数学やコンピューターサイエンスのような分野にも広がってきています。例えば、この NII にいる研究者も普通に arXiv にプレプリントを送るようになってきましたし、PLOS ONE に投稿することなども増えています。そのように広がってきているので、研究者の行動を変えるいい一つのサンプルではないでしょうか。

坊農先生の発表にあったのは割と古典的な研究者像で、物理の先生方は、例えば Progress of Theoretical and Experimental Physics（PTEP）などでは、まず arXiv にプレプリントを送って、それを査読者が見ながら査読して出版することが普通に行われていると聞いています。

●**土屋**　人によっては、arXiv と SCOAP[3] の両方に金を払うのはダブルディッピングである、同じものを見

るのに二度払っている気がすると言います。だから、そのあたりの整合性は保証しなければいけないと思います。別に両方それなりに意義があることは認めるのですが、そのように言う人に対してどう答えるかは準備しておくべきでしょう。

また、プレプリントサーバーとポストプリントのデポジットはやはり意味が違っていて、先ほど安達先生の説明にあったように、査読を経ていないものが載っていることが前提なので、「これはタダでもいいだろう」と誰でも思うわけです。それに対して、一定の品質保証と、場合によってはコピーエディティングを経たようなものに対してアクセスするときには対価を取っても仕方がないということで、今までサブスクリプションのモデルがあったのですが、その後で、さらにそれがオープンになってくるということであれば、一定の制約は付くでしょう。その一定の制約を author's final 何とかというものでごまかしているのだと思います。それが 1 点目です。

2 点目としては、特にデータアーカイブの話になったときに重要なことで、機関リポジトリはデータを載せるのに便利だろうと言うのですが、データの種類は非常に多様で、分野ごとにメタデータが大幅に異なることは想像がつきます。異ならないようなメタデータのモデルをつくろうとすると、Dublin Core のようなことになってしまうので、分野ごとに機能を十分に保証できるデータを保存するか、どこかのように商業的にデータを預かり公開できるような場所をつくらなければいけないと思います。

そのときには、機関リポジトリはそういうところに全く太刀打ちできない可能性があると思うので、サブジェクトごと、機関ごとという役割分担も重要な区別になると予想されます。あくまでコメントですが。

●**山本**　ありがとうございます。もうだいぶ時間が押し迫りつつあるのですが、グリーンに関してはどうなのでしょう、始末をつけた方がいいのでしょうか。

方向性としては、全てがオープンアクセスになれば皆さんハッピーになるのではないかと思いますが、まだまだ過渡期というか、サブスクリプションで電子ジャーナルを提供するスタイルで、「うちはブランドがあるからそれでやっていけるよ」という出版社、ジャーナルもまだかなり残っていると思います。そういうところに対して有効なのかどうかは分かりませんが、引き続き圧力のようなことはやっていかなければいけない気はします。

●**土屋**　圧倒的にまだサブスクリプションの方が多いことは間違いないので、オープンアクセス論文が何パーセントかは微妙な数字の違いがありますが、多くても 20%ぐらいと見ています。

●**尾城**　日本の機関リポジトリ全体で、グリーン OA の対象となる論文の 6%程度しか捕捉できていないという粗い試算があります。

●**土屋**　それは重要な数字で、グリーンはやめた方がいいというのはそれに尽きるのです。機関リポジトリを使ってオープンアクセスの話をするのは、もうやめましょう。機関リポジトリはもともとオープンアクセスなものなので、それは当然、保持しなければいけないとしても、機関リポジトリがオープンアクセスであることによって、それを使って質が保証された研究成果のオープンアクセスを実現するという十数年来の考え方は、もういいかげんやめてしまった方がいいのではないかという極めて謙虚な提案です。

●**安達**　オープンアクセスはこれから未来に向かって行う話で、最近、メジャーな大学でもビッグディールをやめなければいけないという窮状になっている中で、どのようにアクセスを保証していくか、そのようなシステムをどうデザインするかということです。

グリーンやゴールドに選択的に対応するということでは無理で、全てに対して何らかの対応策を取らないとプラクティカルには駄目なわけです。それをどうう

まく組み合わせてやっていくかが問題だと思うので、「ゴールドかグリーンのどちらがいい」と言って解決する話ではないと思います。

●**尾城**　もし、機関リポジトリを本気でグリーンオープンアクセスの受け皿としてもう少し活用していこうというのであれば、何をすればできるのかを本気で考えなければ駄目だと思うのです。今までのようにやっていても、結局のところは紀要論文と学位論文しか入ってこないので、それはやはり考えなければいけないと思います。フロアにも機関リポジトリ担当の方が結構いると思うのですが、何かその辺でご意見を聞かせていただければと思います。

●**フロア 4**　一橋大学の寺島と申します。リポジトリ担当者ではないのですが、その仕事を見ていて感じているところがあります。

先ほどおっしゃったように、人文系の学問の場合、リポジトリに入れてくださいというご要望は多くて、ご協力いただいているのですが、先生方はそれがオープンアクセスという動きと関わっていることはあまり意識されていないのが現状です。

それで、こちらからどんどん説明してご理解いただくのがまず大事だとは思っていますが、同時に、先生方に加えて大学内の各関係部署のご理解を得なければいけません。それも APC の問題などがあまり知られていなくて、それをどこが払うという話にまで本学では至っていません。他の大学のお話を聞いても、図書館の方は結構危機感を持っていて、「APC、どうする」という話をたくさんされているのですが、大学の上層部には「それは図書館でやるのではなくて、別の部署でやる」という反応をされることが多いようです。では、別の部署でやった結果、どのようになっているのかというのが今のところよく分かっていなくて、図書館も主体的に動けないのが現状だと思っています。

ですから、土屋先生のお話にもあったとおり、このままだと図書館はどんどんやることがなくなってしまうと何となく感じているので、それについてどのように積極的に動いていったらいいのかと悩んでいます。

図書館がやるのではなくても、他の担うところが現れて頑張ってくれれば、「まあ、しょうがないのかな」という気がするのですが、担ってくれる動きもあまり見えないので、図書館が頑張って説明していくしかないのではないかと思っています。そのあたりについて、どう頑張って説明していったらいいかコメントいただけたら、ありがたいです。

●**山本**　状況がオープン化に向けて大きく変わりつつある中で、そういうことを心配して学内に説明できる人を見渡すと、やはり図書館員が一番よく悩んでいる気はします。大学全体としてはそういう仕事の需要は必ずあると思うので、そこに飛び込んでいけるかどうかの気がしますが。どなたかコメントはありますか。

●**安達**　私はずっとこの 30 年ぐらい歯がゆい思いをしてきました。アメリカの図書館の人は、図書館でコンピューターを使いだして、電子ジャーナルが出てきたのに合わせて仕事をつくっていったのです。例えば 20 年ほど前は、「私はシステムズライブラリアンです」と言ってコンピューターの活用にのりだしてきました。最近では「データキュレーションをやります」と言っています。「データキュレーションを日本はどうやっているのか、ディスカッションしたい」と言われて、困ってしまった経験もあります。彼らは自然に、学術コミュニケーションをサポートするための仕組みを総合的に考えて、問題解決のための提案をしているのに、日本はそうではないという大変歯がゆい思いをしてきたのです。

この状況を大胆に改革していく必要があるし、それが必要なのだと大学の中で言う必要があります。図書館の人が自分のポジションを維持するのだったら、そのような仕事があるとアピールしていくことが一つの道だと思うのです。

日本は遅れているからある意味で楽なのです。ヨー

ロッパやアメリカの図書館が何をやっているかを見て、日本でどうするかと考えればいいからです。先頭を行くと自分で考えなければいけなくなるのでとても大変です。ゴールドオープンアクセスにどのように合理的に対応するかを考えることは、いいエクササイズだと思います。

●**土屋**　あくまでカウンターバランスで発言しますが、アメリカのライブラリアンは実際にはひどい人たちばかりです。ICOLC などでの経験から言うと、彼らは基本的にアメリカのメーンストリームのライブラリアンから見た場合には傍流の人たちで、ライブラリアンとしての資格を持っていないような人だらけです。実際のオンラインジャーナルが出てきたときに対応できたのは、もちろんライブラリアン出身の人もいましたが、そういう人たちが寄せ集めでいろいろやったという部分が結構あるのです。私はむしろ安達先生とは違って、図書館の方には期待しません。

●**荘司**　国立大学は違うかもしれませんが、私立大学の職員はジョブローテーションがあって、図書館員は、図書館員の募集にアプライしてきた人間ではありません。つまり、職務給ではなく職能給で動いていて、図書館だけではなく、明日は会計課に行くかもしれないし、学部の事務所にいるかもしれないという、いろいろなセクションに行く職員なのです。ですから、逆に言うと大学全体のことを考えなさいと言われていて、安達先生がおっしゃったような 10〜20 年先の大学全体の資質や研究力をアップするためにはどうしたらいいかということは、図書館員としてではなく、考えなければいけない義務のようなものも負わされていたりするのです。

　アメリカの図書館の方たちは、仕事が決まっているので、その仕事がコンピューター化などでなくなるとレイオフされてしまいます。ですから、考えなければいけないのです。何かつらそうだなというのは向こうの人と話していて分かるのですが、そのままいると要らなくなってしまうのです。イングリッシュのカタロガーの人などはレイオフされてしまっていて、例えばスタンフォードではほとんどいなくなってしまったようです。その仕事だけでは駄目だとなると、自分が生き残るために次のトレンドを考えて発表して、他の大学に移って違うことをやるというライブラリアンの方たちがいます。制度的には日本とそのような違いがあるということは少しだけ言っておきたいと思います。

●**山本**　ありがとうございます。さまざまな観点が示されて、変化のときなのでチャンスはありそうではないかということが見えてきたところで、パネルディスカッションを締めたいと思います。長時間どうもありがとうございました。

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
海外動向調査

■ SPARC Japan セミナー報告
概要
参加者から
企画後記
ドキュメント
（講演・パネルディスカッション）

■ SPARC Japan 活動報告

CC BY

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

CC BY

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

CC BY

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）

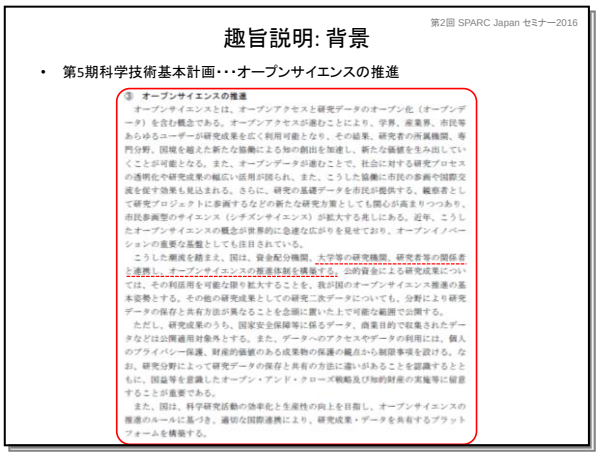
能勢　正仁
1998年に京都大学理学研究科で博士（理学）取得後、米国ジョンズホプキンス大学でポストドクトラルフェローとして3年間研究を行う。2001年帰国、現職。専門は、超高層物理学、地球電磁気学。主な研究テーマは、地磁気変動・脈動、内部磁気圏の高エネルギー粒子ダイナミクス、サブストーム、地磁気指数など。最近は、科学データへデジタルオブジェクト識別子を付与する活動にも積極的に関わっている。

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）。

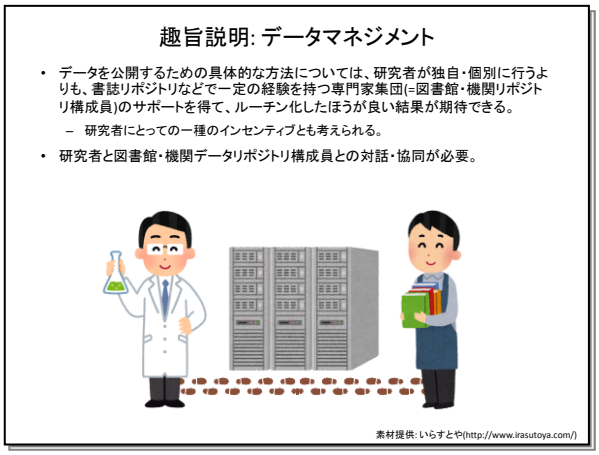
従って、研究者と図書館・機関データリポジトリ構成員との対話・協同が今後は必要になってくるのではないでしょうか。

本セミナーの狙い

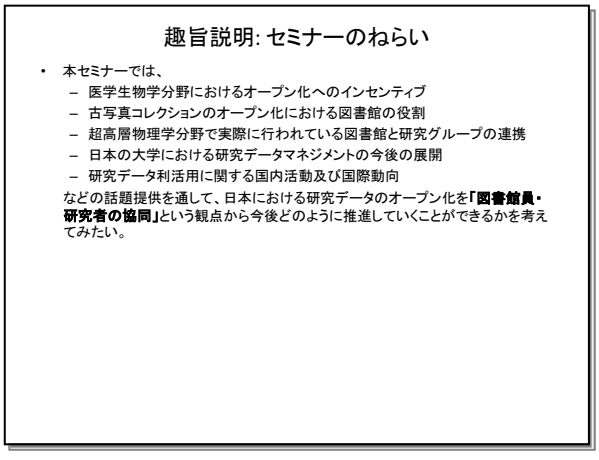
以上のことを受けて、このセミナーでは図 4 の五つのテーマについて各先生方から話題提供いただき、日本における研究データのオープン化を「図書館員・研究者の協同」という観点から、今後どのように推進していくことができるかを考えてみたいと思います。

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 2）



（図 3）



（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）

CC BY

講演要旨

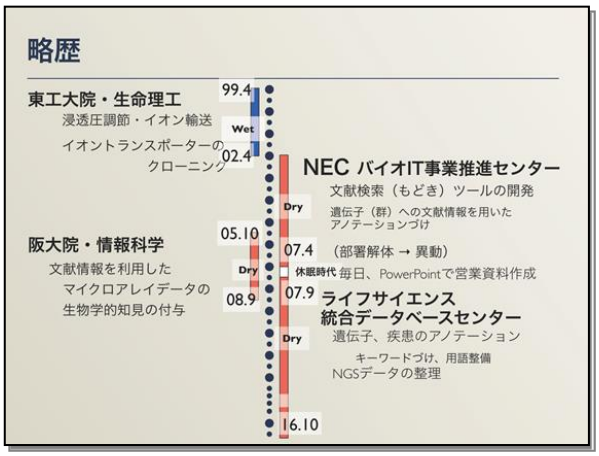
医学生物学分野では、文献情報が MEDLINE に、 塩基配列情報が GenBank にと 1970 年代からデータベースとして保存されており、インターネットの普及とともに誰でも広く利用可能となった。現在は米国 NCBI、欧州 EBI、日本の DDBJ によって、これらに加え遺伝子発現や化合物情報なども公共データベースとして蓄積・運用がされている。このように自然に生命科学データが集積されたのは、出版社が投稿規定内に実験データを公共データベースに登録することを義務化した背景がある。近年は、機器の発展やコストの低下により、診断などヒトを対象にした研究も広くなされるようになり、場合によってはデータを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



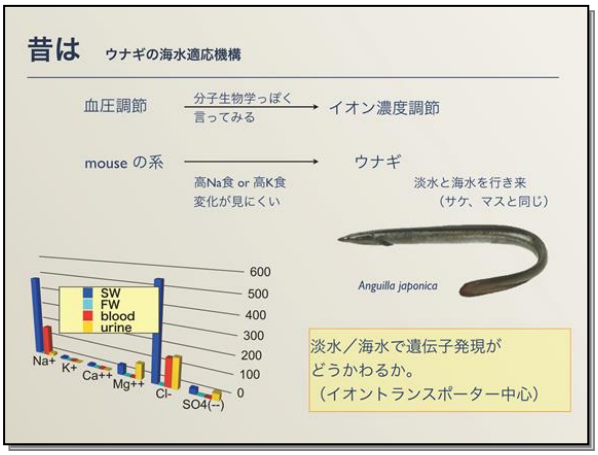
（図 1）

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

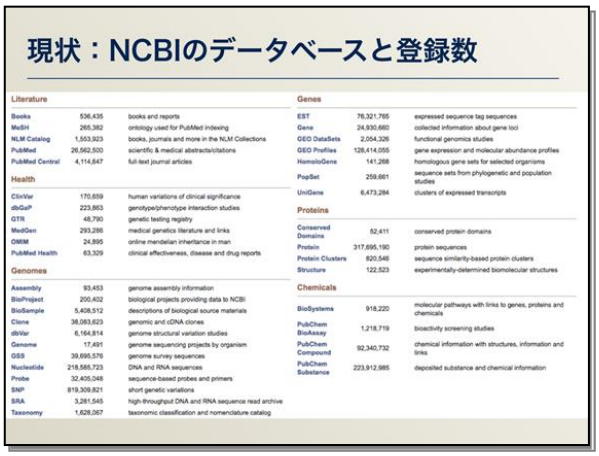
私は生命科学分野にいるので、そのデータベースの実情を最初にお話ししたいと思います。

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。

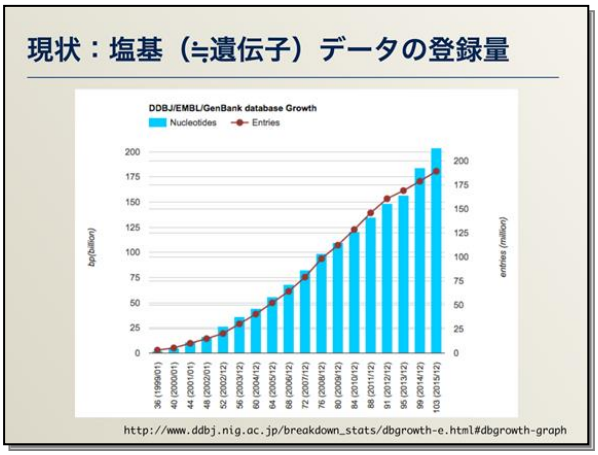
オレンジのバーは文献です。この後、PubMed が出てきますが、1960 年代からずっと集められています。疾患データベースの OMIM はジョンズ・ホプキンス大学で四十数年ぐらい、McKusick 博士が、「病気はこのような感じです」とカタログしていたのを、最近はインターネット時代なので全部デジタル化して、NCBI などから検索できるようにして、この中に一応入っていたのですが、最近は予算が切れて、またジョンズ・ホプキンス大学に戻りました。

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。

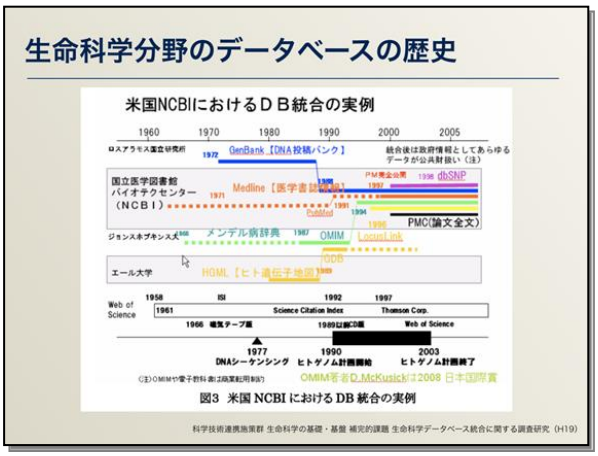
アメリカの NCBI、ヨーロッパの欧州バイオインフォマティクス研究所（EMBL-EBI）、日本の DDBJ は、国際塩基配列データベース（INSDC）というコンソーシアムのようなものを組んでいます。例えば、NCBI に入ったデータも大体は DDBJ から検索でき、逆もしかりで、DDBJ に集まったものが NCBI から検索できるという形になっています。

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）。

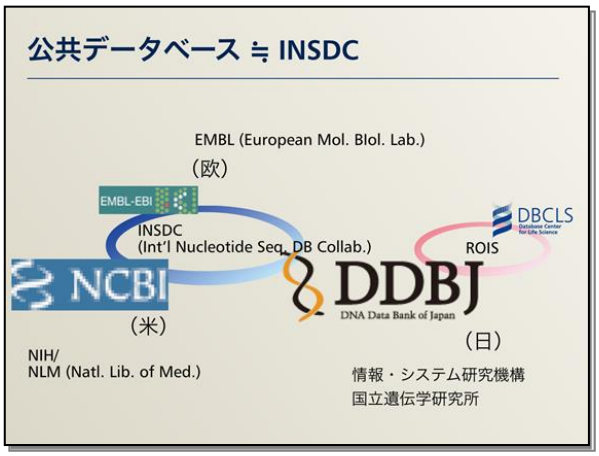
私は DBCLS にいるのですが、この 3 局で集めたデ



（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。

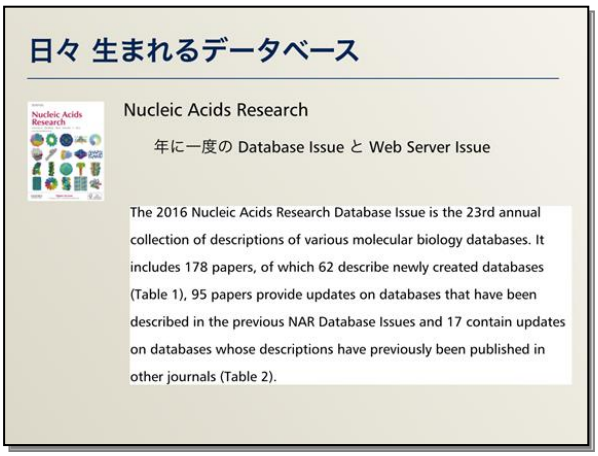
主要な生命科学データベース

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。

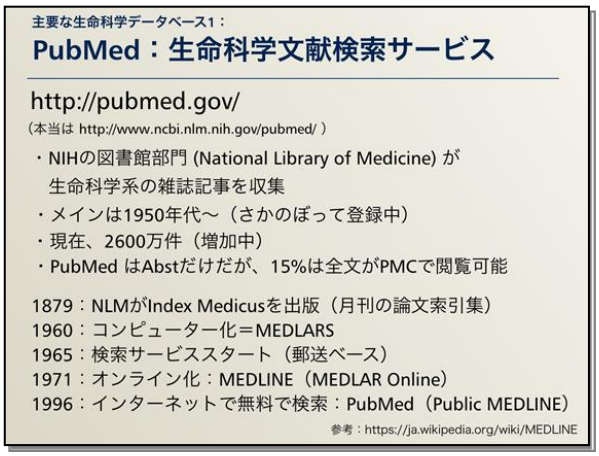
まず、PubMed について述べます。図 9 は NCBI の PubMed で、実際の検索結果の画面です。がん、糖尿病などのキーワードを入れると、関連する文献のリストが出てきます。詳細画面には、文献のタイトル、著者、掲載雑誌、発表日などの情報、Abstract（要約）が書かれています。バナーをクリックすると PDF を取れます。

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。

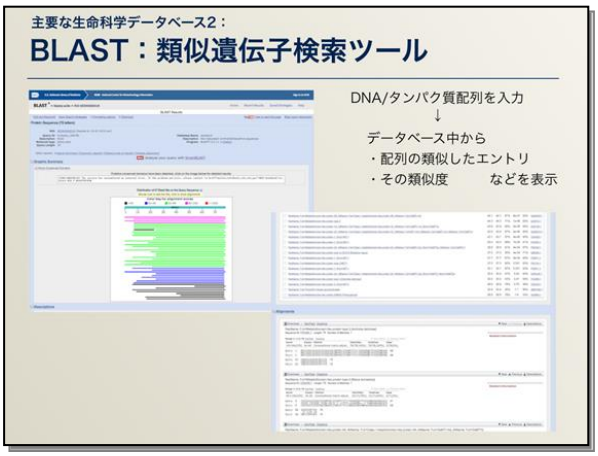
文献が MEDLINE に登録されると、PubMed でそれを検索できるようになります（図 12）。同じように、遺伝子の塩基配列が GenBank、EMBL、DDBJ などの DNA のデータベースに登録されると、BLAST で検索をかけられるようになります。BLAST は DNA の塩基配列情報がオープンでないと検索できないのです。結局、BLAST も PubMed も全てオープンデータの上に成り立っているツールだということです。

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。

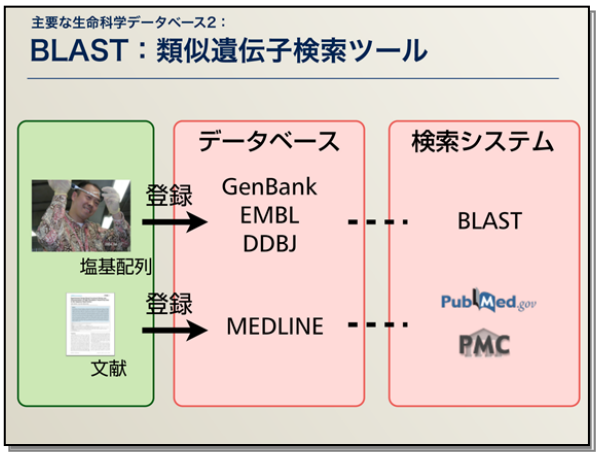
公共データベースにデータが集まる理由

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

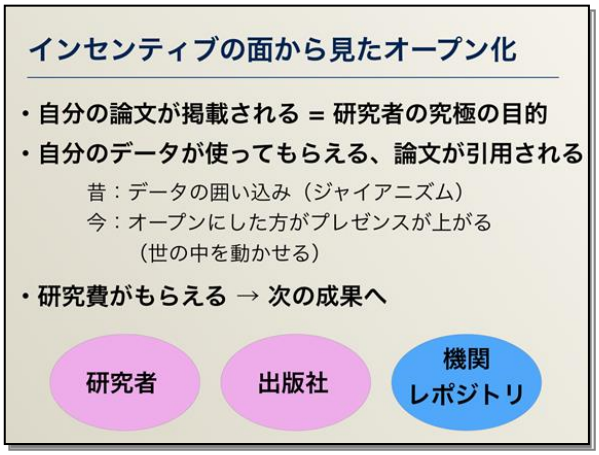
データをオープン化するときは、公共データベースに登録する、自分でデータベースを作成するという手段もありますが、例えば「Scientific Data」「GigaScience」などのデータジャーナルにサブミットする、または機関リポジトリを利用するという方法もあると思っています。

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）

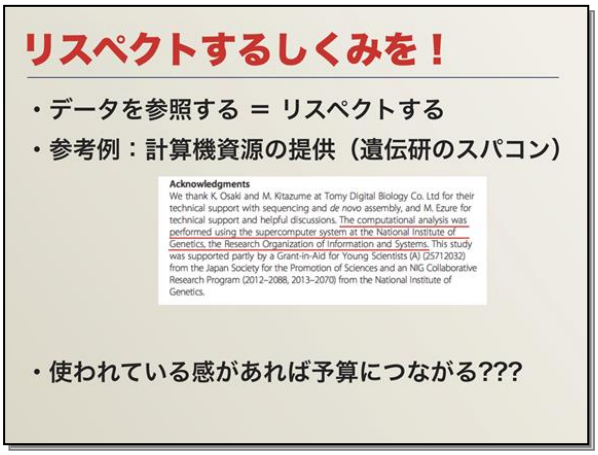


（図 16）

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。

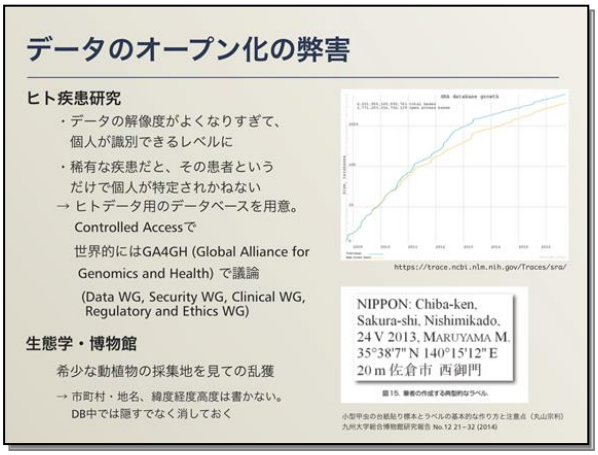


（図 17）

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）

CC BY NC ND

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

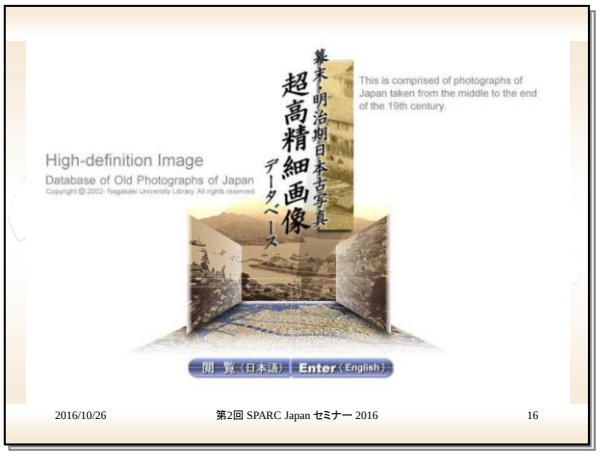
次に構築したのが「幕末・明治期日本古写真超高精細画像データベース」です。2003 年公開で 501 点を収録していました。画像をコンピューター画面上で 5～10 倍に拡大しても鮮明に見ることができます。当時の古写真コレクションは全 5,416 点でしたが、それから 501 点（長崎が 201 点、その他が 300 点）の特に画質の優れた画像を選んでつくったデータベースです。

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。

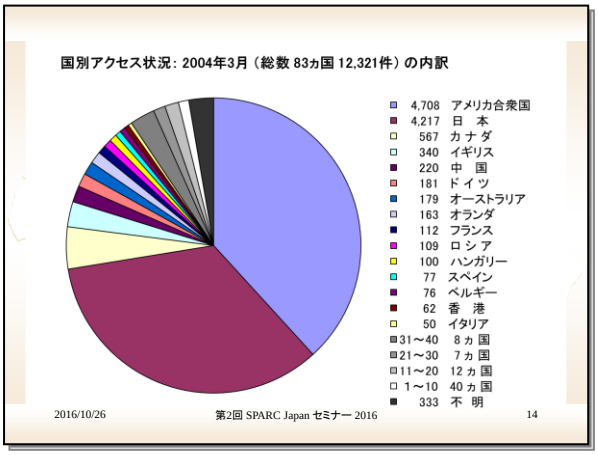
図 4 が最初の高精細画像データベースです。図 5 は 2011 年以降のトップページです。

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

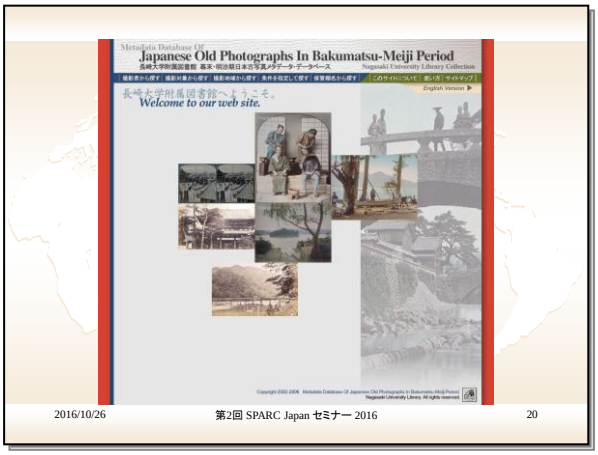
3 番目は「幕末・明治期日本古写真メタデータ・データベース」です。これは初代データベースの後継です。目立つ点は、同一標題を古写真の画像に与えることにより、同一主題の写真の通覧と比較を可能にしたことです。また、高精細画像データベースに収録されている写真については、該当のレコードへのリンクを貼っています。さらに、メタデータを Dublin Core に準拠させて、機関リポジトリに登録し、リポジトリ側からリンクを貼っています。同じように科学研究費補助金（研究成果公開促進費）で構築しました。

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。

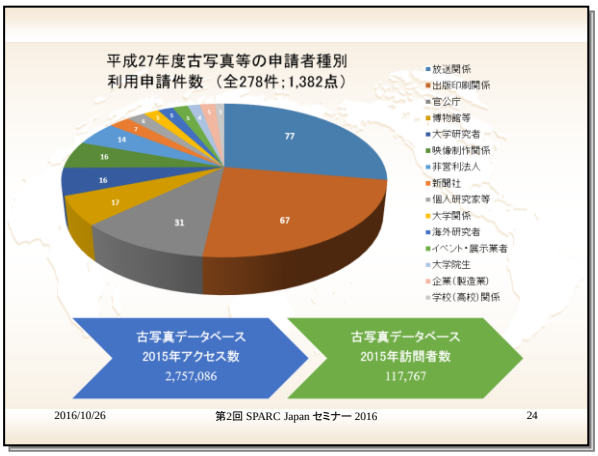
3-1.研究者・学芸員・図書館員のアプローチの違い

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。

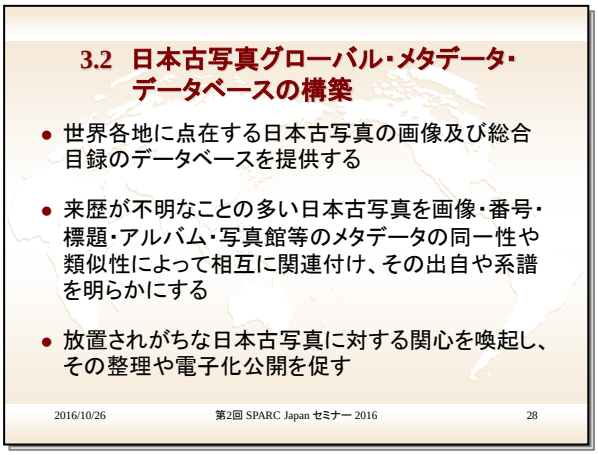
3-2.日本古写真グローバル・メタデータ・データベースの構築

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。

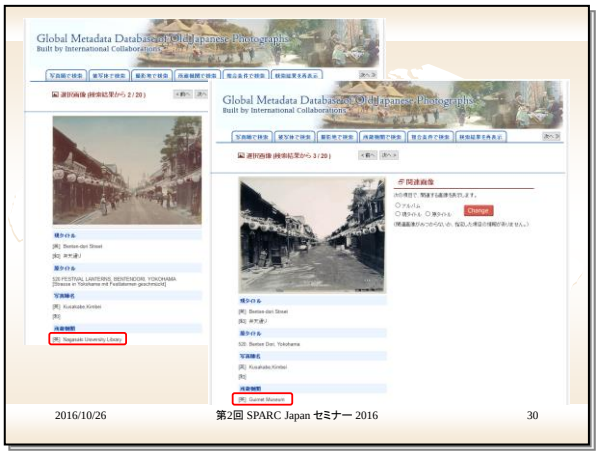
実際に取り組みはじめており、例えば、図 14 は横浜弁天通りの写真です。左が長崎大学、右がギメ東洋美術館に所蔵されています。少し時期が違いますし、彩色の状態も違います。

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。

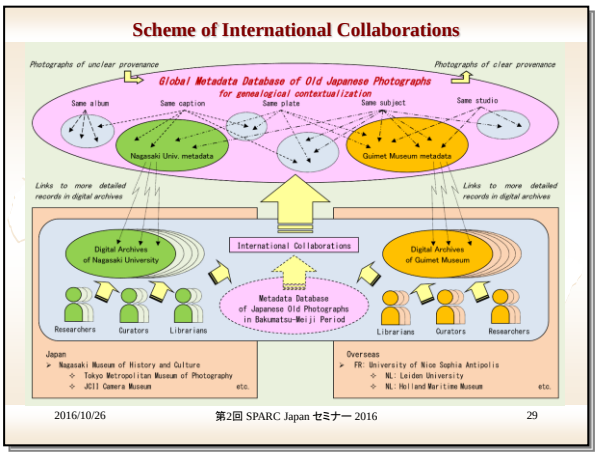
今日お話したデータベースについては、検索エンジンで「古写真」を検索すると、「幕末・明治期日本古写真メタデータ・データベース」が一番上に出てくると思います。そこに参考文献も置いています。他のデータベースを見たいという方がいらっしゃれば、「長崎大学電子化コレクション」を見ていただければと思います。



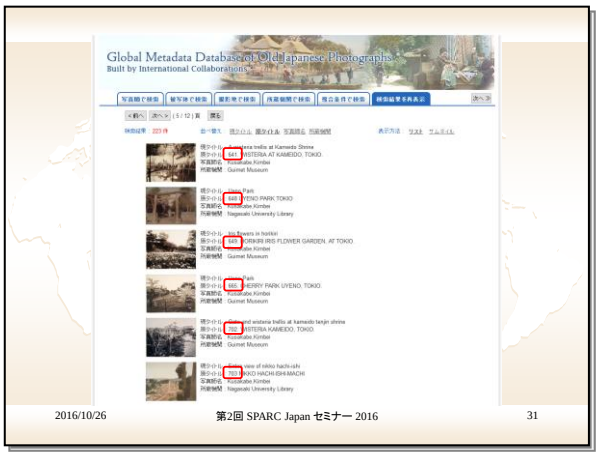
（図 12）



（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）

CC BY

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。

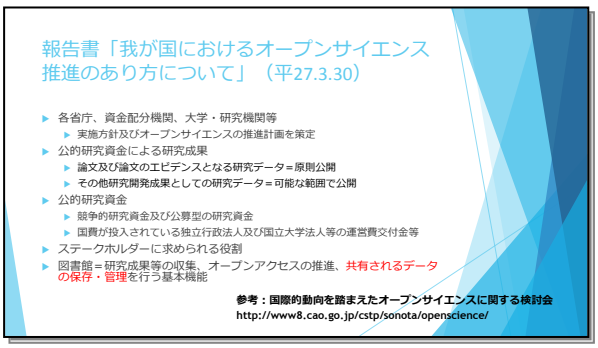
これを受けて、学術情報委員会で審議のまとめが出されました。そこから大学図書館の役割を幾つか抜粋して図 2 に書いてあります。論文のオープン化、研究データのオープン化、研究成果の散逸等の防止、人材育成です。今回、私の発表内容に関係するところしか赤字にしていませんが、研究データのオープン化に関しては、論文のエビデンスとしての研究データの公開に機関リポジトリを活用するようにというようなことが書かれています。

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。

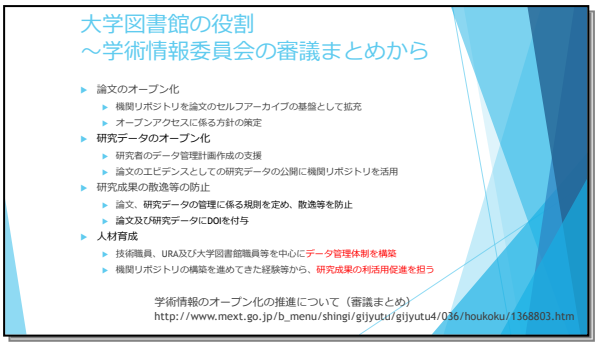
大学図書館はこれらの潮流を受けて、どのような動きをしているのかというと、2016 年 6 月に国立大学図書館協会の総会があり、そこで「国立大学図書館協会ビジョン 2020」が策定されました。その中で重点領域とされたのが、重点領域 1「知の共有：＜蔵書＞を超えた知識や情報の共有」、重点領域 2「知の創出：新たな知を紡ぐ＜場＞の提供」、重点領域 3「新しい人材：知の共有・創出のための＜人材＞の構築」の三つです。今回のデータの話に絡むのは、特に重点領域 1 と重点領域 3 です。

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）

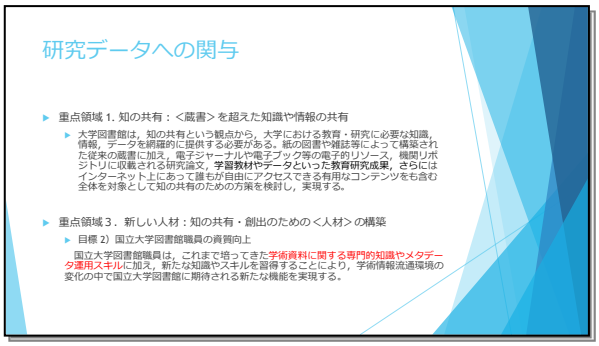


（図 2）

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。

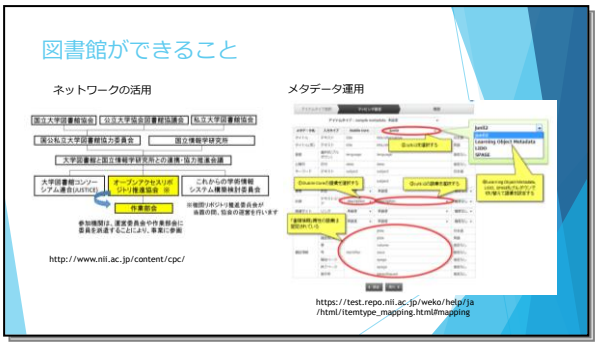
一つ目は概要です。今回の実験を通じて探りたかったことの一つは、研究者へのインセンティブの可能性です。研究者、特にメタデータ作成業務従事者の負担軽減、図書館が持つ流通面での知見の提供ができるかをこの取り組みを通じて考えたかったのです。

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。

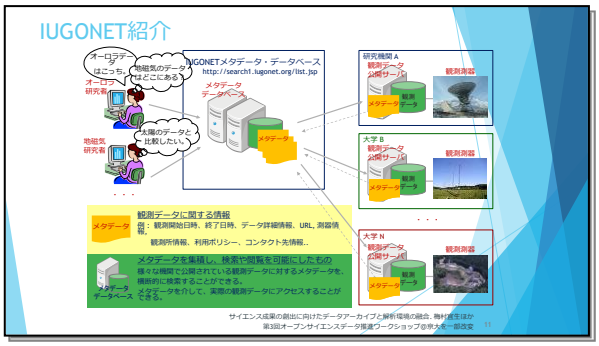
IUGONET 紹介

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。

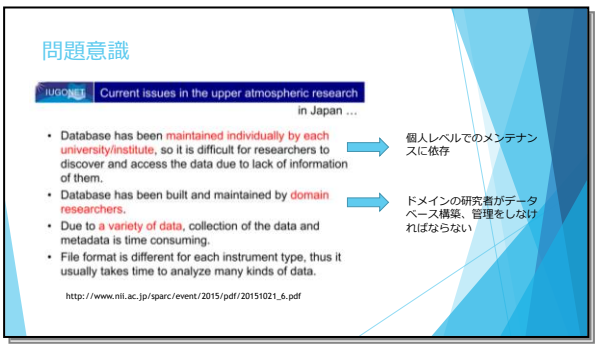
もう一つは、ドメインの研究者がデータベース構築、管理をしなければいけないことです。研究者は専門分野であるドメインの研究をしたいのですが、データベースの構築はドメインとはやや離れます。研究者は素人ではないので、ご自身で何とかしてしまうのですが、負担が大きいのは確かです。

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。

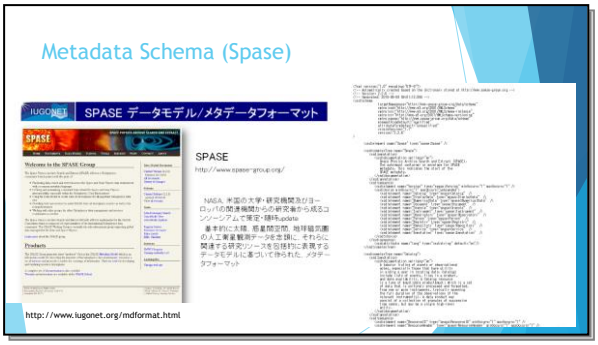
IUGONET で使っているメタデータスキームは SPASE というものです（図 7）。なぜか JAIRO Cloud に搭載されています。SPASE は NASA やアメリカの研究機関から成るコンソーシアムで作成しているメタデータスキーマで、太陽・惑星間空間・地球地磁気圏の人工衛星観測データを念頭に置いたメタデータフォーマットです。ご存じない方がほとんどだと思うのですが、私も半年前まではあまり分かっていなかったので、ここではこのようなものだと見ていただければと思います。

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。

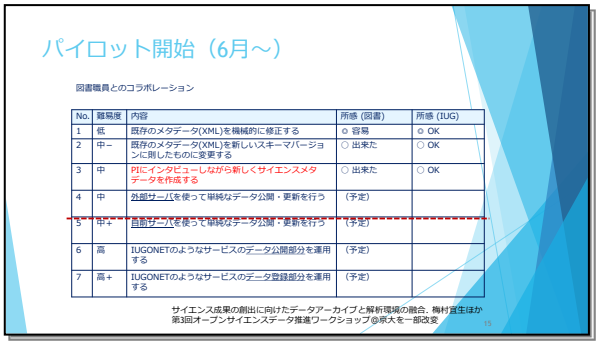
赤い線は業務境界線で、今は 4 番と 5 番の間に引いています。5 番以下は、国内の図書館の中で行っている例、自前サーバーを使って公開したりする例もなくはないのですが、図書館員の業務負荷が非常に大きくなります。また、機関リポジトリとの競合の問題もありますし、自前サーバーではなくクラウドにしたらどうかという話がすぐに出てくるので、共通化するのはこのあたりまでが一番いいのではないか、と IUGONET の担当者と話しています。



（図 6）



（図 7）

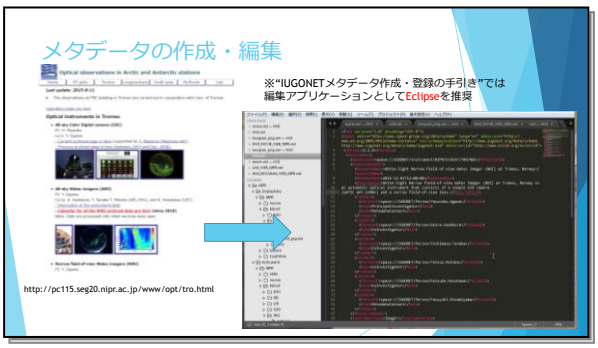


（図 8）

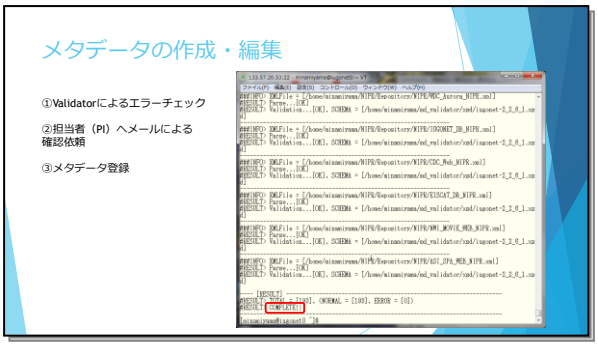
実際の作業はどのようなことをやっているかということで、メタデータの作成・編集です（図 9）。今、メインに行っている作業ですが、（前提として）ウェブ上に既に研究者が情報を載せています。「このようなことをやっています」というようなページをつくって、そこにデータを置いているのですが、それをスキーマに落とし込む作業をしています。IUGONET のメタデータ作成・登録の手引きは、Eclipse というソフトの使用を推奨していますが、私は自分で使いやすいテキストエディタで XML を直接編集してメタデータをつくっています。情報源をメタデータに起こすという作業は、通常の図書館の作業（目録や、リポジトリのメタデータ登録）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）

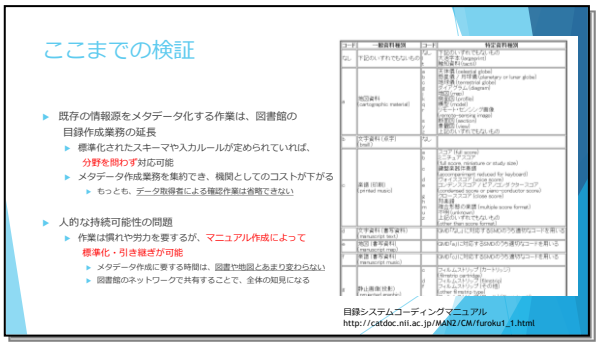


（図 10）

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 11）

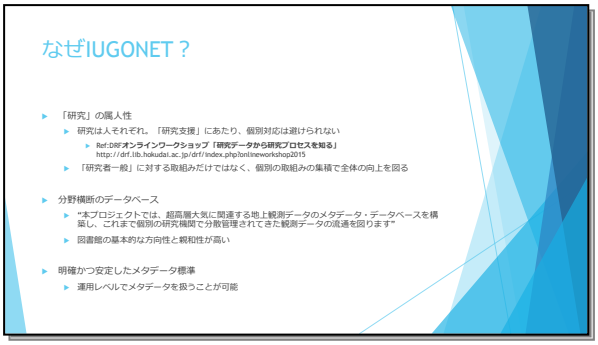
話をしたかったのです。右側に目録システムで取り扱ってきたコンテンツの表をあえて持ってきましたのも同じ意味ですが、地図・楽譜・静止画像など書籍ではないものであっても、最低限のレベルのメタデータは何とか作成できるということが言いたかったからです。

（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。

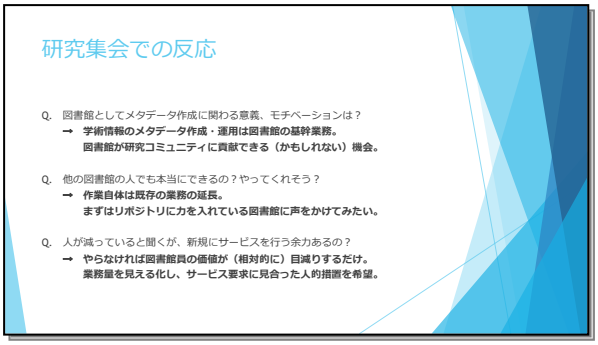
二つ目は分野横断のデータベースということです。IUGONET が分野横断のデータベースであったことが、やりやすさの一つでした。超高層物理という specific な分野ですが、その中でもさらに細かくデータの分野が分かれています。それを横断させるようなデータベースをつくっているということで、図書館は最終的な集合体として全部を大きく横断させるのが目的ですから、基本的な方向性として親和性が高く、図書館の業務としては触りやすいと感じていました。三つ目は明確かつ安定したメタデータ標準です。研究でメタデータを扱うときは、「このようなメタデータがあれば便利だ」という考えが研究を進めていく段階でどんどん出てく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

CC BY NC

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

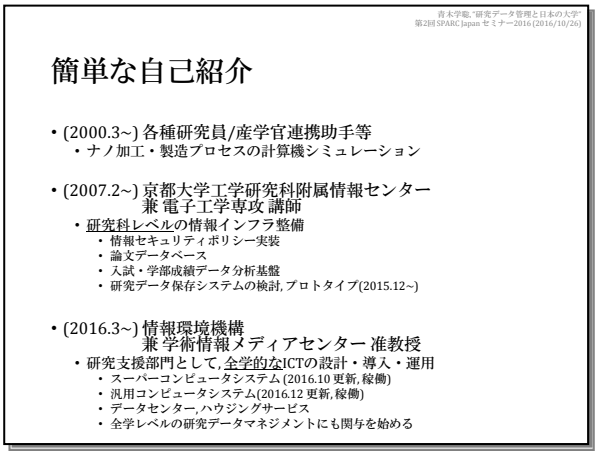
博士（工学）。ナノスケール加工・計測技術のシミュレーションを中心テーマに、2000年より各種研究プロジェクト研究員等として活動。2007年2月より工学研究科講師。教育研究活動と並行し、同附属情報センターにて研究科内の情報セキュリティ、情報インフラ整備、データ分析業務を担当。2016年3月より現職。 大学全体にわたる研究者、研究プロジェクト支援のためのICTシステムの計画、設計、運用に携わる。

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。

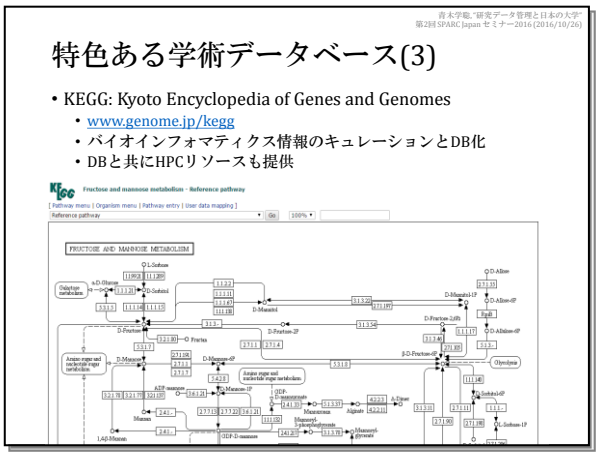
京都大学はやはり大きい大学で、しかも各部局の独立性が高いです。いろいろな小さい研究ユニットやセンターを持っていて、そもそもデータを集めて公開することが存在意義だというセンターもたくさんあるので、非常にたくさんの学術データベースがあります。

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。

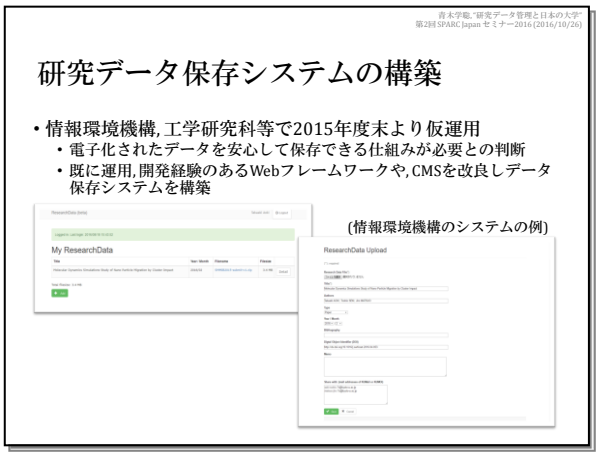
続いて 2015 年 7 月に、京都大学は、年度内に研究データの管理方法、保存方法を部局ごとに制定するように求めました。現在、2016 年度は、部局ごとに決定した管理方針ないしデータ保存計画を基に行動しているはずです。

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。

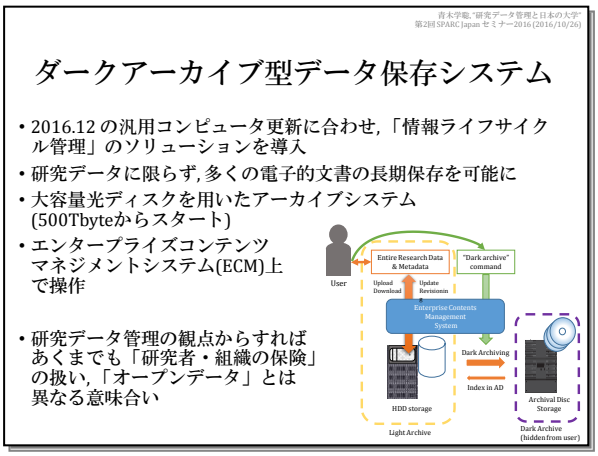
また、困ったことに、ほとんどのデータが死蔵されるものです。ここに収められたデータは何か研究不正が疑われたときに、ここに研究データがあると出すためのものだからです。だから、実はそれほど力を入れて構築したくないシステムで、できるだけ安く、それでなお皆さんに信頼されるシステムにする難しさも抱えています。研究公正とコンプライアンスだけのために IT システムを構築をしなさいと言われると、非常にやりにくい ICT システムになってしまうという事例になってしまっているのではないかと考えます。

2-3.オープン（データ）サイエンス促進のため

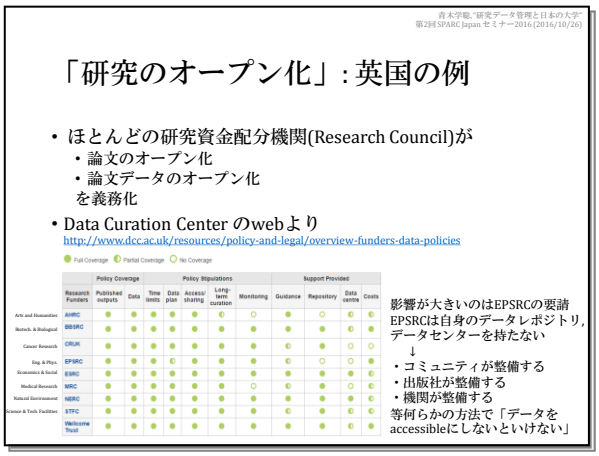
オープン（データ）サイエンスとは、今まで一部の研究者が自発的に実施してきたオープンデータ的な活動から広がり、研究者がデータを集めて、保存して、公開することを当たり前の作業にしていく運動だと私は認識しています。これに対して世界はどうなっているのか、あるいは日本国内はどうするべきかという話をします。

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。

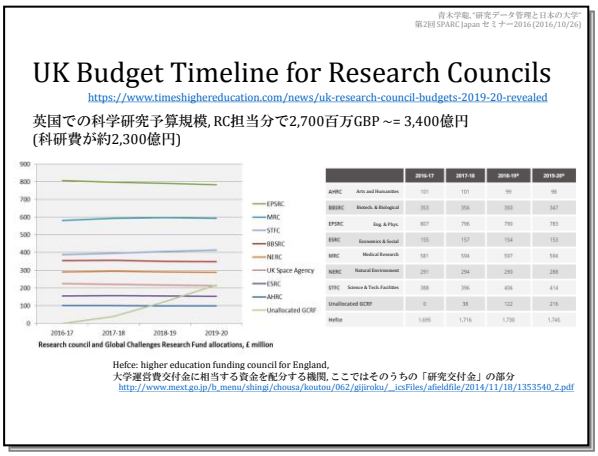
英国で一番インパクトが大きかったのは恐らく、英国工学・物理科学研究会議（EPSRC）から、工学と理学の分野はデータオープンアクセスにするように要請があったことでしょう。EPSRC はデータをオープンにするように言っていますが、データを置く場所は準備しておらず、自分たちで準備するか、コミュニティが準備しているリポジトリを使うようにと言っています。だから結局、工学・理学系の先生たちが研究をして、論文を出版したときに、そのデータにアクセスできるようにするためには、自分たちが所属している研究コミュニティがリポジトリを整備するか、出版社のデータジャーナルに投稿するか、自分自身でやるしかないのです。

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。

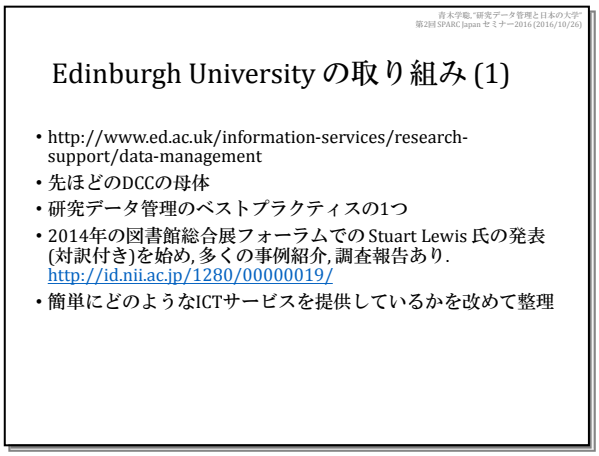
それに対して、英国の各大学はどうしているのでしょうか。一番先進的な取り組みをしているといわれているエジンバラ大学は、研究データマネジメントについて集中的に研究している Digital Curation Centre を生み出した母体になっている大学であり、研究データ管理を具体的に実践している大学の一つです（図 8）。

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）

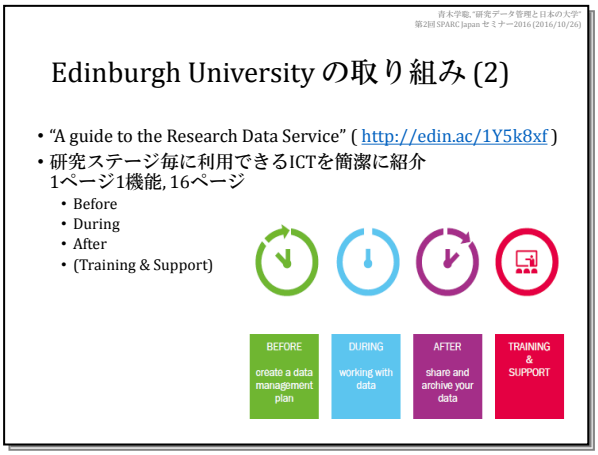
う四つの視点でエジンバラ大学では何ができるか、何を提供しているかをまとめています。1 ページ 1 機能で 16 ページ、ぱらぱらと読んで、なるほどと思える内容です。

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）。

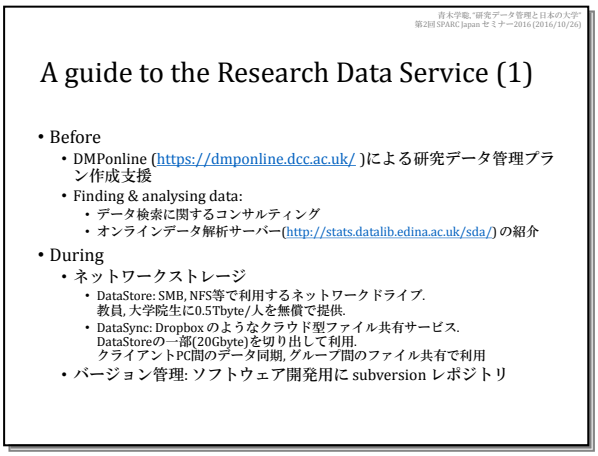
エジンバラ大学の経済・社会学系では、オンラインデータをその場で解析する仕組みも提供しています。このあたりはそれほどインパクトはないのですが、一番研究者にとって魅力的に移るのは、研究をしている最中にデータをどう扱うか、どう保存するかということです。そこでエジンバラ大学では、教員・大学院生に対して、0.5Tbyte、500Gbyte ぐらいのネットワークファイルストレージを使っていいという話をしています。当然、全員が使っているわけではないですが、最大これぐらいまでは自由に使っていいということ、だと推測しますが、少なくとも、自分たちの普段の生活で出てくるデータは全て大学が提供するネットワークドライブ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。

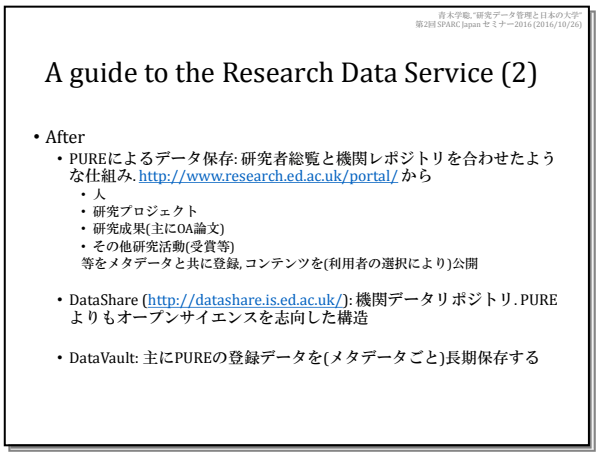
これとは別に、もう少しディテールにこだわった研究のデータセットを公開する機関データリポジトリも持っています。



（図 9）



（図 10）



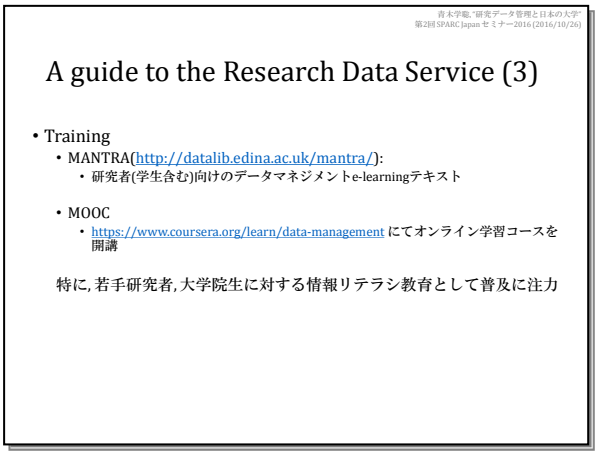
（図 11）

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。

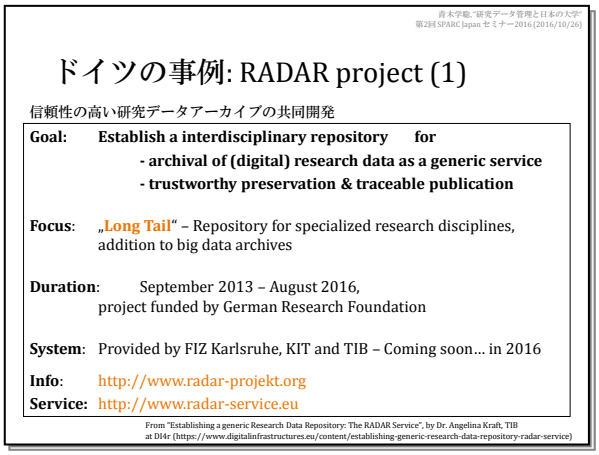
全員がやらなくてはいけないとなった場合は、非常にロングテールなものをサポートしなければいけません（図 13)。よく頑張っている、非常に頻繁にアップデートするデータリポジトリだけでなく、その後ろの方にある、何に使われるか分からないけれども残しておきたいデータまできちんとカバーしなくてはいけません。そうすると、可能な限りメンテナンスフリーで、汎用的な仕組みをしっかり入れなくてはいけません。ということは、できるだけコストを下げるという視点がどうしても必要になってきます。こういうものは果たしてあるのか、ないのかということを、これから各機関で整備しなさいと言われたら、どうしても考えなくてはいけないことに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

2-3-2. ドイツの事例

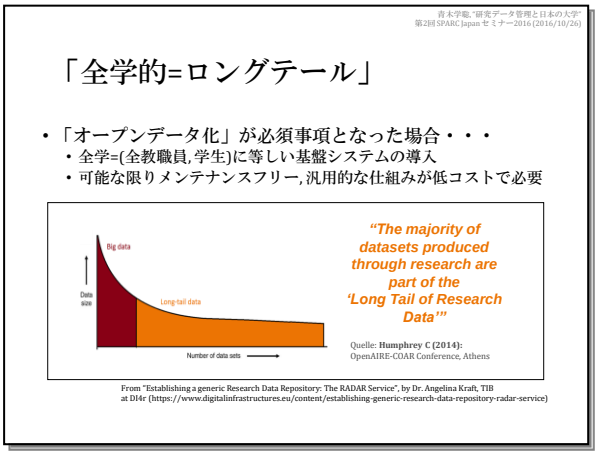
それに対応する形として、ドイツでは、RADAR project という、研究データアーカイブのシステムを共同開発して、それを各機関へ提供するというプロジェクトがあります（図 14)。ステージとしては、研究を始めていて、共同作業が終わり、研究が終わったとき



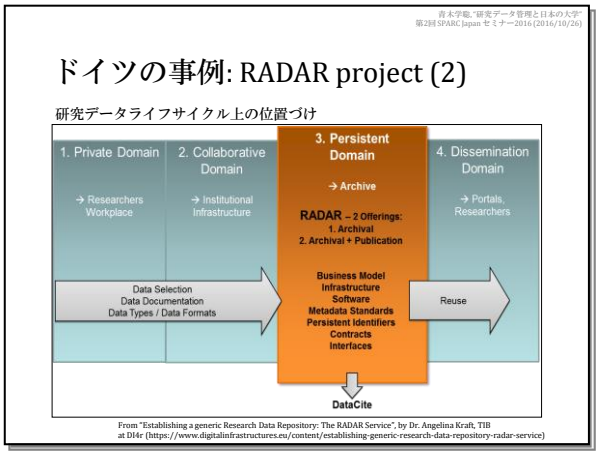
（図 12）



（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。

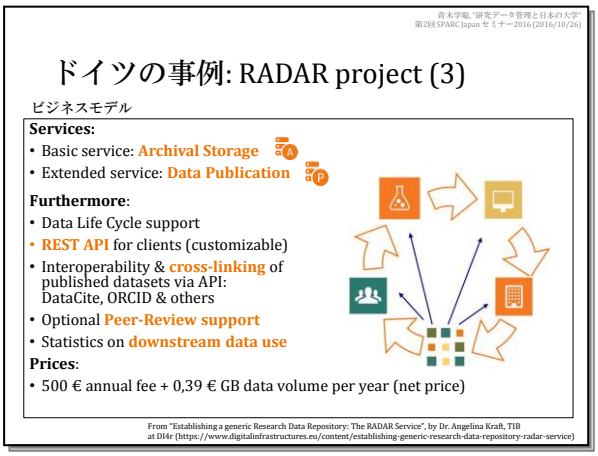
3.京都大学は何をするか

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。

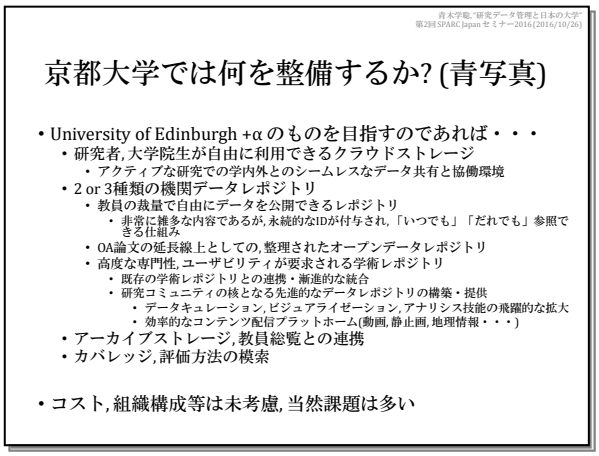
二つ目は、2～3 種類の性格の異なる機関データリポジトリです。データをアーカイブする場所、ためておく場所として必要になります。1 種類目は、教員の裁量で自由にデータを公開できるリポジトリです。先生たちが何かの機会で、このデータは面白いと思ったとき、あるいは、どこかの学会で発表して簡単なペーパーにしたので見てほしいというときに、それに永続的な ID、DOI かまたはそれに類するものを自発的に振って公開できる仕組みです。そこにはいわゆるキュレーションなどの作業はありません。つまり研究者が自発的に公開するドキュメントに対して永続的な ID を付けられる仕組みを持ったリポジトリです。

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。

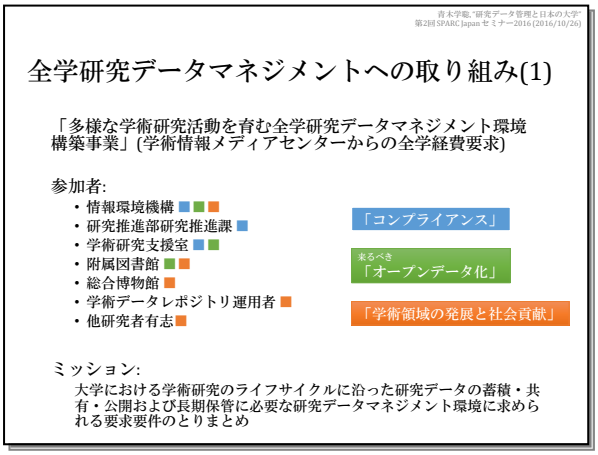
いろいろなデータベース、リポジトリが出現、整備された際には、結局、うまくいったのか、いかなかったのかということは絶対に言われると思います。パフォーマンスをどう評価するかは考えておかなければいけませんが、全くアイデアがない状況です。幾つかの研究でオープンアクセスは経済的価値を生んだというレポートは出ていますが、各機関がオープンアクセスにしてどれくらいの価値があったかということについては、何とも言い難い、評価のしようがない部分があるかもしれません。ここは非常に難しいところだと思います。

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。

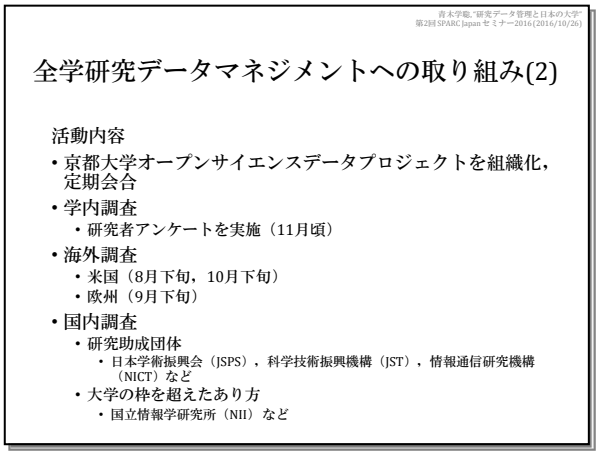
4.全学研究データマネジメントへの取り組み

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

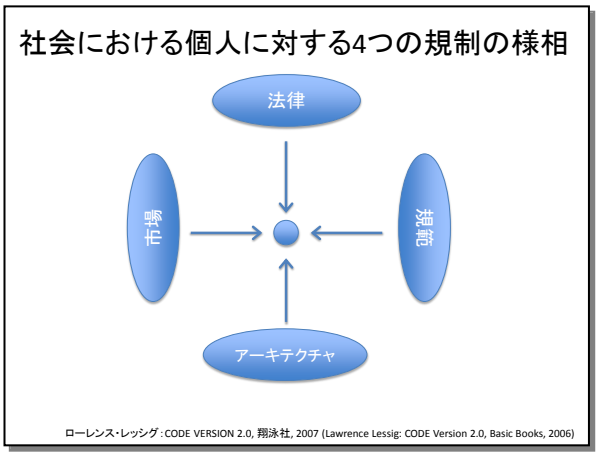
研究データに関する利活用に関する関心が近年、国内外で高まっている。本講演では、オープンサイエンスの流れの理解とその上での研究データ利活用の枠組みについて概要を説明する。その上で、国内及び国際的な動向について概観する。国内では DOI の RA(登録機関)であるジャパンリンクセンターが 2014 年に行ったデータ DOI 実験プロジェクトを契機に分野横断的なつながりができ、それが研究データ利活用協議会の発足につながった。国際的には RDA(Research Data Alliance)が 4 年前より活動を始めており、funder、研究機関、出版社等を巻き込んで、横断的なつながりを形成している。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

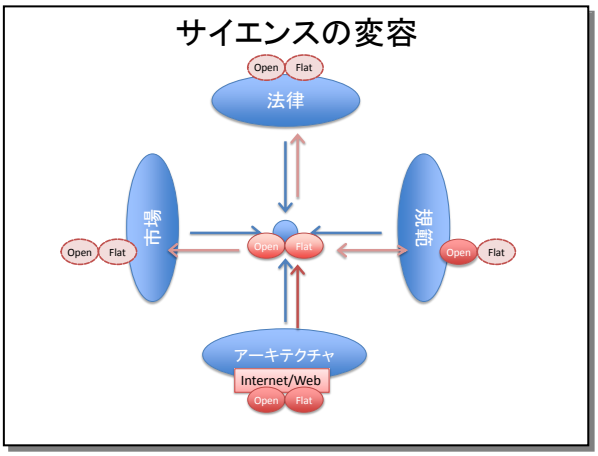
ネットワーク、コンピューターのアーキテクチャがどんどん変わりつつあり、それが直接・間接にわれわれに影響を与えているのが今のオープンサイエンスの世界です。図 3 の一番下が、ソフトウエア、コンピューター、ネットワークの世界で何が起きているかです。その世界から生まれてきたのがクリエイティブ・コモンズ、Web2.0 などです。

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。

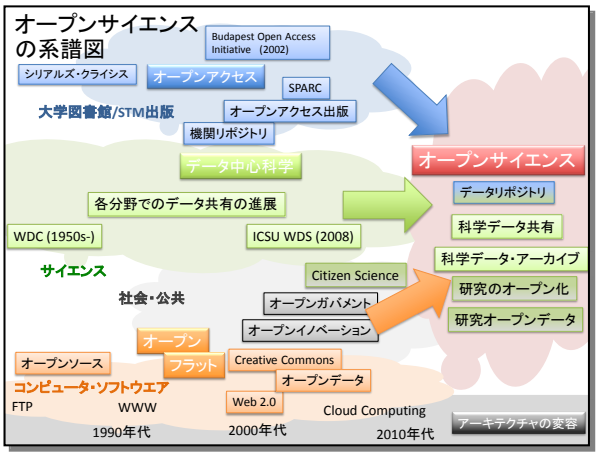
出版物も今までは紙で出版するのが良かったのですが、インターネットの発達によってウェブで出版することで、オープンとのつながりができてきたというのが一番上の青色です。

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）

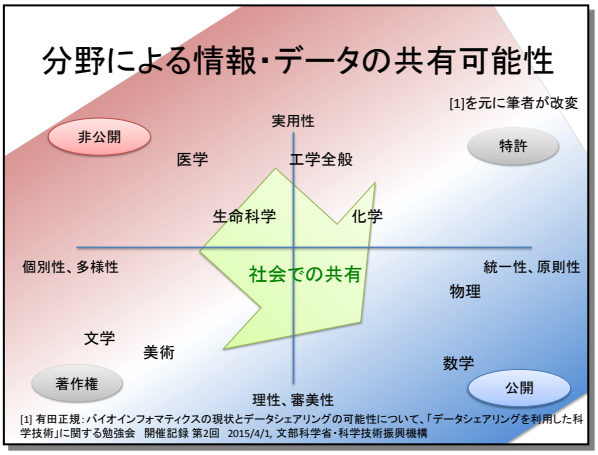
ういう状況で、まだら模様がオープンサイエンスの今だということです。

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。

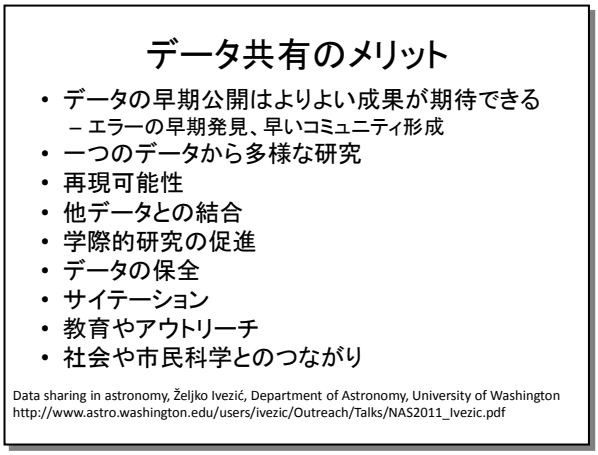
それを無理やり図にすると図 4 のようになります。二つのラインはかなり主観的に分けています。個別性・多様性を探求する学問なのか、統一性・原則性を探求する学問なのか、実用性を探求する学問なのか、理性・審美性を探求する学問なのかをマップしようとする試みです。例えば、理性・審美性を下に置いて、実用性を上に置いたときに各学問がどこにあるか。数学や物理は、より統一性や理性を探求するようなものなので、右下でしょうか。逆に工学は比較的上の方です。文学・美術は個別性・多様性を尊びつつ、ある種の理性・審美性を求めるので左下としています。



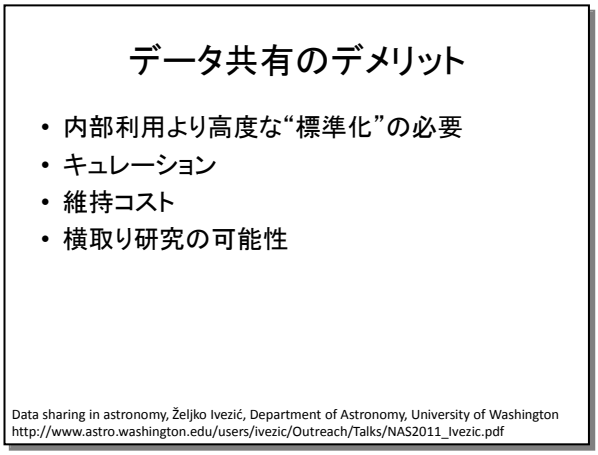
（図 4）

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）



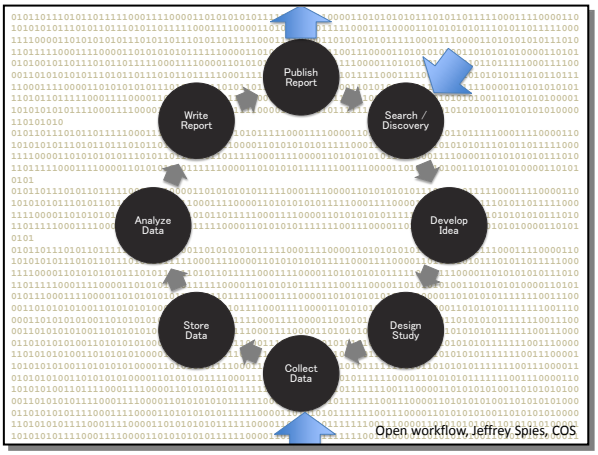
（図 6）

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。

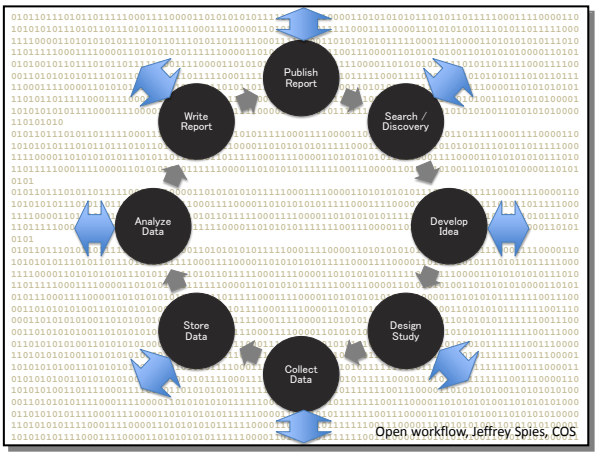
研究データ共有の枠組み

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）

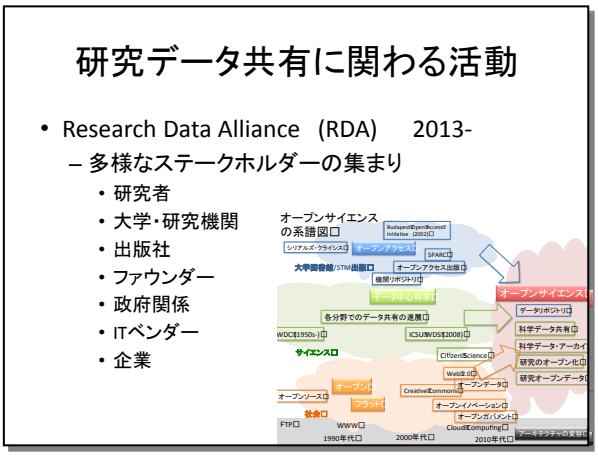
RDA は、研究データ共有と交換の障害を減らすためのインフラストラクチャーとコミュニティ活動の発展と世界的なデータドリブンイノベーションの加速に焦点を当てた国際的な活動です。皆が共有できるインフラをつくりたい、それを支える人の活動をつくりたい、プラス、データドリブンイノベーション、社会インパクトを与えることなど、かなりスコープを広く取っています。

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。

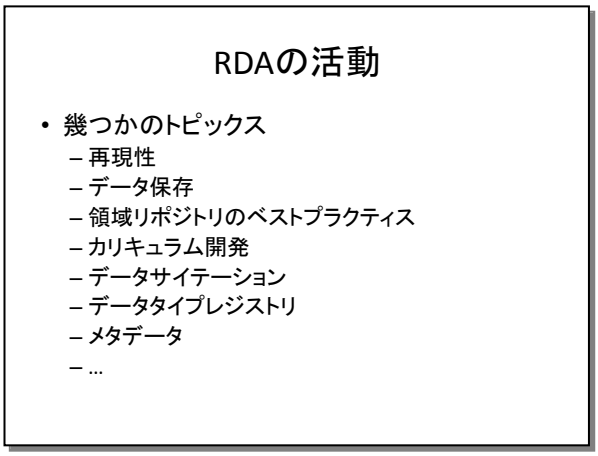
ここはトップダウンの組織ではなく提案ベースで、人が集まってインタレストグループとワーキンググループをつくり、リコメンデーションなどのアウトプットを出します。ワーキンググループでは 18 カ月で何かアウトプットを出すというプロセスだけが決まっていて、あとは自分たちで提案し、人が集まれば承認されるという制度です。

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）

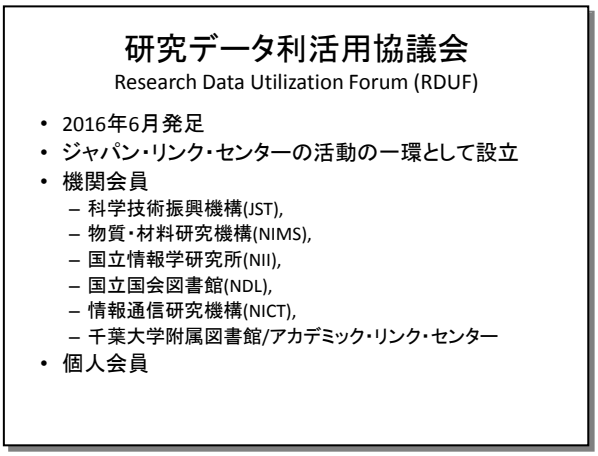
り小さな事務局があるだけです。ワーキンググループ・インタレストグループが勝手に集まって、勝手にオンラインで議論して、プレナリーミーティングのときにサマリーを報告する仕組みになっています。

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。

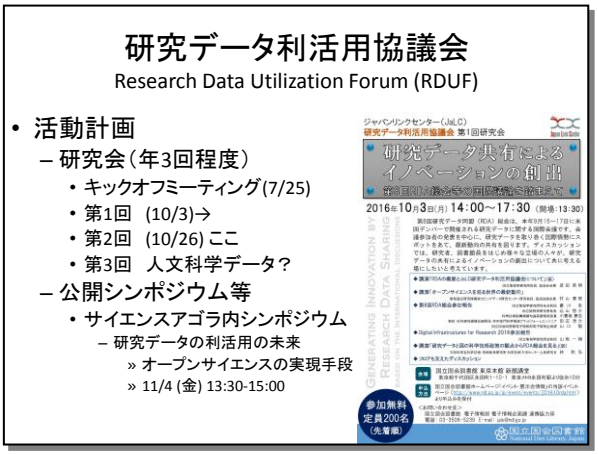
目的は、研究データに関する多様なセクター、特に実務者を集めることです。あまりお偉い方が顔を合わせる場所ではなく、実際にデータを取り扱う人たちが集まる場をつくりたいのです。集まって、研究データの共有と公開に関する課題を共有します。分野を超えると全く問題が違う、でも実は同じ問題もあるかもしれないということを共有したいのです。それを挙げるだけでも十分に価値があります。他には、うちではこういう技術を使っているということを共有できるとうれしいです。

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

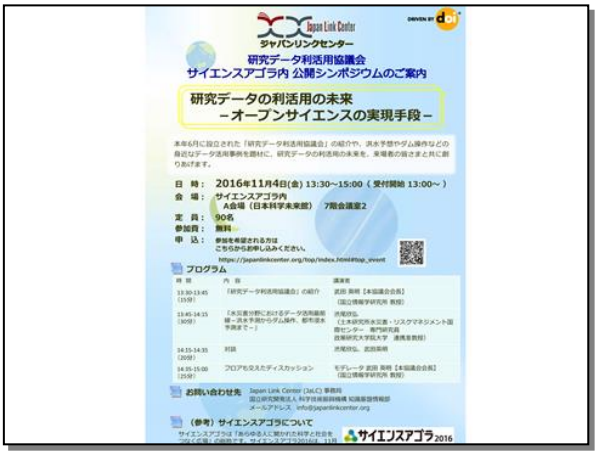
ています。

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。

SPARC Japan NewsLetter
ISSN 1883-826X
NO.32 2017 年 4 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。今回は国立情報学研究所実務研修（CERN）についてご報告します。

CONTENTS

■ SPARC Japan 活動報告
SPARC Japan 運営委員会
国立情報学研究所実務研修（CERN）

■ SPARC Japan セミナー報告
概要
参加者から
企画後記
ドキュメント
（講演・パネルディスカッション）

■ SPARC Japan 活動報告

CC BY

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/

国立情報学研究所実務研修（CERN）

第 4 期及び第 5 期 SPARC 事業の達成目標の一つとして「オープンサイエンスへの活動スコープの拡大」を掲げており、2016 年に高エネルギー物理学分野の情報サービスに係る国際連携協定を締結しました。その具体的活動として国立情報学研究所実務研修生を CERN に派遣しましたので報告します。

- 物理学分野のデータベースに関するデータキュレーション（国立情報学研究所実務研修）

  NII、KEK（高エネ研）、CERN の 3 機関の国際連携協定に基づき、国立情報学研究所実務研修の一環として、以下を実施しました。

  研修期間：平成 28 年 3 月 11 日（火）～平成 29 年 3 月 2 日（木）

  所属機関・氏名：京都大学・大村　明美

  研修場所：CERN（欧州原子核研究機構）

  http://www.nii.ac.jp/hrd/ja/jitsumu/h28/index2.html

■ SPARC Japan セミナー報告

CC BY

第 3 回 SPARC Japan セミナー 2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

2017 年 2 月 14 日（火） 国立情報学研究所 12F 会議室 参加者：118 名

本セミナーでは、今年度のセミナーの総括として、これまでのセミナーの内容を受けつつ、オープンサイエンスのあり方とその方向性について様々な面から検討しました。研究データをテーマとした研修の報告、トレーニングツールの紹介もあり、オープンサイエンスに係る企画や実践の参考となりました。

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/2016/20170214.html)

CC BY

概要

今年度の SPARC Japan セミナーでは，年間テーマの「オープンサイエンス時代の文献とデータの流通 ：科学的知識創成の新たな標準基盤へ向けて」に対し，第 1 回セミナー「オープンアクセスへの道」では，グリーンオープンアクセスとゴールドオープンアクセスの関係性や担うべき役割について議論がなされ，国内研究者の APC 支払額の把握と SCOAP³ モデル発展の必要性などが確認された。

第 2 回セミナー「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」では，データ・サイテーション，データ・ジャーナル，データ・オーナーシップなど研究者へのインセンティブのあり方や，それを支える研究データ管理の具体的方策（オープン化にかかるコストとのバランス，研究データ管理に対する報酬等），などが議論された。

第 3 回セミナーでは，上記を受けて，一年間の SPARC Japan セミナーを総括し，オープンサイエンスを「しなければならないこと」「すべきこと」「したほうが利益があること」「せざるを得ないこと」など多角的に再考することで，オープンサイエンスの先にある科学的知識創成の新たな標準基盤について考えてみたい。

＊ 標準基盤 ： 学術研究（＝科学的知識創成）を支援する基盤は，それを構築するプロセスにおいて，多分にボトムアップな活動によって構築され，いずれは利用者に意識されることなく学術研究の環境として位置付けられていく。このようなプロセスが，インターネットスケールで起こっているのが現在であり，地域や学術分野を越えて，学術環境のスタンダードを確立する活動が芽吹いている。本セミナーではこのような活動がもたらすスタンダードを標準基盤と定義する。

参加者から

（大学/図書館関係）

・東大の小野先生の発表でのアンケート結果がとても興味深かったです。

（大学/研究者）

・北本先生の講演はディープラーニングの最新動向がきけて大変興味深かった。Dekker 氏の発表はヨーロッパの現状がよくまとまっていてよかった。尾城氏の発表は実践的で役立ちそうである。ぜひ使ってみたい。

（企業/図書館関係）

・情報がオープンになるスピード、びっくりしました。

（その他/図書館関係）

・今までのオープンサイエンス、オープンデータのセミナーやフォーラムは概念や理想の話あるいは実現の道程は遠く、どうしたものかと思う話ばかりでしたが、今回は実際の現場に沿った話がきけて前向きな気持ちになれました。

・従来のオープンデータいけいけ！の勢いから制限共有、制限公開のほうへ流れが変わったのかと思わせる講演がいくつかあり、パネルディスカッションで小賀坂氏や小野氏のお話を聞くまでは訝しく感じていました。けれども、最後にお二人が語っていたようにオープンサイエンス、オープンデータを否定するものでも制限共有ありきでもなく、研究者にオープンデータが浸透し納得のいく形で進めていくことが大切で、やるからにはオープン化の結果に責任を持つ、見通しをつけておかなければいけない、との話に、今まで以上に重大さが身にしみました。また勢いに流されずに落ち着いていかなければと思いました。

（その他/その他）

・著作権とデータの関係について明確になった。DMP の具体例が紹介され、イメージがつかめた。

企画後記

CC BY

1 年間を総括する今年度第 3 回の SPARC Japan セミナーの主査という大役を、無事果たすことが出来ました。これも、登壇者、運営委員、事務局、企画 WG、ご来場の皆さまのお力添えがあってのことだと思います。この場をお借りして皆さまに御礼申し上げます。SPARC Japan セミナーは研究者と図書館職員が協働し、企画・開催しています。今後も様々な立場から学術情報流通に関わるホットな話題をご提供できるセミナーであり続けることを期待します。

梶原　茂寿

（北海道大学附属図書館）

今回は twitter 係として、セミナーの内容の情報発信を担当しました。「再考」という重いテーマでしたが、数歩先を進んでいる欧州の実際、データ共有、DMP やライセンスの現実の姿など、地に足を着けて「再考」できたのではないかと思います。企画担当でしたが、お話を伺いたかった方のお話を伺えて役得でした。

小野　亘

（東京学芸大学）

深層学習、計算機科学の分野では、arXiv が最新成果を出す主戦場になっており、プログラムソースや結果の公開・引用が 1 日単位で行われていることは驚きでした。スピードがオープンサイエンス推進の原動力になっていることは、現在ではある特殊な分野での出来事かもしれませんが、一つの要素として念頭に置いておいて損はなさそうです。また、オープンサイエンスの中でファンディング機関が果たす役割は大きいが、データマネジメントプランに関する現実的な実施方

針・評価軸策定は、日本だけでなく欧米でも手探り状態であることも分かりました。今後は、データの提供者である研究者を大きく巻き込んでいく方策を考えていかないといけないですね。

能勢　正仁
（京都大学大学院理学研究科）

本年度の最後を締めくくるセミナーは第 1 回の論文のオープンアクセス、第 2 回の研究データ共有の議論を踏まえつつ、総括として新しい研究プラットフォームの様子をさぐり、具体的に何ができるかを考えるというものでした。欧州を中心としたオープンサイエンスのビジョンから DMP 義務化や作成支援まで、幅広く採り上げられ、研究者の本音も飛び交うディスカッションコントロールが中々に大変でしたが、様々なステークホルダが集まるようになった SPARC Japan セミナーの醍醐味として今後も対話の場が提供できればと思いました。

林　和弘
（科学技術・学術政策研究所）

本セミナーでは twitter での発信を担当しました。第 1 回、第 2 回を受けての総まとめとなる今回、議論が収斂する方向に向かうかと思いきや最先端の話題が次々と飛び出し、改めてこのテーマの奥深さと関係者の本気度を実感しました。国内でも着々と事例が積み重なっている今、来年度の展開が一層楽しみです。

南山　泰之
（国立極地研究所）

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

開会挨拶／概要説明

梶原　茂寿
（北海道大学附属図書館）

梶原　茂寿
2014年より、北海道大学附属図書館本館に勤務。学術システム課システム管理担当で、図書館情報システム及びネットワークの管理と機関リポジトリを担当。2015年、2016年SPARC Japanセミナー企画WGメンバー。

第 3 回 SPARC Japan セミナー2016「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」にご参加いただき、誠にありがとうございます。

第 3 回セミナーで目指すもの

今年度の SPARC Japan セミナーの年間テーマは、「オープンサイエンス時代の文献とデータの流通：科学的知識創成の新たな標準基盤へ向けて」でした。

第 1 回セミナー「オープンアクセスへの道」では、文献に焦点を当て、グリーンオープンアクセスとゴールドオープンアクセスの関係性や担うべき役割について議論がなされ、国内研究者の APC 支払いの捕捉、SCOAP³ モデルの発展の必要性などが確認されました。

第 2 回セミナー「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」では、データの流通に焦点を当て、データサイテーション、データジャーナル、データオーナーシップなどの研究者へのインセンティブの在り方、それを支える研究データの管理の具体的な方策などについて議論がなされました。

そして、この第 3 回セミナーでは第 1 回・第 2 回のセミナーを受けて、1 年間の SPARC Japan セミナーを総括し、オープンサイエンスについてさまざまな方面から再考してみます。オープンサイエンスの先にある、科学的知識創成のための新たな標準基盤について、皆さまと共に考えたいと思います。

セミナーの流れ

本セミナーでは、午前に二つの基調講演を行い、午後は五つの発表を行います。その後、パネルディスカッションということで、長丁場になりますが、皆さまにも活発に発表、質問などしていただけるようになっていますので、ぜひご参加いただければと思います。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

欧州から見たオープンサイエンス

Ron Dekker

(欧州委員会研究・イノベーション総局

(European Commission (DG Research & Innovation)))

CC BY

講演要旨

オープンサイエンスは政治的な課題として重要になっている。2016 年に欧州理事会でオランダが議長国となっている間に、28 加盟国はオープンサイエンスに関する EU 競争力担当相理事会の結論を採択し、オープンサイエンスに関するアムステルダム行動要請(仮訳、Amsterdam Call for Action on Open Science)を表明したオープンサイエンスの会議があった。欧州委員会はヨーロッパオープンサイエンスアジェンダを策定し、オープンサイエンスクラウドやオルトメトリクスといったテーマでいくつかの専門家グループを設置した。また EU 加盟の数カ国では国家的にオープンサイエンスポリシーや戦略を制定している。その他、大学や出版者、資金配分機関等のステークホルダーもまた市民科学を含むオープンサイエンスの活動に関わっている。同時に、例えば学術情報流通の方法といったことへの重要な変更は先行者不利益によって阻害され、大きな財政的な再分配を必要とする。

そしていま、われわれはどのような状況にいるのか。どのようにオープンサイエンスエコシステムの変化を誘導できるか。どのようにして出版への配慮をしつつ、オープンアクセスへの移行を触媒し変化を起こすことができるか。さらに、研究データシェアリングを促進するためには何が必要とされるか。

本講演では、ヨーロッパの状況と上記の課題について述べるとともに、可能な解決策についての論議を呼び起こしたい。また出版や研究データについて主に語る一方で、オープンサイエンスの他の側面やイノベーションを含めた一般的な科学と社会の関係についても触れたい。

Ron Dekker

マーストリヒト大学で経済学を専攻し、労働市場研究のキャリアをスタートした。1995年にティルブルフ大学に入り、研究の軸足をデータマネジメントに移した。これはオランダ科学研究機構(Netherlands Organisation for Scientific Research (NOW))での業務のきっかけとなり、1997年にはデータ部門の責任者となった。その後、社会科学部と中央計画・研究所部のチームコーディネーターとして勤務した。2007年NOW研究所ディレクター。2013年にはオランダ高等教育研究ITイノベーション機構(仮訳、Dutch IT-innovation organisation for Higher Education & Research (SURF))でディレクターを務め、2014年にオランダのEU議長国としての準備のためにオープンサイエンスのプロジェクトリーダーとして教育・文化・科学省 (Ministry of Education、 Culture and Science) に出向した。2016年、欧州委員会研究・イノベーション総局オープンサイエンス担当加盟国出向専門家(仮訳、Seconded National Expert on Open Science at the European Commission、 Directorate-General Research & Innovation)となった。2017年3月から、Consortium of European Social Science Data Archives (CESSDA)ディレクターとなる予定。CESSDAはヨーロッパ最大のインフラ（いわゆるESFRI （研究インフラ欧州戦略フォーラム）Landmark）のひとつで、本拠地はノルウェーのベルゲンである。

私は欧州委員会に出向している立場なので、欧州委員会を代表して発言するわけではありません。今回は欧州委員会のオープンサイエンスに関する動向を踏まえ、オープンサイエンスとは何であるか、欧州のオープンサイエンスポリシーについて、出版物とデータのオープン化に焦点を絞ってお話しし、それからオープンサイエンスの展望について述べたいと思います。

オープンサイエンスとは何か

オープンサイエンスとは、科学を変化させるものです。アイデア・概念、分析、データ収集、出版、レビューという研究サイクル全体が変わります。この状況は、デジタル化によって生じています。生産されるデータの量は指数関数的に増加しており、透明性が求められるようになっています。例えば、心理学では公表された結果の約半分に再現性がありません。従って、説明責任を満たした研究を実現する新しい方法を見つけなければなりません。また、科学は、企業を含む社会全般とより良いつながりを持ち、特に貧困と疾病という大きな社会的課題に取り組む必要があります。

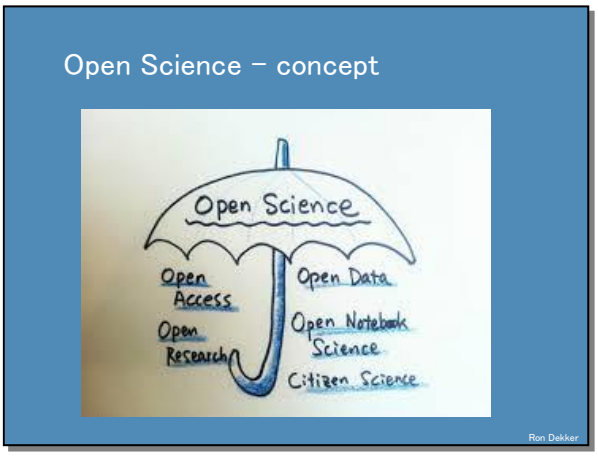
オープンサイエンスとは、データ、出版、ソフトウェア、オープンノート（研究公正）、市民が積極的に科学に参加する市民科学から構成された、包括的用語と見ることができます（図 1）。オープンサイエンスを一つに定義することは難しいですが、私は「オープンサイエンスとは、あらゆる種類の科学的知識は、発見と同時ぐらいに、早いタイミングで公開されなければならないという考えである」というマイケル・ニールセン（Michael Nielsen）の定義が好きです。直ちに共有しなかったり、壁を維持したりする理由はいろいろあるでしょうが、核心はできるだけ早く公開することです。

現在、既に利用可能なオープンサイエンスのツールがたくさんあります（図 2）。この内側のサイクルは、データ収集、分析、出版、レビューといった伝統的な科学です。外側のサイクルはオープンサイエンスで、内側のサイクルに比べてより外界との相互作用があります。オープンサイエンスとは、このように科学をつないでいくものなのです。

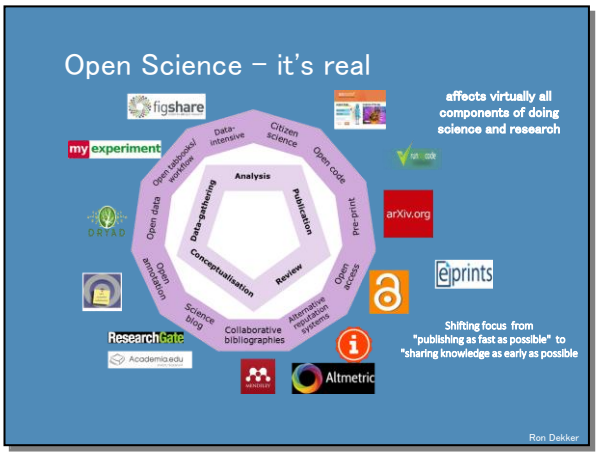
欧州のオープンサイエンス政策

2014 年に欧州委員会は、研究者、出版社、資金提供者といったオープンサイエンスのステークホルダーとのコンサルテーションを開始しました。主なトピックは、出版、データおよび研究インフラでした。

図 3 は、コンサルテーションで出てきたオープンサイエンスの主な促進要因です。同時に促進の障壁となるものも挙げられました。その一つが品質保証です。それは、直ちに公開された場合に、どのように品質を保証するのか。データを共有したいとき、いかにクレジットを提供するのか。データと知識を共有するのに十分なインフラはあるか。研究者とそれ以外の人々は利点を認識しているか、ということでした。そこで私たちは、データと出版物について多くの優先事項を設



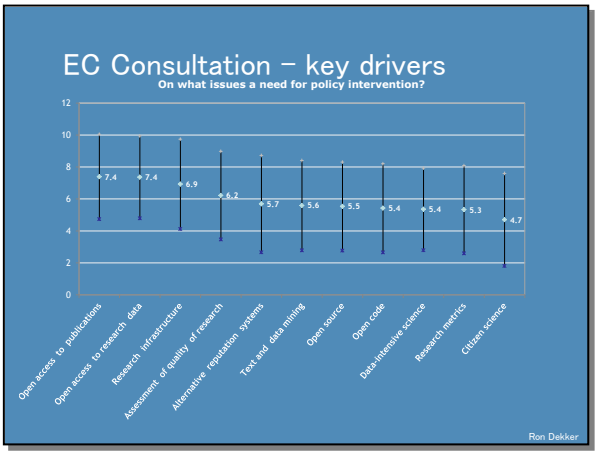
（図 1）



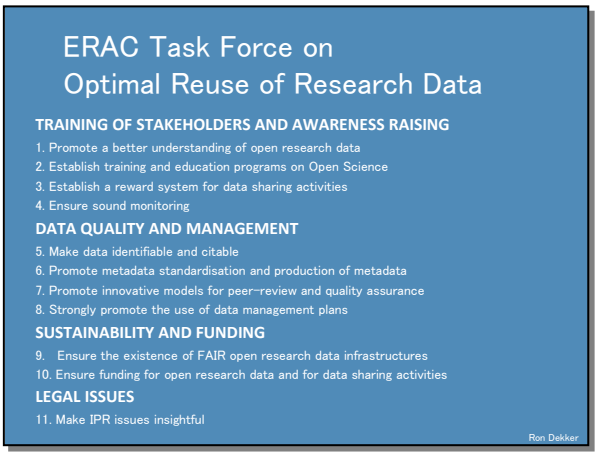
（図 2）

定しました。これは欧州委員会委員のカルロス・モエダス（Carlos Moedas）だけでなく、オランダ教育・文化・科学省副大臣のサンダー・デッカー（Sander Dekker）にも取り上げられました。

昨年、オランダは欧州理事会の議長を務め、オープンサイエンスを優先させることを決定しました。まず簡単に議長国オランダの成果を紹介し、欧州委員会の優先する政策について話します。私たちは、議長国時代にデータと政策に関する成果物をつくることに決めました。ERAC Task Force on Optimal Reuse of Research Data（研究データの最適な再利用に関するタスクフォース）は、データトレーニング、データ管理、持続可能性、IP 問題などに関する数多くの活動と勧告に取り組みました（図 4）。データの最適な再利用を追跡することはとても複雑です。私たちはここで、あえて「オープンデータ」という言葉は使いませんでした。



（図 3）



（図 4）

「オープンデータ」という言葉には、データは直ちに公開されるべきだということを示唆する面がありますが、データの中には機密性があるものや、そのデータを最初に使用する権利があるものもあるからです。そうした場合には、自分に再利用の権利があることを明らかにしなければなりません。

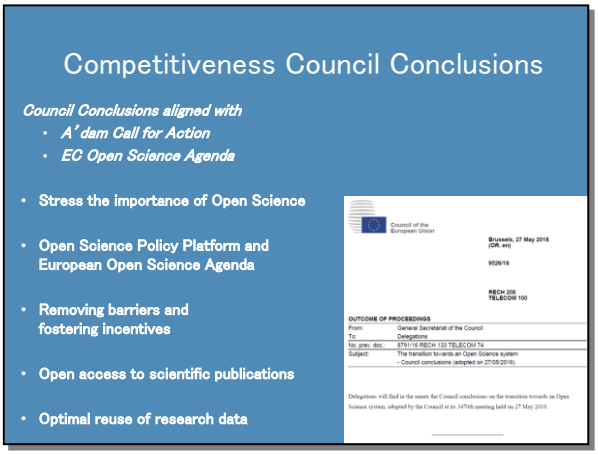
これらを実現するため、アムステルダムで会議を開催し、オープンサイエンスに関するアムステルダム行動要請（仮訳、Amsterdam Call for Action on Open Science）を表明しました（図 5）。そこで決まった目標は、一つ目は、2020 年までに科学出版物の完全なオープンアクセス化を達成することです。二つ目は、データ共有のための新しいアプローチを準備することです。三つ目は、データ共有の際に人を認識し、クレジットを与える新しい報酬と評価のシステムをつくることです。東ヨーロッパの現在のシステムでは、データの共有に対する報酬はありません。権威ある雑誌に出版すれば報酬が与えられますが、教育、データの共有、知識の共有などに対しては報酬がありません。この報酬制度を変えていかなければ、共有するインセンティブがありません。四つ目は、少し秘密めいていますが、つまり知識の共有と監視システムです。率直に言うと、国や大学には出版の費用について何らアイデアがありません。購読の契約費用は知っていますが、例えば、ゴールドオープンアクセスの論文掲載料（APC）については知らないのです。従って、これらのデータを収集し、出版コストに関する情報を共有する必要があり

（図 5）

ます。他にも、オープンサイエンスに関するアムステルダム行動要請には解決の方向性を提示する 12 の推奨事項があり、各ステークホルダーは自分がやるべきことに取り組みます（図 6）。

2016 年 5 月、欧州理事会の全 28 加盟国は、EU 競争力担当相理事会の結論を採択しました（図 7）。全加盟国が「オープンサイエンスを望んでいる」と表明したことは、強力な政治声明でした。この結論の一つ目は、オープンサイエンスの重要性です。二つ目は、ステークホルダーの代表で構成される Open Science Policy Platform を設立することです。このプラットフォームは、出版社、資金提供者、大学、応用研究の従事者、若手研究者からなる 25 人のメンバーで構成されています。欧州委員会は、オープンサイエンスを進める方法についてアドバイスを求めました。三つ目は、オープンサイエンスが進むべき方向として、できるだけ早く結果をオープンにし、テキスト・データマイニング（TDM）を行い、著作権を保持するということです。これは、有料アクセスかまたは全くアクセスがなくなってしまう、という現状から抜け出す一つの解決策になり得ます。四つ目は、出版物のオープンアクセスです。全ての加盟国は、2020 年までに出版物をオープンアクセス化することに合意しました。2020 年までのオープンアクセス化ということは非常に重要でした。なぜなら、政治声明で期限を示すことは重要なことだからです。五つ目は、研究データを、より一般的に最適に再利用できるようにすることです。データは公共財でなければなりませんが、この意見はデータを扱う全ての研究者の間で共有されているわけではありません。

（図 6）



（図 7）

欧州では、現在、強力な政治声明を出しています。これは全ての加盟国間で合意しています。私たちはオープンアクセスとオープンサイエンスに関する国家政策を実行に移すために、ヨーロッパオープンサイエンスアジェンダ、ハイレベル専門家グループ、オープンサイエンスポリシープラットフォーム、Horizon 2020 をつくりました。

アムステルダム行動要請と同様に、ヨーロッパオープンサイエンスアジェンダは、障壁の除去、インフラを提供するインセンティブをいかに与えるか、オープンサイエンスをノーマルサイエンスにすることを目的としています。オープンサイエンスは、できるだけ早く、知識を生産し共有するための一般的な方法とならなければなりません。

このアジェンダでは、報酬システム、品質および影響の測定、将来の出版モデル、探索可能・アクセス可能・相互運用可能・再利用可能（FAIR）なオープンデータ、欧州オープンサイエンスクラウド、研究公正、市民科学、オープンエデュケーションおよびスキルの八つの重要な問題に焦点を当てることが決められました。品質に関しては、現在のジャーナルのインパクトファクターは、研究や論文ではなく、ジャーナルの品質を測定するためのプロキシです。欧州オープンサイ

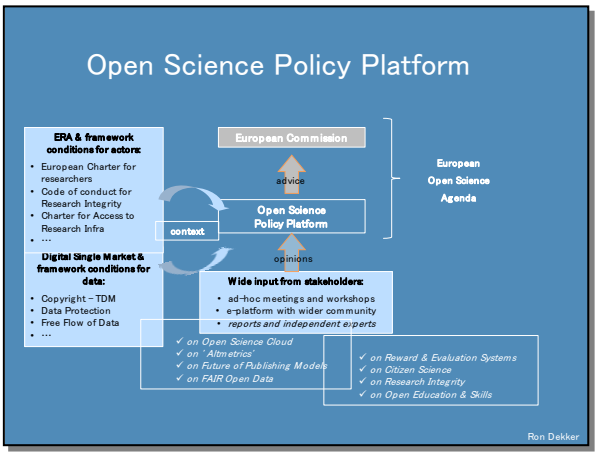
エンスクラウドは、データと全ての研究成果に対するインフラとしての役割を果たしています。研究公正に関しては、結果を再現可能なものにすることが重要です。欧州委員会は、これらの項目に関するハイレベル専門家グループを既に設置し、または今後設置する予定です。

また、欧州委員会は、オープンサイエンスポリシープラットフォームを導入しました。このポリシープラットフォームは、委員会に助言する 25 人のステークホルダーから構成され、ヨーロッパオープンサイエンスアジェンダの八つの項目を、欧州委員会がオープンサイエンスへの移行を促進し、触媒することを推進できるよう、包括的に把握することが期待されています。

チャートを使って説明します。まず左は入力側です（図 8)。オープンサイエンスポリシープラットフォームはワーキンググループを持ち、ステークホルダーから意見をもらうことができます。そしてプラットフォームは、最終的に欧州委員会に助言するだけでなく、結果を機関に還元します。ワーキンググループのステークホルダーがプラットフォームの方策に合意すれば、それは全てのステークホルダーによって容易に実施されることになります。

また、欧州委員会は、リサーチプログラムの中でオープンサイエンスに取り組んでおり、プログラムの中でも EU 第 8 次研究・イノベーション枠組み計画「Horizon 2020」が最も重要なものです。その中では、出版物をグリーンまたはゴールドでオープンアクセス化することを義務付けています。できるだけ早くデポジットし、領域ごとに決められた 6〜12 カ月間のうちに利用可能にする必要があります。ゴールドオープンアクセスの費用は、助成金から支払うことができます。同じことがデータにも当てはまります。助成金受領者は、データをリポジトリに保管し、アクセスを提供する必要があります。私たちは FAIR 原則にのっとって、出版物だけでなく、データも共有したいと考えています。

データ管理計画（DMP）のパイロットも行いました（図 9)。これは若手研究者に、他人に有益なものを生み出すことへの意識づけをするためです。パイロットへの参加は任意であり、本パイロットに関わる人の約 3 分の 2 が参加しています。本パイロットの適用範囲外のプロジェクトから、さらに 10%が自発的に参加しました。このパイロットについては、Horizon 2020 の全てのプログラムをカバーするように拡張することにしました。研究によって生成されるデータ、どのようにしてデータを提供するかといった手順は、むしろ簡単でシンプルであり、研究者はデータキュレーションとセットアップに関わる情報を求められます。これによって、研究者の、他の人にとって価値のあるものをつくりだしているという意識を高め、データを研究者のラップトップから引き出し、機関リポジトリや国家・各専門分野のデータベースに取り込むのです。この DMP は非常に重要で、データを共有するための必要条件です。



（図 8）

（図 9）

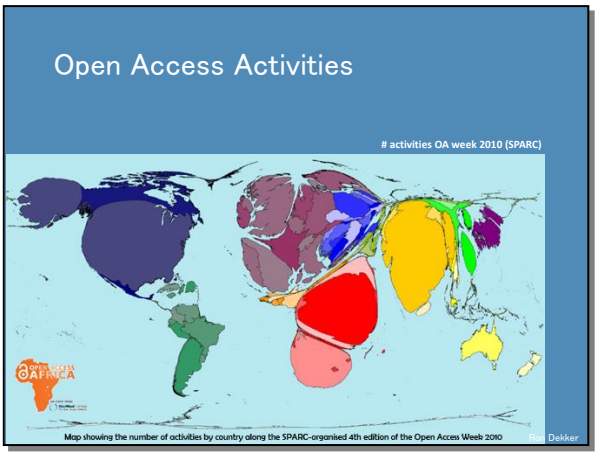
データと出版物のオープンアクセスに関する欧州委員会の政策活動をご紹介します。私たちは政治声明、ポリシー、ハイレベル専門家グループ、ステークホルダーから成るプラットフォームを有しており、これらを補助金交付規程に反映しています。

現在、ほとんどの国がオープンアクセスに関する政策を導入しています。実際に G7 では昨年、声明を発表しました。それは同時に G20 の目的でもあります。オープンサイエンスは、今度欧州で開催される G7 会議の議題になる予定で、報酬制度とクラウドインフラに焦点が当てられます。この議題は、日本と欧州委員会が共同で作成したものです。

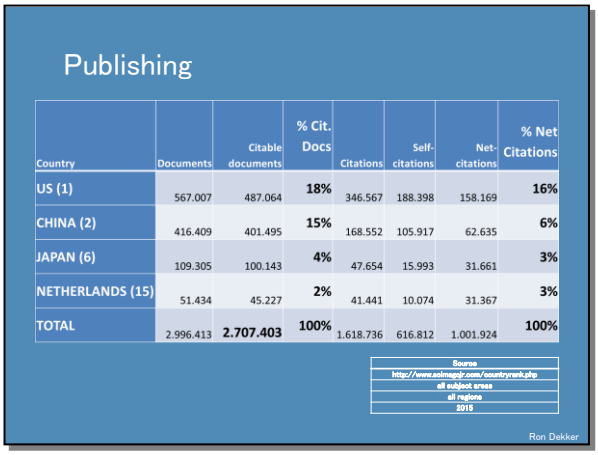
学術出版物のオープンアクセス

出版物とデータについてに移ります。誰がオープンアクセス政策に関与しているかを考えるため、まず私は SPARC による OA Week 2010 の図 10 を使って、オープンアクセス政策が活発な国を紹介します。オープンアクセスは多くの国々で議題になっています。しかし、2050 年に予測される人口動向を見ると、アジア諸国への大きな移動があります（図 11）。人々が住んでいるところは、すなわち、科学、イノベーション、経済成長の可能性があるところです。アジアで、科学を社会にもっとうまくつなげることができれば、科学の枠を超えて大きな可能性があることは明らかです。

出版に話を戻すと、毎年約 300 万件の出版物があります（図 12・13）。90%が引用可能で、その約半数が引用されており、自己引用を差し引くと、270 万件の論文につき約 100 万件が引用されていることになります。つまり、3 件のうち 1 件のみ引用されているので、インパクトファクターは 0.3 です。ところが、論文によってはインパクトファクターが 30 も 40 も付与

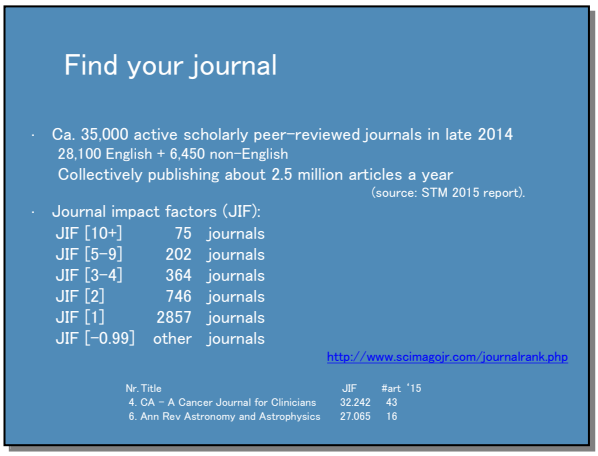


（図 10）



（図 12）

（図 11）



（図 13）

されます。これはつまり、多くの人々が引用を全くされていないということを意味します。これらの論文は誰のために出版されているのでしょうか。誰とこの知識を共有しているのでしょうか。データを見ると、引用率がオランダと日本で同じぐらいであることが分かります。

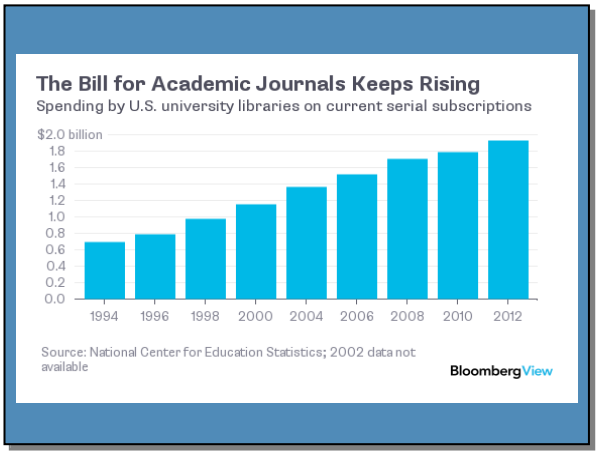
3 万 5 千種のジャーナルがあるので、ジャーナルを見つけることは難しいでしょう。さらに、これら 3 万 5 千種のうち 1%だけが 5 以上のインパクトファクターを持ち、3〜4 のインパクトファクターを持つものも 1%です。オランダでは、研究者は 2 以上のインパクトファクターを持つジャーナルに掲載することが推奨されていますが、これは全研究ジャーナルのわずか 4%にすぎません。

その関連性を述べると、インパクトの高いジャーナルに掲載したい場合はレビューを書かなければなりません。これが最も多く引用されるからです。しかし、天文学と物理学を例に取ってみると、2015 年にアクセプトされた論文は 16 件しかないので、その 16 件の中の一つでなければならないということです。この中に入れば論文は引用されるのですが、これが科学でしょうか。私たちは論文をこのようなジャーナルの一つに掲載するために、公的資金を費やしているのでしょうか。私には、これらのジャーナルに掲載されること自体が目標になってしまっているように思えてなりません。

出版社は、IP アドレスの助けを借りてマーケットを細分化し、インターネットを最大限に活用しています。多くのジャーナルを購読できることが魅力のパッケージ購読がありますが、これにより、一つの購読をやめることが、全てのジャーナルの購読をやめることになるため、難しくなっています。ジャーナルのインパクトファクターはまた、外に出ることが非常に難しい経済的ロックインをもたらします。ですから、他のステークホルダーらはジャーナルのインパクトファクターを品質の代わりとして使用するのです。これは、リサーチカウンシルで助成金を決定する際に関係します。大学のランキングを決めるときにも関係します。ここには先行者不利益があります。そのシステムから出ることを決めたら、もはや大学のランキングに入ることはできないでしょう。それは政治的または金銭的不利益をもたらすかもしれません。

従って、有意で肯定的な結果を得なければならないために、リスクの高い研究を行わず、権威あるジャーナルに掲載することにインセンティブが働きます。有意ではなく否定的な結果しか出なければ、トップレベルのジャーナルには掲載されません。これは出版バイアスにつながります。一例として、臨床試験では、公表されやすいのはポジティブな結果が出た試験ですが、インターネットで検索すれば、ネガティブな結果が出た試験や、有意でなかった結果が多く見られます。これでは私たちは情報のほんの一部を取得するだけになってしまうので、この報酬制度を変える必要があります。一方、現在のシステムは研究者と出版社には利益をもたらしています。研究者は出版社に著作権を与え、出版社はそれによってビジネスを行う機会を得て、優れた論文を提供してジャーナルの価値を高めています。出版社は研究者の評判を高め、その評判は研究者が大学内で助成金や威信を得るのに役立ちます。

これは win-win であると言えるかもしれませんが、誰かがお金を支払う必要があります（図 14）。各誌の価格上昇やジャーナルの数の増加によって、上昇し続けるジャーナルのコストを払わなければならないのは、図書館や大学です。このことは他のモデルを考えてみ



（図 14）

る理由の一つです。

私たちはなぜオープンアクセスに移行したいのでしょうか。第一に、インターネット技術を利用すれば、コピーや配布は簡単で安価です。また、資金提供者はより多くの利益を望んでいます。それは、出版においてのみ言っているのではなく、知識や成果を共有するということも意味しています。そのインパクトは、科学的なインパクトだけでなく、社会やイノベーションに対するインパクトまで拡大しなければならないのです。

重要な原則の一つが、公的資金を受けた研究の結果は公的に利用可能であるべきだということです。資金提供に値する研究は分かち合う価値があるはずです。科学を社会とつなげたければ、出版物や成果物へのアクセスを改善する必要があります。人々は今、専門分野内の成果ばかりに焦点を当てているので、アクセスの改善は科学に役立つかもしれません。エルゼビアのSTM Digestは、各専門分野で何が起こっているかを説明するためのものです。これは一般の人が使用するものではなく、研究者が他の分野について学ぶために使用されています。従って、これは科学が前進する助けとなっています。私たちは、科学の公正性と信頼の向上に取り組むことができるのです。

ゴールド、グリーン、ハイブリッドなどさまざまなバリエーションのオープンアクセスがあります。しかし、オープンサイエンスには多くの誤解があります。公表する義務はありません。それは特許と相反するものではありません。特許を取りたいのであれば、伝統的ジャーナルであろうとオープンジャーナルであろうと、まず特許を取ってから出版する必要があります。また、ピアレビューのプロセスにも相違ありません。どちらにもピアレビューがあります。オープンアクセスには悪質なジャーナルもありますが、それは伝統的なシステムでも同じです。一部のジャーナルでは既存の論文をコピーして購読モデルにすることさえあります。

つまり、われわれがオープンアクセスを望むなら、政策が必要です。お伝えしたとおり、ほとんどの国は既にオープンアクセスに関する政策を持っており、多くはグリーンに関してですが、一部はゴールドに関するものです。恐らく最も見事な政策はゲイツ財団のものだと思います。それは1ページで書かれ、五つの項目からなります。オープンアクセスで直ちに公開し、データを共有し、ゲイツ財団がコストをカバーして、公正な価格を支払うというものです。この方針には例外がなく、2 年間浸透に費やした後、発効しています。従って、出版物のオープンアクセスについては、私たちはどの方向性に進むべきかもう分かっているのです。

研究データの最適な再利用

研究データの最適な再利用については、さらに少し複雑になります。少なくとも欧州では、出版は一つの出版モデルから新しい出版モデルへの移行期間にあります。

研究データは項目の組み合わせです。欧州委員会には三つの柱があります。一つ目はインフラを管理し、コンテンツと連結する欧州オープンサイエンスクラウド（EOSC）です。研究インフラ欧州戦略フォーラム（ESFRI）は大量のデータをつくる大規模な研究インフラであり、コンテンツの一部です。二つ目の柱は、高性能コンピューティング（HPC）ネットワーキングソフトウェアを提供する欧州データインフラストラクチャー（EDI）です。これは、大規模な欧州 HPC、ネットワーキング、ソフトウェアを持ち、これらのサービスと活動を組み合わせて一つのサービスに集約するものです。つまり、サービスとしてのインフラです。研究者として、私はストレージと計算機能が必要ですし、同僚とつながる必要もあるので、これらはサービスとして提供されて然るべきです。三つ目の柱はアクセスを拡大することで、それには中小企業、産業界、政府が含まれます。

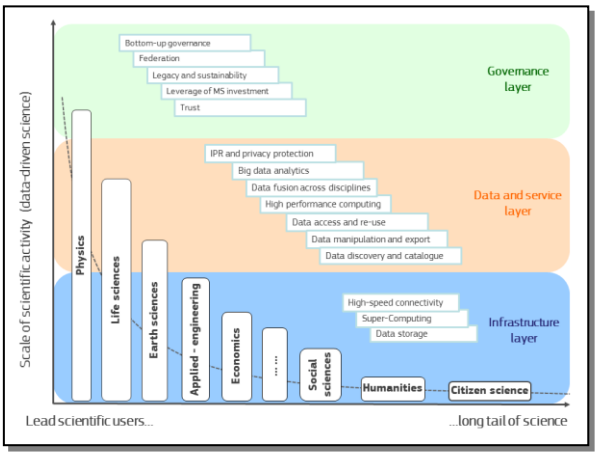
EOSC に焦点を当てると、委員会の前議長はかつて、これは欧州のものではなく国際的なものであると言いました。一部のデータはクローズドなので完全なオー

プンではありません。また、科学のみならず、イノベーション、公共的なものでもあります。これはクラウドでありながら、現実のものであり、設備を提供する地上のデータサービスです。研究者がデータを保存、管理、再利用するための環境をつくり出しています。私たちは、大学や研究所に、そして国家的または欧州レベルでスーパーコンピュータとともに多くのインフラが既に存在していることを認識しています。課題は、既存および新しいインフラといかに連携するかということです。ガバナンスの問題とこれらの施設との連携については、昨年 10 月に発表された EOSC レポートに記載されています。

項目の一つを見ると、図 15 はガバナンスの設定、データまたはコンテンツと新しいサービスの提供、またはインフラの稼働についてのものです。これはこのクラウドの挑戦です。オランダの DANS 研究所の試みは、データに対する一種のミシュランガイドを提供することです。私たちは、FAIR 原則の最初の三つの部分（探索可能・アクセス可能・相互運用可能）のそれぞれに対して五つの質問をし、成績をつけることができます。平均を取ることで、一般的な成績が分かります。これはデータを知らせる最初の方法です。私は、課題の一つはデータを発見可能にすることだと思います。たくさんのデータの中から、どうすれば見つけることができるのでしょうか。

私はクラウドは既に存在していると強調したいと思います。例えば、アメリカ国立衛生研究所（NIH）の NIH Commons においてです。アメリカ国立科学財団にもクラウドがあり、マイクロソフトとアマゾンには商業クラウドがあります。私が印象的に思うのは、ほとんどがアメリカの例であるということです。彼らはまず、NIH が運営している 600 万ドルのクラウドのようなパイロットから始めています。欧州ではもっとトップダウンで行っています。私たちはガバナンスについて考え、始める前に全てを整えたいと思っています。しかし、このアプローチを取れば、日本には既にバックボーンと設備があるため、有利なスタートをすることができます。アメリカがクラウドに投資し、アメリカの研究がクラウドに投資すれば、国家レベルの投資になるでしょう。欧州がクラウドに投資したいと思ってプロバイダに進めば、国際的な投資になるでしょう。また、これらのサービスを提供するために国内または欧州のプロバイダを調整し準備する必要があります。このために、このパイロットが必要なのです。

先ほど述べたように、データは見つけることができなければなりませんが、再利用するためにはデータは信頼できるものでなければなりません。DANS 研究所と日本の World Data System（WDS）による「Data Seal of Approval」は、データが良好なフォーマットで見つけ出され、信頼できるものであるという保証を与えるツールを提供します（図 16）。また、データセットの識別子も提供します。データを作成した研究者にクレジットを与えたいならデータを参照できるようになっていることが大切です。



（図 15）

（図 16）

しかし、データを共有し最適に再利用するには、共有を推進するための信頼を勝ち取る文化的な変化が必要です。このためにはデータを参照できる必要があるため、識別子を利用すべきです。しかし、共有するデータを生産する者に対して報酬やインセンティブも与えなければなりません。また、データ生産者とユーザーの認証も必要です。理想としては、データを提供する、使用する際に自分が何者かを知らせるために、データにアクセスするためのある種のシングルサインオンがあったらいいでしょう。特に生命科学や社会科学では、機密データをどう扱うかに注意を払わなければなりません。一部のデータはインターネット上で公開することはできませんが、それでも研究には関係してきます。われわれはそれにどのように取り組んでいけばいいのでしょうか。データインフラについて結論付けると、その方向性はあまり明確ではありません。私たちは、共有のためのインセンティブ、報酬の仕組み、識別子が必要であることは分かっていますが、これらをどう実現するかはまだ明確になっていません。

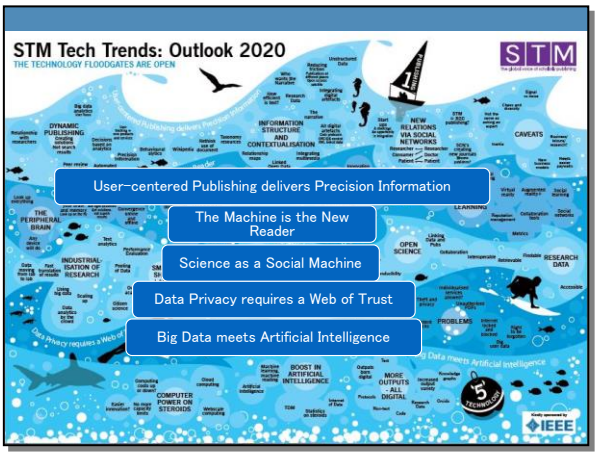
オープンサイエンスの展望

オープンサイエンスに関するビジョンについては、まず出版物から話を始めて、次にデータに移りたいと思います。国際 STM 出版社協会が出した「STM Tech Trends 2015」では、「The Article in a Hub and Spoke Model（ハブ＆スポークモデルにおける論文）」と「Data as first class Research Object（第一級研究対象としてのデータ）」とあります。データは研究成果であって、共有する必要があるということになったのです。

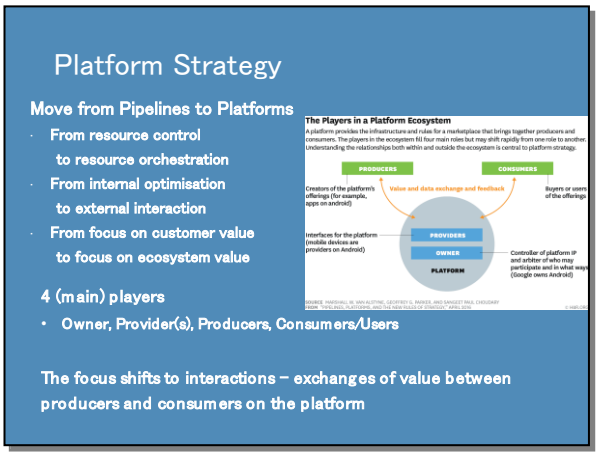
「STM Tech Trends: Outlook2020」は、ビッグデータと、データと人工知能の組み合わせに関するものです（図 17）。機械は、論文やデータを読む新しい読者になるかもしれないし、研究者を助けることになるかもしれません。そのためには、正確な情報を提供することが非常に重要です。私は 1 年間に発行される 300 万件の論文の中から 300～400 件の論文しか読むことができないので、選択の際に何らかの助けが必要です。以前は 1～2 誌だけしか読めなかったものが、たくさんの雑誌が出てきました。関連論文が他誌に出たりすることもあり、フィルタリングが必要です。

出版物に追加された価値はフィルタリング機能です。例えば、私の 15 歳の息子はコンテンツに対してお金を支払おうとはしませんが、インターネットで音楽や映画を見つけるときには助けが欲しいかもしれません。そのために彼はソーシャルネットワークを利用します。一方、研究では、関連情報を見つけるのに助けが必要です。出版物において、読者は関連する論文を探したいと思っていますが、著者はできるだけ多くの読者にアプローチしたいと考えるのは明らかです。出版物の供給と需要という市場があるのです。

新しい方向は、プラットフォーム戦略を持つことです。図 18 は、戦略を持ちたければ異なった考えをしなければならないと言う「ハーバード・ビジネスレビュー」の論文に基づいています。私たちは焦点をシフ



（図 17）



（図 18）

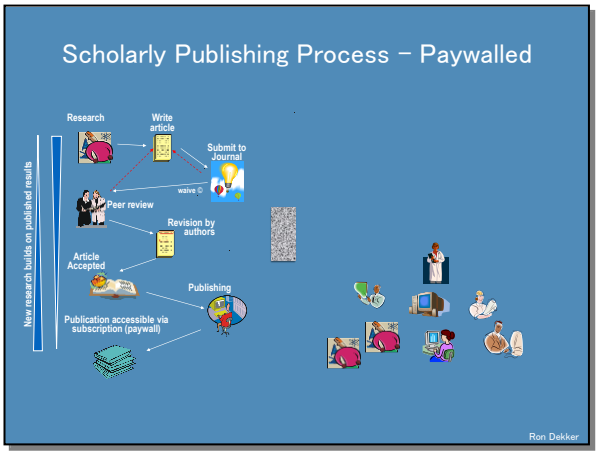
トしています。それは限界コストに関するものではなく、プラットフォーム上の生産者とユーザー間の相互作用と価値交換に関するものです。これらのプラットフォームには、所有者、プロバイダ、生産者、ユーザーがいます。一つの大きな例は Apple です。彼らは iPhone と iPad というプラットフォームを所有していて、これは電話のためでなく、そこでしか使えないアプリケーションを用意しています。ユーザーを見つける App Store と iTunes のことです。Apple の成功は、このつながりをつくったことにあります。

もう一つの例はソニーの PlayStation です。PlayStation 4 では、月会費を払えば、オンラインで世界中の他のプレイヤーとプレイできます。このように、プラットフォーム上の付加価値は重要です。私は、ユーザーと生産者が良好なつながりを持つことが、出版物の方向であるべきだと思います。

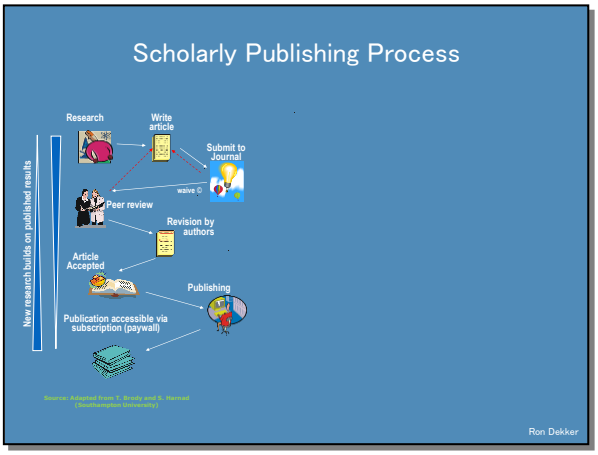
第 3 回 SPARC Japan セミナー2015 で、図 19 を見たことがあると思います。「101 Innovations」は、利用可能な全てのツールを記録したものですが、現在はオープンサイエンスを行うときに利用可能なツールの数であるため、「400+ Innovations」と呼ばれるべきです。

図 20 は伝統的な出版プロセスを示しています。論文を書く、提出する、却下される、別のジャーナルに再度提出する、査読される、拒否されるか改善する、論文が受け入れられる、発表する、これがジャーナルに掲載されるということです。ペイウォールモデルでは、科学界の外の人々は見ることができず、購読していなければ論文にアクセスするために 30 ドル支払わなければなりません（図 21）。グリーンオープンアクセスでは、エンバーゴ方式とプレプリント方式によって利用可能です（図 22）。従って、私たちはプロセスの最後を変えるゴールドオープンアクセスを考え出し

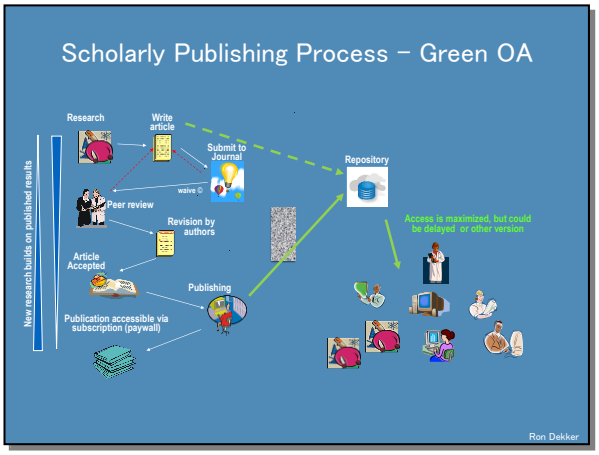
（図 19）



（図 21）



（図 20）



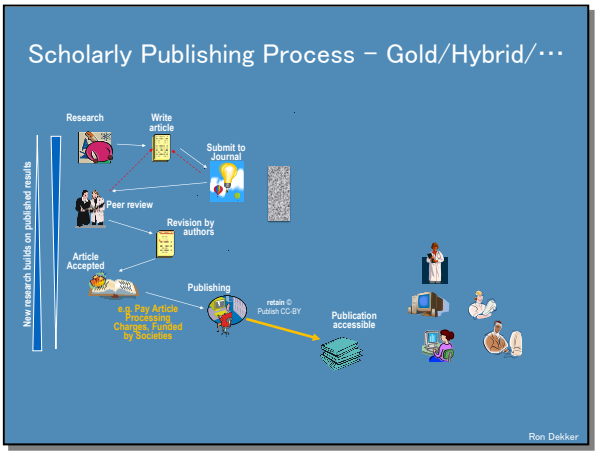
（図 22）

ました（図 23)。私たちが支払いをし、誰でも利用できるようにします。このゴールドモデルは標準になるでしょうか、それとも他のモデルが出るでしょうか。

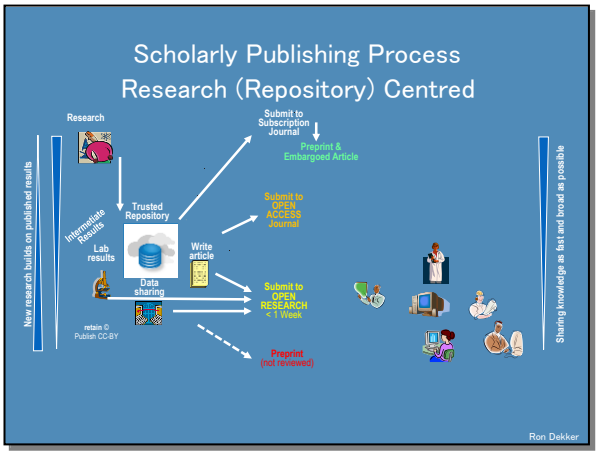
出版ではウェルカムトラストによる新しいアイデアがあります（図 24)。これは「Open Research」と呼ばれ、論文執筆後 1 週間以内に論文が発表されるものです。その後、論文が公開の場で査読され、結果を利用できるようになります。また、これは出版物に加えて検査結果やデータにも適用されます。どのようにしてデータが調査されるかは分かりませんが、これによって、データセットの品質を保証する新しい方法が見つけられると思います。最初に出版し、次に品質評価を行うというように、物事の順番が変わってきています。さらに進むと、研究を再び途中の段階に戻すことができます（図 25)。研究して、それを公開するかどうかを決定します。公開する場合、伝統的なジャーナル、ゴールドオープンアクセス、またはオープンリサーチに送ることができます。私の場合は査読を選ばないで、プレプリントで投稿します。そうして、アクセスを得ることができます。これは研究者にとって非常に簡単な方法です。

この方法は、オープンアクセスをさらに容易にするため、資金提供者にとってもとても簡単です（図 26)。このモデルは、約 750 ドルという FAIR な価格で査読を受けられます。これは新しい方法かもしれませんが、いまだ既存のジャーナルシステムの中にあります。

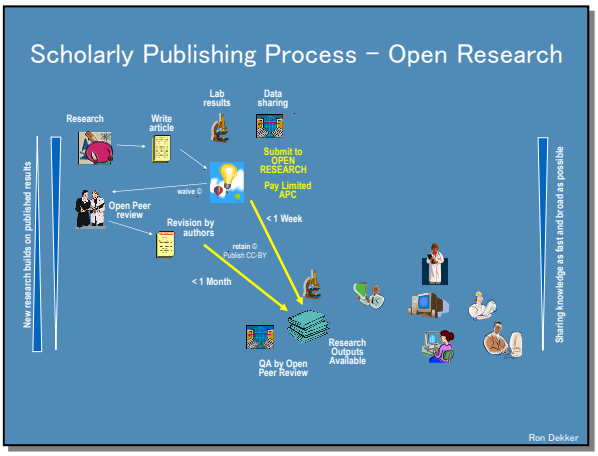
別の動きはプレプリントの中に見いだすことができます。資金提供者が識別子を取得するプレプリントを喜んで認め、プレプリントに関する助成金を容認すると言えば、若手研究者にとって助けとなります。2 年間の博士研究員の助成金を受けている場合、発表されるまでに 1 年かかる論文を提出する時間がありません。



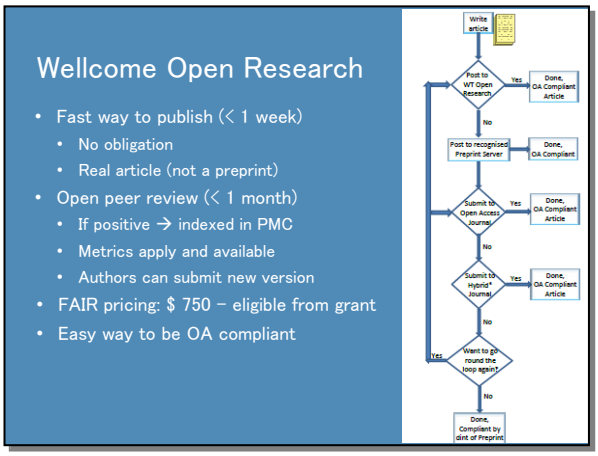
（図 23）



（図 25）



（図 24）



（図 26）

このプレプリントでタイムスタンプを取得した場合、助成金に対する新しい申請でそれを使用することができます。これによって出版や知識の共有もスピードアップします。私たちはプレプリント用の信頼できるリポジトリのシステムを持つことになるでしょう。このように、出版における知識共有の動きを追うことで、将来の方向性を見いだすことができます。

データに関しては、皆さんの力が必要です（図 27)。データを再構成するために必要な労力を考慮しても、データには莫大な価値があります。研究者がデータを準備し共有することに労力を割けば、他の誰かが、論文や新しい知識を生産するのに多くの時間を節約して収入を得ることになります。従って、データ共有に関する市場をどのように設定したらいいかというのが大きな疑問の一つになります。私たちにはインフラがあるので、パイロットプログラムも行うことができます。私たちは、全ての学問分野の国家政策を立てるべきか、国際的にそれぞれの学問分野ごとに政策を立てるべきか、あるいはその両方を行うべきかということについて議論することもできます。

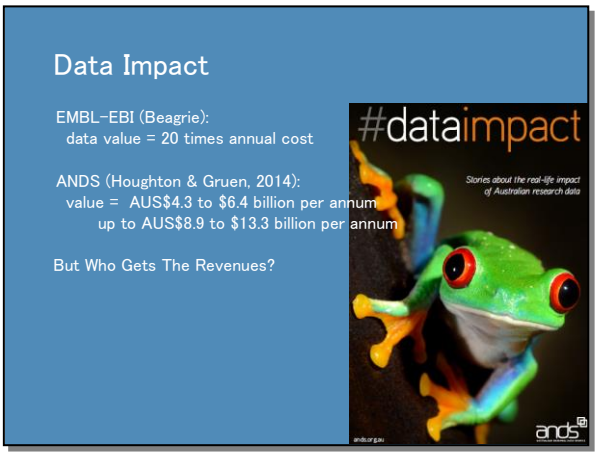
市場を創出するに当たって重要な win-win の一つの例は、Structural Genomics Consortium（SGC）です（図 28)。SGC は、創薬研究に関する大規模な官民連携による研究団体であり、創薬ターゲットとなるヒトのタンパク質の 3 次元構造を同定・解析する研究を行っています。製薬企業は、競合同士であっても、対立するそのときまではこのコンソーシアムで協働しています。

このコンソーシアムでは、信頼できる組織を設立し、10 の研究グループを持っています。組織はそこに調査したいと考えている上位 3 種のタンパク質を伝えます。その中から、興味を持っている参加団体が多かったトップ 30 のタンパク質について、どの団体がそれに興味を持っているか、それについてどのような取り組みをしているかは明らかにせずに研究しています。参加団体はこれに同意しています。これは win-win です。多くのデータと研究能力を持っているのですから、需要と供給、生産者とユーザーをマッチさせることが、データに対する課題です。

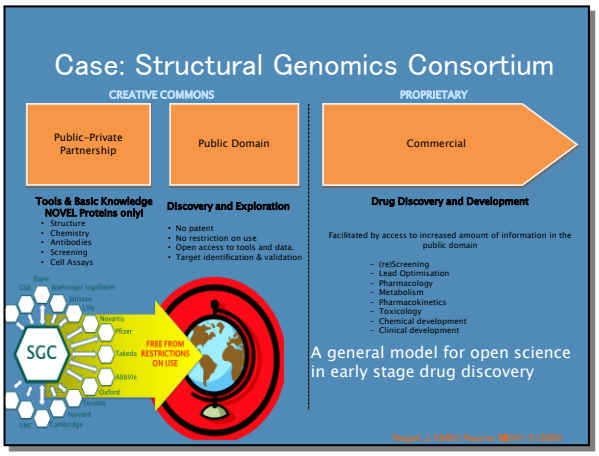
膨大なデータの中から、価値あるデータを探すことができるようにしなければなりません。これが基本です。図書館は、（データと研究者の）仲介役を果たすことができるかもしれません。図書館同士はつながっているため、他の図書館で何が起こっているかを知っています。データに関する質問がある場合、図書館員は手助けすることができます。

結論として、私たちはオープンサイエンスの新しい時代の一部とならなければなりません。私たちはより多くの人々にアプローチし、科学だけでなく社会にも大きな影響を与えたいと考えています。それは二度手間や出版バイアスを防ぐことができます。科学、社会、イノベーションに影響を与えるためには、将来使うためにデータを保存する必要があります。

課題は、いかに協力し、イノベーションを生み出し、共有するかということです。私は日本に来られてとて



（図 27）



（図 28）

もうれしいです。私が日本を知るようになったきっかけは、伝統的な映画「楢山節考」、伝統的な企業でどのようにイノベーションを起こすかについて説明した『知識創造企業』という本です。私たちをオープンサイエンス、データの再利用、出版物のオープンアクセスに導いてくれるのは、伝統の融合、過去の理解、イノベーションと変革への勇気だと考えています。最後に、南アフリカ共和国のことわざをご紹介して終わります。「早く行きたいなら、一人で行きなさい。遠くに行きたいなら、一緒に行きなさい」。

●**深貝**　横浜国立大学の深貝と申します。SPARC Japan の運営委員をしています。オープンサイエンスについて、このところ日本でも非常に活発に議論しています。歴史を振り返れば、近代の入り口の科学革命のころ、グーテンベルクの活版印刷を通じて、こもっていた局所的な知識がオープンになって、そして科学者が振る舞い方を変えていきました。それと似たようなことが今起きはじめています。ネットワークを使って、今までの印刷物ではない情報が飛び交うようになった中で、人々の振る舞い方が変わってきているのです。

電子ジャーナルが出たというのは、単に印刷物が電子版で出ただけですから、ある意味で伝統的です。しかし、最近は、人々がネットワークを使って、知識を交換し合っています。知識を交換することによって、多くの人が知識の生産の場に関わり、アイデアがひらめくということになりつつあるのです。それを一層加速しようというのがオープンサイエンスだと思います。

ただ、その場合のバリアは、学術の成果が誰かに帰属し、その人がインパクトファクターによって評価を受けるという方向に加速していることです。そうすると、科学者・研究者は、高い評価を受けるために、知識を狭めた上で成果を挙げることに走りがちです。しかし、広い目で見て人間の知識を多く豊かなものにしていくためには、オープンにしていった方がいい。そのような揺らぎの中にあるわけです。

その中で、公的資金を使って得られた成果をオープンにするということを今やっていますが、それだけで十分なのかどうか、お考えをお示しいただけるとありがたいです。

●**Dekker**　現在は、印刷物をインターネットで利用できるようにするのではなく、そもそも成果を最初からデジタルで出すようにしなければならない、そういうパラダイム転換の最中にあると言えます。私はウェルカムトラストの「Wellcome Open Research」が好きなのですが、その理由は伝統的な論文と中間結果の両方に焦点を当てているからです。そこでは重要なデータを公開し共有したいと思えば、することができます。ジャーナル「Science Matters」では、研究コミュニティにとって重要であると思われる場合は、中間結果を公表することができます。私たちは、伝統的な出版を越えて、成果の新たな発信方法を見つける必要があるのです。

これらの成果を測定する際は、ジャーナルのインパクトファクターを超えた新たな指標が必要になります。資金提供者は、代替物や新しい成果の発信方法に報酬を与えるべきです。イノベーションまたは社会的議論に重要な貢献をした場合も、資金提供者や大学によって報酬が支払われるべきです。研究者に対する独自の報酬システムを持つ大学や医療機関もあります。資金提供者は、出版物の数で評価することをやめるよう、所属する委員会のメンバーに求めるべきです。代替案としては、意味のないインパクトファクターを積んでいくのではなく、査読者がトップ 3 またはトップ 5 の出版物を精査することが挙げられます。異なるものを比較しているのですから、分野を超えてインパクトファクターを比較することはできません。DORA イニシアチブは、研究を評価するために定性的な情報を使用すると述べています。それが進むべき道であり、また、私たちは、プレプリントや他の公開方法の価値を認めるべきです。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

ディープラーニングとオープンサイエンス

～研究の爆速化が引き起こす摩擦なき情報流通へのシフト～

北本　朝展

（国立情報学研究所）

CC BY

講演要旨

講演者はこれまでオープンサイエンスの意義を利便性、透明性、参加という 3 軸から考えてきたが、これだけでは科学研究にオープン化がどう貢献するかを理解しづらいという問題があった。そこで「スピード」という新たな軸を導入することで、オープンサイエンスの意義を再考してみたい。その背景にある仮説は、研究のスピードが極限まで高速化すると、情報流通もそれに追従して高速化せねばならないため、情報流通の妨げとなる「摩擦」を取り除く方向に進化して、結果的に科学研究がオープン化するというものである。こうした仮説を検証するのに最適なフィールドが、人工知能および機械学習の一分野である「ディープラーニング（深層学習）」である。この分野ではいったい何が起こっているのか？この例外的な研究分野から得られる教訓がどれほど一般化できるかは未知数であるが、これまでの動向を分析することで、オープンサイエンスの一つの可能性を先取りしてみたい。

北本　朝展

1997年東京大学工学系研究科電子工学専攻修了。博士（工学）。現在、国立情報学研究所コンテンツ科学研究系准教授、総合研究大学院大学情報学専攻准教授、情報・システム研究機構人文学オープンデータ共同利用センター準備室長。画像データの分析を中心に、データ駆動型サイエンスを人文科学や地球科学、防災等の分野で幅広く展開する。文化庁メディア芸術祭アート部門審査委員会推薦作品、山下記念研究賞などを受賞。オープンサイエンスの展開に向けた、オープン化や超学際的研究コラボレーションにも興味を持つ。

私の研究分野は情報学で、画像処理や画像データベースの研究をしていたのですが、その後、データ駆動型サイエンスに発展し、気象情報、地球環境情報、人文科学情報などの分野で研究を行うようになりました。この流れで、最近はオープンサイエンスの問題にも関わっています。

1.オープンサイエンスの背景

1-1.オープンサイエンスへの収束

オープンサイエンスとは、「オープン」という言葉をてこにして、サイエンス（研究）の方向を変えるものだと考えています。共通する部分は、今よりもよりオープンにしようという方向性です。いろいろな方向に、いろいろな運動があるのですが、それを「よりオープンに」という一語で束ねるとどういう世界が見えるかという話になると思います。

ただ、個々の活動で「オープンサイエンス」の意味は異なるので、単一の定義は困難です。よく私が言っているのは、オープンサイエンスは同床異夢であるということです。オープンにしたいという思いは共通しているのだけれど、どんなオープンを夢見ているかという認識を共有しないと、実は違う夢を見ていたとい

うことがよくあります。私は、図 1 にある、オープンアクセス、オープンデータ、データマネジメント、シチズンサイエンス（市民科学）、研究の再現性など、透明性・参加・協働・共有の視点で動いている多様な活動は、全て「オープンサイエンス」というキーワードでまとめられると考えています。

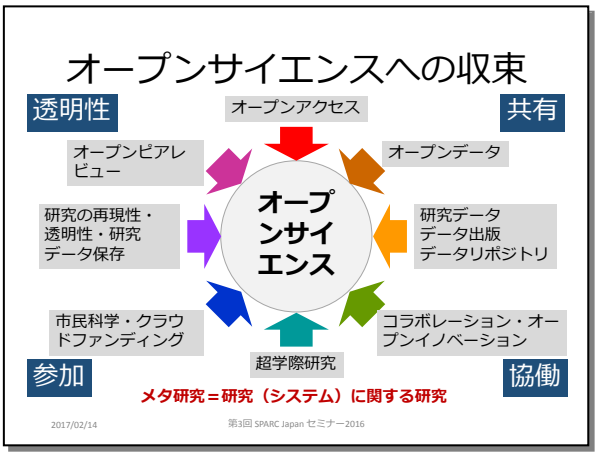
「オープン」には三つの側面があると考えています。一つ目は、他者が使えるということです（再利用）。外部の人が研究成果を自分の目的に再利用できます。これにはオープンデータやオープンアクセスなどが当てはまります。

二つ目は、他者が検証できるということです（透明性）。外部の人がエビデンスを検証し、正当性を判断できます。政府の政策を検証するオープンガバメントも同じような目的を持っていますが、他者の行った研究を再現・検証することも非常に重要なテーマになってきています。

三つ目は、他者を受け入れることです（参加）。研究のプロセスに他者を巻き込むことで、今まで得られなかったような成果を生みます。これにはオープンイノベーションや市民科学が当てはまります。

1-2.制度を分析する四つの視点

オープンサイエンスを分析するときには、法、規範、市場、アーキテクチャの四つの視点があると考えています。この四つの視点は、Lawrence Lessig が書いた“Code: And Other Laws of Cyberspace”（邦題『CODE ―インターネットの合法・違法・プライバシー』）という本で挙げられていて、Lessig は制度をどうやって変えていくかをこの四つの視点から考えることができると主張しています。



（図 1）

日本語で言うと、法は「しなければならない」、規範は「すべきである」、市場は「した方が利益がある」、アーキテクチャは「せざるを得ない」と言い換えられます。このような道具を使って制度を変えていく方法は、オープンサイエンスでも有効ではないかと考えます。

例えば、法によるオープン化は、資金提供機関が決めるルールを変えることによってオープン化を進めていくというものです。データ管理計画なども基本的にはそうで、ルール変更が重要な役割を果たします。例えば、資金提供機関が、どれだけオープン化したかで研究評価するというルールになれば、研究のやり方が変わっていきます。ただ、これは上からの押し付けで副作用も大きいので、本当にそれを適用していいのかという見極めは非常に重要です。ゆっくり慎重に進めていくことが求められます。

規範によるオープン化は、そもそもサイエンスはオープンであるべきであるという考え方です。これは分野によっては一般的な考え方です。例えば、地球科学では地球全体のデータがないと研究できないので、シェアしなければ研究が進まないという状況があります。こういう分野は、データ共有が不可欠なので、オープンな文化が発達したと言えます。

それだけではなく、世代の差もあります。上の世代より若い世代の方がもともとオープンな文化に生まれているので、共有に抵抗がありません。

ただ、こういう規範によるオープン化は、文化圏が違うと同じ文化を共有できないので、全分野で共有することはなかなか難しいです。

市場によるオープン化は、「した方が得だ」というものです。オープン化によって報酬が得られるのであれば、みんなオープン化するだろうということです。研究成果をオープン化すると例えば引用が増加するな

らば、オープン化した方が得だからオープン化しようという考えを持つ人も出てくるでしょう。ただ、これは報酬と裏腹に損失への不安があります。他者に成果を横取りされるのではないか、報酬は労力に見合うのか、こういう損失への不安があると、よほど報酬が確実でないとなかなかオープン化できなくなります。

アーキテクチャによるオープン化は、Lessig の本でも情報社会特有の方法であると述べられているのですが、何らかのプラットフォームの上で作業すると、必然的にオープンに誘導されてしまうというものです。例えば、何かツールを使っていて、ワンクリックでオープン化できる機能があれば、オープン化へと誘導されてしまうかもしれません。プラットフォーム上の見えないルールで、いつの間にか誘導されるというのがこの方法の特徴です。

また、苦痛の軽減という視点も重要です。オープン化は大変だけれど、有償のサービスを使えば簡単にできるとなると、そういうプラットフォームの上に乗ってオープン化を進める状況も当然起こり得ます。こういうものは良くも悪くも企業にとってはベンダーロックインのチャンスであり、そうした利益がオープン化を進めていくという未来もありえます。

このように四つのオープン化は方向が全く異なりますので、オープン化の話題ではどのオープン化に焦点を合わせているのかを明確にすることが有益ではないかと思っています。

1-3.研究データのオープン化

研究データの問題についても触れたいと思います。私は研究データを三つの種類に分けています。研究資源データ、論文付属データ、研究過程データです。

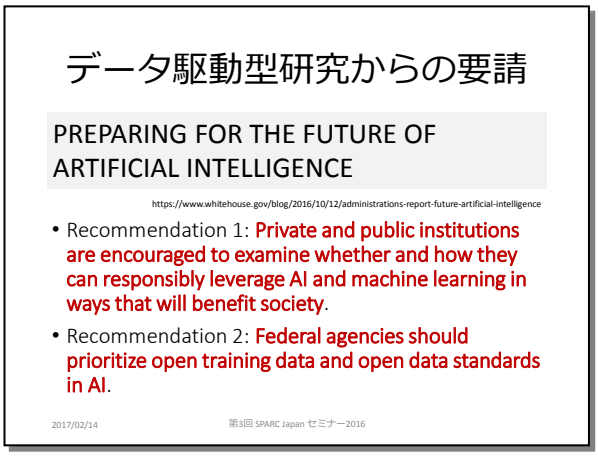
研究資源データは、研究の入力となるデータです。観測データや、評価用データセットなど、これを使って研究するというものです。これは再利用が目的ですから、再利用しやすいオープンデータになっているかという点が評価基準になります。

論文付属データは、研究の出力となるデータです。典型的なのは論文のエビデンスとなるデータです。これは、再利用も可能ではありますが、それよりも透明性、すなわち研究が再現できるかの方が重要ですので、再利用とは違った視点でのオープン化が求められます。

研究過程データは、研究の入力と出力の間で生み出されるデータです。例えば、日々の研究活動のエビデンスとなるようなデータです。このデータは透明性のために重要で、研究不正防止のためのデータ保存の義務化などもこれに関連しますが、基本的にオープン化が目的ではなく、クローズドに長期保存し、本当に必要があれば限られた人に対してオープンにするものです。

私が関心を抱いているのは、研究資源データをどうやって再利用のためにオープン化するかということです。このニーズは、データ駆動型研究が広まるにつれてますます高まっています。図 2 は、「Preparing for the Future of Artificial Intelligence」という、2016 年 10 月にホワイトハウスから出たものです。ホワイトハウスのウェブサイトは最近大幅に変わりましたので、今アクセスできるかは分かりません。ここでは、マシンラーニングや AI を発達させるには、データをオープンにすることで研究を活発化させることが重要であると強調されています。この AI とマシンラーニングが今後のオープンデータの一つの焦点になります。

データについて、考えるべきはライフサイクルです。ライフサイクルとは、データが誕生してからたどるプロセスを段階の遷移として表すモデルです。研究者は



(図 2)

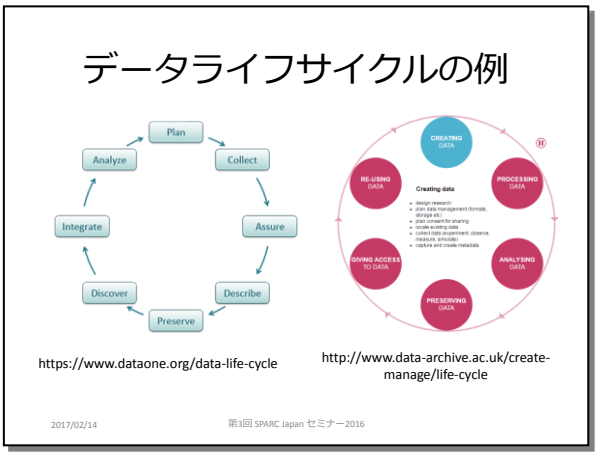
全体に関わりますが、そのうちの一部にはライブラリアンも関わります。例えば、データ管理計画などはライブラリアンが関わることが期待される段階です。

例として、Plan、Collect、Assure、Describe、Preserve、Discover、Integrate、Analyze という段階があって、これがぐるぐる回るのがデータライフサイクルです（図3)。個々の段階で何をするかということが重要になるのですが、左のサイクルと右のサイクルをご覧になって分かるように、データライフサイクルはそれぞれ違います。ここには示していませんがもっとたくさんの図がありますので、唯一の正解があるというよりは、プロジェクトに適したライフサイクルは何かを考え、その中で誰がどこの部分を担当するかを考えることが重要だと思います。

1-4.ライブラリアンの役割変化

こういう状況の中で、ライブラリアンも役割が変化していくことが考えられます。従来のライブラリアンは本を扱っていました。本はこれ以上加工されない最終の生成物であり、これを利用者（読者）のために整理していればよかったのです。しかし、データ管理計画など、データを扱わなければいけないとなると、データライブラリアンが必要だという話になります。

データは、本のような最終生成物ではなく、変わりうるものです。また、データにどんなメタデータを付けるのかという話になったとき、メタデータに決まったフォーマットがあればいいですが、ない場合は作者のところに行って、このデータをどのような目的でつくったのかインタビューしながらフォーマットを決めるプロセスが入ります。本で言えば、作者に作品の意図を聞いて、それを記述するというような、編集者や出版者に近い役割を果たさなければなりません。ですから、従来の最終生成物だけ扱っていればよかった時代と、役割もかなり変わってくるのです。



（図 3）

ライブラリアンは、データに標準的な価値を与えて使えるようにすることが目的ですが、一方でキュレーターは、データに付加価値を付けることが仕事です。データライブラリアンが整理した段階で、標準的なメタデータが付いて誰でも使えるようにはなりますが、そこにさらにデータキュレーターが関わり、このデータにはどんな価値があって、どういう目的に使えるかを売り込む。そんな人がいると、データの利活用も進むでしょう。このようにライブラリアンの役割変化が今後進んでいくと考えています。

現段階でデータライブラリアンに求められるスキルの一つ目は、データライフサイクルに関する知識と、DOI などの識別子を与えてデータをアクセス可能にすることです。

二つ目は、メタデータを付けないとデータにアクセスできないので、メタデータ規格や語彙に関する国際標準との相互運用性などに関する知識です。

三つ目は、データは置いておくだけでは使われないので、リポジトリに入れるなど、データの共有や利活用を促進する方法に関する基本的な動向の把握です。

四つ目に、上級編として、研究者コミュニティとの議論やエスノグラフィ的観察に基づき、データ管理システムを設計し構築する能力です。これができるデータライブラリアンは大変ありがたいと思います。

ただ、10 年後を考えると、こういう仕事のかなりの部分は AI に代替される可能性があります。今からデータライブラリアンになろうという方は、10 年後のキャリアを目指しますから、そのときまでに代替される知識を学んでしまったら、人生戦略上、非常にまずいです。ですから、10 年後どうなるかを予想しな

がらデータライブラリアンに必要なスキルを考えなければいけません。

メタデータの付与についても、例えば、アクセスが目的であれば、項目や品質の考え方も変わりますし、AI 向けメタデータであれば、人間が付与するかどうかも分かりません。ですから、AI との関わり方は必ず考慮すべきポイントになります。この場合、データライブラリアンには二つの道があると思います。一つは、AI を使いこなす側に立って、データライブラリアン兼データサイエンティストというスーパーマンになる道です。もう一つは、AI の下で働く道です。現実世界のダーティデータを AI が使いやすいクリーンデータに整理して提供する。スーパーマンになるのもいいですが、下で働くというのも必ずしも悪いことではありません。今の世の中も実際にはそうなっていて、Google のためにデータをきれいにしてあげるというのは、みんながやっていることです。ですから、データライブラリアンになるのであれば、10 年後も維持できる価値を考えなければいけないというのが前半部分のお話です。

2.オープンサイエンス再考

AI が出てきましたが、後半は AI の研究の世界で起きていることの話です。

2-1.オープン化の第四の軸「スピード」

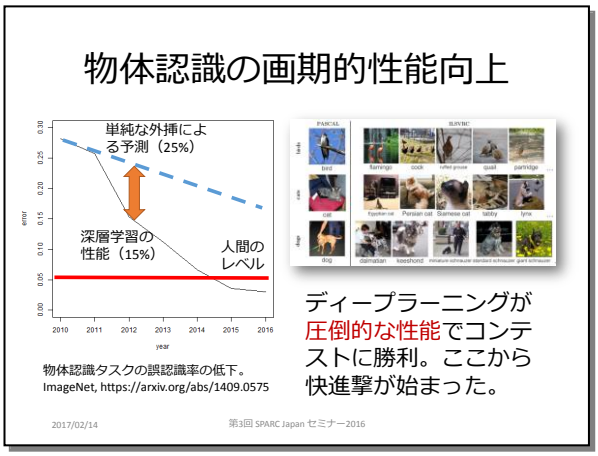
これまでオープンサイエンスの意義は、「再利用できる」「透明性がある」「参加できる」といった三つを考えてきました。ただ、これだけだと研究者にはあまり実感がなく、これだけではオープン化が進まないという感じがありました。そこで、オープン化の意義として、「スピード」があるのではないかという点を今日はお話したいと思っています。研究のスピードが極限まで高速化すると、情報流通もそれに追従して高速化しなければいけないので、情報流通の妨げとなるような摩擦を取り除く方向に情報流通プラットフォームが進化し、結果的に科学研究がオープン化するというような仮説です。その例をディープラーニング（深層学習）という分野の現状から検証したいと思います。

2-2.ディープラーニング登場

皆さん、ディープラーニングはお聞きになったことありますか。AI という言葉はニュースで日々お聞きになっていると思います。今、第 3 次人工知能ブームと人工知能学会会長も言っていますが、第 3 次人工知能ブームの中心的存在がこのディープラーニングで、技術的には「ニューラルネットワーク」と呼ばれるものです。ニューラルネットワークの中で特に層が多いものが「ディープ」と呼ばれます。原理は昔から知られていたのですが、ビッグデータとアルゴリズムの改良で画期的な性能向上があり、それが衝撃を与えて人工知能ブームにつながっています。

画期的な性能向上は、まず画像認識の分野から始まりました（図 4）。画像だけを与えて、それが何かを認識する物体認識タスクのコンペティション「ImageNet Large Scale Visual Recognition Challenge（ILSVRC）」が 2010 年から始まりました。緩やかなスピードで誤認識率が改善していくと多くの人が思っていたところ、2012 年には誤認識率がいきなり 25%から 15%まで下がり、ディープラーニングが圧倒的な性能で優勝しました。みんながその結果に驚き、その後は一斉にディープラーニングに走り、今は人間よりも性能が高くなっています。

ここからいろいろな分野にディープラーニングの適



（図 4）

用が始まったのですが、私が一番好きなのは囲碁のソフトウエアの AlphaGo です（図 5)。人間のチャンピオンも AlphaGo に負けました。過去のデータはもちろん学んでいますが、自己対戦によって戦略を深化させ、人間とは異なる戦略を用いて勝利したことが非常に衝撃的でした。

私も感動して、AlphaGo 観戦ツイートをまとめて Togetter ページをつくりました。私は囲碁というゲームがプレイできますので、この勝利がどう衝撃的かということも具体的に分かりましたし、この勝利は今までとはちょっと違うと感じました。この AlphaGo をつくっている会社が DeepMind です。この会社、今は Google の傘下にありますが、何のプロダクトも公表されていない時点で数百億円相当の金額で買収されたとして話題になりました。この会社が今、ディープラーニングでは重要なキープレイヤーになっています。

まずディープラーニングで面白いのは、オープンソース競争が始まっているという点です。ディープラーニングの最先端のライブラリを各社が競ってオープンソース化しています。Google の TensorFlow、Microsoft の CNTK、Facebook の Torch、他には日本でも Preferred Networks という会社が Chainer というライブラリを公開して頑張っています。

本当に最先端のライブラリは知的財産の塊なのに、それがどんどんオープンソース化されているのはなぜでしょうか。やはりオープンソースに人がおびき寄せられることを狙っているのでしょう。オープンソースのコミュニティができ、協力者が増えれば、そこから創出される価値も増えます。例えば、それを使ったライブラリや可視化ソフトなどいろいろなものが出てくれば、全部自社で開発しなくてもよくなります。まさにオープンイノベーションですが、そこでは競争すべき領域と協調すべき領域があります。差別化できる部分は守りつつ、外部の力を使えるところは使うというのが今の状況です。

（図 5）

では、Google がオープンソース化したからといって、みんなが Google と同じことができるかというと、それはありません。ソフトウエアは確かにオープンですが、例えば AlphaGo を実行するのに、数十億円のコンピューター資源が必要になるなら、誰でもできることではありません。また画像認識でも、数億枚の画像データを使わなければいけないとなると、そんなデータを持っているところは限られます。だから実際は同じことはできないのですが、少なくともソフトウエアについてはオープン化がどんどん進んでいます。

このようにオープンソース化が急速に進み、論文の実験コードなども再利用や再現性の観点からオープンソース化され、誰でも試せるようになりました。共通基盤データも、例えば先ほどの物体認識タスクの実験データがオープン化されているので、誰でもソフトウエアをつくれば性能が比較できます。さらに応用分野がどんどん広がっているので、他の分野の研究者や一般の人々も大挙参入しています。ですから、一刻も早く成果を公表しないと、自分の成果が古くなってしまうのではないか、という恐怖を皆さん持っています。スピードへの強烈な圧力がかかるという、他の分野にはない独自の状況が生まれているのです。

2-3.arXiv の利用

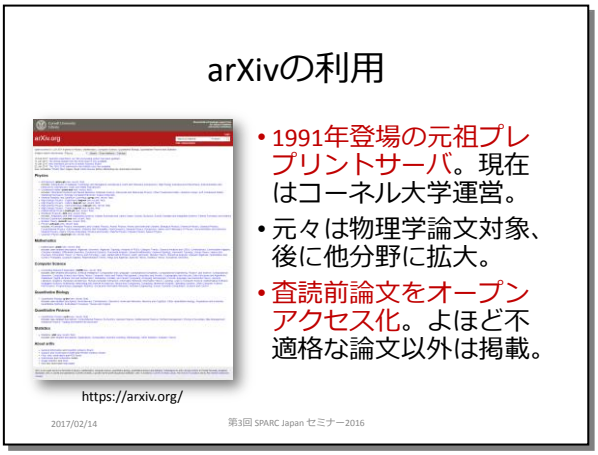
そこで登場するのが arXiv です（図 6)。なぜ今さら arXiv なのだと疑問を持たれる方もいらっしゃると思います。これは 1991 年登場の元祖プレプリントサーバーで、登場してからもう 25 年以上たっているからです。現在はコーネル大学が運営しています。実は今、この arXiv が大きく変わっているのです。

arXiv への投稿数はずっと上昇しています（図 7)。2010 年ぐらいから加速しているように見えます。分野別の投稿状況を見ると、もともと high energy physics が多かったのですが、割合としては減ってきました。math も多いですが、cs（コンピューターサイエンス）が近年になって割合を増やしてきています。これはディープラーニングの盛り上がりと同期した変化です。

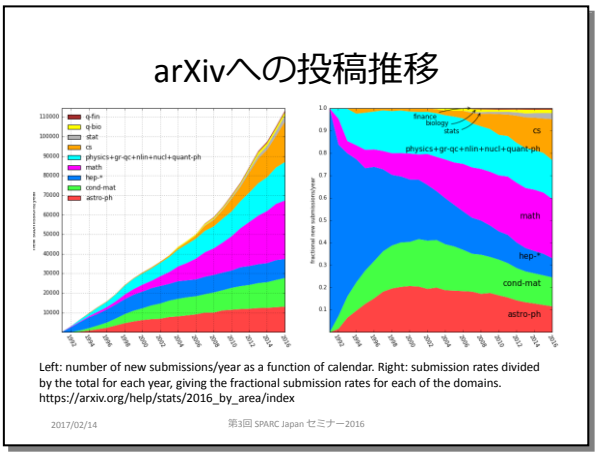
ディープラーニングの分野では、この arXiv が主戦場になってきています。AlphaGo をつくった DeepMind のサイトには Publications というページがあり、134 件（雑誌 10 件、arXiv 系 34 件、著名国際会議 90 件）の出版がリストになっています。雑誌は「Nature」などです。著名国際会議のリストもすごいのですが、そのような国際会議や「Nature」に混ざって、arXiv が同格でリストに並んでいます。別に彼らがそれらを区別している感じはありません。実際、DeepMind が出している論文は、arXiv で終わってしまうものもあります。つまりレビューは受けていない。arXiv にまず最新の成果をどんどん流す文化があり、その中で必要なものだけ査読を受けているような印象を受けます。

その arXiv がどう主戦場かを見てみましょう。例えば図 8 は 2016 年 10 月 10 日に arXiv に投稿された、東京大学の松尾豊先生らの論文です。この論文を、10 月 11 日に arXiv に投稿された論文が引用しています。この論文を書いたのが DeepMind の人なのです。DeepMind の人は arXiv を毎日毎日チェックし、1 日前の論文でも引用して投稿しています。まさに引用の爆速化が起こっています。ネットの世界から見れば、Twitter なら 1 分前でも引用できますので、1 日前の情報を引用すること自体は別に珍しくないと思います。ただ、従来の学術情報出版の基準から見ると、考えられないようなスピードで引用が進んでいると言えます。

さらに、最近進展している分野なので、2016 年や 2015 年の論文が必然的に多くなります。そうなると、論文のほとんどが arXiv にあるので、arXiv を読んでい



（図 6）



（図 7）

（図 8）

れば基本的な動向は追えることになります。よほど古いものでない限り arXiv を見ていればいいということになれば、過去論文データベースの重要性が低下してきます。

皆さんにとって心配なのは、arXiv に投稿することが二重投稿にならないかという点でしょう（図 9）。二重投稿と arXiv がどう整理されているか。ディープラーニングの中心的な国際会議である Neural Information Processing Systems（NIPS）は、“Prior submissions on arXiv.org are permitted.”と明言しています。他の会議では、International Conference on Machine Learning（ICML）も“e.g., in arXiv”と書いています。一方、Computer Vision and Pattern Recognition（CVPR）という会議は、arXiv.org への投稿は出版ではないからオーケーと言っています。この「出版ではない」というのは微妙な線引きですが、レビューされてないから出版ではないという解釈を取っています。ただ NIPS では数語で済むところ、CVPR は数パラグラフを使って arXiv がなぜ出版でないかを説明しており、この長さの違いが文化の違いを反映していると考えられます。

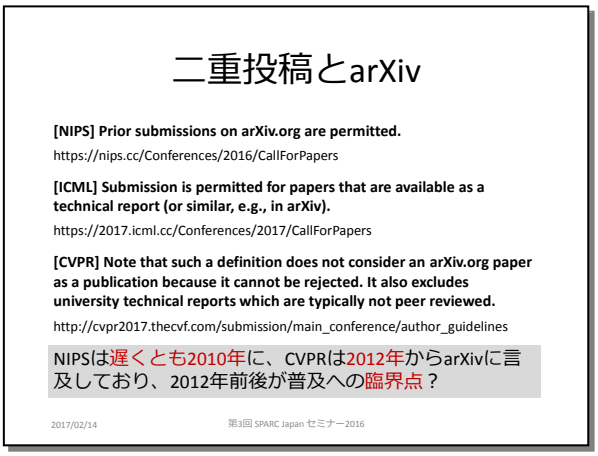
NIPS は遅くとも 2010 年にはこう書いているので、昔からそうであったと言えます。一方 CVPR は 2012 年からこのような注意書きが入りました。ですから、2012 年前後が普及への臨界点だったのではないかと思います。

さらに、図 10 は 2016 年 12 月 9 日に出たプレスリリースで、「なお、本研究成果は 2016 年 12 月 5 日に『Computing Research Repository』に公開されました」と書いてあります。この「Computing Research Repository」というのは arXiv です。つまり、arXiv に投稿したのでプレスリリースしますというものです。

このプレスリリースを行った大阪府立大学の黄瀬浩一教授および岩村雅一准教授は知り合いだったため、個人的にメールを出してプレスリリースの背景について幅広く意見交換させて頂きました。特に、通常のプレスリリースは査読済み論文が出版されるタイミングが多いのに対し、査読がない arXiv への投稿の段階でプレスリリースを行うことへの考え方をお聞きしました。すると、「慎重にすべきという意見もあったが、速報性を重視してこのタイミングでリリースすべきと個人的に判断した」とのことでした。

このような成果の迅速な公開は、実は Ingelfinger Rule と呼ばれるルールに反します。これまで査読論文に関連する内容の公表を遅らせてきたのは、プレプリントやプレスリリースを論文の公開前に行ってはいけないというルールのためでした。しかし速報性を重視する圧力が高まっており、そうも言っていられない状況になってきています。

ディープラーニングの会議で、International Conference on Learning Representations（ICLR）というものがあります。ここは面白くて、著者が arXiv に論文を投稿し、そのリンクをプログラム委員会に連絡することで、査読が始まるというプロセスを採用しています。つまり論文を最初に arXiv に投稿しなくてはいけない



（図 9）

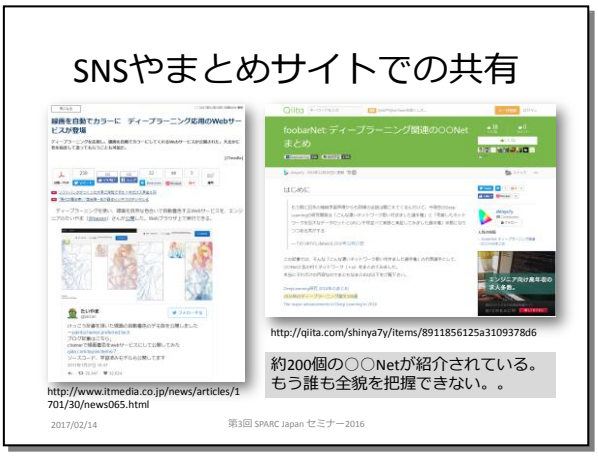
（図 10）

のです（図 11）。

最近はさらにオープンピアレビューに移行しています。OpenReview というシステムを使い、投稿した瞬間に論文が公開されるだけでなく、その後のレビューも公開されることになります。ICLR のサイトには、“The friction in our publication system is slowing the progress of our field.”、つまり friction（摩擦）があるようなシステムは問題であり、私たちは新しいやり方に変えるのだという宣言が書いてあります。私は本報告のテーマ「摩擦なき」に決めたとき、実はこの言葉を知らなかったのですが、やはり摩擦というアナロジーに行き着くのだなと感慨を覚えました。

このような急速な進歩は、SNS やまとめサイトでどんどん共有され、どんどん広まっていきます（図 12）。この「ディープラーニング関連の○○Net まとめ」には、○○Net というものが約 200 個紹介されていて、

（図 11）



（図 12）

もう誰も全貌を把握できないという状況です。このようなソーシャルな方法で、情報共有が加速度的に進んでいます。

これはある意味、市場によるオープン化であって、オープンソース、オープンデータにする方がお得だという動機からオープン化が進んでいきます。オープンアクセスリポジトリも、最初に「出版」して、その後に査読に行く世界になってきています。ソーシャル化で研究成果は SNS で迅速に共有され、公開サービスもいろいろな人が立ち上げ、日々どんどん進化している状況です。さらに、その公開サービスは、多くのユーザーがすぐに試すので、迅速性や頑健性も日々ネット上でチェックされます。そういう時代になってきました。

Michael Nielsen は、邦題『オープンサイエンス革命』という本を書いた人です（図 13）。実はこの英語のタイトルは“Reinventing Discovery: The New Era of Networked Science”です。「オープンサイエンス」とはタイトルでは言っていないのですが、オープンサイエンスに関することも論じています。

この本で描かれているオープンサイエンスの世界が、ディープラーニングでかなり実現してしまっているのではないかというのが私の感想です。いろいろな形でオープン化が進んで、知識が共有されています。面白いのは、Nielsen が 3 冊目に書いた本がディープラーニングの本なのです。両方ともネットワークという点で、2 冊の本には多くの共通する面があるようです。

（図 13）

法によるオープン化も着々と進んでいます（図 14）。ゲイツ財団は、財団の助成を受けた論文の、財団の OA 方針に対応していないトップジャーナルへの掲載は認めないこと、ウェルカム財団はプレプリントを助成金申請に記載してもよいことを表明しています。

3.摩擦なき学術情報流通へのシフト

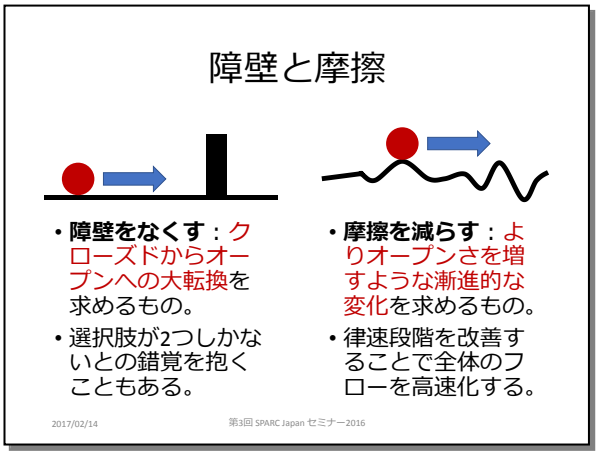
このような動きの中で起こっていることは、摩擦なき学術情報流通へのシフトだと思います。そこで考えなければいけないのは、我々が本当に欲しいものは何かということです。「人々が欲しいのは、ドリルではなく穴である」という有名な言葉があります。ドリルを買いに来た人がいて、ドリルがないときに、別のドリルを勧めるのではなく、そもそもなぜドリルが必要なのかを聞き、穴が必要なのだったら穴を開ける別の機械をお勧めすればよい。目的と手段は何かと考えることを忘れてはいけないという言葉です。

私は昨年の SPARC Japan セミナーで障壁について話しました。障壁をなくすのは壁を取るという考え方で、時としてクローズドからオープンへの大転換を求めることにもなりますが、摩擦を減らすというのは動きをスムーズにするということで、そうした漸進的な変化も必要ではないかと考えています（図 15）。

摩擦の代表例として査読があります。arXiv への投稿も、要は査読の問題をどうするかに集約されます。査読があることで公表が遅れる、不当な査読がある、査読者を見つけるのが大変で持続的でないといった問題があります。

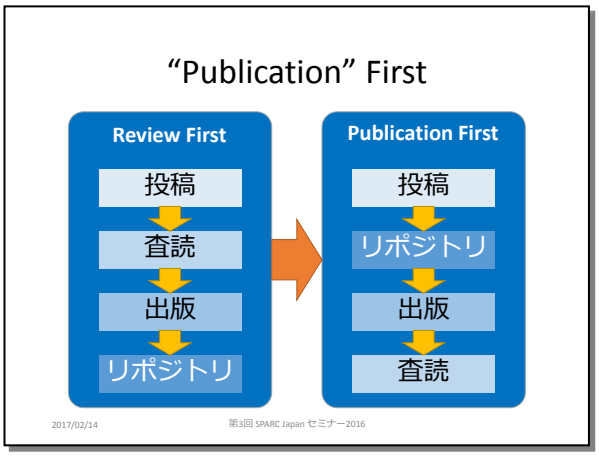
これまでの査読はゲートキーパー、つまり優れた論文だけを選別する門番の役割を果たしていました。これは確かにスペースが有限であれば有効な方法です。紙面や時間枠の制約があるので、門番があなたは通す、あなたは通さないと選ぶ方法が有効でした。しかし今はもうネットとなってスペースは無限なので、むしろ利用者の注意力が有限であるという点が制約になってきます。利用者にとって重要なものを、既に出ているものからどうやって拾ってくるか、ということが大きな課題となります。

そのとき、一つの予測として、選ぶという行為が付加価値サービスになる可能性があります。誰にとっても、自分が読むべき論文を選んでくれるとありがたい。これはある意味では、出版物のオープンアクセスが、著者支払いモデルから購読モデルに戻るとも言えます。



（図 15）

（図 14）



（図 16）

つまり、あなたのために論文を厳選してくれる付加価値サービスを購読するのです。これを私はSubscription model 2.0と呼びたいと思います。ただこのような状況では、どのリポジトリにデポジットするかが流量勝負となるかもしれません。そうなるとarXivのようなリポジトリに比べ、機関リポジトリなどは難しくなるかもしれません。いずれにしろPublicationが最初、後で査読というモデルに移行する可能性が高いと思います（図16）。

出版は人類の知の共有が目的で、評価は人類の知の継承が目的だと思います。ですから、オープンな出版と、多様な評価を組み合わせることで、出版と評価を分離できればいいのではないか。最終的に欲しいのは門番ではなく、この人の言うことは信用できるというブランドではないかと考えます。

4.まとめ

オープンサイエンスの重要な駆動力としてスピードがあるというのが本報告の主張です。ディープラーニングはスピードへの圧力が強烈な特殊事例かもしれませんが、決して孤立事例ではないと考えています。物事は臨界点を過ぎると急速な変化が起きるものであり、そういう時期を既に迎えた分野としてディープラーニングを捉えることもできます。本セミナーは図書館の参加者が多いと思いますが、この未来に図書館が果たせる役割は何か、次のアクションをぜひ考えていきましょう。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

JST におけるオープンサイエンスへの対応（DMP 導入試行をはじめとして）

小賀坂　康志
（科学技術振興機構）

CC BY ND

講演要旨

JST は 2013 年に研究成果論文へのオープンアクセス方針を公表して以来、国内外動向をふまえて研究成果へのアクセス・流通の促進（いわゆるオープンサイエンス対応）について検討・対応を行ってきた。論文については引き続き環境整備を進めると共に、研究データについても、2016 年には RDA 総会を誘致開催（東京）し、また戦略的創造研究推進事業の一部の研究課題について、データ管理計画（DMP）を導入する等、対応を進めている。本講演では、JST におけるオープンサイエンス対応の状況について報告する。

小賀坂　康志

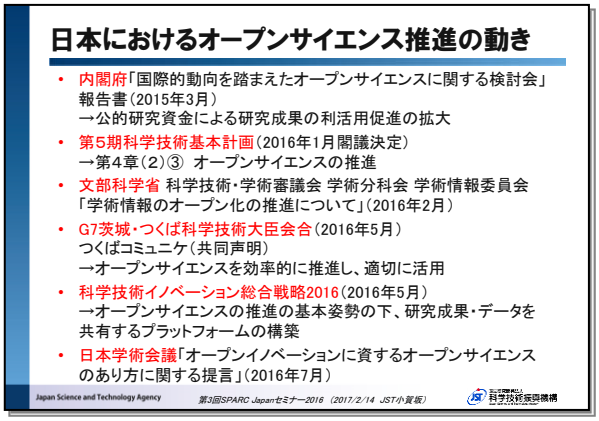
学習院大学大学院自然科学研究科博士後期課程修了。博士（理学）。JSPS海外特別研究員（NASA/GSFC）等を経て、2000年名古屋大学大学院素粒子宇宙物理学専攻助手。X線天文衛星を用いた高エネルギー天体物理学研究及びX線望遠鏡開発に従事。2009年JST入職、科学技術振興調整費、A-STEP、再生医療、CREST・さきがけ等のファンディング事業を経て、2015年10月より現職で情報事業（J-STAGE、JaLC、researchmap等）を担当。

科学技術振興機構（JST）の戦略的創造研究推進事業では、2016 年からデータマネジメントプラン（DMP）の大規模な導入を始めています。国内では大規模導入は恐らく初めての事例なので、その経緯も含めて皆さんに共有できたらと思います。

1.概況（JST におけるオープンサイエンス対応）

JST には三つの機能があります。一つ目は、研究開発戦略の立案や提言といったシンクタンク機能です。二つ目は、科学技術イノベーションを興すファンディング事業です。具体的には、基礎研究、シーズの実用化研究、産学連携、国際共同研究をやっています。三つ目は、基盤事業です。具体的には私がいる情報事業、あるいは科学コミュニケーション、理数学習支援です。

この JST の立場でオープンサイエンスの推進をするというときに、踏まえなければならないのは、各種の政策文書あるいは提言です（図 1）。上から順に内閣



（図 1）

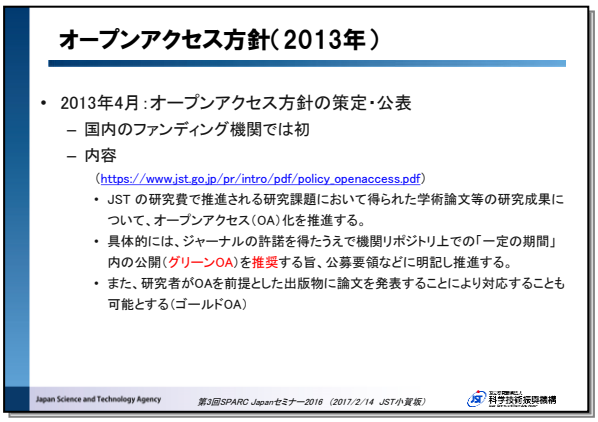
府、今年度から始まっている第 5 期科学技術基本計画、文部科学省、果ては、昨年日本で開催された G7 科学技術大臣会合、科学技術イノベーション総合戦略、日本学術会議の提言、こういうところでオープンサイエンスということが盛んに言われるようになってきています。研究資金配分機関である JST としても、この対応を考えていかなければなりません。

最初にこれまでどんなことをやってきたのか簡単にご説明します。振り返ると、2013 年、JST は国内のファンディング機関としては初めてオープンアクセス方針を策定・公表しました（図 2）。JST の研究費で推進された研究課題において得られた学術論文等の研究成果については、オープンアクセス化を推進することを宣言しています。具体的な内容としては、グリーンオープンアクセスを推奨し、ゴールドオープンアクセスについても可とします。可とするというのは、それに必要な経費を研究費から支出してもよいという意味です。このような方針を打ち出したところです。

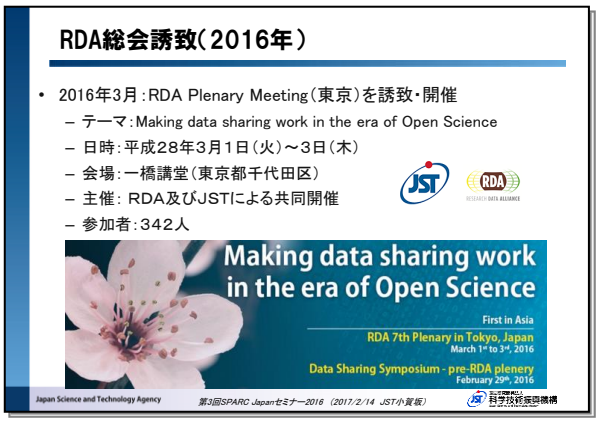
続いて、ジャパンリンクセンター（JaLC）としての活動ですが、2015 年に研究データに DOI を付与するという取り組みを始めています（図 3）。ここでは研究データについて関心のある機関やプロジェクトが集まって DOI 登録をする実証プロジェクトを行い、この結果について得られた知見を基に、「研究データへの DOI 登録ガイドライン」を公表しました。

昨年はリサーチデータ・アライアンス（RDA）の plenary meeting を東京に誘致・開催しました（図 4）。おかげさまで 340 人という多数のご参加を頂いたという結果でした。

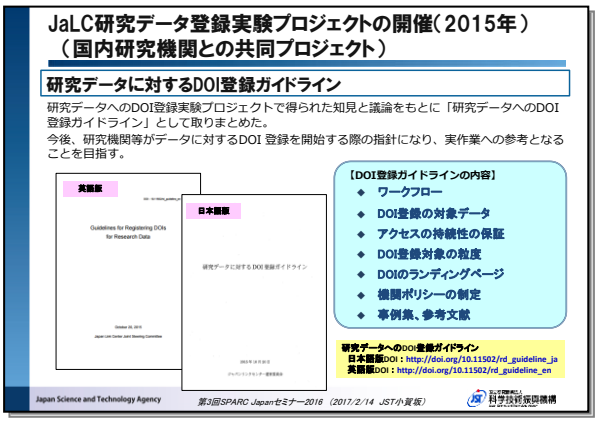
また、これも昨年ですが、研究データ利活用協議会を設置して、研究データのさらなる利活用に向け、国内主要機関や多くの参加者を得て、種々、議論を行っています（図 5）。



（図 2）



（図 4）



（図 3）

（図 5）

2.戦略的創造研究推進事業における DMP の導入

このような経緯を踏まえて、本題である、戦略的創造研究推進事業における DMP の導入についてご説明したいと思います。

まず、戦略的創造研究推進事業についてごく簡単に申し上げます。趣旨は、科研費のような純粋基礎研究とは異なり、国が定める戦略的な目標の達成に向けて基礎研究を推進するというものです。研究体制は、複数の研究者から成るネットワーク型の研究所を構築し、そこにプログラムオフィサーたる研究総括を配置し、研究領域の運営をするというスタイルです。

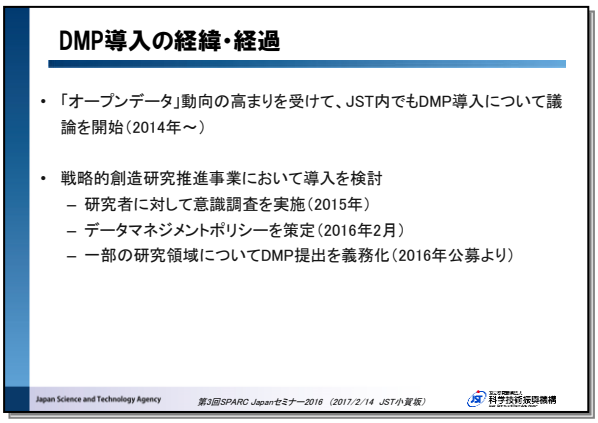
DMP の導入に向けての経緯として、オープンデータの動向の高まりがあり、これに対して JST 内でも議論しなければならないという機運が高まりました（図 6)。これが 2014 年ごろの話です。

このような背景を受けて、戦略事業においても DMP 導入を検討してきましたが、まずは対象となる研究者に対して意識調査を行うところから始めました。これが 2015 年のことです。

これに基づいて、2016 年 2 月に事業としてのデータマネジメントポリシーを策定しました。その中で、DMP に書き込むべき事柄などを定めています。

それで、いよいよ 2016 年の公募から、一部の研究領域について、DMP の提出を義務化しました。

それでは、この経緯を少し順を追ってご説明したいと思います。

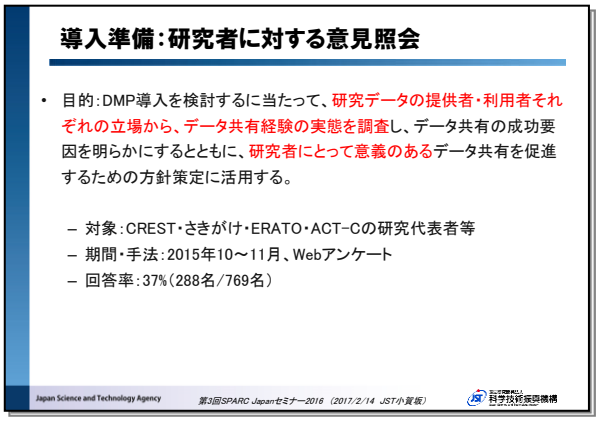


（図 6）

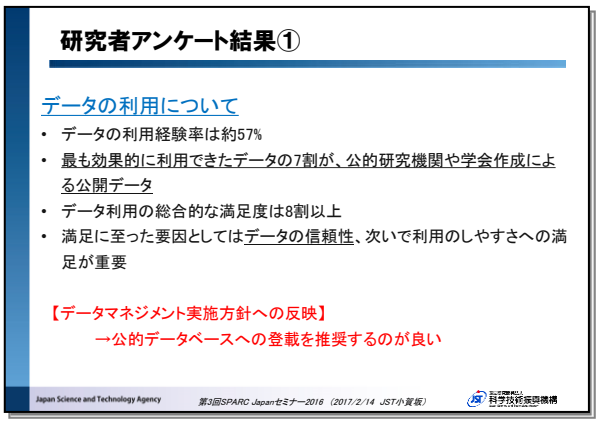
2-1.導入準備：研究者の意見照会

まず、DMP の導入に先立って、研究者に対する意見照会を行いました（図 7)。対象研究者に対してデータを提供する側、あるいはデータを利用する側のそれぞれの立場から、データを共有した経験の実態を調査し、それを基に、研究者にとって本当に意義のあるデータ共有を促進するための方針を策定したいと考えたのです。期間は 2015 年 10～11 月、ウェブアンケートで行い、対象となった研究者は 769 名、これに対して 4 割弱の 300 名弱の回答を得ています。主な項目三つについてアンケート結果をご紹介します。

最初は、データの利用の経験について聞いたところ、「他者のデータを利用した経験がある」という割合は 6 割弱でした（図 8)。その経験者に、「最も効果的に利用できたデータは何であったか」と聞いたところ、その 7 割が、公的研究機関や学会が作成した公開データでした。また、利用者の 8 割以上は「満足」であったという回答でしたが、その「満足」に至った要因と



（図 7）



（図 8）

して一番大きなものは「データの信頼性が高かったこと」という意見がありました。

いろいろ論点はあるのですが、データの利用者がこういう感想を持っているということなので、これを踏まえて、事業として、データの登録先は公的データベースをできるだけ使うように推奨することがよかろうと考えるに至りました。これもデータマネジメントポリシーに書き込んであります。

他方、今度はデータを提供する側としての経験を聞いたところ、約半数の研究者が「データを提供した経験がある」と答えました（図 9）。提供の結果、「コラボレーションがうまくいった」という回答が半数強でした。何をもって成功と思ったかについては、「共同研究に進展した」「相手側の研究が進展した」が多かったです。

その他、いろいろありますが、こうした回答を踏まえて、特に私どもが重要だと思ったのは、利用者からのフィードバックが得られるような仕掛けを組み込むことです。そのため、出典を明記すること、データと論文情報のひも付けをしっかりすることも方針として書き入れました。

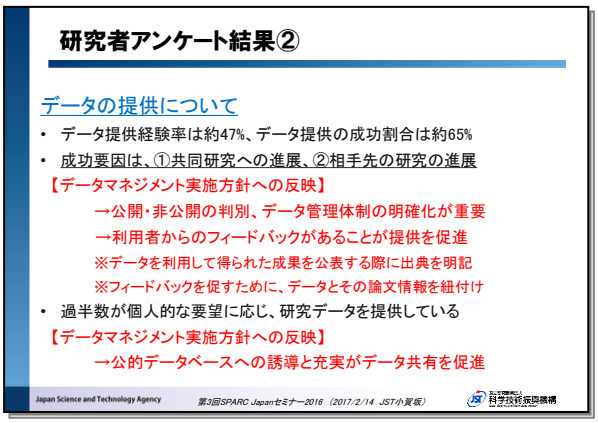
また、データを提供した経験を有する研究者の過半数が「個人的な要望に応じて研究データを提供した」と回答しました。これは逆に、個人的なコネクションがないと研究データの共有に至っていないという現状があると解釈し、パブリックなデータベースにできるだけ登録を促すことが重要であろうと考えました。

「データを提供したことがない」という場合、その理由を聞いたところ、「相手側からの要望がない」「義務負担」「自身へのメリットが不在」「管理責任が重い」「論文発表」「特許出願」「守秘義務契約に抵触するという懸念」などがありました（図 10）。これはわれわれも想像がつくところではあります。特に、若手層のデータ共有の機運を高めていくことも大切であろうということで、まずは将来のある若手研究者に DMP の作成をしっかり促すことが重要であると考えました。

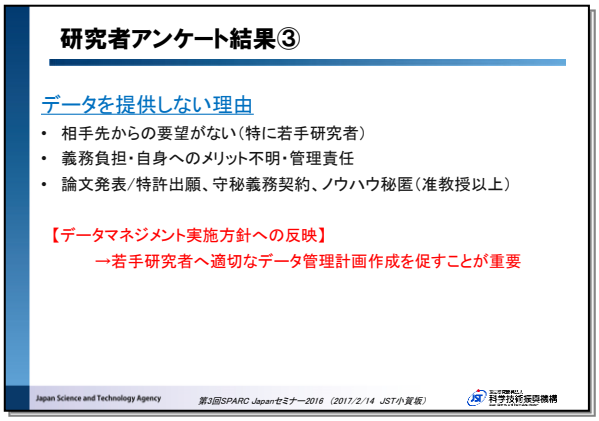
2-2.データマネジメントポリシーの策定

こうした調査を踏まえてデータマネジメント実施方針を立て、その中で DMP の記載内容を図 11 のとおりに定めました。これが出たのが 2016 年 2 月、ちょうど 1 年ぐらい前だと思います。

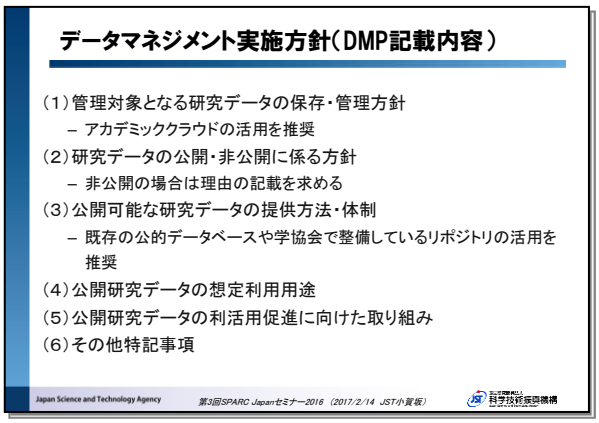
内容としては 6 項目あり、1 点目は、管理対象となる研究データの保存・管理方針を定めることです。JST としては、国の方針である、アカデミッククラウドの活用を推奨することを明記しています。



（図 9）



（図 10）



（図 11）

2 点目は、研究データの公開・非公開に係る方針を定めることです。非公開の場合には理由の記載を求めることにしています。ここで一言付け加えると、そもそも DMP は、JST がそれ以外にデータ取り扱い方針を定めていない場合は、研究者が自分の意思でもって決めることですから、公開せずともよいわけです。しかし非公開の場合は、なぜか、ということは書いていただきたいという思いが込められています。

3 点目は、公開可能な研究データがあった場合、その提供方法や体制を定めていただくことです。JST としては、既存の公的データベースや学協会で整備しているリポジトリの活用を推奨するとしています。

その他は、公開研究データの想定利用用途、公開研究データの利活用促進に向けた取り組み、その他特記事項の記載を求めています。

2-3.公募要領への記載

公募要領には、例えばこのように記載されています（図 12）。「平成 28 年度以降に新たに設定された研究領域で採択された研究代表者は（中略）『データマネジメントプラン』を研究計画書と併せて JST に提出していただきます。また、上記方針に基づいてデータの保存・管理・公開を実施していただきます」。このように求めることにしました。

ここでお気付きだと思いますが、欧米の方針と違い、まず、採択された研究者に求めることにしています。これには大きな理由があります。欧米では応募時に求めることが一般的であろうと思いますが、それは採択審査のときの評価基準にすることを意味しているのです。しかし、実際にはそれほど審査に重く使われているわけではないと聞いています。私も NSF や DOE のオフィサーと話をしましたが、その辺は悩ましいということです。

われわれも DMP が出たところで、それをどのように評価軸として設定し、どれぐらいウエートを置けばよいのかということについては、知見がありません。また、評価を行う研究者も、そのことに必ずしもリテラシーがあるとは言えないと思います。従って、現時点では採択審査の基準にはせずに、採択後に研究計画書の一部として出していただくことにしました。

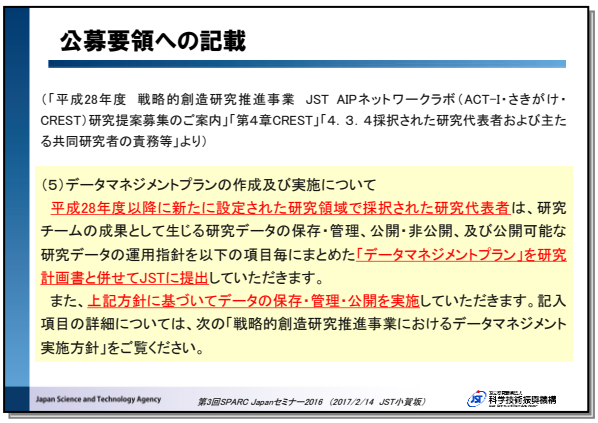
2-4.DMP の導入・提出状況及び研究者からの反応

公募の結果はもう出ており、採択課題も決定しています（図 13）。DMP 提出の対象となった研究課題は 79 課題あります。一つのプロジェクトが複数のデータ項目についての管理方針を定めてくるので、提出されたデータ項目は全部で 230 件ぐらいありました。

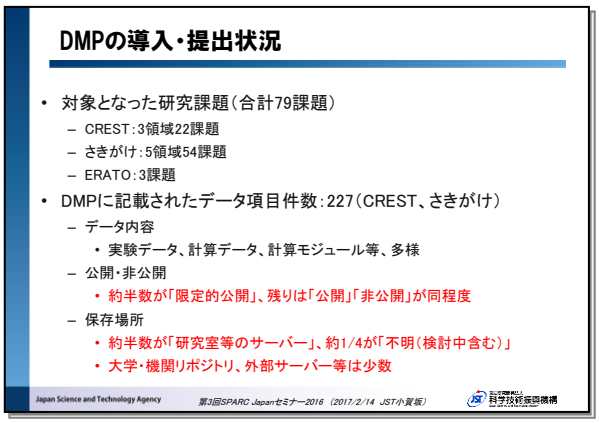
その内訳を少しここにブレークダウンしてあります。データの内容は、実験データ、計算データ、あるいは計算モジュールそのものであったり、多様でした。

公開・非公開に係る方針については、「非公開」「限定的な公開」「公開」の 3 段階で記載を求めているのですが、約半数が「限定的な公開」という回答でした。残りは「公開」と「非公開」という回答が半分ずつぐらいでした。

保存場所については約半数が「研究室等の自前のサ



（図 12）



（図 13）

ーバー」と回答しています。残る半数のうちのさらに半分、全体の 4 分の 1 が「検討中」も含む「不明」という回答でした。翻って、「大学」「機関リポジトリ」「外部サーバー」等は、ごく少数の、1〜2 割程度という回答でした。確かにそうであろうと思う反面、少々期待外れな面はあったと言わざるを得ません。

もう少し所感的なことを申し述べると、全体的な傾向として、公開の範囲を限定的にするという回答が半分ぐらいあったのですが、その公開の範囲は、おおむね研究チーム内、研究領域内、CREST なら CREST の事業の中の研究者ないしは共同研究者内でした（図 14）。

限定的な公開には 2 種類あるといわれています。一つは「制限共有」といって、共有グループをかなり明示的に限定した者の中での共有です。これに対して、「制限公開」という考え方があります。これは基本的に公開なのですが、メンバーシップを限定するということです。例えば、利用申し込みに対して要件を満たしていれば利用を許すという形です。そういう観点からすると、回答はほぼ「制限共有」にとどまるというレベルであったということです。

しかしながら、公開すべきかどうかについては、それほどかっちりとした答えを持っていないのではないかという印象を持っています。というのは、要請があれば「非公開」から「限定的公開」へ、あるいはそこから「公開」する用意もあるという回答もあったからです。

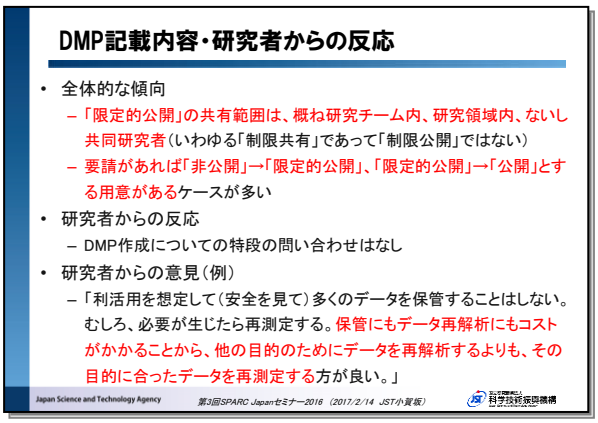
一つ面白かった意見は、厳しいといえば厳しいのですが、「利活用を想定して（安全を見て）多くのデータ保管をすることはしない。保管にもデータ解析にもコストが掛かるので、他の目的のためにデータを再解析するよりも、むしろ、その目的に合ったデータを再測定する方がよい」というものでした。これも容易に想像がつく意見なのですが、こういうご指摘に対して、なぜデータ保管をするのかという説明をわれわれはしていかなければいけないのだろうと思った次第です。

そういうわけで、DMP の作成はおおむね順調に進んでいます（図 15）。うまくいった理由は、過剰なボリュームを求めなかったこと、事業側であらかじめ記載例を作成したこと、場合によっては JST 側の担当者が作成を支援したことだと考えられます。今後は、データの保管場所についての問題、共有や公開する場合の支援なども必要になるでしょう。この方針は平成 29 年度も続くので、次第に導入する研究課題の数は増えていくと思っています。

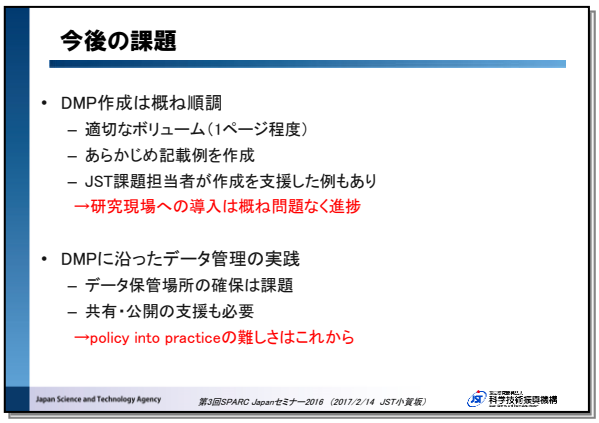
3.JST のその他の取り組み

最後に、今後、JST のオープンサイエンス対応として何を考えているか、簡単にご紹介します。

一つはオープンアクセスに関するもので、CHOR Pilot Project です（図 16）。CHOR とは米国にベースを置く出版者の連合体ですが、CHORUS という取り組みをしています。CHORUS の詳細は申しませんが、JST と千葉大学は CHOR と組み、この取り組みが日本でワークするかどうかについてのパイロットプログラムを走らせていて、近々、結果が出ようとしています。



（図 14）

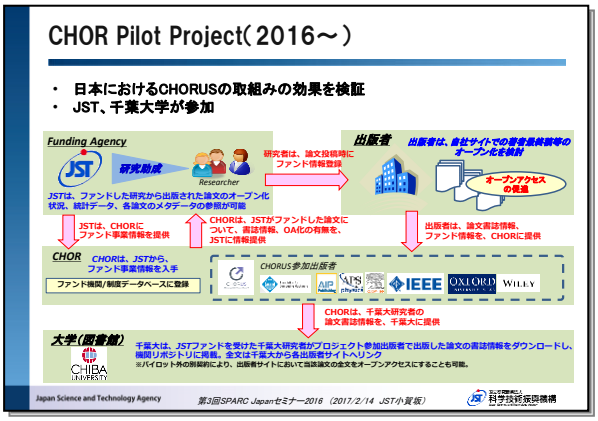


（図 15）

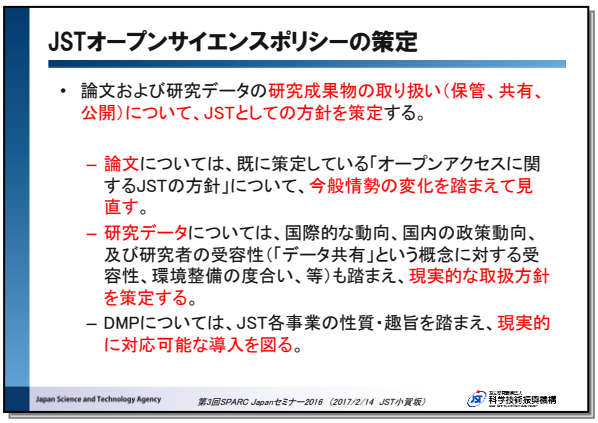
うまくいけば、ファンディングエージェンシーとしてこの取り組みを維持していくことも考えています。

また、少しテクニカルな話になりますが、オープンサイエンスの動向が高まると、インターオペラビリティの確保の観点から、早晩、persistent identifier（PID）をどう管理するのかということが課題になってくるかと思います。端的に言うと、グローバルな PID のシステムとドメスティックなシステムとの相互乗り入れをする機能が早晩必要になるのではないかと考えていて、JST というよりは JaLC として、こういうシステムが必要ではないかと検討しています。

もう一つ、JST としてのオープンサイエンスポリシーの策定を進めています（図 17）。論文については、既に策定しているポリシーの見直しを行って再策定しようと思っています。研究データについても、国内外の動向を踏まえ、研究者の受容性についても十分考えた上で、現実的な取り扱い方針を策定したいと考えています。



（図 16）



（図 17）

DMP については、今、JST の一部の事業に導入していますが、全事業に対して導入を図ることを検討しています。その際には JST は多様な事業があるので、各事業の性質・趣旨を踏まえ、現実に対応可能な導入を図っていくことを考えています。そう遠くない将来に、JST としての統一方針をお示しできるものと思います。

●**古賀**　天理大学の古賀と申します。ご紹介いただきましたポリシーあるいはプランの中で、データ保存の永続性を保証するための手段は今後どこまで考えられるのか、お聞きしたいと思います。その手段としては、システムリプレースやお金を継続的に提供することも考えられるかと思いますが、データ永続保存の取り組みに関して、考えをお聞かせいただければと思います。

●**小賀坂**　データそのものが失われないように保管するという観点と、公開の場合はデータを検索可能な状態に置くという観点の 2 点があると思います。前者については、国の方針として NII のアカデミッククラウドの整備も進むので、この中で国としてセキュアな環境が整っていくことになっています。

一方、検索可能性については、今のところは論文に振られている DOI を研究データにも適用するのがよいのではないかということで、各国で取り組みが進んでいて、日本でも JaLC がそういう取り組みをしました。こういうものが整備されていくのではないかと思っています。もし考えに抜けがあれば、どなたかコメントいただければと思います。

●**林**　NISTEP の林です。DMP がうまくいった理由として、記載例を用意したという話と、担当者が作成を手伝ったという話がさらっとありましたが、もう少し詳しく具体的に、どういう人材がどういうふうに何を手伝ったか教えていただけますか。それがつまり、

DMP を支援する人材の一例であるように見えたのです。その観点からもう少し踏み込んで、SPARC Japan セミナーという場でもあり、ライブラリアンがそういうことを支援しようとした場合に役立つようなことを知りたいので、実際に何が起きたかお聞かせ願えますでしょうか。

●**小賀坂**　JST のプロジェクト担当者が研究計画書を見て、この研究であれば、例えば NMR をやると言っているので、このデータが出てくるのではないかと推し量って、これは厳しそうだから限定公開かなということを担当者の考えでもって書いて、「先生、こんなものでしょうか」とお示しします。先生はそれでよければ「それでいいよ」とおっしゃるし、そうでなければ「ここは違う」とおっしゃいます。行為としてはそういうことをやった担当者もいたということです。

なぜそういうことをする、ないしはできるかというと、基本的に JST の課題担当者は、研究領域の設定の段階からコミットして、研究分野の専門的内容に踏み込んだ検討をしてプログラムの設計をしますし、出てきた研究計画書について、きちんとそれで研究が実施されるかというところまで踏まえて中身を見て、研究者とやりとりをするので、その中で研究内容についても学んでいきます。特段、何かのレクチャーを受けたということではなく、OJT でできるようになっていくというのが本当のところではないかと思います。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

材料科学分野における データ利用のライセンスの考え方 -図書館からの視点

小野寺　千栄
（物質・材料研究機構）

CC BY SA

講演要旨

データには著作権がないため、利用に際してはライセンスが重要となる。本発表では、物質・材料研究機構で 2016 年に開催した 3 回の連続セミナーをもとに、データ利用におけるライセンスの考え方について整理し、データ利用の実例について紹介する。また、図書館はそこでどのような役割を担うことができるのかを考える。

小野寺　千栄

物質・材料研究機構 図書館員。大学図書館等での勤務を経て、2015年より現職。研究情報の受発信を含めた図書館業務全般に携わっている。

今日は材料科学分野におけるデータ利用のライセンスの考え方について、図書館からの視点でご紹介します。まず、物質・材料研究機構（NIMS）についてご紹介した後、NIMS で開催した「オープンサイエンスと著作権」のセミナーについて、次に NIMS においてデータがどのように扱われているのか具体的な事例をご紹介し、最後にデータ利用について図書館がどのように関わることができるのかを皆さんと一緒に考えられればと思っています。

1.NIMS の概要

NIMS は国内で唯一の物質・材料科学の専門機関です。2016 年 10 月には理化学研究所、産業技術総合研究所と共に、特定国立研究開発法人に選定されました。組織のミッションとして、物質・材料科学技術に関する基礎研究および基盤的研究開発、研究開発成果の普及とその活用の促進、NIMS の施設および設備の共用、研究者・技術者の育成およびその資質の向上の四つを掲げています。職員数は約 1,500 人で、そのうちの約半数が研究に関わる職員です。私はそこで図書館司書として働いています。

図書館では、研究に関わる職員へのアドボカシー活動として、セミナーを開催しています。2016 年は「オープンサイエンスと著作権」について 3 回の連続

セミナーを開催しました。NIMS のデータ利用についてご紹介する前に、データの著作権についての考え方を整理するため、まず、このセミナーの概要をご紹介します。

第 1 回は 2016 年 1 月に開催しました（図 1)。講師は、内閣府の真子博さん、文科省の著作権分科会の委員でもある末吉亙弁護士です。約 100 名の参加者のうち、NIMS 内部からは研究系職員の約 6%に当たる 48 名の参加がありました。真子さんにはオープンサイエンス政策の背景と今後の展開についてお話しいただき、末吉弁護士にはオープンサイエンスに必要となる著作権、特に論文と研究データの著作権についてお話しいただきました。

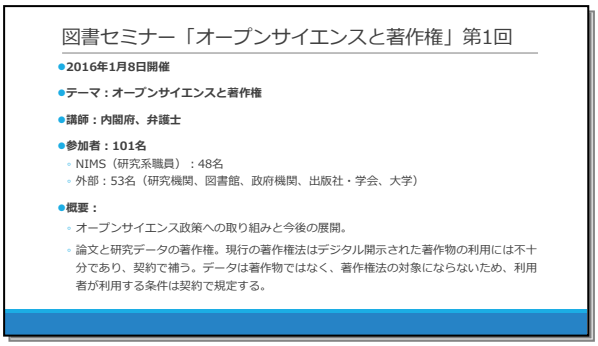
講演では、(1) デジタル著作物を利用するには現行の著作権法は不十分であり、足りないところは契約で補うことができる、(2) 著作物とは、思想または感情を創作的に表現したものであり、データは著作物には該当せず、著作権法の対象にはならない、(3) そのため、第三者がデータを利用する際の条件は、契約で規定することになる、とお話しいただきました。この第 1 回の反響を受けて、第 2 回のセミナーを開催しました。

第 2 回はデータ共有を切り口に開催しました（図 2)。講師は出版社、研究者、弁護士の 3 名でした。出版社からは、ネイチャー・パブリッシング・グループが発行しているデータジャーナル「Scientific Data」のエディターVarsha Khodiyar さんにお越しいただきました。研究者としては、この後、NIMS の事例でもご紹介する高分子データベースのデータ収集について桑島功さんに、弁護士として第 1 回に引き続き、末吉弁護士にお話しいただきました。

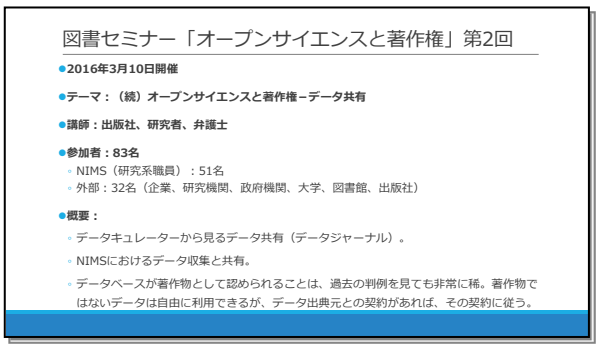
Varsha さんからは、データ共有のプラットフォームとしてのデータジャーナルについてお話しいただきました。末吉弁護士からはあらためて、データは著作物に当たらないこと、そのためデータを収集して提供するデータベースも著作物として認められることは過去の判例を見ても非常に稀であり、データの利用は、データ提供元と利用者との契約によって規定されるということをお話しいただきました。

最後となる第 3 回のセミナーでは、材料・科学分野においてもデータ駆動型研究が推進されている背景から、テキスト・データマイニングや人工知能といった次世代の知的財産権を取り上げ、弁護士と出版社から講師をお迎えしました（図 3)。

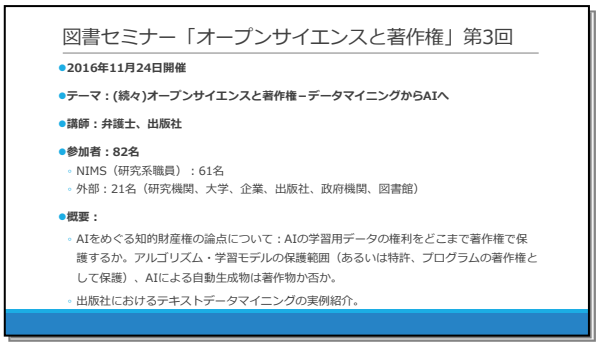
まず、内閣府の次世代知財システム検討委員会の委員でもある福井健策弁護士からは、人工知能をめぐる知的財産の論点として、三つのポイントをお話しいただきました。一つ目は、人工知能の学習用データの権利を著作権で保護するかどうか、保護する場合はどこ



（図 1）



（図 2）



（図 3）

までを保護するのか。二つ目は、アルゴリズム・学習モデルの保護範囲をどうするか、特許やプログラムの著作物として保護するのか。三つ目は、人工知能による自動生成物を著作物とするかどうかということです。

出版社からは、エルゼビア社のヴァイスプレジデントの Anders Karlsson 氏にお越しいただき、エルゼビア社が提供している研究者向けのテキスト・データマイニングについて紹介していただきました。

3 回のセミナーを通じて、(1) 論文は著作物として著作権法で保護されるが、データは著作物ではないということ、(2) だからといって、何をやってもいいのではなく、契約によって条件を設定して公開できるということ、(3) 人工知能が絡んできて、知的財産権は今まさに議論が進んでいるところだ、と整理できました。

2.NIMS におけるデータ

データの著作権について整理できたところで、NIMS におけるデータ利用の実例を、作成者、利用者、提供者の三つの立場からご紹介します。

2-1.Case1

Case1 はデータの作成者としての関わりです（図 4）。研究データのメタデータ付与については以前から取り組まれていましたが、データ駆動型研究ではデータ共有がますます重要になるため、メタデータ作成の効率化・共有化が進められようとしています。

その一つが、計測データのメタデータ記録システムの開発です。NIMS 内の計測系研究グループで検討ワーキングが組織され、その議論から開発されたシステムです。研究者の立場からさまざまな工夫が凝らされています。

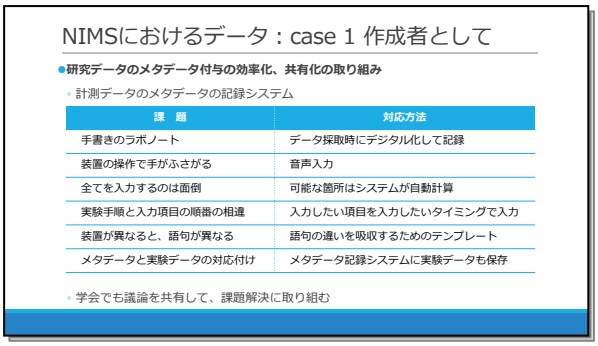
まず、計測データのメタデータを記録するときの課題として、実験時に手書きのラボノートしか記録していない場合、別途デジタル化作業が必要となります。そこで、計測装置を制御する PC からも利用できるよう、ブラウザベースの入力インタフェースによって、データ採取時にデジタル化して記録できるようになっています。

また、装置の操作で手がふさがるため、音声入力にも対応していたり、全ての項目を入力しなくても済むように、可能な箇所はシステムが自動計算してくれるようになっています。

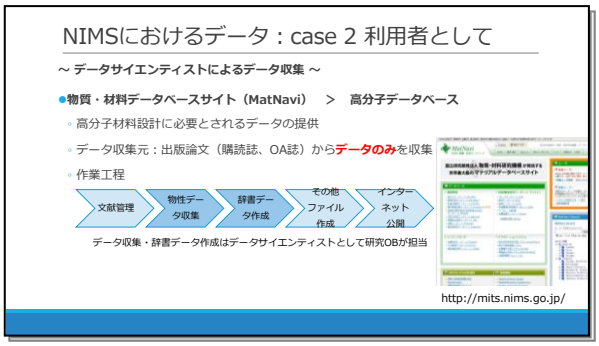
実験手順と入力項目の順番が違っていると画面を行ったり来たりしなければならないので、入力したい項目を入力したいタイミングで入力できるようになっています。また、装置が異なると語句が異なるため、その違いを吸収するためのメタデータテンプレートが用意されています。そして、メタデータと実験データ（生データ）を確実に対応付けるため、メタデータ記録システムの中に実験データも保存できるようになっています。メタデータ付与については、組織内だけでなく、学会でも議論が共有され、課題解決のため分野全員での取り組みが進められています。

2-2.Case2

Case2 は、データの利用者としての関わりです（図 5）。NIMS では MatNavi というサイトで、材料開発に



（図 4）



（図 5）

役立つさまざまなデータベースを公開しています。MatNavi はユーザー登録すれば無料で利用できます。今日は MatNavi で提供しているデータベースの中から、高分子データベース「PoLyInfo」のデータ収集についてご紹介します。

PoLyInfo は高分子設計に必要なデータを提供するデータベースです。データの収集元は、購読誌を含む出版論文で、その中からデータのみを収集しています。論文を収集元としている理由は、データの所在が明らかなこと、膨大な数が確実に出版されて公開されていること、必要とする情報が含まれていること等が挙げられます。

作業工程は、まず収集対象の文献を選定します。高分子分野は年に約 3,000 論文が出版されているそうですが、スクリーニングによって 600～700 文献まで絞り込みます。次に、その文献から物性データを収集して、その物性データから構造検索などを可能にするための辞書データを作成します。その他の必要なファイルを準備して、高分子データベースとして公開しています。このうち、データ収集と辞書データ作成については、データサイエンティストとして、高分子研究 OB や有機化学研究 OB が担当しています。Case2 では、データサイエンティストによるデータ収集についてご紹介しました。

2-3.Case3

Case3 も利用者としての関わりです（図 6）。マテリアルキュレーションという手法によって、新しい材料探索を研究している方の例をご紹介します。それには多くのデータを取得する必要がありますが、ボーンデジタルではない論文のデータも必要だったため、スキャンした論文 PDF のグラフから数値データを抽出するソフトを開発されました。同様の無料ソフトもあったそうですが、使い勝手が良くなく、効率的にデータ取得できなかったため、自分の研究用途に合うソフトを開発したところ、企業からも自社の古いデータをデジタル化したいというニーズがあり、特許を取得してライセンス契約することになったそうです。

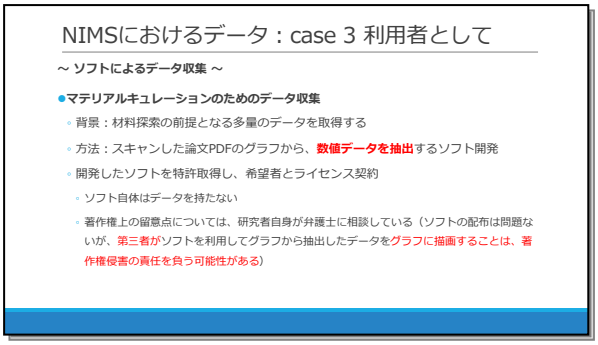
ソフトを提供する際の著作権上の留意点については、研究者自身が弁護士に相談しており、グラフからデータを抽出することは著作権法上の問題はないが、抽出したデータをグラフに描画することは、著作権侵害の責任を負う可能性があるため、ソフト利用者にもそのように伝えているそうです。Case3 ではソフトによるデータ収集についてご紹介しました。

2-4.Case4

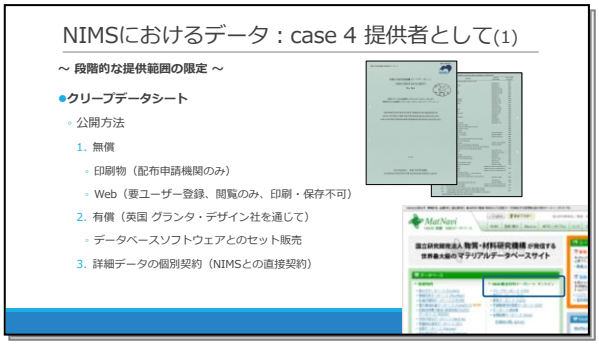
Case4 は提供者としての関わりです（図 7）。MatNavi で提供しているデータシートのうち、クリープデータシートがどのように提供されているかをご紹介したいと思います。

その前にクリープとは何か、ごく簡単に説明します。クリープとは、物体に一体の荷重を加え続けることでゆがみが生じる現象のことをいいます。NIMS では高温下で金属を引っ張り続けたゆがみを測定していて、長い試験では約 40 年引っ張り続けて、世界最長記録を更新しました。

このクリープデータを「無償」「有償」「詳細データ



（図 6）



（図 7）

の個別契約」という三つの方法で提供しています。無償提供は、印刷物とウェブの MatNavi です。印刷物は、配布希望の申請をした機関のみに配布していて、申請の際には利用目的も明記するようになっています。これはデータを営利目的で利用されないように管理するためで、過去には配布を拒否したケースもあったそうです。ウェブの場合はユーザー登録により、利用を管理しています。ウェブ利用においては、閲覧のみで印刷や保存ができないようになっているため、無償提供の場合、デジタルデータは一切提供していません。

有償提供のデータは、無償提供しているものと基本的には同じですが、イギリスのグランタ・デザイン（Granta Design）社が開発しているデータベースソフトウエアとセットで有償販売されています。さらに詳細なデータが欲しいというユーザーとは、提供内容を個別に契約しますが、これは NIMS との直接契約になります。

提供方法によって、利用を認める範囲は異なっています（図 8）。まず、無償配布の場合、データをプロットしたり、データ解析結果を公表することは可能ですが、データを営利目的で利用することは認めていません。二次利用をコントロールする目的もあって、利用は登録制となっています。

有償販売の場合、その範囲はグランタ・デザイン社との契約に基づくことになります。詳細データを個別に契約する場合も、その範囲はケース・バイ・ケースですが、データそのものを開示しないこと、第三者へ提供しないこと、データを基に解析した結果は公開してもいいという 3 点は、原則として共通しています。

Case4 では、データ提供の範囲を段階的に限定している事例についてご紹介しました。

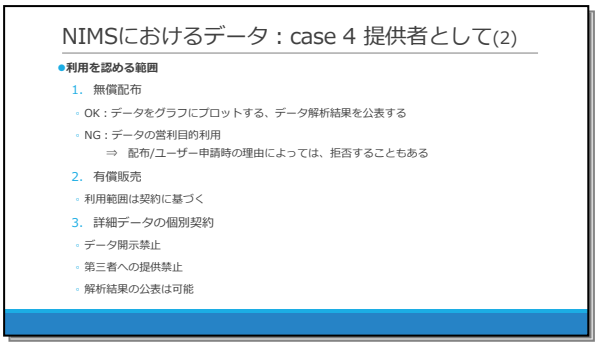
2-5.Case5

最後も提供者としてのケースです（図 9）。NIMS では情報統合型物質・材料開発イニシアティブ（MI²I）を進めています。これは JST のイノベーションハブ構築支援事業の採択課題の一つで、データ科学をこれまでの物質・材料科学に融合させることによって、新しい物質・材料科学の研究を加速する取り組みの場となることを目指しています。

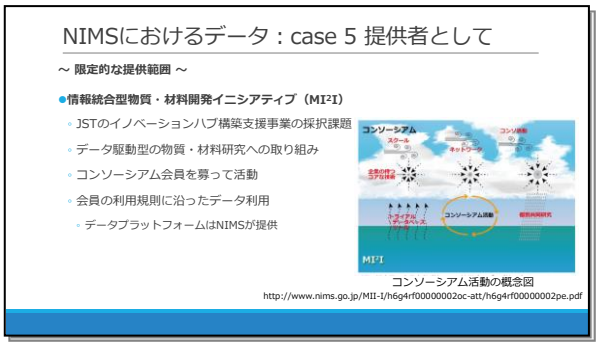
活動の根幹を成すものの一つに、コンソーシアム活動があります。活動の概念図が右下の図です。データベースやツールを会員が使用することで、会員が持つコアな技術が拡大し、コンソーシアムで提供されるスクールや研究者のネットワークが技術の開発を促進します。技術が成長することで、雨粒のようにイニシアティブの中に落ちてきて、個別の共同研究を開始する契機になることを目指しています。

このため、研究コミュニティの発展に向けて、会員間でできる限りのデータ共有を目指しており、データ共有のためのプラットフォームを NIMS が提供しています。Case5 では、会員の利用規則という契約による、限定的なデータ提供の事例についてご紹介しました。

いずれのケースでも著作権法の範囲内とそうではないところを明確に区別して、著作権法の及ばない範囲については、契約でその利用範囲をコントロールしていることがお分かりいただけたかと思います。



（図 8）



（図 9）

3.図書館業務とデータ

ここまで NIMS の研究者とデータとの関わりをご紹介しましたが、この中には図書館は関わっていません。データを扱うにはその分野の専門知識も必要ですが、図書館員がその必要な知識を習得するのを待っている時間的な余裕はありません。そこで、図書館員が今あるスキルでできることは何か、三つ挙げられるのではないかと思いました（図 10）。

まず、図書館の伝統的業務である資料の保存と管理のスキルを生かして、研究者にデータの保存と管理に必要な情報を伝え、それを通じて図書館員もデータの扱いを学ぶということです。これについては、この後、尾城さんから詳しくご紹介いただけるかと思います。

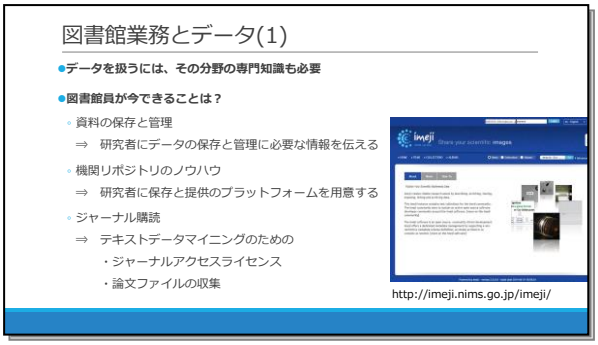
次に、機関リポジトリのノウハウを生かして、研究者に保存と提供のプラットフォームを用意し、その対話の中で図書館員も必要なプラットフォームについて理解を深めるということです。NIMS では、文献をセルフアーカイブするリポジトリとは別に、マルチメディアデータをセルフアーカイブする imeji というリポジトリを研究者に提供しています。imeji ではメタデータ項目を研究者自身が設定できるので、拡張的なメタデータ管理ができるようになっていますが、先にご紹介した研究者自身が開発したメタデータ記録システムと共同するということも考えられると思います。

そして、ジャーナル購読のノウハウの利用です。データ駆動型研究では、どれだけのデータを集められるかがポイントとなるので、テキスト・データマイニング用に購読とは異なるジャーナルアクセスのライセンスや論文ファイルの収集が必要となります。NIMS でつい最近、そうした事例があったのでご紹介したいと思います。

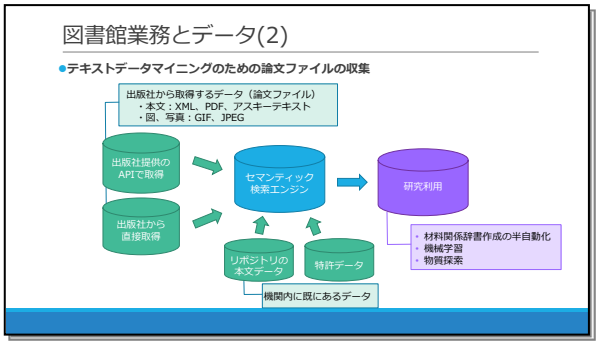
出版社とのコネクションを持っているのが図書館なので、データマイニングのための出版社とのライセンス契約を図書館を通じて行いましたが、出版社によって提供可能な範囲はさまざまでした（図 11）。まず、購読ジャーナルしか認めないというところ、研究目的で NIMS 内だけで使うなら提供できるというところ、研究用途として実験的に無償提供するところ（学会）など、さまざまでした。マイニング用のデータなので、論文 PDF のように美しく成形されている必要はなく、こちらの希望としては本文データと図や写真のデータだけを入手できれば足りるのですが、出版社によって、API 経由だったり、出版社サーバーから直接取得だったりと、それぞれ可能な方法でデータを収集しました。

取得したデータを知識ベースとして、セマンティックな検索エンジンに投入しますが、データをよりリッチにするために、機関内で既にあるデータ（リポジトリの本文データや特許データ）についても、このセマンティック検索エンジンに投入します。それを材料関係辞書作成の半自動化や、機械学習、物質探索に利用することを予定しています。

契約交渉を通じてあらためて感じたのは、ライセンス契約の範囲設定の難しさです（図 12）。出版社から論文データを購入する場合、テキスト・データマイニングの成果を機関内にとどめておくのか、それとも第三者に提供するのかで、料金が大きく変わってきます。また、ソフトウエア会社との契約でもセマンティック検索エンジンにより生成した辞書データをソフトウエ



（図 10）



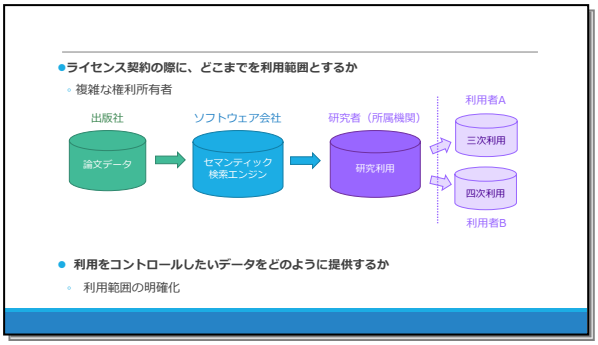
（図 11）

ア会社に提供するかどうかで、やはり料金が変わってきます。ソフトウエア会社に辞書データを提供するということは、出版社のデータを第三者に提供することになるので、出版社が第三者への提供を認めていない場合はできません。

NIMS では今回、研究で利用するために NIMS 内だけで限定利用することを前提として、細部については、その都度確認しながら進めるということで、出版社からデータを購入しました。NIMS の中だけでの利用、つまり、二次利用までのライセンスですが、研究成果を公開して、三次利用あるいは四次利用ということを考えると、権利の所有者はどんどん複雑になって、ライセンスの内容も、都度、協議するというわけにはいかなくなることが想定されます。

今回のまとめとして、材料科学分野ではオープンにできるデータとそうでないデータがあり、オープンにできない、つまり、利用をある程度コントロールしたいデータの提供方法の検討ポイントは、利用範囲を明確にすることではないかと思っています。

ライセンスについてはまだまだ勉強中で試行錯誤しているところなので、もし、そういうセミナーを開催される際には、ぜひ教えていただけるとうれしいです。



（図 12）

●**角田**　JAMSTEC の角田と申します。著作権法をよくよく調べられていて、よく分かりました。ありがとうございます。ただ、そもそも著作権法をいくらいじったところで、多分、データそのものをどうこうするのは全く別の話になってしまうのではないかというのが、私がまた別途、弁護士さんに聞いたときの感じでした。

著作権法というのは、中身が事実であろうとシミュレーションであろうと全く関知しないし、中身が面白かろうがつまらなかろうが、陳腐だろうが独創的だろうが、それも全く関知しません。中身をどう表現するかという表現法が独創的かどうかで保護するというのが著作権法の本来の精神なので、データそのものを著作権法でどうこうしようとしても、それは本質的に無理で、データをどう表現するかに関して、びっくりするような表現方法、独創的な表現方法があるのであれば、それは著作権法の保護の範囲内になるかもしれません。

では、論文はどうなのかと皆さん思われるのですが、もし論文が本当に定型化されて書かれていて、しかも、どの AI でも、体裁、表現が同じような論文が書けるということであれば、論文の中身がいくら独創的であっても著作権法では保護されないというすごい状況が起きるので、データそのものをどう保護するかということになった場合には、もう著作権法とは全く関係なしにライセンス契約なり何なりを考えるのがいいのではないかと思った次第です。

●**小野寺**　おっしゃるとおりで、データは著作権で保護される範囲ではないということは、NIMS で開催したセミナーでも弁護士の先生が繰り返しおっしゃっていたことです。データというのは客観的な事実を示したものなので、そこに独創性や創造性は関わらないということを学んだところです。

●**角田**　全くそのとおりなのですが、データにはシミュレーションデータというものもあります。人間が本を書いたり、絵を描いたりするときも、事実をそのまま写すというものと、シミュレーションしたり、頭の中にあることを事実に基づかずに表現するということもありますよね。そういうこともありますのでということを一言付け加えさせてください。

●**小野寺**　ありがとうございます。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

研究データ共有の理想と現実，そして実践可能性
～地球環境分野の研究基盤に関する意識調査から～

小野　雅史

（東京大学地球観測データ統融合連携研究機構）

CC BY SA

講演要旨

「研究データの共有は科学の発展に貢献する良い考えだ」という理念については、研究者のみならず、政策立案者、資金提供者、市民等、多くのステークホルダーが基本的には同意するはずである。しかし、実際には、多くの研究データが、共有または公開される状況に至っていないのが現状である。このオープンサイエンスの理想と現実の間には、どういった課題が存在するのだろうか？これを明らかにするために、我々は地球環境情報分野の現場の研究者を対象として意識調査を実施した。そこで、本講演では、この意識調査の結果をもとに、我々がオープンサイエンスを前に進めるために実現可能な取り組みについて、考察した内容を紹介する。

小野　雅史

東京大学地球観測データ統融合連携研究機構（EDITORIA）の特任研究員。DIAS（データ統合・解析システム）という研究データ基盤を中心とする事業「地球環境情報プラットフォーム構築推進プログラム」に参加している。過去に、地理情報標準ISO/TC211の仕様検討委員、GEO（Group on Earth Observations）のオントロジータスクチーム、Belmont Forum E-Infrastructures and Data Management Collaborative Research ActionのData sharingグループのメンバーとして活動。

本日のセミナーのテーマは「オープンサイエンス再考」ですが、私の報告のタイトルにある「研究データ共有」は、オープンサイエンスの要素の一つです。研究データ共有の理想、広義にはオープンサイエンスの理想といったときに、皆さんは恐らくこういうものをイメージされると思います。「研究者がみんなデータを提供して、それをみんなでシェアして、それが科学の発展を促進し、ひいては社会への貢献に資する」。これは非常にいい考えです。みんないいことだと思っていると思います。でも、その一方で現実はどうなっているのでしょうか。

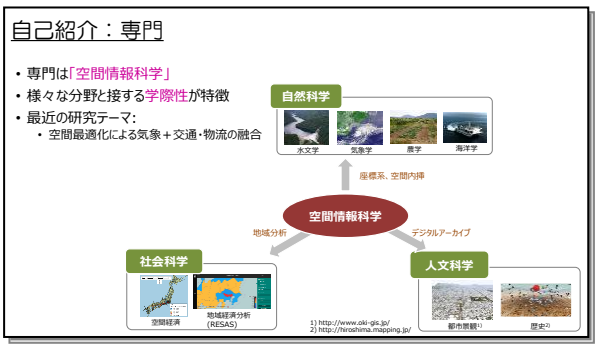
実はこれをとても簡単に知る方法があります。例えば、誰でもいいので研究者を捕まえて、こういう質問をしてみてください。「先生、あなたが持っている研究データのうち、何パーセントぐらいを他の人とシェアしていますか」。これは言うまでもなく、非常に歯切れの悪い答えが返ってくると思います。これはオープンサイエンス推進派といわれている先生でも、そんなに変わりはないと思います。つまり、こういうことなのです。みんないいことだと思っているのに現実はそうなっていない。ですから、この理想と現実の間にどういう空白があって、その隙間を埋めるために何をしたらいいのか、そのためのヒントになるような話を今日はできたらと思っています。小賀坂さんや小野寺

さんのお話とも非常に符合する部分があるので、思い出しながら聞いていただけたらと思います。

自己紹介

私の専門は空間情報科学です（図 1）。この分野の特徴は、自然科学だけでなく、社会科学、人文科学などいろいろな分野と接することが多い学際性です。ある分野とある分野を融合させるということも比較的やりやすくて、最近の私の研究テーマは気象と交通・物流の融合です。研究とは全く関係ないのですが、その他、色、調味料といった民間資格を持っています。私は料理が結構好きなので、家に調味料が 80 種類以上あって、それを組み合わせて毎回新しいカレー粉をつくる、そんなことをしていたりします。一見これは全然関係ないように聞こえるかもしれませんが、分野の融合のように、異質なものを組み合わせて何か新しいものをつくるのが好きな人間、それが私です。

こうした私自身の特性を生かしながら、これまでさまざまな異分野の融合関係のプロジェクトを実施してきました（図 2）。



（図 1）

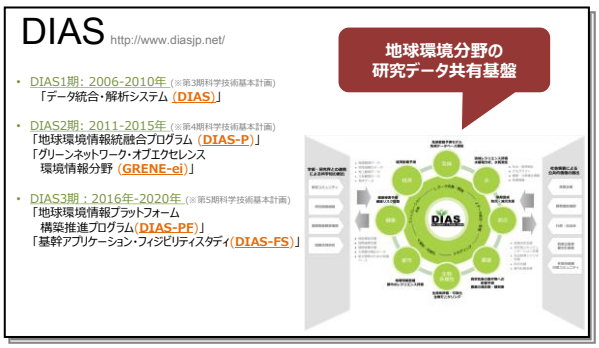
（図 2）

一例として、ISO/TC211 の geographic information standards に関する仕事、国内の地球環境データ統融合の Data Integration & Analysis System（DIAS：データ統合・解析システム）というプロジェクト、DIAS の国際版のような Group on Earth Observation（GEO：地球観測に関する政府間会合）のオントロジー等のタスクグループでの仕事、ベルモント・フォーラムのデータシェアリンググループのメンバーとしての活動などを行ってきました。

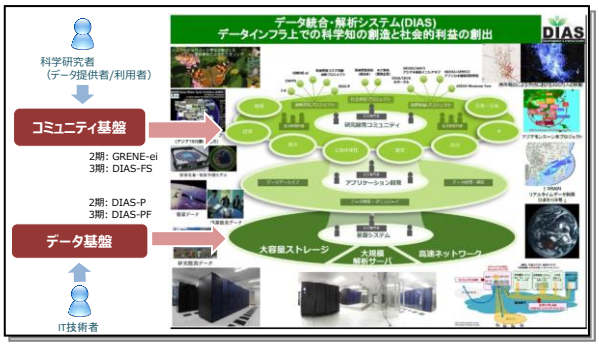
Data Integration & Analysis System（DIAS）

今でも私のメインのプロジェクトは DIAS です（図 3）。DIAS とは、地球環境分野の研究データ共有基盤のことです。DIAS という名前が使われだしたのは、第 3 期科学技術基本計画の 2006 年からで、ちょうど今年から DIAS 第 3 期に入りました。

この DIAS の特徴を一つ挙げると、過去の SPARC Japan のイベントで、北本先生からもお話があったと思いますが、コミュニティ基盤とデータ基盤の二つのレイヤーを持っていることです（図 4）。このデータ



（図 3）



（図 4）

基盤を支えるのは主に IT 技術者が中心で、コミュニティ基盤の中核となるのが、いろいろな分野のサイエンスの研究者です。このコミュニティ基盤の上で地球環境に関する課題を取り上げ、その課題であれば私はこういうデータを持っている、そのデータを使わせてくれれば私はこういう分析ができるというような議論をしながら、研究を進めているという枠組みになっています。

このような形で研究を進めていくことに力を入れているのは言うまでもなく、その他にも力を入れている領域があり、その一つが教育です（図 5）。東京大学での講義、サマースクールなどの企画、アジア地域でフィールドスタディなどを行っています。

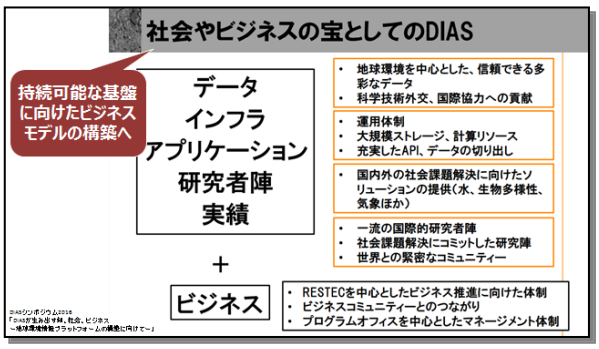
他に産学連携もしています。図 6 は河川のダム操作支援システムの事例です。例えば、東京大学がモデルや手法を考案して、電力会社は電力データを出して、日本工営などの建築コンサルはシステムのオペレーションをするという枠組みで共同作業を行います。

このように民間企業との仕事経験を蓄積していて、今期の DIAS の第 3 期で力を入れている領域がビジネスモデルの構築です（図 7）。なぜ研究基盤がビジネスなのかと思うかもしれませんが、キーワードは「サステナビリティ」です。国家財政もこれから緊縮していく中で、やはりデータ基盤は自分で独自に継続していくことを考えなければいけないのです。European Open Science Cloud でもサステナビリティの議論はいろいろ出ていて、コマーシャルサービスの利用も挙げられていると思うのですが、今、DIAS でもそちらの検討に力を入れています。非常にチャレンジングな課題ですが、いろいろ考えているところです。

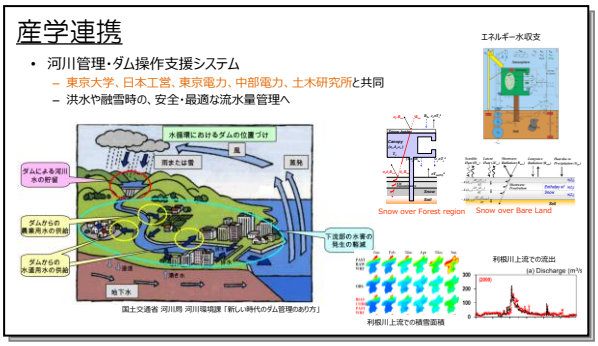
ただ、ビジネスモデルをつくるといっても、別に DIAS のメンバーが MBA を取りに行くというような話ではなく、ポイントになるのは人だと思っています（図 8）。つまり、DIAS の考えにどれだけの人が乗ってきてくれるか、それがポイントになると思うので、潜在ユーザーがどういうところにいるかを調査しながら、リーチする活動をしています。

このように、いろいろな活動を続けているのですが、当然、全てが順調にいくわけではありません。中にはデータ利用者側のニーズとデータ提供者側のシーズが

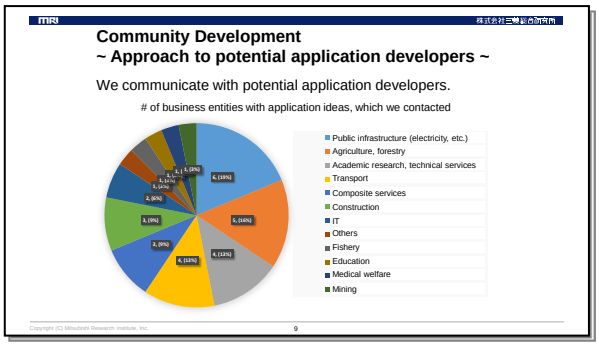
（図 5）



（図 7）



（図 6）



（図 8）

うまく折り合わないといったことが出てくるという問題があります。

データ共有とデータ公開

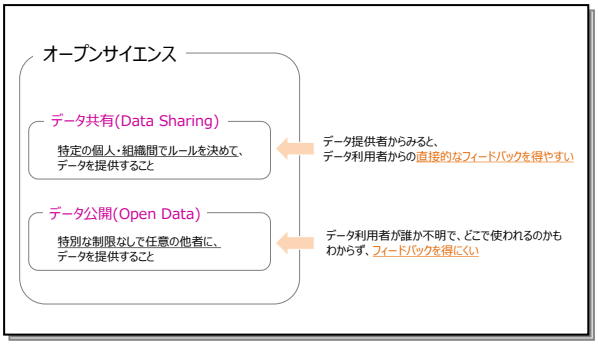
問題の詳細に入る前に、データ共有とデータ公開の二つのコンセプトについて説明します（図 9）。

データ共有（data sharing）の定義は、特定の個人・組織間でルールを決めてデータを提供することです。一方、データ公開（open data）とは、特別な制限なしで任意の他者にデータを提供することをいいます。

これはオープン・データ・イニシアティブや、ベルモントでも同じような意味で使っていたので、国際的にも通用する定義だと思います。オープンサイエンスというのは、広い意味でこの両者を包含する概念であると考えます。

こうして比較してみたときに、ひょっとすると皆さん、データ公開の方が制限がない分、いいことなのではないかと見えるかもしれませんが、一概にそうとは言えないのです。なぜなら、データ提供者の視点から見ると、利用者からのフィードバックを得やすい枠組みは、どちらかというとデータ共有の方だからです。

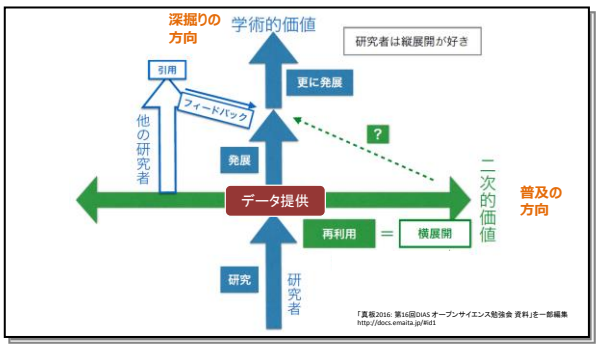
これは少し考えれば分かると思うのですが、特定の範囲でデータだけではなくルールもシェアしてデータを提供し合うという枠組みなので、相手の姿もイメージしやすいですし、コミュニケーションが取りやすいのです。一方で、データ公開の方は、任意の第三者なので要は相手が誰かも見えにくい、あるいはどこで使われるのかも分からない、そもそも連絡が取れるのかも分からないというような問題があって、フィードバックを得にくい枠組みです。



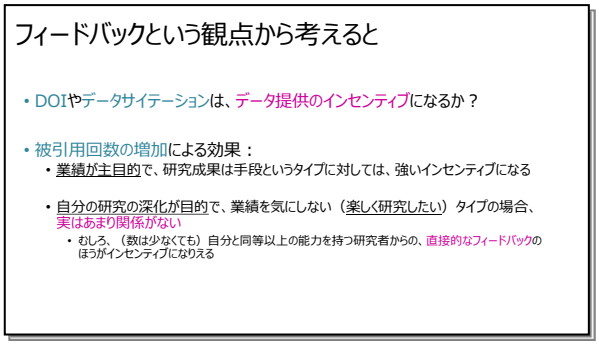
（図 9）

このフィードバックを得られるかどうかは、研究にとって非常に重要な要素です。図 10 は、DIAS の勉強会で国立環境研究所の真板英一さんがベースをつくられた図です。平たく言うと、研究には普及の方向と深堀りの方向の二つの軸があるということです。尖った研究者で、一つの研究テーマをどんどん深掘りしていくタイプは、縦軸の方向に進みます。でも、オープンサイエンスの基本戦略は、どちらかというと横軸の普及の方向性です。

つまり、オープンサイエンスは、オープンにしてみれば誰かが使ってくれて、何か面白いことが起こるかもしれないというような割に楽観的な推論に基づいているのです。確かに誰かが面白い使い方をしてくれるかもしれないのですが、一つの研究テーマを深掘りするタイプにとっては、それは直接的にはあまり関係ないのです。ですから、そういうタイプの人にはどちらもフィードバックが非常に重要で、これが不可欠になります。こうした観点で見ると、DOI やデータサイテーションはインセンティブの本丸といわれていますが、



（図 10）



（図 11）

これがインセンティブとして効く相手と効かない相手がいるということも分かってきます（図 11）。

例えば、業績が主目的で、研究成果は手段というタイプに対しては、強いインセンティブになります。その一方で、自分の研究を深掘りするのが目的だったり、他の人がどう思っているかは関係ない、私は楽しい研究がしたいのだというタイプに関しては、実はあまり関係ありません。そういうタイプにとっては、自分の分野で、かつ、自分と同等以上の能力を持つ研究者からの一言という方が、強いインセンティブになるのです。このような議論をわれわれ DIAS でしてきたので、DIAS はデータ公開よりはデータ共有の方をアイデンティティとしてやってきたという経緯があります（図 12）。

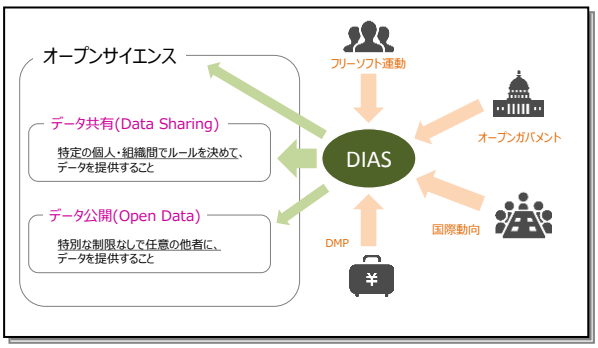
意識調査の実施

このような中、オープン化に対するプレッシャーがだんだん強くなってきて、その方向性も考えなければいけない、対応していかなければいけないという状況になったので、今後の指針を得るために調査を企画しました（図 13）。この詳細については「情報管理」11 月号に掲載されているので、そちらをご覧いただければと思います。今日はその調査の一部と、そこから分かったことを中心にご紹介します。

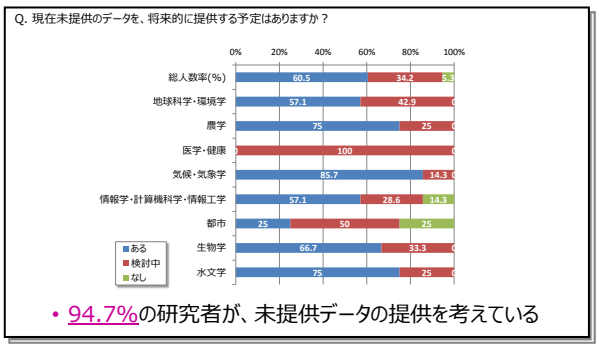
まず、研究者はそもそもデータの提供に反対しているのかという素朴な疑問があります。それに対する答えが図 14 で、要はそんなことはないという結果でした。実に 94.7%の研究者から、まだ出していないけれど、提供の予定がある、あるいは検討しているという回答が得られています。

次に、そんなに考えてくれているのなら、なぜそれを出してくれないのかという疑問が沸き起こると思います。

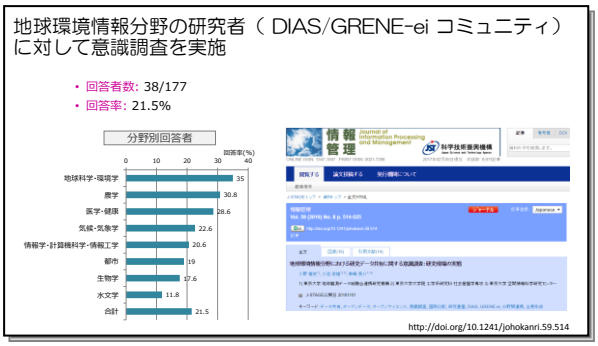
それに対する答えは、「時間がない」ということがやはり一番強い理由になるかと思います（図 15）。この理由が「インセンティブがない」という理由の倍近くを占めています。ここから、仮にインセンティブとして、DOI、データサイテーションが進んだとしても、何らかのサポートがないと思うとおりにデータの提供は進まない可能性があるということが見えてきます。



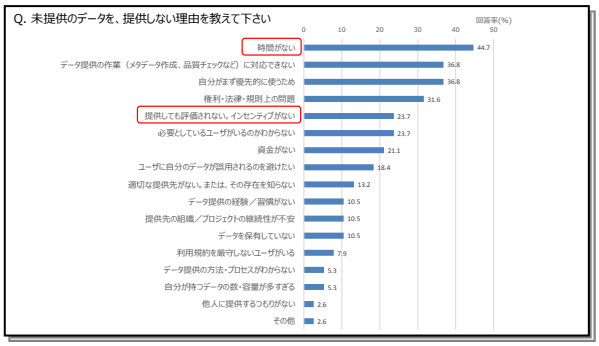
（図 12）



（図 14）



（図 13）



（図 15）

誰にだったら提供してもいいのかという疑問に対する結果は、図 16 です。青い方がメタデータに対する答えで、赤い方がデータに対する答えです。上の方に行くほど公開の範囲が広めで、下の方に行くほど狭めという整理の仕方をしています。やはりメタデータに関しては別に任意の人に見せてもいいけれども、データに関しては、「関係する分野だったらいいけれど、ちょっと公開は」という結果が見えてくると思います。

続いて、いざ提供するときに、データ共有あるいはデータ公開するときの条件はどういうものがふさわしいかという質問をしました。

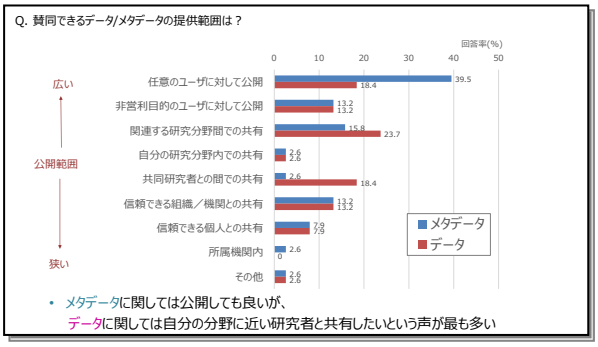
図 17 が、データ共有の条件の結果です。左側が私の調査の結果、要は国内の結果で、右側が私が調査の参考にした DataONE という国外のプロジェクトの結果です。青いバーが、データ提供側としてこういう条件を入れてほしいという結果です。赤いバーが、データ利用者として、その条件であればのんでもいいという結果です。

この結果から面白いことがいろいろ分かりました。三つだけポイントを言うと、一つ目に、国内の結果でも国外の結果でも同様に、データ共有の条件として、そのデータを使ったということを論文で引用すること、データ利用規約を守ることがトップの方に来ています。

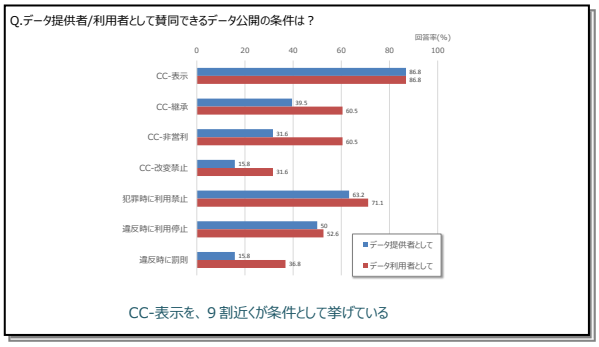
二つ目は、国内の結果では、データの提供者よりも利用者として受け入れ可能な条件が多い、つまり、データを使わせてくれるのなら、その条件は少々厳しくてものむという結果が出ているということです。

三つ目に、プロジェクトへの参加機会を提供するという条件で、国外の結果と国内の結果で顕著に差が出たことです。つまり、あなたのプロジェクトに参加させてくださいというデータ共有の条件を、海外では重視している、国内では重視していないという結果です。これはどういうことかというと、データ共有をコラボレーションのチャンスとして見る向きが海外では強く、国内では弱いという結果なのではないかと解釈しています。

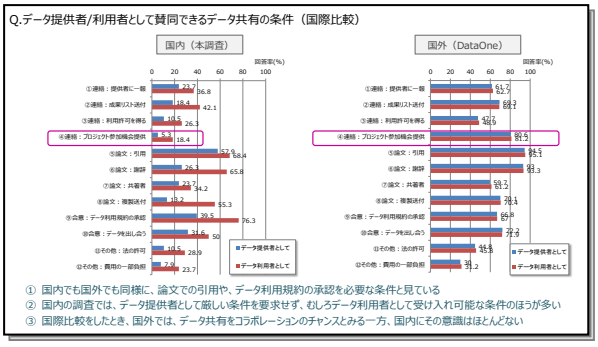
続いて、図 18 はデータ公開の条件の結果です。クリエイティブ・コモンズ・ライセンス（CC ライセンス）は表示がマストだという回答が 9 割近くあることがまず分かりました。



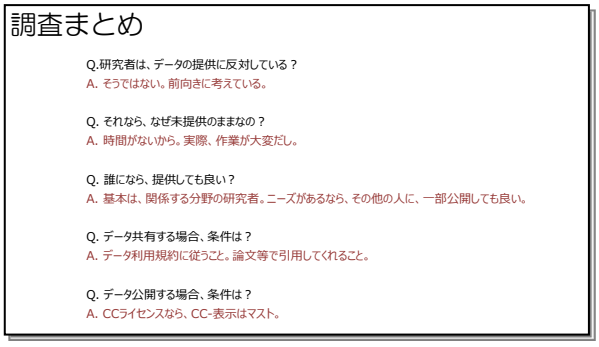
（図 16）



（図 18）



（図 17）



（図 19）

CC ライセンスは、それを守らなかったときも特にペナルティがなく、フリーライディングが起きやすいという問題が指摘されていました。ではペナルティとしてどういうものが適切なのか、私が個人的に興味を持ったので、そのような質問を入れてみました。それによると、罰則まではいかないけれど、利用停止処分はしたいという結果が見えました。

私の調査を FAQ 形式で簡単に整理すると、図 19 のような結果になりました。この調査をやってみて、出てきた結果の全てをまだ消化しきれていないところがあって、それに対する具体的なソリューションがすぐには出てこない状況です。しかし、これが研究者の現在の考えであるということはきちんと受け止めて、今後何をするかを考えていく必要があると思っています。

私たちはこれから何をすればいいのか

これから何をすればいいのか、考えたことを、簡単にご紹介したいと思います。

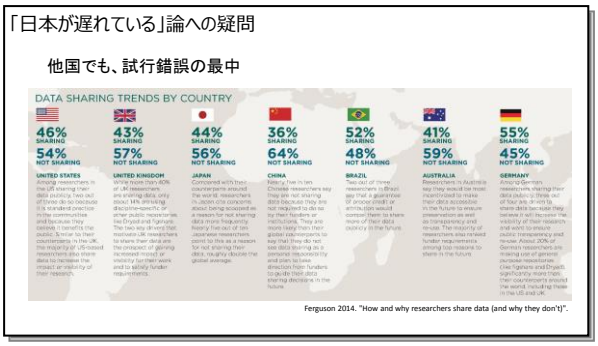
まず大前提として、焦って間違った方向に行ったり、強制力のあるルールをすぐ採用したりしない方がいいのではないかと思っています。最近、特に気になっているのは、オープンサイエンス系の議論だと、とにかく海外は進んでいて日本は遅れているから早くやらなければいけないという論法を使う人が結構いることで、これに私は違和感があります。

というのは、私の経験からも、確かに海外の方がいろいろなチャレンジ、トライアル、新しいコンセプトが出てはいるのですが、実際に全ての研究者の実態を調査すると、全分野に進んでいる人も遅れている人もいて、その統計結果を出すと、そんなに変わらないのではないかという印象があります。

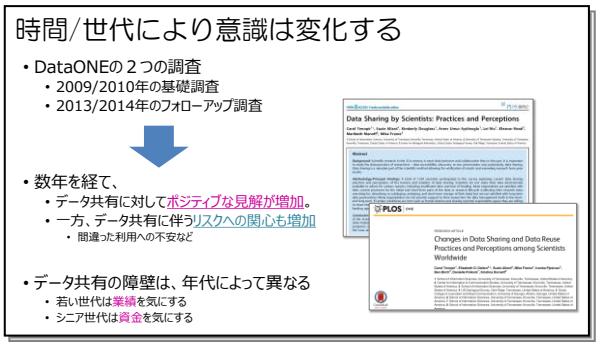
それを裏付けるかのような調査結果も出ていて、例えばイギリスはオープンデータですごく進んでいるというイメージがあるかもしれませんが、実際には日本と、sharing と not sharing のレートはほとんど変わらないのです（図 20）。ですから恐らく、海外でも進んでいる人は進んでいるけれど、全体としてはそんなに変わらないのです。日本でも海外でも同様に試行錯誤の最中であるというのが実態に近くて、国別に進んでいる、進んでいないというよりは、ライフサイエンスのように進んでいる分野もあれば、全然まだ進んでいない分野もあると言った方が、私にとっては実感として腑に落ちる感じがします。

また、焦って何かをやらない方がいいということを裏付けるかのように、時間がたてば、やはり意識は変化するということが分かってきています。DataONE が基礎調査とフォローアップ調査を行い、数年たって研究者の意識がどう変化したのかを調べた有名な論文では、やはり研究データ共有に対してポジティブな見解がだんだん増えていったという結果が出ています（図 21）。ただ、その一方でリスクへの関心も増加しました。これは私なりに解釈すると、最初にいいことばかりを言い過ぎて、現実の方があまり追い付いていないということがあったので、この落差を見て、本当に大丈夫なのかという不安が増加したのではないかと思います。

ですから、何か結果を出そうと焦ってやるよりは、北本さんからもスピードの話がありましたが、（物事



（図 20）



（図 21）

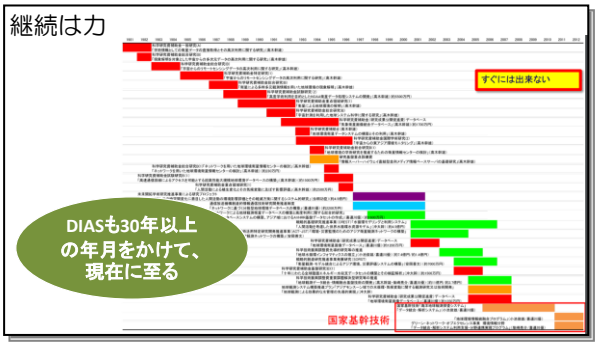
を進めるときには）適切な速度というものがあるので、そこを見極めながら進めていく必要があるのではないかと思います。

この論文は他にも面白いことが書いてあります。やはり年代によって考え方が違って、若い世代は業績を気にする、シニア世代は例えば教授だともう地位は安定しているので業績はあまり気にせずにプロジェクトで予算が取れるかどうかを気にするということで、インセンティブは年代によってもだいぶ違うのではないかということが分かりました。

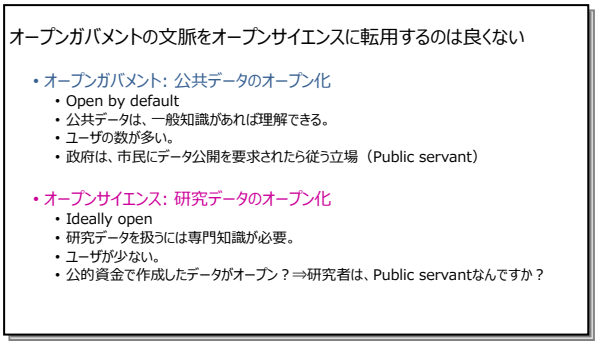
ということで、焦ってやるよりは継続の方が重要です。DIAS もこの名前で活動しているのは 2006 年からですが、30 年以上かけて今のところまで来ています（図 22）。

オープンガバメントとオープンサイエンス

その他に、オープンガバメントの文脈をオープンサイエンスに持ち込んで、例えば政府標準利用規約に従って研究データをオープンにしろと言うポリシーメーカーの方がたまにいるのですが、この二つは全く違う概念だということを説明しておこうと思います（図 23）。



（図 22）



（図 23）

例えば、公共データの人口統計データを考えたときに、これは別に特別な知識がなくても使えます。その一方で、研究データは誰が使うのか。先ほど河川の例を出しましたが、水循環の分布型流出モデルのチューニング化された初期値で使うパラメーターセットを誰が使うのだということです。こういうものを第三者にオープンにすることは、あまり意味がないのではないか、その辺のバランスも考えながら進めなければいけないのではないかと思います。

もう一つ重要な論点として、立場の違いもあると思います。例えば、民主主義の基本原理から言って、市民がデータを公開しろと言えば、政府はパブリックサーバントなので出さなければいけないのです。でも、公的資金で作成したデータがオープンという論法は、国に言うのなら分かるのですが、研究者に言うのは適切なのか、私の方でも違和感が残っています。研究者はパブリックサーバントなのかということです。

これは突き詰めて考えると非常に深いテーマだと思います。やはり市民は間違えることがあるのです。そういうときに、こちらの道が正しいのだと誰が言うのか。その役割を負っているのが研究者ではないか、だから研究者はパブリックサーバントというよりは、ファクトに対するサーバントであるべきだと思っています。いくら国や市民が言ったとしても、それを全て鵜呑みにするのは、民主主義を正常に回すために良くないのではないかということを考えています。

データ・マネジメント・プラン（DMP）

私も海外で実際につくられているデータ・マネジメント・プラン（DMP）はどんなものか見てみたのですが、文章で説明して書いてあるものから、データの細かいスペシフィケーションを書いてあるものまで、まちまちでした（図 24）。これをつくることにどのくらい実効性があるのかという疑問が出てきて、きちんと検証した方がいいのではないかと思いました。しかし、小賀坂さんの報告を聞くと、その辺は慎重に丁寧にされているということなので、安心しました。逆に、

日本のアドバンテージは、いったんルールが決まると丁寧にやることだと思うので、国内で丁寧な作業、仕事をして、日本はうまくいっているという事例を見せられたらいいのではないかと考えています。

私たちはこれから何をすればいいのかをまとめると、すぐに結果を出そうと焦らない、やるべきことは継続する、やること（やったこと）は有効だったかどうか検証することではないかというのが私の考えです。

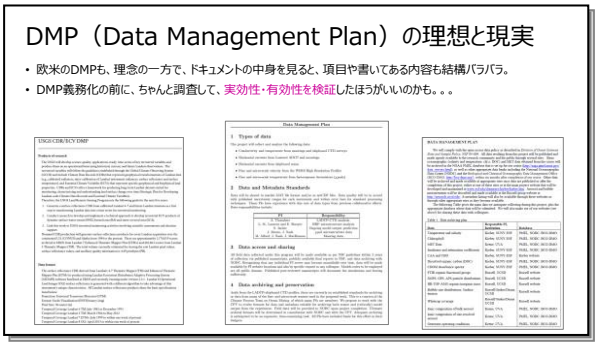
文献とデータの比較

最後に、SPARC Japan セミナーは図書館の参加者が多いイベントで、図書館は、文献管理で得たノウハウが研究データに転用できるのか悩んでいる方が多いのではないかと予想したので、文献の世界とデータの世界を少し比較してみました。

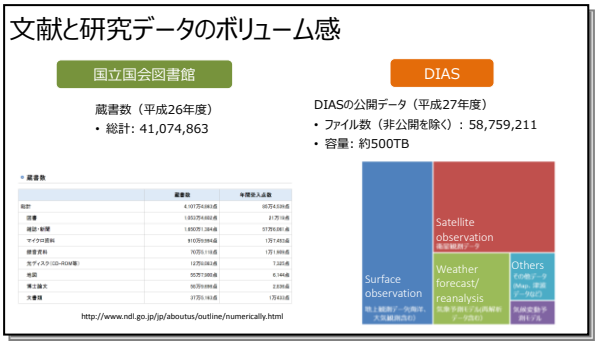
まず、文献と研究データのボリューム感は、国立国会図書館の今の蔵書数が約 4,100 万であるのに対して、DIAS の公開データが昨年度時点で約 5,800 万でした（図 25）。DIAS の本体は非公開領域なので、それを入れると数倍以上に膨らむかもしれませんが、このぐらいのボリューム感です。これだけのボリュームがあると、メタデータをどうやって付ければいいのかが悩ましいところで、適切なメタデータの粒度に悩んでいるところです。皆さんも、国会図書館のメタデータをもう一回ゼロから付け直すと言われたら、かなりうんざりで、それは無理だと思いますよね。そういうデータの世界で悩んでいるのがわれわれだということです。メタデータの粒度問題について本当に悩んでいるので、もしいいアイデアがあれば教えてほしいです。

図 26 は大きさの比較です。世界一大きい本は通称『オバママニア』という本で、厚さが約 34cm だそうです。世界一大きい画像データはアルプス山脈の図で、約 46TB あるらしいです。

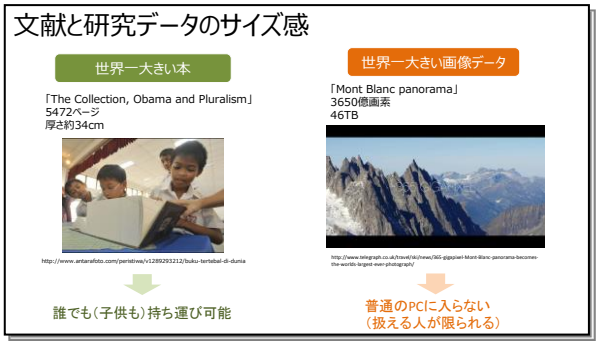
言いたいことは、データの方が大きいとかそういうことではなく、大きい本は子どもでも何とか持ち運べるサイズである、でもデータは移動させるだけでも特別な環境がないとできない、こういう違いがあるということです。この違いは結構大きいのではないかと思っています。



（図 24）



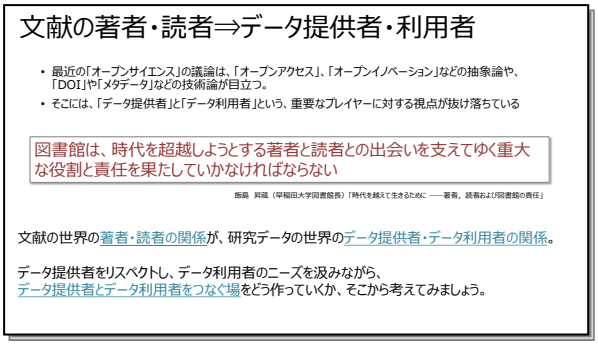
（図 25）



（図 26）

まとめ

最近のオープンサイエンスの議論は抽象論、技術論に引きずられている印象があります（図 27)。そこにはデータ提供者とデータ利用者という重要なプレーヤーに対する視点が抜け落ちていると感じています。データ提供者と利用者は、文献の世界で言う、著者と読者の関係です。その著者と読者を歴史を越えてつなげていくという役割が図書館にあるのではないかと思っています。ですから、データ提供者とデータ利用者をどうつなげていくか、そこのアーキテクチャをどうつくっていくかという視点で今後を考えていっていただければうれしいです。データキュレーター、データライブラリアンの役割もそこにどう関われるかがポイントになるのではないかと考えています。



（図 27）

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

研究データを用いたサービスの調査・企画

A 班：機関内研究情報統合システム（IRIIS）の企画構想
B 班：オープンデータ利活用総合基盤システム「ODEN」の企画

大向　一輝（国立情報学研究所）
田村　峻一（滋賀医科大学附属図書館）
梶原　茂寿（北海道大学附属図書館）

講演要旨

国立情報学研究所では教育研修事業の一環として、様々な機関に所属する図書館員が協力し、学術情報流通に関する諸課題の解決を目指す学術情報システム総合ワークショップを毎年開催している。今年度は 8 名の受講者が「研究データを用いたサービスの調査・企画」をテーマとして約 6 ヶ月に渡る検討を行った。本講演ではワークショップの概要と目的、その効果について述べる。参照：平成 28 年度 カリキュラム及び講義資料/成果物(http://www.nii.ac.jp/hrd/ja/ciws/h28/curritxt.html)

大向　一輝

1977年京都生まれ。2005年総合研究大学院大学博士課程修了。博士（情報学）。2005年国立情報学研究所助手、2007年同助教、2009年同准教授。セマンティックウェブやソーシャルメディア、オープンデータの研究とともに、学術情報サービスCiNiiの開発に携わる。株式会社グルコース取締役。著書に『ウェブがわかる本』（岩波書店）、『ウェブらしさを考える本』（丸善出版）がある。

A 班：機関内研究情報統合システム（IRIIS）の企画構想

当班では、研究データを含めた研究資源や研究に関する情報の管理を、機関単位で集約的に行うことを想定し、「機関内研究情報統合システム（略称 IRIIS）」を企画構想した。企画にあたっては、機関内の情報を網羅的に収集できること、および簡便に業務が行えることを重視した。今後このような仕組みが標準化し普及することで、より多くの機関が研究データ管理に取り組みやすくなることが望ましい。

田村　峻一

2013年、滋賀医科大学に図書系職員として採用。以来、図書受入業務及び機関リポジトリ業務を担当している。日本医学図書館協会認定資格ヘルスサイエンス情報専門員（基礎）。平成28年度学術情報システム総合ワークショップ受講者。

B 班：オープンデータ利活用総合基盤システム「ODEN」の企画

B 班では、散在するオープンデータのメタデータを一元的に集約し、ワンストップでの検索を可能にするオープンデータのメタデータデータベース「オープンデータ利活用総合基盤システム － ODEN」の構築を企画した。本講演では、そのコンセプトと概要について紹介する。

梶原　茂寿

2014年より、北海道大学附属図書館本館に勤務。学術システム課システム管理担当で、図書館情報システム及びネットワークの管理と機関リポジトリを担当。2015年、2016年SPARC Japanセミナー企画WGメンバー。

研究データを用いたサービスの調査・企画

大向　一輝（国立情報学研究所）

私は、日本の論文を検索する CiNii や、大学図書館の蔵書データを検索するサービス、博士論文を検索するサービスといった、国立情報学研究所（NII）の情報サービスの企画設計を担当しています。

研究者としては、オープンデータの話でも時々出てきますが、リンクト・オープン・データと呼ばれるようなデータフォーマット、データの表現方法に関する研究をしています。でも今日はその話は一切出てきません。

NII の教育研修事業

NII は多面的な組織で、情報学に関する研究の場所であり、大学共同利用機関として、大学のインフラを提供しています。そのインフラの中にネットワークサービスと学術情報サービスがあり、さらに学術情報サービスの中にはオンラインサービスの提供と人材育成もミッションとして入っているので、今日はそこの部分についてのお話になります。

NII の教育研修事業では、大学図書館の方々を対象として提供しているプログラムがたくさんあります。これまでの成果を図 1 のようにサイトに載せています。

現在行っている研修のメニューが図 2 に並んでいます。一番上は講習会で、システムの使い方をレクチャーしたりするのですが、下に行くほど専門性が高まって、専門研修では、例えば目録の取り方の一番難しいところについて研修したり、リポジトリ事業についてもシステムの使い方を覚えていただかなければいけないので、その研修を行います。

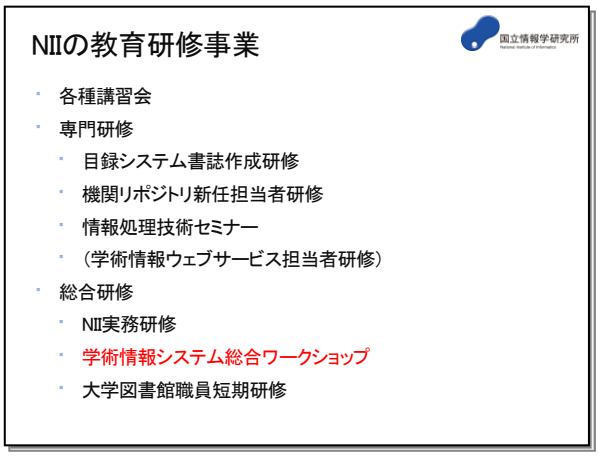
さらにレベルが上がっていくと、最終的には三つの研修（NII 実務研修、学術情報システム総合ワークショップ、大学図書館職員短期研修）があります。NII 実務研修は、NII に半年や 1 年来て、机を並べて普通に仕事をしていただくものです。完全に OJT で、何が研修なのかよく分からないのですが、そこで全国の大学図書館に共通する課題について取り組んでいただきます。大学図書館職員短期研修は、数日間にわたってレクチャーを受けて学ぶものです。

この後、学術情報システム総合ワークショップの成果発表が 2 件あるので、私はその前座として、この後の発表はどういうコンテキストの中でつくられてきたものかについてお話しします。

学術情報システム総合 WS とは

大学図書館、学術情報流通をめぐる問題は日々刻々と変わっており、その一つがオープンサイエンスになると思います（図 3）。もはや特定の部署で頑張ったら解決する、組織で頑張ったら解決するという問題は、皆さんの能力が高いのでとっくに終わっていて、残っ

（図 1）



（図 2）

ているのは横断的に取り組まないといけないものばかり、だからこそ非常に難しく感じるのです。そういう問題をみんなで連携・協力しながら解ける力を身に付けてもらうことがこの研修の目的です。この研修では、全くレクチャーは行われません。約 6 カ月間、ただ放っておかれて、いつまでに報告書をつくって持ってくるように、と指示される、極悪な研修スタイルを取っているのです。3 回ぐらいは実際に集まるのですが、その間は基本的にはオンラインでやりとりします。

この研修は 4 年目を迎え、今年も 8 名の方に、どういう研修をするかを知りながら来ていただいているはずなので、非常にありがたいことです。

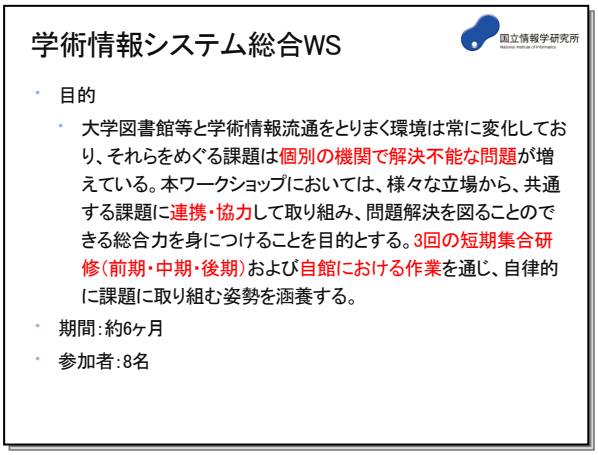
6 月に 1 回初顔合わせをして、次に 2 カ月後に会いましょうと解散して、また 2 カ月ぐらいたって集まって、最後に報告書を書くというスケジュールです（図 4)。日本中の図書館員の方なので、その間は基本的にはじかに会うことはほとんどなく、オンラインのツールを使ってやりとりをします。ビデオチャットを使う場合もあれば、われわれが提供しているタスク管理システムの場合もあります。メールで連絡を取ると、議論発散というか、日常の仕事と混ざってしまうので、ツールを使ってきちんと管理しながらやってもらえるようにすることだけは支援します。こういうスタイルで、今年も半年間無事に最終成果まで持ってきていただきました。

テーマは、毎年ころころ変えています（図 5)。初回は、デジタルアーカイブ的な、どこにどんなデジタルリソースがあって、それをどうやって活用するべきかを調査していただきました。

次の年は、CiNii を中心にログを実際に見て、そのユーザーがどんな振る舞いをしていて、その振る舞いに基づいてどんなサービスをさらに提供するべきかを考えていただきました。

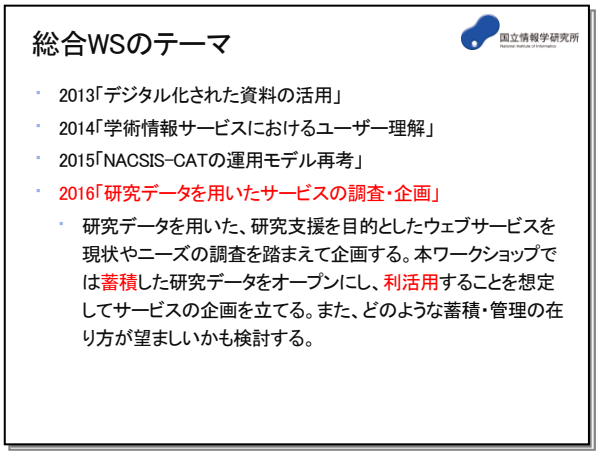
去年は、今もホットな話題ではあるのですが、総合目録システムをどのように新しく変えていくべきかということを、データを探りながら考えていただきました。

今年は、オープンサイエンスに係る「研究データを用いたサービスの調査・企画」というテーマにしました。研究データのオープン化、オープンサイエンスは、理念が分かると非常に大事なような気がするし、制度がどんどんつくられていて、それに沿って人々が行動する、そういうところの議論は非常に先行していると



（図 3）

（図 4）



（図 5）

感じます。では実際に、研究者、あるいは図書館員など学術情報を支える人たちはどうなるかというと、情報サービスを普段使用する中で、その理念や制度に従うことになっていくはずです。北本先生の報告では Lessig の四つの力に触れられましたが、規範、法、市場などがあるけれど、やはりアーキテクチャ、何かをせざるを得なくなる、何かを仕向けるものが、実際の生活の中で一番関わるところであって、そこのシステムをもう少し深掘りして考えないと、実際にオープンサイエンスをするということが何なのか理解できないのではないかという個人的な疑問もありました。

それで、そこにフォーカスして、別に制度を決めてもらうのではなく、情報サービスを企画すること、サービスの姿を描き表すことをテーマにしました。

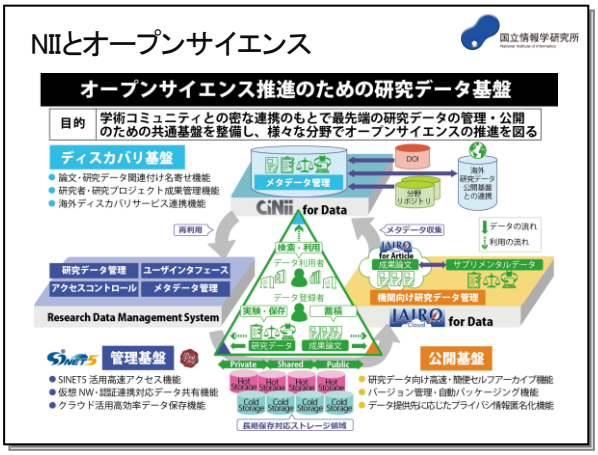
これは架空の設定ではなく、NII の中でもこれからオープンサイエンスはかなり強く打ち出していくことになっています。図 6 は古いもので、単語はアップデートされたものがこれから出てくると思うのですが、NII の中で「三角の図」と呼ばれているものです。

NII が提供する情報サービスには三つの要素が必要であると考えられています。一つ目は情報管理で、研究者が普段、データをシェアリングしながら研究することをサポートするためのものです。二つ目は、何かの成果を公開して世の人に使ってもらうための基盤です。三つ目は、そういうものをもっと集めて検索可能にし、発見可能にする基盤です。この三つがそろって情報サービスによる支援だということでやろうとしていて、これに来年度からかなり力を入れていくので、これをやるための研究員やスタッフを今まさに募集しています。興味がある方はアプライしていただけると大変うれしいです。

学術情報システム総合 WS のより大きなテーマ

こういうこともあって、情報サービスの中身をつくっていただくということを研修プログラムにしました。そこで、ある考え方を受講者の方にお願いしています。それは、現状について調査・分析することは大切ですが、現状に引きずられてはいけない、オープンサイエンスは試行錯誤の真っ最中であって、積み上げベースで次のサービスを考えるというよりは、図書館としてはこういうふうに関わりたいということを明確にしながらサービスを描き出していくべきであろうということです。図 7 は、研修にも参加していただいた岡本真さんが描いた図を引きました。まずビジョンを立てた上で、その中でわれわれがやるべき仕事、それを実現する手段を書き下していくことで、システムの仕様にしていってくださいということを言いました。

また、単に理念ベースで書き下すだけではなく、仕様書のスタイルで書き起こしてください、画面の設計はきちんとやってくださいということを言いました。やりたいこと、願望だけを書くのではなく、要る機能、要らない機能をきちんと選別してものにしていくこと、願望をたくさん書いて、それに合わせて絵を描くと大変な量になって、そもそもマンパワーとして無理なの



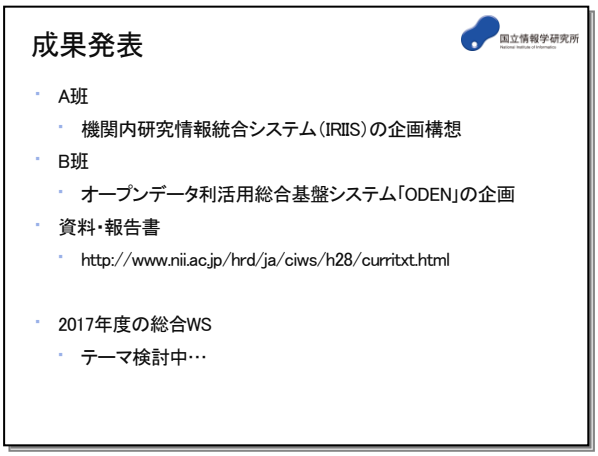
（図 6）

（図 7）

で、ある種の制限をかけながら考えることを指示しました。

これから二つのチームの発表がありますが、それぞれ全く出自、バックグラウンドの異なる人たちで考えたことです。コラボレーションに慣れていくことが、今後の図書館による問題解決にとって非常に重要な基盤になるのではないかと思って実施しました。

まだ来年度もこのスタイルは続けるのですが、テーマは検討中で、もう一回オープンサイエンスを扱うのか、違うテーマを扱うのか分かりませんが、ぜひ興味がありましたらご参加ください。資料・報告書等は全て図 8 に載っています。



（図 8）

A 班：機関内研究情報統合システム（IRIIS）の企画構想

田村　峻一（滋賀医科大学附属図書館）

ワークショップの A 班のディスカッションの内容について、発表させていただきます。

1.班員紹介とワークショップの流れ

まず、私も含めて A 班の班員を紹介します（図 1）。田村、野間口、堀、柏倉という 4 名が A 班に分かれて作業を行いました。私と野間口さんは、滋賀医科大学と京都大学の図書館員として勤務しています。滋賀医科大学は医科系の単科大学で、京都大学は総合大学ということで、少し毛色の違う大学になっています。残りのお二人は、堀さんは自然科学研究機構国立天文台研究力強化戦略室で、URA の研究評価の仕事をしています。柏倉さんは海洋研究開発機構（JAMSTEC）の地球情報基盤センターで、地球シミュレータ研究成果リポジトリというリポジトリの管理業務に当たっています。このように、A 班はバックグラウンドの異なる 4 人が集まったので、大変だったところもあるのですが、お互いに自分とは違う世界のことを知れて、大いに勉強になりました。

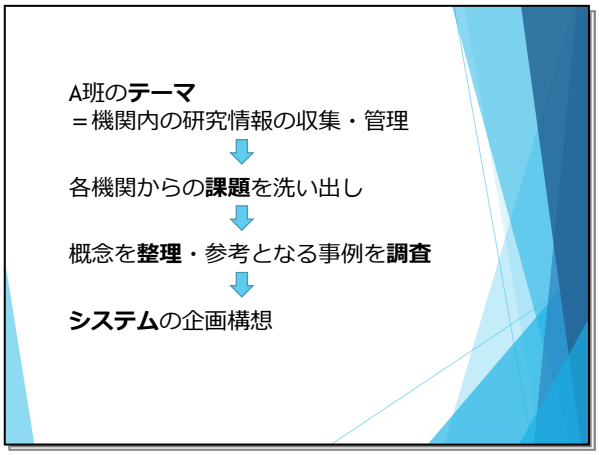
なぜこの 4 人で集まったかというと、自分の機関内の研究情報を収集し、管理する仕事をしていることが共通していたからです（図 2）。よって、この共通テーマからディスカッションをスタートさせました。各機関で感じている課題を洗い出し、その後、概念を整理して、参考となる事例を調査、最終的にシステムの企画構想という流れでした。各段階について説明していきます。

課題の洗い出しでは、各機関で感じている課題を持ち寄りました（図 3）。いろいろありましたが、大きく集約化と簡便化という二つのベクトルに整理しました。

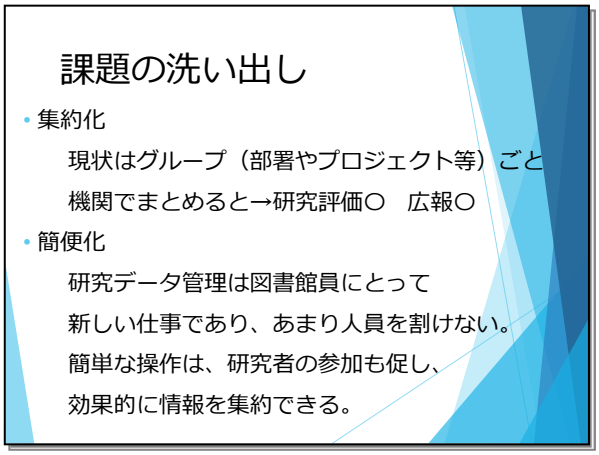
集約化については、現状だと、研究グループごとに研究成果の出どころがばらばらになっていて非常に探しづらい、そのため、これを機関でまとめれば、研究評価や分析に当たって便利であり、外部に対する広報としても良いということです。

簡便化については、研究データ管理は図書館員にと

（図 1）



（図 2）



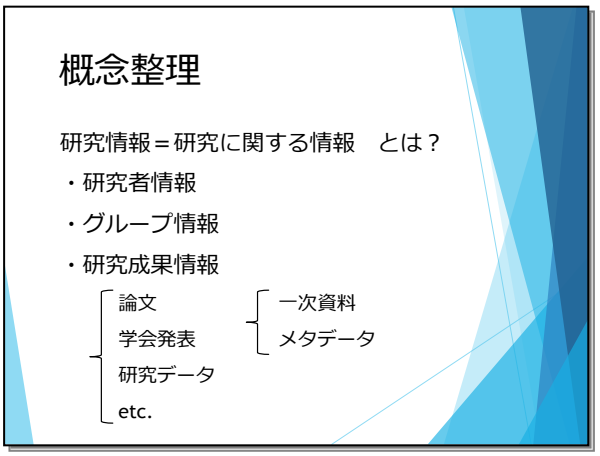
（図 3）

って新しい仕事で、それを導入するに当たっては、簡単な操作で行えることを重視したいと考えました。職員も入力・作業をしますが、研究者にも作業に参加してもらって、みんなが使いやすい、みんなで使っていけるようなシステムをつくった方がいいという話になりました。

「研究情報」という言葉を先ほどから使っていますが、これは研究に関する情報全般を指しています（図 4）。これを細かく分析すると、研究者に関する情報、研究のグループ（研究者の集まり）に関する情報、研究成果に関する情報に分かれます。研究成果情報はさらに、論文や学会発表や研究データなどに細かく分かれます。研究成果情報は、一次資料（論文の PDF ファイルそのもの、研究データそのもの等）とメタデータから成るという別の観点からの分類もあります。

「データ」という言葉、「情報」という言葉を安易に使うと、かなり議論が錯綜してしまったので、このようにまとめて考えました。

図 5 は参考事例調査の一つの大きな例です。OAIS 参照モデルという、研究成果が生産者によって生産され、それを保存計画に基づいて運用し、消費、アクセスに至るまで管理して計画していくというモデルを見て、この四角形に当たるようなシステムをつくろうかということなりました。

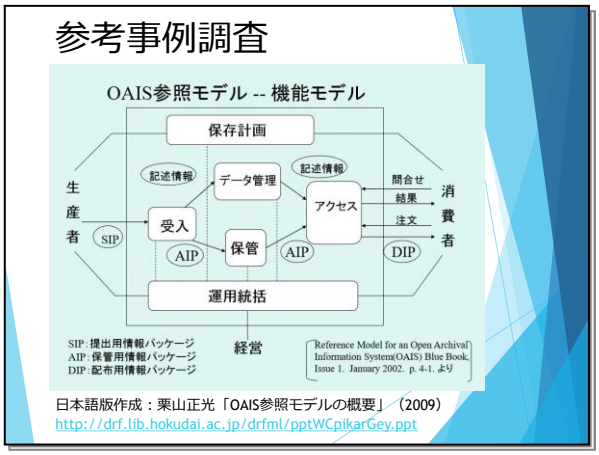


（図 4）

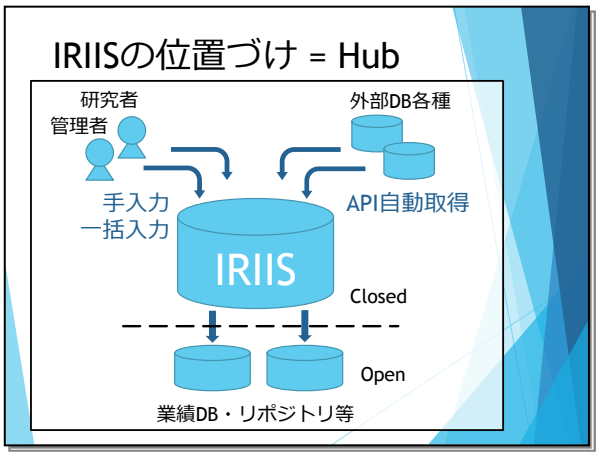
2.IRIIS の概要

そしてできたシステムが IRIIS（Institutional Research Information Integrated System：機関内研究情報統合システム）です。最初は I がもう一つ少なくて IRIS と呼んでいたのですが、某図書館システムと同じ発音になってしまうので、Integrated の I を 1 個足して、統合するシステムだとしました。われわれが実務を担当している大学や研究機関など、一つの機関内で完結して使用するシステムを想定しています。

IRIIS がどういう構成になっているか紹介していきます。IRIIS はハブとしての位置付けを持っており、複数のソースから情報を取ってきて、いったん蓄積します（図 6）。ユーザーは研究者と管理者の 2 種類を設定しています。研究者は研究者で、管理者というのは、図書館員や URA のような事務職員を設定しています。



（図 5）



（図 6）

外部のデータベースからも自分の機関の研究に関する情報と思われるものを自動的に引っ張ってきて、それをプールします。プールされた情報は、クローズドのまま保持するものもありますが、オープンにできるものはオープンに切り替えて、それをまた別のサーバー、リポジトリにはき出していくというシステムです。

3.IRIIS の四つのテーブル

IRIIS の中身は主に四つのテーブルから成っています（図 7）。先のスライド（図 4）で研究情報を分類した名称を使っていますが、それぞれ研究者情報テーブル、グループ情報テーブル、研究成果情報テーブル、さらにデータ保存スペースという機能を付けて、大きく四つのエンティティから成るシステムを構築しました。以下、それぞれについて画面イメージと業務フローの概念図を示して説明していきます。

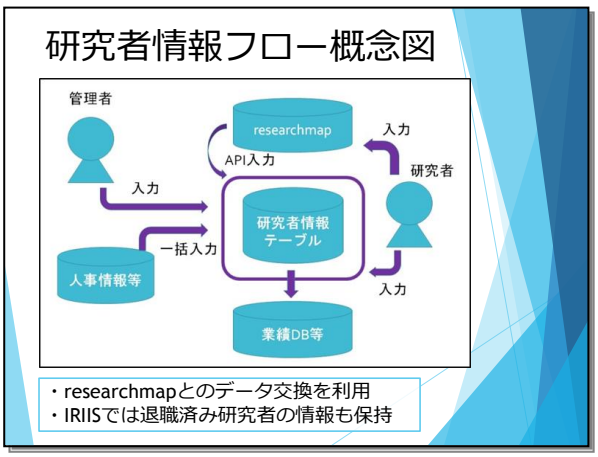
（図 7）

（図 8）

3-1.研究者情報テーブル

図 8 のような画面のイメージ図をつくりました。研究者に関する情報が表示されています。タブで研究成果の画面にも行きますし、後で述べる個人スペースの画面にも行くことができます。

研究者情報は researchmap とのデータ交換が自動的に行われます（図 9）。これはもう滋賀医科大学等でも既にやっていることなのでイメージはできると思います。人事情報など、機関内の他のデータベースで持っているような情報も、TSV や CSV ファイルなどを使って一括で取り込めるようにすると楽になるのではないかと考えました。人事の情報はカレントで存在している研究者の情報ですが、退職済みの研究者についても学術情報としては保持しなければならないので、退職済みフラグのようなものを使って情報を保持するというイメージです。



（図 9）

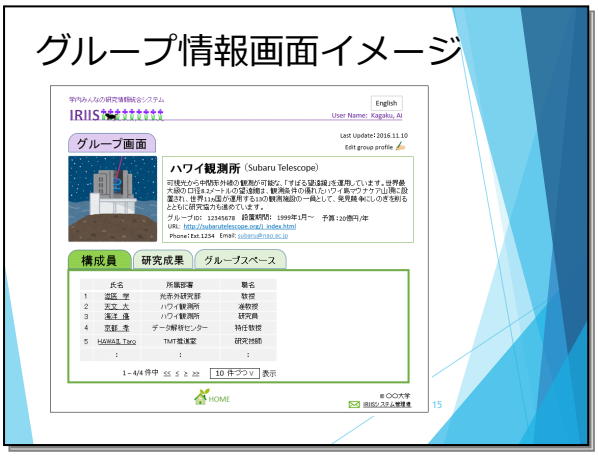
3-2.グループ情報テーブル

グループ情報画面も同じような画面になっています（図 10）。グループ独自の情報として、誰が所属しているのかということと、設置期間があります。グループの情報を独自に研究者とは別に持つことで、グループ単位での研究成果の集約に有効活用できるのではないかと考えました（図 11）。

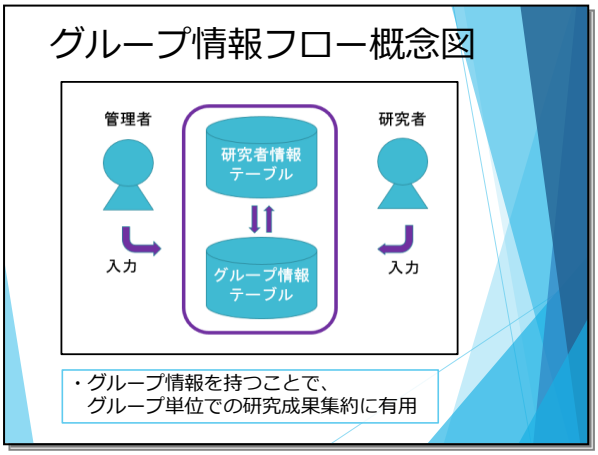
3-3.研究成果情報テーブル

研究成果情報については、API 等で取り込んできたデータは、そのまま取り込みデータとして持って、これは確かに私のデータである、この機関のデータであると承認されたものが次の段階に行きます（図 12）。そのうち公開できるデータには公開フラグのようなものを立てて、という 3 段階のレイヤーにデータを分けています。

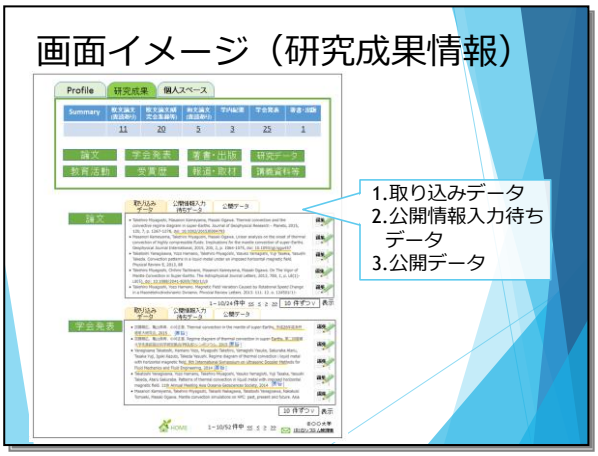
図 13 は研究成果情報のフロー図です。外部から取ってきて、管理者や研究者が内容を成形して、アタッチファイルを付けられるようなものは付けて、それをためて、機関リポジトリ等のサーバーに移せるようにしておきます。研究データについても文献と似たような感じで、DataCite を例として考えていますが、きちんとこれができるかどうかは、今後見ていかなければいけない課題だと思っています。



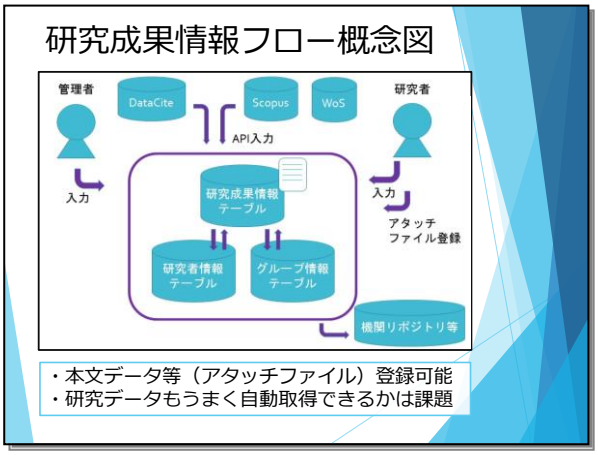
（図 10）



（図 11）



（図 12）



（図 13）

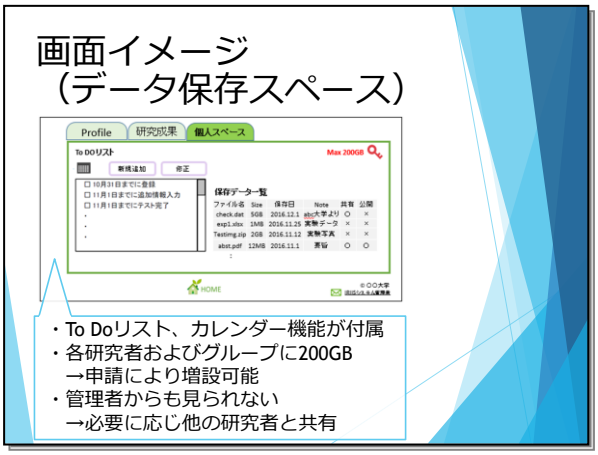
3-4.データ保存スペース

研究者が研究に使うファイルをアップデートして保存しておけるデータ保存スペースもつくりました。それに付属的に To Do リストやカレンダーの機能も付けて、Dropbox、Backlog のようなプロジェクト管理ツールを参考にしたような設計になっています（図 14)。

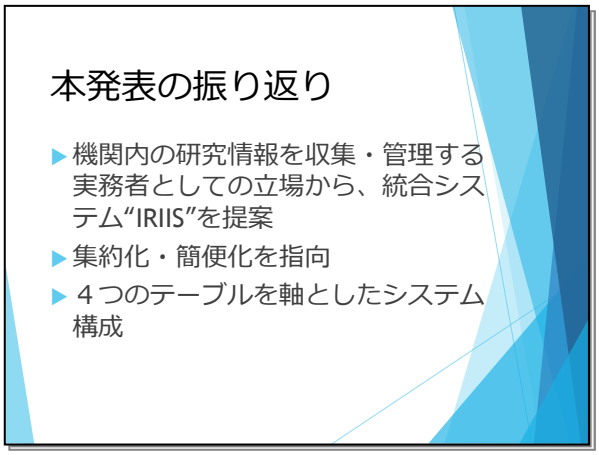
研究中のデータは保存スペースに保存されますが、研究期間が終わったら、これは公開できるようになったという通知が管理者に行って、管理者がメタデータを入力して、同じようにデータリポジトリのような別のサーバーに移動できるシステムを考えました（図 15)。

図 16 は本発表の振り返りです。今後の展開としては、本システムの IRIIS は仮想の大学に導入するという設定で構想したので、抽象的なモデルのようになっていますが、実際に各機関で導入するとなると、機関ごとの環境や予算状況に合わせて、具体的に協議することが必要だと考えられます（図 17)。連携する外部のシステムも、これから変化していくものが多くあると思うので、それに合わせてつくり替えていくことが必要だと思います。

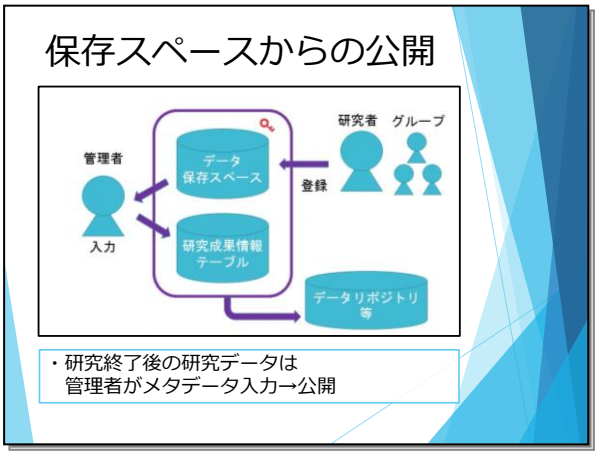
今後、この IRIIS のようなモデルが標準化し、広く普及することで、多くの機関が研究データ管理に取り組みやすくなる環境が整うことが望ましいと考えています。



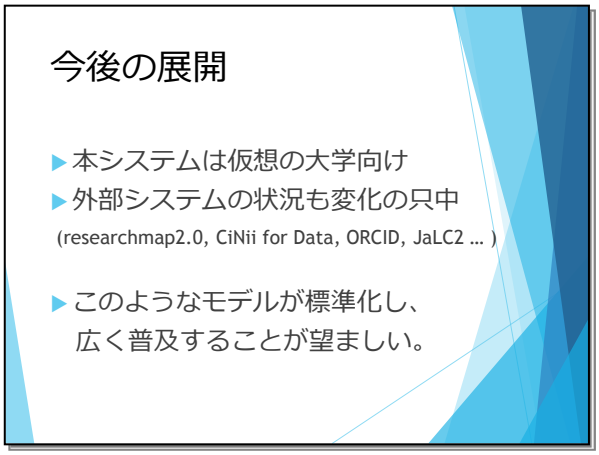
（図 14）



（図 16）



（図 15）



（図 17）

B 班：オープンデータ利活用総合基盤システム「ODEN」の企画

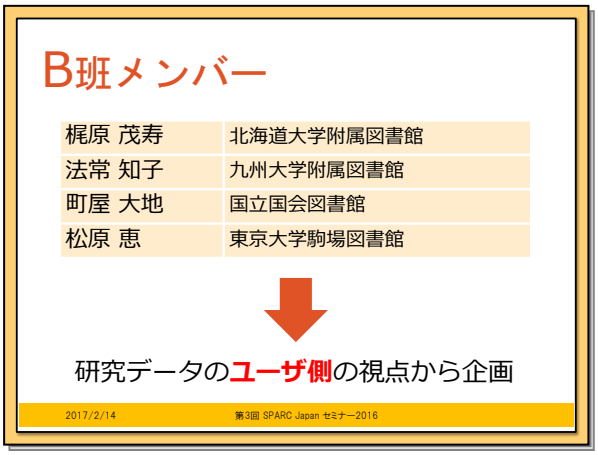
梶原　茂寿（北海道大学附属図書館）

B 班は、オープンデータ利活用総合基盤システム「ODEN」の企画について説明します。

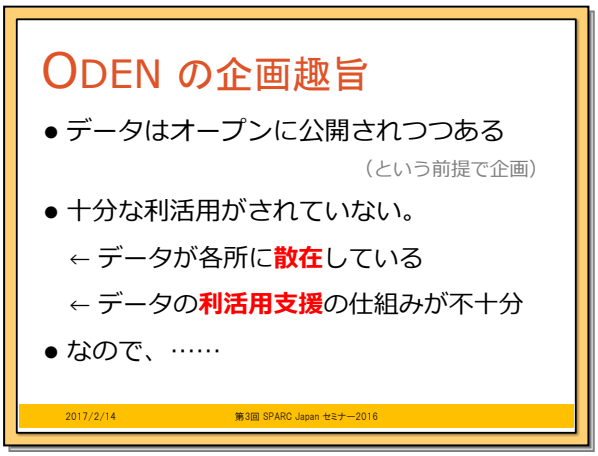
ODEN の企画・開発

B 班のメンバーは、北海道大学の梶原と、九州大学の法常さん、国立国会図書館の町屋さん、東京大学駒場図書館の松原さんの 4 人です（図 1）。われわれの班は A 班とは違って、研究データのユーザー側からの視点で、どのようにオープンデータを使っていくかを考えた企画となりました。

この企画は、データはオープンに公開されつつあるということが前提です（図 2）。データは既にいろいろなドメインやコミュニティで持っていて、オープンにされているものもあるけれども散在していて、その利活用を支援する仕組みがまだ不十分な状況があることから、オープンデータ利活用総合基盤システム（Open Data Enhancing Network）、頭文字を取って、愛称「ODEN」の開発を企画することにしました。



（図 1）



（図 2）

基本コンセプトは、「散在するオープンデータのメタデータを一元的に集約し、研究データのワンストップで検索を可能にするためのメタデータデータベース」です。利用を促進するための機能として、Amazon の「このデータを使った人はこんなデータも使っています」というレコメンド機能や、ライトユーザーに向けたゲーミフィケーション要素を導入しています。また、どのようなレベルの利用者であっても利活用しやすいように、ユーザーが交流・情報交換できるフォーラムや、ユーザーによる自発的な情報の蓄積を可能とするツール&マニュアル機能も準備しています。

ワークショップの最初の集まりで示された説明資料の中に、「サービスの企画をする際には、VMSO（Vision、Mission、Strategy、Objectives）を定めよう」ということがありました（図 3）。われわれもそれに沿って今回の企画をしていきました。われわれが考える ODEN のビジョンは、「DATA for Your HAPPINESS 研究データを解き放ち人々を自由にする！！」です。

そして、そのビジョンを実現するために、「あらゆ

平成28年度学術情報システム総合ワークショップ
岡本真「サービスの企画」資料より (CC BY 2.1 JP)
http://www.nii.ac.jp/hrd/ja/ciws/h28/1-02.pdf

（図 3）

るオープンデータが見つけやすく、使いやすい、場［コミュニティ］を実現する。そのことにより、全社会の研究・イノベーションを加速させる」というミッションを掲げています。

システムのイメージ

図 4 は ODEN のシステムのイメージです。単純に発見の可能性を上げる検索のための部分と、見つけたオープンデータを使っていただくための部分の大きく二つに分かれます。

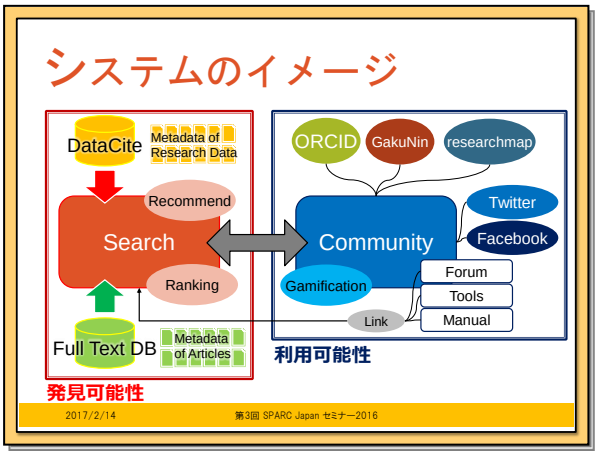
図 5 は画面遷移イメージです。検索エンジン、ODEN のトップ画面、基本検索画面、詳細検索画面があります。そこからこちらで収束しているメタデータにたどり着いて、そのメタデータは基本的に DataCite のスキーマを利用して連携しているという形です。そこからデータのユーザーのところに飛んだり、ユーザープロフィールなどもあるのですが、そこでレベルやメダルといったゲーミフィケーション要素の管理もできます。また、フォーラムやツールマニュアルなども用意しています。

次に、発見可能性のための機能について詳しくご説明します（図 6）。検索機能を充実させており、研究データのメタデータを対象とした検索が可能、引用論文の抄録・キーワードからの検索も可能であることがコアな機能です。

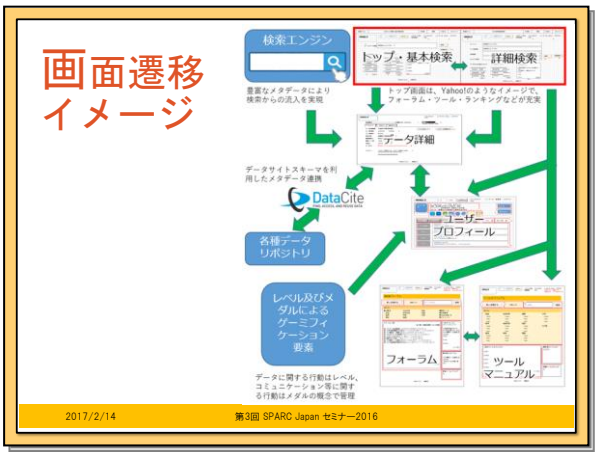
また、メタデータがどれだけアクセスされたかというアクセス回数順、ユーザーによるメタデータの評価順のランキングを示す機能があります。さらに、「このデータを使った人はこちらのデータも使っています」というレコメンド機能や、マイページで今までに使ったデータの照会もできる機能を考えました。

利用可能性を高めるための機能としては、フォーラム機能、ツール＆マニュアル機能、ゲーミフィケーションがあります（図 7）。フォーラム機能では、ユーザーが相互に「こんなデータを探しているのだけれど、何かいいデータはないですか」という情報交換や交流をすることで、自助的な能力育成につながると考えています。DIAS などが既にこういうものを実現していると思います。

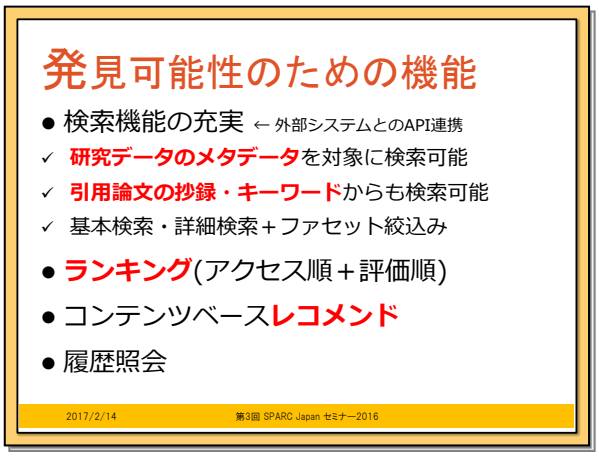
ツール＆マニュアル機能では、ユーザーがナレッジを蓄積し、ある一定以上のレベルにあるユーザーや研究者が自分でツールやマニュアルをつくって、ここに提供することができるようになっています。



（図 4）



（図 5）



（図 6）

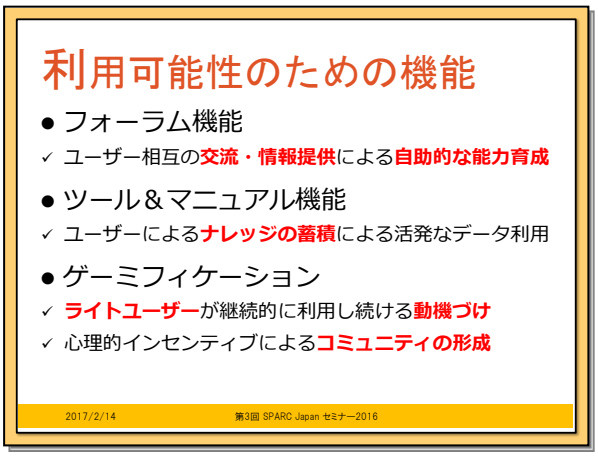
ゲーミフィケーションについては二つの機能があって、一つはレベルです。例えばデータにアクセスした回数など、データに関する行動を起こすとレベルが少しずつ上がっていくようなことを考えています。もう一つはメダルです。メダルは ODEN の具になっていて、フォーラムに 3 回参加すると大根がもらえる、といった具合になっています。

まとめ・今後の展望

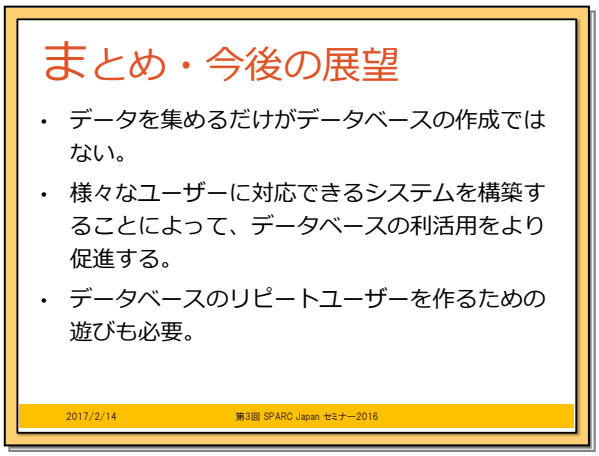
データを集めるだけがデータベースの作成ではありません（図 8）。研究者やデータを扱っている人だけに対応しているようなシステムでは、オープンアクセスの促進につながらないのではないかということから、さまざまなユーザーに対応できるようなものを考えていきました。そして、データベースのリピートユーザーをつくることも大事なので、ライトな遊びの部分も必要だと考えて、ゲーミフィケーションを取り入れました。

あらゆるオープンデータが見つけやすく、使いやすい場［コミュニティ］を実現する。そのことにより、全社会の研究・イノベーションを加速させるということを考えながら今回の企画を行ったということです。

ワークショップ成果物に仕様書と報告書が載っているので、A 班のものも、B 班のものも、ぜひそちらを見ていただければと思います。



（図 7）



（図 8）

CC BY

●**大向**　半年かけて、それぞれ 4 名の班で、仕様書をつくって提案をしていただきました。私は中身にコントロールはかけないようにやってきました。サービス全体の姿もそうなのですが、その所々に表れる細かいディテールが重要だと考えています。

例えば、そもそも研究データはメタデータを付けて検索できたらそれでいいのかということに対しては、どちらの提案からも鋭く突っ込みが入っています。私自身は、例えば研究の途中でデータが欲しくなったら、データ検索エンジンにアクセスして、そこにキーワードを入れて、リストを見て、クリックするというような振る舞いが本当にあるのか、非常に疑問に思うわけです。欲しいデータがあれば、欲しいと分かっているものであれば、そうやって探すかもしれませんが、そうではないものをデータ検索エンジンで探すということは、今のところ自分の研究プロセスの中にはありません。それよりも、例えば論文のデータ、プロジェクトの何か、日本であれば科研費のデータベース、プロジェクトベースのデータベースを見て、そこで使われたデータということで初めてアクセスする可能性があります。

そういうことを考えると、これまでの図書館的に、データに対してきちんとリッチなメタデータを付けて、それをデータベースに入れれば、オープンサイエンスの支援ができたとは到底言えないというのがこの二つの班の提案から見ても非常に思うところです。

どちらの報告にも、かなり人ベースということが出てきていました。自分のソーシャルメディアのプロフィール画面のようなものがあって、それにひも付いた形でデータが載っている、そういうことが割と自然に出てきていることを考えてみても、やはりメタデータでは済まない、図書館なら図書館で、人のネットワークのファシリテーションまで手を出さないと、研究データアクセスに対する支援になっていないという可能性が見えてくると思います。

これからわれわれとしても、何かしらの形でそのシステムのプロトタイピングを進めていかなければいけません。直接的なメタデータアプローチは、ゼロでは全くサービスにはならないですが、そういうエッセンスをどう取り込んでいくか。そして、まだ顕在化していないニーズ、情報要求がどこにあるのかということを、架空の設定ではあるけれども考えていただくことによって、少し浮き上がった部分があるのではないかと考えました。

第 3 回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

RDM トレーニングツールの紹介

尾城　孝一

（東京大学附属図書館）

CC BY NC SA

講演要旨

世界各国で公的研究資金による研究成果について、論文だけではなくエビデンスとなる研究データの保存や共有を求めるポリシーを制定する例が相次いでいる。適切に研究データを管理するには、研究者自身に加えて研究支援者が研究データ管理に係る知識やスキルを習得する必要がある。本発表では、機関リポジトリ推進委員会の研究データタスクフォースで開発を進めている、研究データ管理に関する研究支援者向け学習教材について紹介する。

尾城　孝一

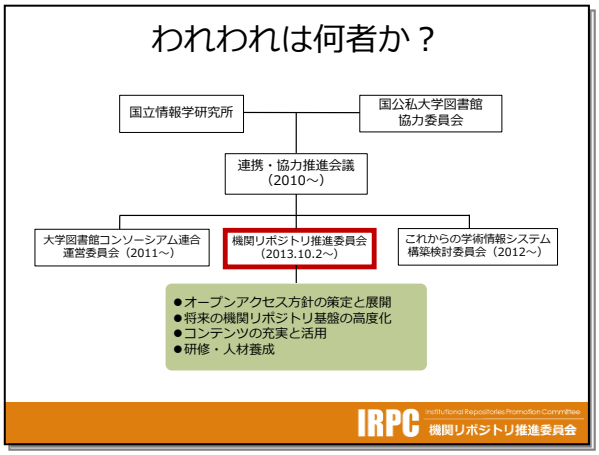
東京大学附属図書館事務部長。1983年1月、名古屋大学附属図書館に採用され図書館職員としてのキャリアを開始。その後、東京工業大学附属図書館、国立国会図書館、千葉大学附属図書館、国立情報学研究所、東京大学附属図書館、大学図書館コンソーシアム連合事務局を歴任。2015年4月より、現職。

今日は、機関リポジトリ推進委員会の研究データタスクフォースが開発した、RDM（Research Data Management）トレーニングツールについて紹介させていただきます。

機関リポジトリ推進委員会の取り組み

国立情報学研究所（NII）と大学図書館の連携・協力の枠組みの中に、機関リポジトリ推進委員会という委員会が置かれており、そこで機関リポジトリ、オープンアクセス、オープンサイエンスに関するさまざまな活動をしています（図 1）。

機関リポジトリ推進委員会には、基幹的な業務のサポートをするための三つの常置の作業部会があります（図 2）。その他、先端的なプロジェクトに取り組むためのアドホックなタスクフォースが六つ設置されており、われわれはその中で研究データタスクフォースとして活動しています。これから紹介する RDM トレーニングツールの開発や、ケーススタディを通じて研究データ管理のノウハウを蓄積していくような活動をしています。



（図 1）

図 3 がそのタスクフォースのメンバーです。私の他に、NII の山地先生、大学や研究所の図書館員の皆さん、URA の皆さんにも入ってもらい、一緒に活動しています。数は少ないのですが、これは今考えられる最強の精鋭部隊だと自負しています。

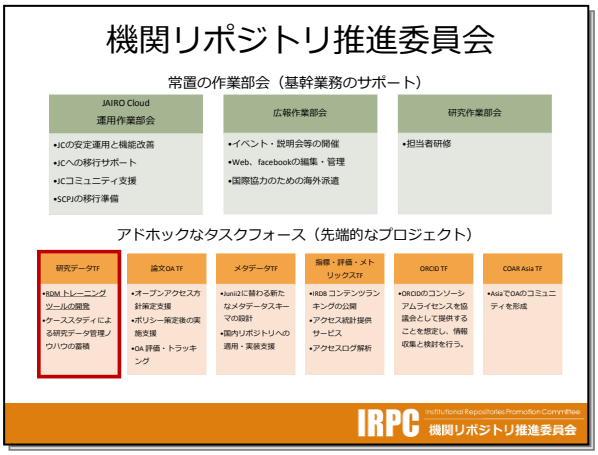
さて、オープンサイエンスなのですが（図 4）、研究データに対して研究者の皆さんはどういう意識を持っているのでしょうか。DataONE というプロジェクトに関連する調査があり、図 5 はその最初の方の調査結果なのですが、1,329 名の研究者のうち、78%が、少なくともデータの一部はアクセス制限のないデータリポジトリに登録したいと考えている、しかし、自分自身のデータに他の研究者が容易にアクセスできていると回答した回答者は 36.2%にとどまっています。データを電子的に利用できるようにできない理由としては、やはり時間がない、お金がない、データの登録先がないといったことが挙げられています。

この調査から、データを公開したい、あるいは研究資金配分機関等のポリシーによって公開しなければならない、けれども実現できないでいる研究者が結構たくさんいるということがうかがえるかと思います。

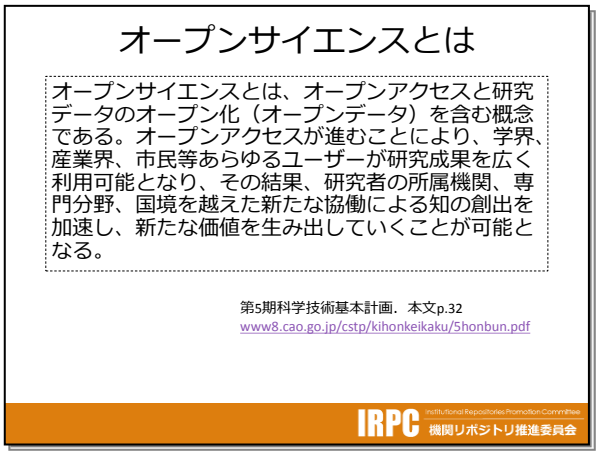
こういう研究者を支援し、オープンサイエンスを進めていくためには三つの基盤が必要だと考えています。一つ目は研究データ基盤（共通システム）、二つ目は研究者を支援する人の基盤（支援スタッフの確保・育成）、三つ目は組織の基盤（コミュニティ）です。今日はこのうち二つ目の支援スタッフの確保・育成につながるトレーニングツールの紹介をします。

RDM トレーニングツールとは

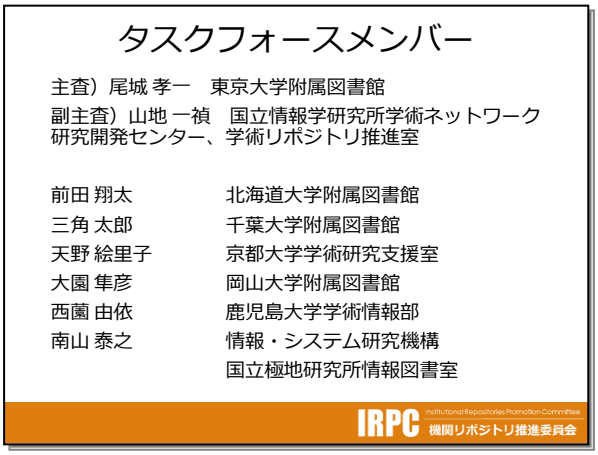
文部科学省が 2016 年 2 月に公開した審議のまとめの中でも、研究データのオープン化に関して、大学等



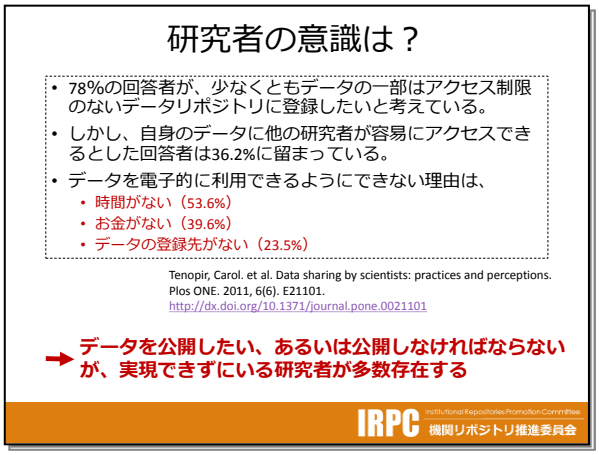
（図 2）



（図 4）



（図 3）



（図 5）

に期待されている役割が幾つか挙げられているのですが、その一つとして、技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資するということが書かれています（図 6）。

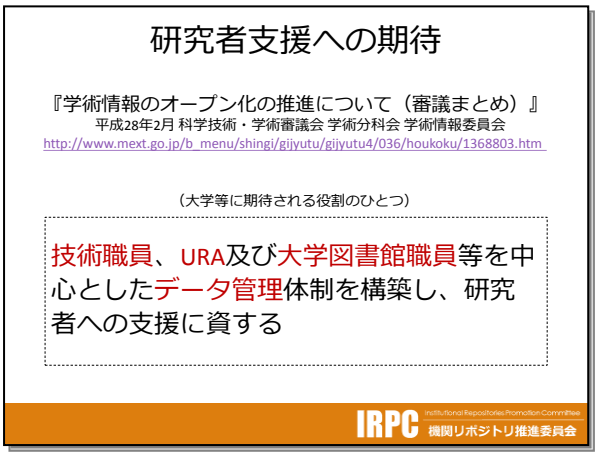
しかし、研究データ管理の支援をやれと言われても、一体何をやっていいのかよく分かりません。これは海外でも事情はどうも同じようで、どういうことをやればいいのか、それが学べるオンラインのツールがたくさんできています（図 7）。その日本版の教材をつくろうというのが、われわれが RDM トレーニングツールの開発に取り掛かったきっかけです。

ツールの目的は、学習者が RDM に関する基礎的な知識を得ること、研究者支援サービスの足掛かりを得ることです。

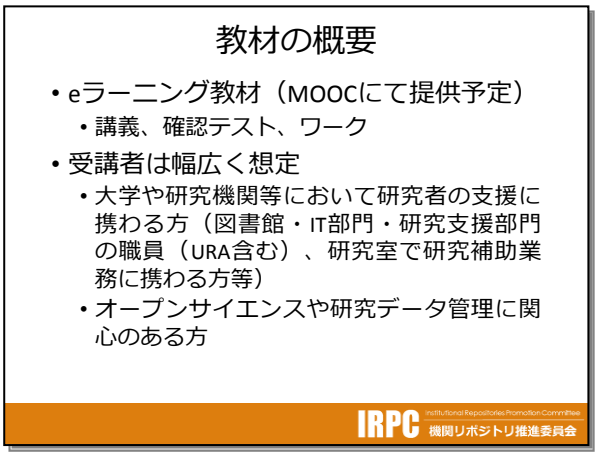
教材の形式は、e ラーニング教材にして、MOOC で提供できるようにします（図 8）。講義、確認テスト、ワークなども適宜挟みこんでいきます。受講者は最初は図書館員を想定していたのですが、せっかくならもっと幅広く、IT 部門の人や、URA といった研究支援部門の人たち、それからオープンサイエンスや研究データに関心を持っている人たちに使ってもらえるようなツールにできればということで、教材の開発に取り掛かって、約 1 年かかってほぼ完成に近づきつつあります。

全体の教材の構成は図 9 のようになっており、7 章から成っています。データの生成、加工、分析、保存、公開、再利用と、データのライフサイクル全般についてひととおり学べるようになっています。

北本先生から、2017 年のスキルと 2027 年のスキルというお話がありましたが、取りあえずこのツールは 2017 年のスキルについて学べるようになっています。

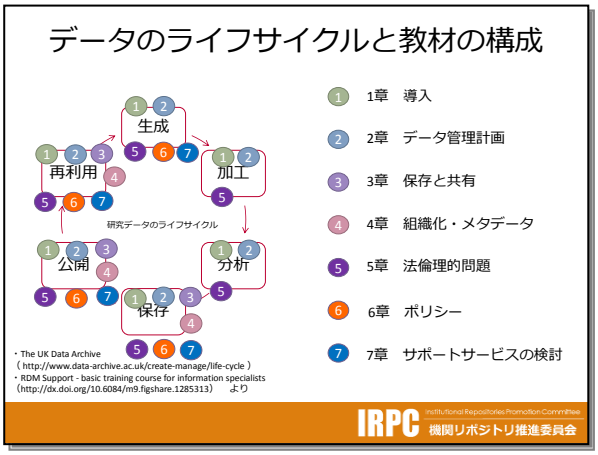


（図 6）



（図 8）

（図 7）



（図 9）

若い人にとっては、これは人生戦略上の損失につながってしまうようなツールかもしれないのですが、まずは基礎的なところから学んでいくのも大事ではないかということです。

図 10・図 11 に、各章のシラバスを簡単に書いてあります。導入、データ管理計画、保存と共有、組織化・メタデータ、法倫理的問題、ポリシー、サポートサービスの検討、このような内容になっています。

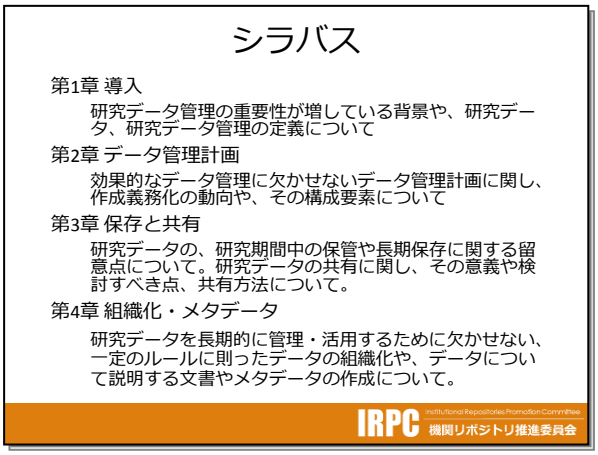
例えば第 1 章では、研究データ管理とは何かという説明があり、それに関するワークなども挟みこまれています（図 12・図 13）。

また、JST の小賀坂さんからデータ管理計画（DMP）のお話がありましたが、DMP のつくり方なども学べるようになっています（図 14・図 15）。

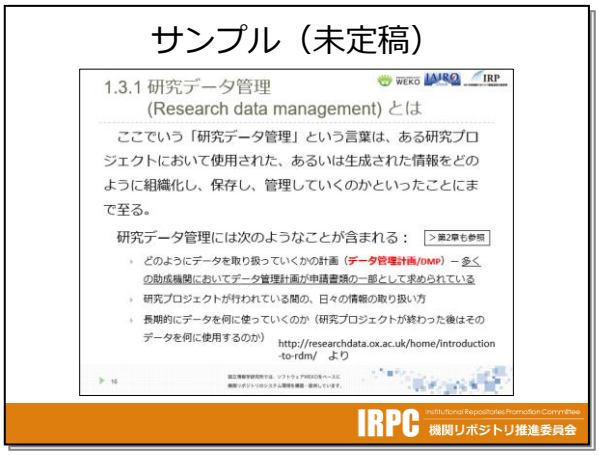
今後の展開

この教材は、まず 3 月末までに完成させて、MOOC 化も併せて完了させて公開したいと考えています（図 16）。その後、4 月から、JPCOAR（オープンアクセスリポジトリ推進協会）というコミュニティが本格的に活動を始めます。今の機関リポジトリ推進委員会の活動も、この協会の活動に引き継がれていきます。われわれ研究データタスクフォースもこの新しい協会の中で、教材をより広く活用してもらって、学んだ人からフィードバックを得て改定して、より良いサービスにしていきたいと考えています。

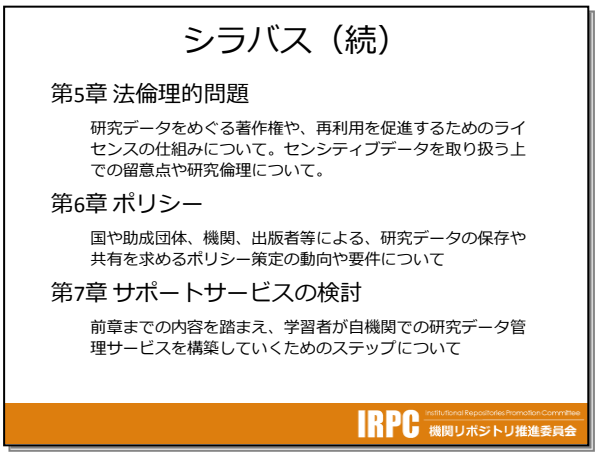
2027 年のスキルについて一緒に考えて、こういう活動を共にやっていきたいという方がいらっしゃいましたら、ぜひ声を掛けていただければと思います。



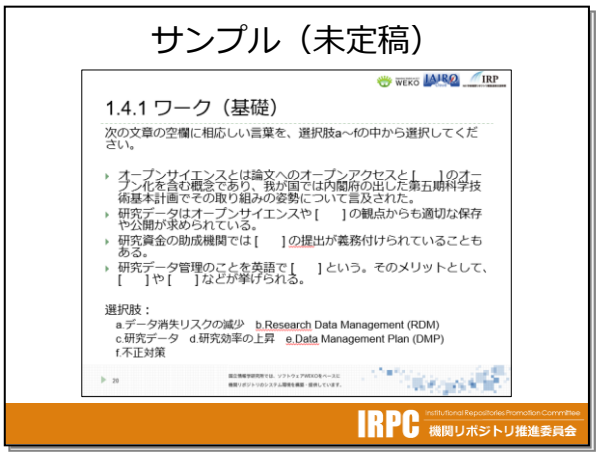
（図 10）



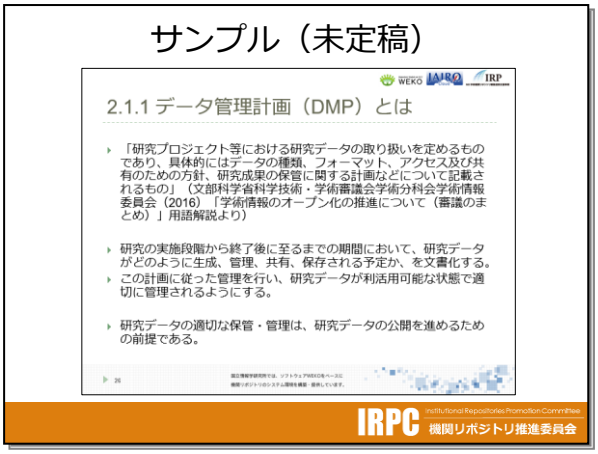
（図 12）



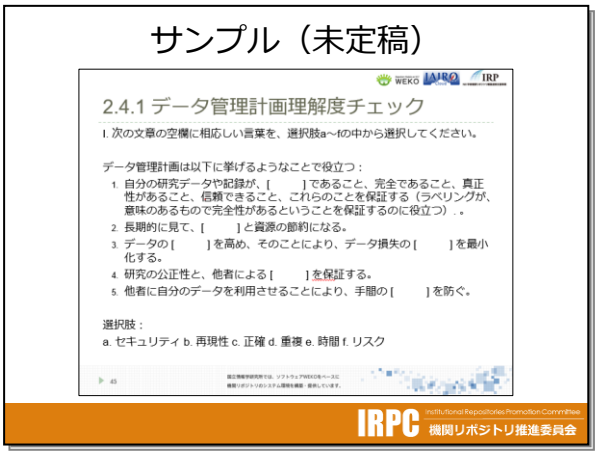
（図 11）



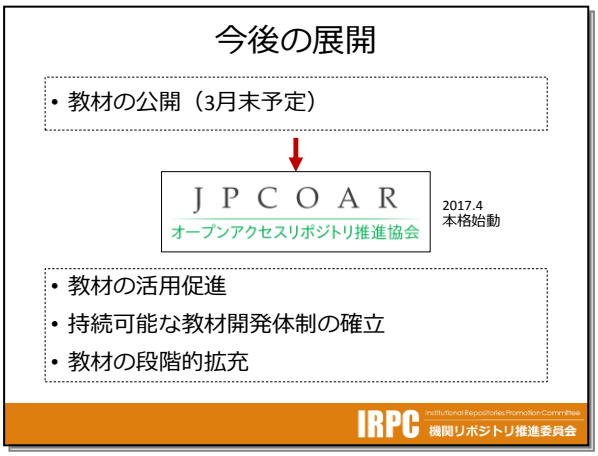
（図 13）


サンプル（未定稿）
2.1.1 データ管理計画（DMP）とは
「研究プロジェクト等における研究データの取り扱いを定めるものであり、具体的にはデータの種類、フォーマット、アクセス及び共有のための方針、研究成果の保管に関する計画などについて記載されるもの」（文部科学省科学技術・学術審議会学術分科会学術情報委員会（2016）「学術情報のオープン化の推進について（審議のまとめ）」用語解説より）
研究の実施段階から終了後に至るまでの期間において、研究データがどのように生成、管理、共有、保存される予定か、を文書化する。
この計画に従った管理を行い、研究データが利活用可能な状態で適切に管理されるようにする。
研究データの適切な保管・管理は、研究データの公開を進めるための前提である。
IRPC
機関リポジトリ推進委員会

（図 14）


サンプル（未定稿）
2.4.1 データ管理計画理解度チェック
I. 次の文章の空欄に相応しい言葉を、選択肢a～fの中から選択してください。
データ管理計画は以下に挙げるようなことで役立つ：
1. 自分の研究データや記録が、[ ]であること、完全であること、真正性があること、信頼できること、これらのことを保証する（ラベリングが、意味のあるもので完全性があるということを保証するのに役立つ）。
2. 長期的に見て、[ ]と資源の節約になる。
3. データの[ ]を高め、そのことにより、データ損失の[ ]を最小化する。
4. 研究の公正性と、他者による[ ]を保証する。
5. 他者に自分のデータを利用させることにより、手間の[ ]を防ぐ。
選択肢：
a. セキュリティ b. 再現性 c. 正確 d. 重複 e. 時間 f. リスク
IRPC
機関リポジトリ推進委員会

（図 15）


今後の展開
• 教材の公開（3月末予定）
JPCOAR
オープンアクセスリポジトリ推進協会
2017.4
本格始動
• 教材の活用促進
• 持続可能な教材開発体制の確立
• 教材の段階的拡充
IRPC
機関リポジトリ推進委員会

（図 16）

第3回 SPARC Japan セミナー2016

「科学的知識創成の新たな標準基盤へ向けて：オープンサイエンス再考」

ディスカッション 「科学的知識創成の新たな標準基盤へ向けて」

林　和弘　（科学技術・学術政策研究所）
Ron Dekker　（欧州委員会研究・イノベーション総局（European Commission（DG Research & Innovation））
北本　朝展　（国立情報学研究所）
小賀坂康志　（科学技術振興機構）
小野寺千栄　（物質・材料研究機構）
小野　雅史　（東京大学地球観測データ統融合連携研究機構）
大向　一輝　（国立情報学研究所）

●**林**　本年度の SPARC Japan のセミナーを締めるこのディスカッションでは、今日の話と、年間を通して掲げたオープンサイエンスの再考という非常に大きなテーマを採りあげます。これをどうまとめたらいいか、まだ私の頭の中には何もないのですが、皆さまと一緒に何かつくり出せればと思っています。

今日は Ron さんに、日本もここまで来ているのだということをお見せしてきたので、あらかじめ Ron さんから、ご紹介した日本の取り組みについてコメントを頂きたいと思います。

●**Dekker**　お聞きした内容は非常に有用でしたし、多くのことを学びました。急いで決断しないことが重要です。RDA の報告書「Riding the Wave」は、何が決定されているかを理解せずに決定を下さないよう警告しています。私は今日、オープンサイエンスに関する日本の状況を概観することができました。日本には既に有用な情報がたくさんありますし、インフラや政府など、いろいろな要素もあります。皆さんは恐らく、パイロットやディスカッションを通してさらに理解を深める必要があるでしょう。

同時に、野心的な目標を設定することが課題です。オランダが EU 理事会の議長国になったとき、私たちは野心的な目標を掲げたと思っていましたが、その95%を達成しました。そのため私はもっと野心的であってもよかったかもしれないと思っています。最初に立てた目標を達成できなかったとしても、それは大きな失敗ではありません。野心を持って、一つ一つの要素をつなぎ合わせていってください。

「オープン」の目指すものはさまざまです。ですから、自身が求めるオープンとは何かを明確にしてください。ユーザーは誰か。誰が論文を読むのか。誰がデータを使用するのか。皆さんはコミュニティを拡大するためにデータを再利用するに当たって、初期採用者やアンバサダー（代表者）を見つけることができます。ピーター・スーバー（Peter Suber）が述べたように、オープンサイエンスの最大の障壁は誤解です。アウトリーチは非常に重要です。

最後に、ODEN やオープンアクセスリポジトリ推進協会（JPCOAR）のような素晴らしい例が幾つか見られました。これらは、前進し、野心的な目標を設定し、構成条件を設定するための優れた取り組みです。皆さんはどんな条件の下で何をしたいですか。それを明確にすることが所有権を管理することにもつながり

ます。特にテキストとデータマイニングを考慮すると、これは重要な問題です。

市場を開放し続けることは重要なことです。少なくともヨーロッパでは、出版はもはや自由市場ではありません。一部の出版社がルールを決めて、図書館がそれに従ってお金を支払うという構造になっています。私たちはデータをオープンにするための市場を維持したいと思っています。そのためには、多くのユーザーと生産者が必要です。生産者とユーザーのつながりを促進するためには、図書館員のような仲介者や、ODEN が必要です。

●**林**　ありがとうございます。ODEN チームには非常に力強いご講評を頂きました。

小野さんから日本もそんなに遅れているわけではないという話がありましたが、日本もいろいろな取り組みを、しかも地に足が着いた形でしているということを Ron さんに示せたと思います。ですから、ここからは国際的にも問題になっていることについて議論していければと思います。

会場から頂いた質問が多様で非常にまとめにくい中で、出だしからハードな話題に行きたいのですが、「小野さんの講演にあった『データのオープン化を強要、義務化するのは良くない』という意見について、パネリストの皆さんの本音を聞きたい」というご質問があります。拙速過ぎるのは良くないということかと思います。小野さんは既に態度を表明されているのですが、他の皆さんはいかがでしょうか。小賀坂さんなどは実はこの件は前から一家言お持ちの方だったので、どうですか。

●**小賀坂**　これはイエスでもありノーでもあります。ご存じのように、オープンサイエンス推進については、欧州はかなりトップダウンで進んでいます。対して、アメリカはかなりコミュニティドリブンで進んでおり、これは方針の問題だと思います。大事なことは、政策として進めたいと思うなら進めればいい、強制もすればいい、ただ、その結果に対してきちんと国や行政が責任を持つべきだ、それに尽きると思うのです。強制的に研究者の振る舞いを変えさせた方がひいては国の産業競争力の強化に役立つと思うならば、そうやればいいけれど、その帰結についてはきちんと責任を取るべきです。

と申し上げた上で私の個人的な考え方は、研究者は自身のやりたいようにやるので、彼らに受容されないポリシーは意味がないということです。JST はずっとその方針でやってきています。今回もその方針を取っているつもりです。

●**林**　この点、他のパネリストで何か申し上げたい方はいらっしゃいますか。大向さん、どうですか。

●**大向**　今日は図書館活動を技術的に支える立場として参画しておりますが、私も技術的なソリューションのないポリシーは勘弁してほしいと思います。もちろん、法律や何かの制度によって固めることは必要だと思うのですが、それを裏付けるテクノロジーがなくて、研究者がメインの課題ではないところに時間を使うことになってしまい、最終的なアウトプットのクオリティを下げることになってしまうのであれば、それは非常に問題です。ポリシーとそれを裏付けるソリューションを一緒に考えたいということが一番思うところです。

●**林**　ということで日本は、どうしても、という枕が正しいかどうか分かりませんが、研究者側からのボトムアップ型になりがちです。でも、ある OSTP（Office of Science and Technology Policy：アメリカ合衆国科学技術政策局）の方の言説を借りると、ボトムアップだけに頼り過ぎるといつまでたっても時間がたつばかりで進まない、だから、carrot and stick の stick が必要なのだとおっしゃっています。今日は既に幾つかヒントはあったと思うのですが、ボトムアップのアプローチをより加速させる手法について、何かパネリス

トの方、あらためて申し上げたいことはありますでしょうか。北本さんなどはたくさんあるのではないでしょうか。

●**北本**　私も、基本的にはオープンにした方が得だという状況で進めたいと思っています。ボトムアップなのですが、ボトムアップの単位がどこかだと思うのです。ある経済学では、個人として最適な行動をすると全体最適にならないという合成の誤謬のような考え方があります。ボトムアップで本当に個人として最適な方向に行くと、全体最適にならないというのは、恐らくオープンサイエンスでも同じだと思います。

ですから、「個人としての最適性をどうぞ追求してください」だと、いつまでたっても進まないかもしれませんが、全体の、例えばコミュニティ単位で最適化するということであれば、要は、個人としての最適性から少し違う方向に進んでくださいということであれば、全体としては良くなる可能性があるので、そういう単位で偉い人が考えないと進まない面があるのではないかと思います。

●**林**　そうですね。リサーチコミュニティの見識のようなものが研究者の集団をリードすると。Center for Open Science の Brian Nosek さんは、オープンサイエンス・フレームワークをつくるときに、あまり先を走り過ぎると人が付いてこない、寄り添い過ぎるとそこに停滞してしまう。だから、マラソンランナーを教育するときに、10m 先にペースが速い人を走らせるように、少し前を走って付いてきてもらうのがよいのだとおっしゃっています。

そういうアプローチが日本でもできるかという話があるかと思うのですが、小野さんは実際に DIAS に関わっていて、今のようなアプローチは絵に描いた餅だといった話など、もう少し具体的なエピソードをご披露いただけますでしょうか。

●**小野**　確かに、おっしゃるとおり、研究者は言わないとなかなか動かない、周りに言われないと自分の好きなことだけをやり続けるという面があると思います。強要するのが良くないと言ったのは、今すぐデータを出せというような言い方は良くないのではないかということです。

小賀坂さんが DMP でされていたように、考えるところから始めろという強制の仕方はありだと思います。取りあえず、公開を前提としてまず考えてくれ、できないならできない理由を考えろという強制の仕方はありかなと思っています。いきなり出せ、とにかくいつまでに出せ、明日出せというのは良くないでしょう。

それでもなかなかやらないようであれば、これも小賀坂さんがおっしゃったことに私も全く同感なのですが、国が責任を持って強制させる、ただ、強制させた結果、うまくいったかどうかを検証する必要があると思います。間違っていたらやめなければいけないし、それでオープンサイエンスが進んだという結果が出たのであれば、それは研究者の側が間違っていたことになります。国が責任を取る、きちんと検証するという前提で強制するならば、それはそれで一つのやり方ではないかと思います。

●**林**　Ron さんに戻してみたいのですが、今のように、日本ではボトムアップ型である中で、政府がトップダウン型で進めるなら、きちんとポリシーに対する責任を担保しなければならないという議論になりました。一方で、ヨーロッパは Horizon 2020 に従って、例えば 2020 年までにオープンアクセスは 100%実現するということを明快にうたっています。既に宣言されたポリシーに対する責任、チェック体制についてはどのようになっているか、お聞かせ願えますでしょうか。

●**Dekker**　オランダが欧州理事会の議長国のとき成功した要因の一つは、政治的バックアップを得た包括的アプローチでした。大学と資金提供者は、自分たちはオープンアクセスを望んでいると述べました。一部の出版社でさえ、私たちは変化が欲しいと言っていま

した。共通の目標を持った包括的な提携があれば、前進することができます。私は、何が求められているかを明確にして、官僚的形式主義を最小限に抑えることに気を使いながら、トップダウンのアプローチを組み合わせたいと思っています。オランダのリサーチカウンシルが直ちにオープンアクセス化したいと言えば、それは誰にとっても非常に明確です。

先週、ある若手研究者が、「全員に適用される限り、研究者は義務的であるかどうかは気にしない」と言っていました。それが義務でなければ、研究者は余分な努力をしたいとは思わないでしょう。誰もが変わるよう誘導されるべきです。科学者も（他の人間と同じように）変化を好みませんが、時には特定の方向へのインセンティブが必要な場合もあります。だからこそ、私たちは政治的声明と、それをしたいと主張する資金提供者の声明を提供しようとしたのです。慣れるには時間がかかります。私たちにはパイロットがあり、再びゲイツ財団の例を挙げると、結果がオープンアクセスで公開されるというのが資金調達の条件です。

●**林**　という状況なので、やはりトップダウンは慎重でありつつ、ある程度は必要になってくると思います。そこを日本のファンディングエージェンシーを含めて、今後それなりにポリシーを定めていく必要があるのではないかと思います。私もどちらかというとそれを後押しする役割で今 NISTEP にいるとも言えます。

これは大枠の話でしたが、個別に見ると小野寺さんへの質問が一番多いです。「既存の研究データ管理ツールではなく、自前でデータ管理をすることのメリット・デメリットを知りたいです」というご質問があるのですが。

●**小野寺**　自前で管理するというのは、既存のもので足りない部分があったから、それを補うために自分たちで開発するというのが背景だと思います。ですから、デメリットについてはないというか、あってもそれを解決するために、コミュニティの中で議論して、そのデメリットも解決する方向に持っていこうとしていると理解しています。

●**林**　強いて言うと、課題はモチベーションの維持だったりしませんか。やりたい人がいるうちはいいのだけれども、そういう人が異動してしまったり、モチベーションが下がったりしたときにどうするか。自前でつくるときは、結構ボランタリーなリソースを使うことが多いと推察されるのですが。

●**小野寺**　人がいなくなったときという問題は確かにあると思うのですが、ご紹介した NIMS の事例では、小さいグループではなく、同じ計測系の研究グループ、組織の中のコミュニティとして活動しているので、ある人が抜けたからといってそこでストップしてしまうことにはならないと思いたいです。

●**林**　コミュニティとして持続性を確保するという感じですか。

●**小野寺**　はい。

●**林**　先ほどの北本先生の、ボトムアップはコミュニティベースでという話と通ずる議論になってきて、狙っていなかったのですがきれいにつながりました。

次に、ライセンス、著作権、ビジネス回りのところを気になさっている方が多く、「NIMS のデータで無償と有償と制限公開がある。その使い分け、違いの基準を機関としてどういう方針でつくられているか」。これは説明があったような気もするのですが、あらためてこの部分についてご紹介いただけますか。でも、これは研究者側が選択しているという感じで、機関として何かそれを推奨する、ガイドをするということはないのでしょうか。

●**小野寺**　そうですね。

●**林**　それは DIAS もそうですよね。出すデータの公開レベルを研究者側が選べると伺っているのですが。

●**小野**　はい、基本的には、データ提供者がどうしたいかというのを中心に決めていただているという状況です。

●**林**　今度は制限公開からビジネスサステナビリティの方に話を変えたいと思います。有償でやる場合に、まず小野寺さんに質問をしたいのが、イギリスの会社と組んでお金を取っているという話だったのですが。

●**小野寺**　お金を取っているのはイギリスの会社です。

●**林**　NIMS にもお金は流れてくるのですか。

●**小野寺**　はい。

●**林**　それは差し支えない範囲で、どのぐらいをカバーするのでしょう。

●**小野寺**　この発表でご紹介するためにインタビューをしたのですが、そこまでは聞いていないので、お答えできません、すみません。

●**林**　その辺が結構クリティカルになってきて、ある質問では、「私の分野ではデータベースを研究者に対しても有償で売る研究者がいて、データ統合を進める上で大きな障害となっています」というご指摘があります。一方で、データベース事業を回すためには、お金をもうける目的ではない範囲で収入を得ることも重要な場合がある、これがデータ研究管理基盤の大きなトピックになっています。

　これについて、小野さん、DIAS 側からとして、確かソフトウエアを販売するなどいろいろなケースがあったと思うのですが、見通しについてはどのようにお考えでしょう。

●**小野**　DIAS 自体もビジネスや自律的な運用に向けて検討は進めていますが、実際にそこまでいっていないので、いろいろ考えながらやっていっている状況です。さまざまな課金モデルが考えられるのですが、先ほどのお話にあったような、データに対してお金を取るというやり方は、ちょっとうまくいかないのではないかと今のところ思っています。

　では、どういうところでお金を取るかとなるのですが、やはりデータをどのように使いたいかによるのです。われわれがやっているのは地球環境関係のデータ基盤なので、これがあったらユーザーも満足だし、世の中の地球環境の課題に役に立つというアプリケーションをつくるために DIAS にあるデータを使って開発するという場合、アプリケーションの開発部分にユーザーにお金を出してもらうようなモデルがあるのではないかと考えているところです。

●**林**　その一方で、事業の持続性が見えない中で、やみくもに研究データを登録することに対する懸念を示す質問もありました。要は、体のいい研究者のバックアップ保存先として使われてしまうのではないかという懸念ですが、それについて何かコメントができる方はいらっしゃいますか。

　まだそういう話よりも、いかにデータを登録するかというフェーズではあるかと思うのですが、早晩、逆転するはずですよね。みんなが登録するようになったら、全部は登録できないとなって、何かを捨てる方の議論になるのですが。今このテーマを扱うのはちょっと難しいでしょうか。

●**北本**　私が DMP で一つ恐れているのはそれなのです。義務化と言って、「では出せばいいのか」と形式的にデータを出すことが広まってしまうと逆効果になりかねません。あまり規制を厳しくするとそういうことが起こり得るので。

●**林**　トップダウンで早くやり過ぎるとそうなってし

まう。

●**北本**　やり過ぎると、では出せばいいのかということになってしまいます。例えば、DIAS だとセレクションはしているのですが、そのセレクションをどこまですべきかという問題があります。figshare などは何でも受け付けていますが、それでよいのかという疑問はあります。そこをどうするのかは結構大きな問題です。ただ研究データの墓場のようにすべきでないとは思います。

●**小賀坂**　似たようなことがかつて知財でも起こりました。昔は基礎研究の研究者は知財取得に対しては無関心でしたが、国がそれではいけないということで、研究者に知財を出しなさいと言う時代がありました。

ところが、その維持経費が莫大になって、支えきれないまで膨れ上がり、逆に何でもかんでも出せばいいというものではないという時代が来ました。大学も、最近ではライセンス可能な特許しか支援しない傾向にあるそうです。今度は逆に、将来化けるかもしれない基本特許すら出せないという寄り戻しが来ています。ただ、多分バランスが取られていいところに落ち着くのではないのかと思っています。

データについては、明らかに保管インフラに掛かる経費とのバランスですよね。多分、あまりアプリオリな答えはなくて、現実的なところで落ち着くのではないかという気がしています。

●**林**　今の議論に関して、Ron さん、ヨーロッパではデータをどこまで持つか、あるいはデータをどうやって捨てたらいいかという議論はどのぐらい進んでいるのでしょうか。それとも、まだ研究者にデータを登録してもらうのが先というフェーズにあるのでしょうか。

●**Dekker**　それは分野によります。物理学や生命科学のように多くのデータを生み出す科学では、データを保管するためのプロトコルがあります。大学には 10 年間データを保管するという行動規範があります。プロセスの一部として作業が終了した後、その作業成果を単純に捨てることができるセクターは他にありません。民間部門でそうしたら解雇されるでしょう。したことについて責任を負うことは、良い研究をすることの一部です。

オランダでは、心理学の分野で大きな研究不正が起こりました。それまで高く評価されていた教授がデータを改ざんしていたのです。データを作成するように求められたとき、彼はそれを渋りました。引っ越しのときに調査結果を紛失したというわけです。彼が最終的にデータを提出したとき、統計モーメントが全て 3 次および 4 次モーメントと同じだったので、これは不正だとみなされました。この事件は、再現性の確保のためにデータを保管しなければならないという議論に拍車をかけました。

再利用のためにデータを保管することの必要性に関する議論はあまりありません。保管はコストのほんの一部です。コストは、研究者とそのアシスタントがメタデータを追跡し記述するために費やされる作業時間に掛かるものです。さらに、IPR（知的財産権）は非常に複雑です。少なくともオランダでは、データの権利所有者は、（一部においては）データを得るための資金を提供した者、つまり資金提供者または大学になります。彼らはデータを追跡し、記録する責任があります。他の研究者がデータを再利用できるように、データを見つけられる状態にすることが課題です。議論は、データを保管すべきかという問題から、再利用に適したデータの品質をどう保証するかという問題に移行していくでしょう。

●**林**　日本は小さなところの議論をしてしまったのではないかという感じになっていますが、話を変えます。今日の話題提供の中で一つ大きなインパクトのある話は、ウェルカムトラストがつくったようなオープンリサーチ・プラットフォーム、ないしはプレプリントサーバーという意味での arXiv の広がりだったと思いま

す。それに関するご質問が何点か来ています。まずは小野さん宛てに、「プレプリントサーバーのような、雑誌を経由しない論文の公開の場があれば、研究者にとって本当にいい世界になると思いますか」という質問が来ています。研究者を代表するわけではありませんが、今日の話を伺っていかがでしょうか。

●**小野**　何とも言えないところがありますが、プレプリントサーバーなりに論文をどんどんアップロードしていくという活動が評価されるようであれば、そういうものに乗る研究者の方もいるのではないかと思っています。でも、大抵の研究者は、それなりに権威があるところをファーストプライオリティとして投稿するのではないかと思います。

最近聞いた話で、PLoS ONE はオープンで比較的投稿しやすいということで、一時期とても盛り上がって投稿数が増えていたのですが、だんだんそれが減って、一時の盛り上がりが下がってきているようです。だから、今プレプリントサーバーのようなものを用意しても、そんなに集まらないような気がします。

●**林**　生々しい本音をありがとうございます。一方で、北本さんのご紹介のように、ディープラーニングの世界ではスピードがすごく大事ということで、極端な表現をすれば、査読なんかしていられない感じですよね。だから、そういうのは新しいディシプリンの中で機能するだけなのか、既存のエスタブリッシュされた研究、ディシプリンにも影響を与えるのか、これがこれから注目されていくことになるのだろうと思います。

その辺の見通しについて、パネリストの皆さんは何か思うことはありませんか。あるいはそろそろ会場の方から何かありましたらお願いします。この話はみんなで考えないといけない、新しいパラダイム、研究パラダイム、評価パラダイムを生み出すためのプラットフォームづくりにつながると思うのです。端的に言うと、旧来のピアレビューをいい意味でどのように捨てていく、変えていくかというような話にもなると思うのですが。Ron さん、どうぞ。

●**Dekker**　それは品質評価を破棄するものではありません。ウェルカムトラストは二つのことを行っています。Wellcome Open Research にはピアレビューされた出版物が含まれますが、大きな違いは、まず発行して、その後、レビューが行われることです。

ウェルカムトラストは、プレプリントを認める多数の資金提供者の連合に加わっています。プレプリントを掲載するためには、プレプリントを保管するためのアーカイブや施設が必要です。幾つかの分野、特に生物学、医学、保健においては、成果物にタイムスタンプを入れることが非常に重要です。それは、導入、分析、データ、結果、結論を含む本格的な論文である必要はないということにもつながります。論文が幾つかの重要な結果に分解されることになります。プレプリントでも、研究コミュニティにとって重要で関連性の高い結果を公表することができるのです。それはまさにエボラ危機の間に起こったことです。他の研究者がこの伝染病を抑える方法について結論を下すことができるように、データと結果ができるだけ早く公開されました。

●**林**　そうですね。結局、とにかく出して、その成果が公開されるタイムスタンプをまず付ける。その後、今までどおりのピアレビューもやって、今までの評価のスタイルも担保する、そういう世界が始まろうとしています。これもまた、ディシプリンによるとは思うのですが、そういう流れの中で、研究機関、大学職員、研究者そのものをどう育成するかという話になると思います。そういう質問も来ています。

「プレプリントに出すのが当たり前の世界、ないしはオープンリサーチ・プラットフォームのような、とにかく出して後からピアレビューをするという世界になるとき、大学や研究機関の図書館員の育成、採用方法を大幅に変える必要があるように思いますが、そのような可能性は日本にあるのでしょうか」という、悲

観的なご指摘も頂いています。

意訳すると、プレプリントサーバーが当たり前の世界、オープンリサーチ・プラットフォームが当たり前の世界、公開までいかなくても研究データのシェアリング、コラボレーションが進むような世界になったときに、大学の職員やその他はこの激動の時代にどう対応していったらいいのかということだと思います。

●**大向**　何が起こるかさっぱり分からない、明日になればポリシーが外的要因によってがらっと変わってしまう、そして最適な情報システムの姿がまるで変わってしまうということは、まだ学術情報分野ではあまり経験しないことかもしれませんが、ウェブの世界で私たちは散々経験しています。Facebook が来た、Google が来た、Amazon が来た。いつの間にか、フィットしなければいけないというわけではないけれど、フィットした方がより生活のクオリティが上がるという状態に置かれる中で、近年、図書館が幾つかある選択肢のうち、どの情報システムを選ぶということぐらいしかやってこなかったことはとても危ないと思っています。だから、自分でつくらないといけないときにはつくらなければいけないのだ、制度もアーキテクチャーも含めてつくる側に回らなければいけないということを、何とかメッセージとして伝えたいと思って、NII の研修プログラムを数年前から始めているのです。

研修プログラムで目指していることは、知識を増やすことではなく、動じないというか、新しいシチュエーションに対して、クオリティが高いかどうかは別として、ソリューションをひねり出して次に行こうと思えるようになることです。それを今鍛えておかないと怖いと思ってやっています。

●**林**　10 年前、機関リポジトリを始めたころに比べると、格段に図書館の方々が研究に対してどうコミットできるかを一生懸命考えています。10 年前だとまだ、「出版した論文の著者原稿を下さい」と言うのが研究者とのコミュニケーションで大事なことだと真面目に議論していました。それはそれでその当時は大事だったのですが、10 年たって、また違うコンテキストで研究データにどう向き合うかを考えています。それを繰り返して、本日も北本さんが 2027 年のデータライブラリアンのスキル要件についておっしゃっていましたが、ライブラリアンが何に対応できるかきちんと考えれば、多少ポジティブな方向に向かうのではないかと思います。

北本先生、何か追加のコメントはありますか。

●**北本**　大向先生がおっしゃったように、ウェブの世界はずっと変わってきたのに、学術情報が変わらなかったので、もう限界というところではないかと思います。ですので、今後大きく変わっても別に不思議ではありません。

先ほど Ron さんもおっしゃいましたが、プレプリントサーバーができたから査読が不要になるということではないのです。何らかの品質管理は依然として必要です。ただ、品質管理の方法がピアレビューだけではなくなる可能性が、プレプリントによって開けます。

自分でチェックすべきものを他人にアウトソースするというか、他のレビュアーのチェック結果を使わせてもらうことによって品質管理が簡単になるというのがピアレビューの仕組みですが、今後はそれを AI が行うかもしれません。あくまで目的は品質管理であって、ピアレビューという手段が欲しいわけではないというのが変化だと思います。

●**林**　そうですね。ピアレビューという仕組みも、手紙で研究情報をやりとりする時代に生まれたレガシーなフレームワークなので、必ずしも様式にこだわる必要はありません。けれども、これも分野によると伺っています。例えば high energy physics の研究でプレプリントがはやったのは、書く人数と読む人数が同じくらいで小規模なため、査読をしなくてもその中身が分からない人はもぐりだという文化があるから進んだからだという話を物理の研究者に聞いたことがあります。

一方、私が専攻した、製薬につながる有機合成化学は、書く人に対して読む人の方が圧倒的に多いので、信頼できる情報が欲しい、誰かによって信頼が担保された情報が欲しいということで、ピアレビューが重要になります。ただ、今の ICT のインフラ、ハードウエア、ソフトウエアがあり、これからデジタルネイティブの人材も増える中で、新しいクオリティコントロール像をどうつくっていくかという話になってくると思います。

さて、これはどこで出していいか分からなかったのでここまで引っ張ったのですが、Ron さんに質問があります。「アメリカではアメリカファーストの政策の下に保護主義的な動きが見えて」、ここからは私が付け加えます。「環境系の研究などは情報を出すことを制限するような動きがあります。さらに、イギリスの EU 離脱もあり、ヨーロッパに所属している国々が保護的に動く恐れがあるのではないでしょうか。そういう中で、Horizon 2020 や European Open Science Cloud を進めていく上で、どのように見通しを立てていますか」というご質問です。

●Dekker　私はアメリカが保護主義であると言ったのではなくて、保護主義的世界が起こりつつあると言いました。しかし、新しい知識を深めたい場合、早い段階で可能な限りオープンにしなければなりません。お金を得たい場合は、ある段階まで知識を保持して、特許またはライセンスを取ればいいと思います。

Structural Genomics Consortium を例に取ると、非常に競争力のある製薬会社ですら、独自で事業を行うには複雑過ぎるし、あまりにも高価になってしまうので、共同で行うことを決めました。資金は大きな要素になります。自分にとって何が良いのか、他にとって何が良いのか、両者にとって良いものは何かを考えて、win-win の状況をつくり出す必要があります。リサーチを行う際、データの再利用は推進力となります。リサーチをすることによって、選挙で生き残ることができます。ヨーロッパでは、カトリック教会の次に大学が最古の組織です。1500 年以上存続しており、恐らく将来も存続するでしょう。

しかし、変化が速くなっています。追求すべき方向が分からない場合、まず 10〜15 の花を咲かせて始めてみましょう。それが私がアメリカで見いだしたものです。試してみて、それが駄目なら、進んでその取り組みを中止するのです。そうすることによって、専門知識を構築していくのです。これが不確実な新しい未来に備えるための最良の方法です。

●林　頼もしいコメントをありがとうございます。これで全ての質問でもないのですが、大体お答えしたことになります。ここからフロアからのご指摘やご質問なども受け付けたいと思います。

●深貝　今までオープンサイエンスという話になると、図書館側はどうやったらいいかという戸惑いがあり、研究者側は成果は自分のものとして上げないと競争に勝てないという思い込みにはまり込んでいました。しかし、今日のお話を伺っていると、図書館あるいは情報を学術的に管理する方々が新しい試みをしつつあって、多いに励みになりました。今の学術研究のスタイルにどっぷり漬かった、「自分の研究データは自分のものである」という人々が変化していく可能性を開くためのプラットフォームをいろいろつくっていただいていると思います。

そういう意味で、今後は実際にデータを集める仕組みを試してみたり、ユーザーとしてデータを試してみたりするといいと思います。具体的には、この面についてもの分かりがいい研究者を頼りにしながら、特に学生と接している教員に頼んで、実際にデータを使ってみます。また、柔軟な発想の院生たちにデータを使ってもらいつつ、データを集めて登録することに参加してもらいます。それから、狭いディシプリンの考え方に染まっていない学部生に、報告にあった ODEN を使ってもらえば、分野の垣根を越えた柔軟な発想がどう生きるかを試してみることができると思うのです。

ただし、図書館の職員だけでそれは無理で、教員に頼まなければいけないので、実験的にそういうデータベースの新しい試みをなされている方々は、ぜひ図書館長を巻き込んでいただくと進んでいく可能性はあります。逆に、弊害が見えるかもしれないけれども、それはそれで貴重なことだと思います。

●**林**　ご指摘ありがとうございます。深貝さんはまだ図書館長でいらっしゃいますか。去年まででしたか。「現図書館長から頼もしいお言葉を」と返そうと思ったのですが。そうやって実例に取り組んでいく、進めていくこと自体が大事であるということですね。今のお話は、FOSTER という Horizon 2020 の取り組みを想起させます。若い人にオープンデータだけを使って研究をさせるという研修プログラムがあると伺っています。

最後に、今日のテーマは「オープンサイエンス再考」なので、今日のディスカッションないしはこの 1 年の SPARC Japan を振り返って、あなたにとってのオープンサイエンスとは何か、何がキーになるかをパネリストの皆さんから一言ずつ頂いて終わりにしたいと思います。

●**小賀坂**　JST は、決まった額のお金で成果を最大化していただくためにやれることをやるという立場ですから、その文脈でオープンサイエンスを推進すべきなのですが、この議論をするときに、ステークホルダーが偏っていると思うのです。今足りていないのは、一つはドメインサイエンティストだと思います。この話にリテラシーのない人たちにきちんと入ってもらわなければいけないのです。もう一つは、リワードシステムに関してなのですが、評価担当者です。評価担当者をどうやって評価に組み込んでいくかをやっていただきたいと私の立場では思います。

●**小野寺**　材料科学の分野では共有できるものとそうでないものがありますが、共有できるものについても、そのデータの持ち主がコントロールできるような、どう使われているかが把握できるような仕組みができればと思っています。学術情報流通は研究者がコントロールできない状況になってしまいましたが、オープンサイエンスを進めるにあたっては、研究者が意図しない方向に行かないようにコントロールできればいいと思います。

●**小野**　私は最初に間違ったメッセージを発してしまったかもしれないと反省しています。「焦ってはいけない」とは言ったのですが、それは出さないという意味ではなくて、やはり出せるデータと出せないデータがあるのです。出せるデータも、まだすぐには出せないから、ちょっと時間を下さいという部分があるということです。

●**林**　エンバーゴの議論になっていなかったですね。

●**小野**　そうですね。だから、それを今すぐ出せと言うのはやめてほしいということです。われわれも徐々にそういうことを考えていきながら、出せるタイミングができたら出していきますということなのです。それを言いたかったのです。

まだ出せないデータはどうしても出せないので、それはちゃんと理由を聞いていただいて、「それだったらしょうがないね」と納得していただきたいということがまずあります。今出ていないデータも、今後実際は出ていくと思います。そこに関してはすごくポジティブに見ています。

年代によって考え方が違うという話をしましたが、先ほど、まだ偏見がない学生と一緒にやったらいいのではないかというお話がありましたが、全くそれに同感で、下の世代に行くほどオープンマインドなのです。データもシェアするし、何だったら個人情報もシェアしているような、そこはやり過ぎではないかという気がするのですが。時間がたてば世代が変わるので、今後どんどんそういうことが進んでいくのではないかと

思っています。時間がたてばいずれ解決する問題があるので、焦らずじっくり構えていけばいいのではないかと思います。図書館の方も、気難しい研究者と一緒にやるよりは、なるべくオープンマインドの人を探して、一緒に活動を続けていった方がいいのではないかと思います。

もう1点、小賀坂さんが、ドメインサイエンティストのオープンサイエンスの議論への参加が足りないとおっしゃっていたのですが、私はその他にも、経済や法律の専門家の参加が特に日本は足りてない気がします。私はベルモント・フォーラムのデータシェアリングに参加していましたが、そこには経済と法律の専門家がいました。経済は、サステナビリティを考えていく上で必要になります。法律は、これも結構重要で、データの利用規約のところで結構研究者は悩んでいるのです。そこがどうしたらいいのか決められないので、出せないという人が多いのです。逆にこういうところを国に支援していただいて、どういうデータの利用規約が研究データにとって必要かを考えていただきたいと思います。

データの法律的なライセンスの話は、われわれも弁護士に相談したのですが、法律も、知財に詳しい人、民事に詳しい人など、分野が分かれているのです。その中でデータの専門家が全然いないので、相談する相手がいないという状況です。それは研究者のレベルではなかなかできないところなので、その辺の支援を考えていただきたいと思っています。

●**林**　大事なポイントですね。

●**大向**　私にはあまり深い見識はないのですが、オープンデータ、オープンガバメント、オープンサイエンスなど、「オープン」の付くムーブメントと長く付き合ってきて思うことは、「オープンは笑ってやること」です。しかめっ面しながら、「オープンにしなければ」と言いつつやることは決してうまくいきません。新しいものを取り入れるときは、取り入れたらこんなことができるようになった、こんな機能があることが分かった、今まで見えていなかったものが見えるようになったという理由で、みんながアダプトしていくというプロセスがあります。オープンかクローズドかは、普通にコスト計算をすると、クローズドの方が良かったとしても、オープンの方が面白いから、楽しいから、便利だからということで乗り越えてきたこれまでの歴史があるのです。特にウェブを見ていると、それが成功要因だと強く感じるところがあります。

具体的にどうというわけではないのですが、オープンサイエンスとはどのようなものか体験してもらうために、オープンサイエンスをするとはどのような感覚なのか知ってもらうために、体験にフォーカスした形で仕組みをつくっていく、動かしていくことにも目を向けてみてもいいのではないかと思います。この議論に限らず、普段思っているのはそういうことです。

●**北本**　私は、本来的に、オープンがいいところもあるし、クローズドがいいところもあると思うのです。それが歴史的な経緯によって逆転しているようなところがあって、必ずしも全体的に一番いい形になっているとは思えない部分があります。ですから、オープンが全てでもないし、クローズドが全てでもないのですが、どちらがいいのかもう一度考えるといいと思います。本当はオープンがいいのに、いろいろな経緯でクローズドになってしまっている部分はオープンにしませんか、ということではないかと思っています。

「研究とは shoulder of giants（巨人の肩）の上に乗って行うものである」というニュートンの有名な言葉がありますが、そういう「巨人の肩」をどうやってつくっていくかが、科学の非常に大きな課題です。それは個人でつくった方がいい場合もあるし、集団でつくった方がいい場合もあります。集団でつくるというのはまさにオープンなサイエンスだと思います。何となく人工的なバリアをつくって、何となく個人でやって、無駄な競争をして消耗するのはあまり賢くないので、集団でつくった方がいい場合に、オープンサイエンス

を取り入れることが非常に重要だと思います。

●Dekker　あるときは個人で、あるときは集団でという考えに同意します。私にとってオープンサイエンスとは、科学が変わる一つの道です。好むと好まざるとにかかわらず、科学は変わっていきます。列車に飛び乗ってそれをフル活用するのは大学や国次第です。私は、日本はオープンサイエンスを実現するための要素の全てを持っていると思います。皆さんは優れたインフラを持っていますし、列車に飛び乗ってそれをフル活用できる明晰な人たちがいます。今までも革新的な新製品を開発してきたと思いますが、今や革新すべきは科学であり、その列車に飛び乗るのです。私ならば政府を待たず、若い人たちに投資します。

●林　ありがとうございます。第 1 回セミナーでは、グリーンやゴールドといった研究論文のオープンアクセスの話をして、第 2 回セミナーでは研究データの話をしました。論文のオープンアクセスについては大体固まったという議論がある中で、今日衝撃的だったのが、グリーンでもゴールドでもないプレプリントサーバーないしはオープンリサーチ・プラットフォームに載せた後に、ピアレビューを行うという仕組みが生まれたことです。このような形で、まだまだ学術情報流通は変わりつつあります。論文の世界ですら、変わりつつあることが今回分かりました。

そして、研究データの方も、インセンティブやビジネスモデルに問題はあるのですが、今日は日本でも地に足の着いたさまざまな試みが繰り返されているという報告がありました。NII では研究データ基盤も来年度から進むという時代になりました。

ということで、いよいよ policy into practice、オープンサイエンスを実践する段階に入ってきました。できることが山ほどあって困るぐらいになっていると思います。お越しいただいているステークホルダーの方々に、科学研究そのものが変わるというビジョンの下、今後 5 年間、10 年間の間にやれることは何かという課題をお持ち帰りいただければ、本セミナーないしは今年度の SPARC Japan セミナーはうまくいったということになるのではないかと思います。

これにてパネルディスカッションを終了させていただきます。どうもありがとうございました。

国際学術情報流通基盤整備事業（SPARC Japan）年報
―平成 28 (2016) 年度―

平成 30 年 3 月

発行　大学共同利用機関法人　情報・システム研究機構
国立情報学研究所　学術基盤推進部　学術コンテンツ課

〒101-8430　東京都千代田区一ツ橋 2 丁目 1 番 2 号
TEL　03-4212-2351
FAX　03-4212-2375
E-mail　sparc@nii.ac.jp