

◆特集◆ 研 究 活 動 を 支 える 図 書 館

大学図書館による 研究データ管理の最前線

●研究力を強化するエディンバラ大学の事例

池 内 有 為

はじめに

1996年の冬、詳しいデータを見たい論文があった。思い切って、つたない英文メールを出してみたところ、すぐにリッチテキスト形式のファイルが送られてきた。研究の世界のオープンさに驚き、感激した大切な思い出である。

2014年現在、筆者が初めてこの論文を読み、データを見たいと思ったとする。しかし、データを入手できる可能性は低いだろう。過去の論文のデータ入手を試みたVinesらの調査によれば、18年前の論文の著者のメールアドレスは31%が無効で、26%は返信がこなかった。そしてデータを受け取ることができたのは、わずか17%だったという¹⁾。

もし、このデータが論文からリンクされ、リポジトリで公開されていたらどうだろう。実はVinesらの調査データは公開されており、論文にはデータを登録したりリポジトリの名前(Dryad)とDOIが書かれている²⁾。このDOIをたどればCC0で公開されているデータと分析に用いたRコードにアクセスできるため、(メールを出す必要もなく)すぐに彼らの調査結果を検証したり、再分析することが可能である。

Vinesらのように、論文に用いたデータやコードを公開する研究者が増えている。そしてデータに容易にアクセスできるような仕組みを整えたり、さらにはデータの引用や評価を行おうとする活動が、国や分野を超えて行われている。本稿は、こうした「オープンサイエンス」の潮流の中で大学図書館が行っている研究データ管理(Research Data Management: RDM)の先進事例を紹介し、日本での実現可能性について検討する。

1 オープンサイエンスと図書館

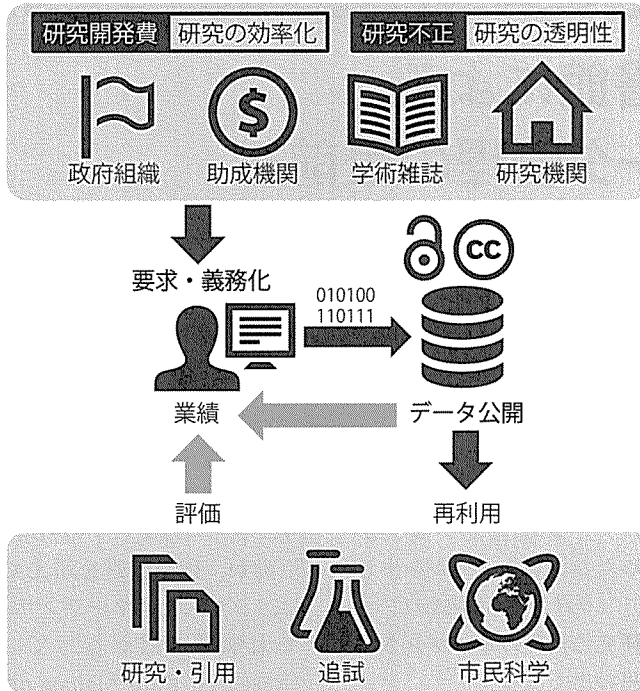
1.1 オープンサイエンスの動向

2009年に0.7ZBだった世界の研究データは、2020年には35ZB(38兆GB)になると予想されている³⁾。この莫大なデータは、各国の研究開発費から生み出される⁴⁾。研究やデータ収集の重複を防ぎ、データを再利用することによって研究のコストを下げるとともに効率を上げ、地球規模の課題に取り組む動きが盛んである。グリッドコンピューティングやクラウド技術によって大規模なデータ共有と解析のための計算資源の分散が可能になっていること、そしてヒトゲノム計画による全塩基配列の解析やアトラス実験によるヒッグス粒子の発見などの成功事例もオープンサイエンスを後押ししている。

一方で、相次ぐ研究不正への対策として、データ公開によって追試を可能にし、研究の信頼性を

いけうち うい：筑波大学大学院図書館情報メディア研究科
 キーワード：大学図書館、研究データ管理(RDM)、研究データ共有、オープンサイエンス、データリポジトリ、機関リポジトリ、メタデータ、データ記述、データジャーナル

図1 研究データ公開とオープンサイエンス¹²⁾



向上させようとする趨勢もある。特に医学分野におけるデータの改竄や捏造は、人類の生命や健康にかかわる極めて重大な問題をはらんでいるため、米国国立衛生研究所 (NIH) や欧州医薬品局 (EMA) を中心に、プライバシー問題や企業利益といった障壁を乗り越えてデータ公開が実現されようとしている。

このような状況を背景として、OECD や G8 といった国際機関や各国の政府組織、助成機関、学会、学術雑誌、研究機関などがデータ公開を義務化している。

こうして公開されたデータは研究者だけではなく、プログラミングやデータ分析技術をもつ一般の人々にも活用されている。弱冠 15 歳のジャック・アンドレイカが、蛋白質データを駆使して膀胱がん検査の方法を開発した⁵⁾ というニュースは日本でも話題になった。こうした大規模データを扱えるデータサイエンティストの育成は、日本を含め各国で盛んであり⁶⁾、学校教育での活用も期待されている。また、データの採取や分析にボランティアとして市民が関わる eBird や Galaxy

Zoo といった市民科学も活発に行われている⁷⁾。

研究データ共有の実現を支える国際機関の活動もめざましい。主な団体として、“障壁なきデータ共有”をめざす研究データ同盟 (RDA)⁸⁾、研究データへの DOI 付与を推進する DataCite⁹⁾、英国の高等教育機関のデータ管理を支援するデジタルキュレーションセンター (DCC)¹⁰⁾、豪州の国立データサービス ANDS¹¹⁾ などがある。DCC や ANDS は英国と豪州の機関であるが、その対象範囲は国内にとどまらず、世界の研究データ共有に貢献している。

1.2 図書館の役割

オープンサイエンスの理念は広く受け入れられるものの、研究者にとってデータ公開は従来にない大きな負担となる¹³⁾。また、被験者のプライバシー保護や、共有したデータが適切に使われないのではないかと懸念もあり¹⁴⁾、なかなか公開は進まなかった。

しかし、2003 年に NIH がデータ共有方針を導入したのを皮切りに、2007 年には英国の

表1 DMPのためのデータ管理原則

DMPの標準的な項目	英国 DCC	豪州 ANU	米国 NSF(1)	米国 NSF(2)
倫理とプライバシー	●			●
予算	●	●		
データ共有とライセンス	●	●	●	●
データストレージ/保存/セキュリティ	●	●	●	●
データの所有者とステークホルダー		●	●	●
責任		●	●	●
データ形式とメタデータ	●	●	●	●
研究の生産物 (products) / 文献		●	●	●

DCC: デジタルキュレーションセンター, ANU: オーストラリア国立大学, NSF(1): 米国科学財団(工学), NSF(2): 米国科学財団(バイオ)

出典: Ferreiraほか Data governance in e-science¹⁵⁾ より作成

Wellcome Trust が、そして2011年以降は米国科学財団(NSF)ほか各国の助成機関が次々とデータ公開を義務化した。研究者は研究資金を獲得するためにデータを管理し、適切な形で公開する必要に迫られている。助成金の申請時には、データストレージやライセンス、データ形式など、表1に示すような項目について記入した「データ管理計画(Data Management Plan: DMP)」を提出しなければならない。そこで研究支援や資料の組織化、機関リポジトリの運営などを行ってきた大学図書館が、その経験を活かしてRDMサービスに取り組みは始めている。こうした業務はデータ(デジタル)キュレーションとも呼ばれ、データライブラリアンやデータキュレーターといったデータを扱う専門職員の求人も行われている。

2 日本の状況と国際競争力への懸念

2014年12月現在、日本の主な助成機関による研究データ公開の義務化は行われておらず、協力依頼¹⁶⁾にとどまっている。また、文部科学省が2014年8月に発表した「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」¹⁷⁾には、「大学等の研究機関における一定期間の研究データの保存・開示」が盛り込まれているが、データの公開までは求めている。日本学術会議から

「オープンデータに関する権利と義務—本格的なデータジャーナルに向けて」¹⁸⁾と題する報告書が公開されるなど、研究データ共有の重要性は認識されていると考えられるものの、海外に比べると動きは鈍い。

一方で、学術情報流通と評価はダイナミックに変わりつつある。研究データ公開のインセンティブとして、公開したデータやコードそのものを業績とみなしたり、論文と同様に引用したり評価したりする仕組みが整えられようとしている。トムソン・ロイター社は2012年にData Citation Indexを公開しており、収録対象のリポジトリに登録されたデータの引用状況を追跡できる。また、研究データを公開している論文は、そうでない論文よりも引用が多いことを示す複数の調査もある¹⁹⁾。

賛否はあるものの、引用は研究評価指標として広く用いられており、たとえばTimes Higher Educationの世界大学ランキングの指標は30%が引用から算出される²⁰⁾。公開したデータをきっかけとして国際共同研究に繋がる場合もあるだろう。このように、研究データ公開は国際的な競争力の向上に繋がる反面、公開しない場合は不利益になる可能性も高い。

エディンバラ大学による 3 研究データ管理 (RDM) ——研究のライフサイクル全般の支援

本章では、RDM の先進事例としてエディンバラ大学の取り組みを紹介する。なお、以下の内容は、スチュワート・ルイスによる 2014 年 11 月の図書館総合展での講演²¹⁾と筑波大学附属図書館でのミーティングを再構成して記述した。ルイスは、同大学のオープンアクセスや RDM を担当する「研究・ラーニングサービス」のリーダーである。

3.1 エディンバラ大学の概要

エディンバラ大学は、1583 年に設立されたスコットランドの国立研究大学である。学生は 33,000 名、教職員は 8,000 名で、人文・社会科学、医学・獣医学、理工学部があり、生命科学や地球科学など 22 の学科をもつ。世界大学ランキングは 17 位²²⁾で、かつてはチャールズ・ダーウィンやアダム・スミスも在籍し、ピーター・ヒッグス博士など 20 名のノーベル賞受賞者を輩出している。

海外の大学図書館の多くは、助成機関がデータ公開を義務化したことを契機として RDM 支援に乗り出した。しかし、同大学はこれに先駆け、2009 年に JISC (現 Jisc) の助成を受けてデータ専用リポジトリ DataShare を公開した。2011 年には英国の大学ではじめて RDM ポリシーを作成し、2012 年 8 月から 2015 年 5 月にかけて「RDM ロードマップ」²³⁾と呼ばれるプロジェクトに取り組んでいる。このプロジェクトによって RDM サービスを完成させる予定であり、130 万ポンド (約 2 億 2,400 万円) の予算を費やしている。また、1.1 で紹介した DCC の本部はエディンバラ大学にある。

3.2 研究データを中心とした支援

エディンバラ大学の RDM 支援は、研究のライフサイクル全般にわたる。すなわち、(1)研究前の資金獲得、(2)研究中のアクティブデータの一時保

存、(3)研究後のデータ公開、(4)非公開データの長期保存、(5)データ目録の作成による業績報告まで支援を行っている。さらに、(6)研究者のニーズを把握して、(2)から(5)のシステム設計に反映させている。この 6 項目について詳しくみていきたい。

(1) データ管理計画作成と資金獲得

英国の助成機関の多くは、助成金の申請時に「データ管理計画」の提出を求めている。そこで DCC の DMP Online を導入することによって、計画書の作成を支援している。DMP Online を使えば、申請する助成機関や所属機関を選択し、表示される項目を次々に入力するだけで PDF 形式の「データ管理計画」が自動的に作成されるため、時間と労力が削減できる。例年、エディンバラ大学では 2,200 件以上の助成金申請が行われるが、2014 年は 11 月までに約 180 の「データ管理計画」がこのツールで作成されたという。

(2) データの一時保存

研究中に使用するアクティブデータの保存や共有のためのインフラとして、ownCloud を導入する。研究者が好む Dropbox とよく似たオープンソースであることから選択したという。スタッフや院生を含むすべての研究者には 500GB の領域が無償提供され、さらに追加領域を購入することもできる (年間 200 ポンド/TB)。

(3) データリポジトリによるデータ公開

2009 年に DSpace をカスタマイズしたデータリポジトリ DataShare を公開し、現在も運用している。2014 年 3 月の時点では、125 アイテム、1,946 ファイル (平均 16 ファイル/アイテム)、合計 76GB のデータが公開されていた²⁴⁾。主題でブラウズすると、臨床試験、形態学、天体物理学、歴史、音楽学、詩歌など 185 種類の多様な分野が登録されている (2014 年 12 月現在)。

しかし、図 2 にもみられるように、データ形式はごく一般的なもので、登録が多い順に plain text (345 ファイル)、x-wav (316)、xml (214)、zip (190)、jpeg (180) と続く。また、1 ファイルあたりのサイズは 0.1MB 以下のものを除くと

図2 データリポジトリ DataShare の簡略レコード画面

**アイテム
簡略記述**

タイトル:	OttoGwadoAyoker_songWriting
著者:	Gwado Ayoker, Otto; Remjisen, Bert
日付:	2014-12-01
Citation:	Gwado Ayoker, Otto; Remjisen, Bert. (2014). OttoGwadoAyoker_songWriting, 2014 [sound]. University of Edinburgh. School of Philosophy, Psychology and Language Sciences. Linguistics and English Language. http://dx.doi.org/10.7488/ds/192
Dataset Description (abstract):	The recording at the center of this item is a Shilluk song. In early 1996, the Shilluk Language Council was divided among itself over the chair, while the literacy team and the Bible Translation team were very active. I was not happy with what the elders were doing. So, I made this song and sang it first in June 1996, at the closing ceremony of 1 & 2 Samuel Workshop in Omdurman. Then I sang it to the leaders at the Unity High School in Khartoum. The song was encouraging because it challenged everybody to contribute to the work being done by the people mentioned in the song.

ファイル

ファイル	サイズ	フォーマット	閲覧	記述
02NOV2014_OttoGwadoAyoker_songWriting.wav	4.342Mb	WAV audio	Download	recorded song
02NOV2014_OttoGwadoAyoker_songWriting.TextGrid	18.05Kb	text/praat-textgrid	Download	annotations (in Praat TextGrid)
02NOV2014_OttoG..._songWriting_metadata.docx	12.42Kb	Microsoft Word 2007	Download	metadata re. the song
02NOV2014_OttoGwadoAyoker_permissions.jpg	385.2Kb	JPEG image		scanned permissions document
OttoGwadoAyoker_questionnaire.docx	13.07Kb	Microsoft Word 2007	Download	questionnaire data re. performer
02NOV2014_OttoGwadoAyoker_songWriting.xlsx	10.08Kb	Microsoft Excel 2007	Download	spreadsheet compilation of information re. performer and performance

ライセンス

このファイルには次のライセンスファイルが添付されています。

Creative Commons

100MB 程度のものが最も多く、最大でも 2GB 程度である。研究中にサイズの大きいデータを扱う場合でも、DataShare に登録するのは最終版のみで、サイズが大きい場合は圧縮するという。

DataShare のメタデータはシンプルで、必須項目は、タイトル (title)、登録者 (contributor [depositor])、助成機関 (contributor.other [funder])、公開者 (publisher)、形式 (type)、日付 (date) である。オプションとして、作成者 (creator)、データ内容 (description, tableofcontents)、言語 (language)、権利 (rights)、分野 (subject) などである。

図2の簡略レコード画面ではタイトル、作成者、日付、引用、データセット記述(抄録)などが、その他の項目は詳細レコード画面で表示される。検索のキーとなる分野は繰り返し可能な非統制語で、分野分類 (subject classification) には高等教育統計局 (Higher Education Statistics Agency : HESA) の JACS code²⁵⁾ を用いている。

また、DataShare には DataCite から取得した

DOI を付与する機能があり、各アイテム、もしくはファイルごとに付与している。このほか、クリエイティブ・コモンズやオリジナルのライセンスなども表示される。

(4) データの長期保存

2015年に研究データの長期保存用アーカイブ (Data Vault) を構築するため、設計を進めている。公開用ではないため、安く低速のストレージとし、長期保存に向けてデータの暗号化や圧縮も行わない。プロジェクトごとにまとまり (buckets) を作り、メンバー以外には非公開として安全性を高める。なお、当該メンバーがいなくなった場合は、部局が管理する。

(5) データ目録と研究者データベース

データ共有の状況にかかわらず、エディンバラ大学の RDM ポリシーや RCUK の OA ポリシーはデータ目録の共有を求めている。そこで2014年11月には、データ目録のシステムである Data Asset Register を稼働した。項目はデータセット

のほか、出版物、助成金、研究設備も含まれている。公開用インターフェイスのEdinburgh Research Explorer (<http://www.research.ed.ac.uk/>)は、日本のresearchmap (<http://researchmap.jp>)のように個人単位で表示される。さらに一覧表示される業績目録から、出版物やデータセットへのリンクが貼られている。

Data Asset RegisterはDataCiteのメタデータスキーマ²⁶⁾を活用しており、必須項目は、ID (Identifier), 作成者 (Creator), タイトル (Title), 公開者 (Publisher), 出版年 (Publication Year), オプションは、分類 (Subject), 協力者 (Contributor), 日付 (Date), 言語 (Language), リソースタイプ (ResourceType), 別ID (AlternativeIdentifier), 関連ID (RelatedIdentifier), サイズ (Size), フォーマット (Format), バージョン (Version), 権利 (Rights), 記述 (Description), 位置情報 (Geolocation) などである。また、DataShareと同様にDOIを付与している。

(6) アドボカシー

どんなに高性能のシステムを導入しても、利用者の要望にあわず活用されなければ無駄になってしまう。そこで、研究者とワークショップや面談を繰り返すことによって、ニーズを把握している。

ワークショップではストレージの仕様などについて議論を重ね、使いやすく有用な機能を備えようとしている。現在は、研究データの長期保存に向けてソフトウェアの保存やマイグレーションについて検討しているということだった。

また、一對一のセッションも行っている。ルイスが例としてあげた、ある結晶学分野の研究者の回答をまとめると以下ようになる。

- X線解析装置を用いて年間400回の実験を行い、合計400GBのビットマップとCIFを作成する
- データはローカルのUSBハードドライブに保存し、DVDにバックアップする
- データはあまり再利用しない
- ケンブリッジ結晶構造データベース (CSD) で共有する

- DataShareやData Asset Registerは不要だが、Data Vaultに過去のデータをアーカイブしたい

研究で作成するデータの種類や保存状況、登録先などは、分野や研究者によって異なるはずである。サービスをニーズに合致させるために、こうした研究者との相互理解が必要だという。

3.4 意識喚起とトレーニング

エディンバラ大学においても、いまだにRDMの重要性やサービスの認知度は低い。そこでRDMサービスを理解してもらうためのイベントや、実践のためのトレーニングを実施している。

(1) RDM サービス開始イベント

2014年9月、RDMサービスの公式な開始イベントとして、“Dealing with Data”カンファレンスを開催した。学長からRDMの重要性について話してもらうことによって、トップダウン式に意識喚起を行ったという。

(2) オンラインコースウェア (OCW) : MANTRA

RDMトレーニングの対象として、研究者、サポートスタッフ、そして図書館員が想定されている。ウェブサイトで公開されているMANTRA²⁷⁾は、RDMの概要、データ管理計画、保存とセキュリティ、データの組織化、統計や地理情報システムといったソフトウェアの操作方法など、実践的な九つのメニューがあり、それぞれの目的にあわせて自習できるツールである。CC BYライセンスで公開されており、この1年間で144カ国からのアクセスがあった²⁸⁾。

RDMポリシーや支援の機会を理解してもらえるように、サポートスタッフにも研究者向けのトレーニングを行っている。同大学のリサーチアドミニストレーター (URA) は、すでに3分の2がトレーニングを完了した。また、図書館員にとって重要なのは、RDMについての知識を深め、その重要性を理解し、教員に対して自信を持って説明できるようになることだという。

3.5 連携によるサービス提供と研究力強化

このように多岐にわたる研究支援を行っているため、関連部署も多い。エディンバラ大学の情報サービス (Information Services)²⁹⁾には七つの部門があり、うち図書館は(1)EDINA (ナショナルデータセンター)とデータライブラリ、(2)利用者サービス、(3)図書館と大学のコレクションの3部門で、その他に(4)IT インフラ、(5)IT アプリケーション、(6)学習・教育・Web、(7)DCCの4部門がある。また、RDM スタッフは、リーダー、プロジェクトマネージャー、データライブラリアン、サブジェクトライブラリアン、IT インフラスペシャリスト、利用者サポートスタッフなどから構成されている。

さまざまな部門やスタッフが協力してサービスにあたり、コミュニケーションを密にすることによって、業務を重複させることなく効率よく研究支援を行っている。研究支援の経験がないスタッフも、協働することによって研究プロセスの知識や自信を深めているという。また、多くの部門が関わることは、学内におけるサービスの認知向上にも役立つだろう。

さらに、サービス全体を俯瞰し、大学全体に貢献しようとする視点も重要である。ルイスはDSpaceの名誉コミッターや、Jiscのリポジトリサポートプロジェクトのリーダーなどを歴任する技術者としての側面も持つが、エディンバラ大学が世界トップランク²⁹⁾の大学であり続けるために研究力を強化することを強く意識している。データリポジトリの最適化だけに注力するのではなく、学術情報流通の変化を捉え、その中で所属する研究者のプレゼンスを高めるためのサービス構築を目指している。

研究者のニーズに合わせて研究を支援し、競争力を高めるRDMサービスは、研究者にとって欠かせないものとなり、ひいては大学における図書館の存在感と必要性を揺らぎないものにするだろう。

4 日本の大学図書館による RDM サービスの展望と課題

4.1 機関リポジトリ (IR) の活用と学内外の連携

2014年12月4日現在、国立情報学研究所 (NII) の「機関リポジトリ一覧」に掲載されたIR数は383に達した³⁰⁾。日本の大学図書館がRDMサービスを開始するならば、研究データの登録先として有望なのはIRだろう³¹⁾。現在、IRのコンテンツは出版物が中心だが、データもわずかながら公開されている。2014年11月の統計によれば、404リポジトリに収録されたコンテンツの合計1,427,675件のうち、データやデータセットは52,418件(3.7%)、ソフトウェアは29件(0.0%)³²⁾であった。

公開したデータの発見可能性を高め、有効活用するためにはDOIの付与が欠かせない。折しも2014年3月31日にはジャパンリンクセンター (JaLC) がDataCiteの会員となった³³⁾。10月からは1年間の「研究データへのDOI付与登録実験プロジェクト」³⁴⁾を開始し、千葉大学附属図書館やNIIなど8機関が参加している。IRへの実装も視野に入っているだろう。

また、文部科学省が8月に発表した「教育研究の革新的な機能強化とイノベーション創出のための学術情報基盤整備について—クラウド時代の学術情報ネットワークの在り方— (審議まとめ)」³⁵⁾には、データ共有を見据えたアカデミッククラウドの必要性や運用ルールに関する記述が盛り込まれている。

こうした基盤を活かしつつ、海外の動向をふまえながら日本の大学図書館によるRDMサービスについて検討することができるだろう。検討に際しては、JAIRO Cloudを提供するNIIやデータアーカイブの経験豊富な国立国会図書館などの諸機関と、また実際の運用にあたっては、学内のURAや研究資金担当、IT部門などと連携することが望ましい。

4.2 データ記述とメタデータ

RDM サービスをはじめようとする図書館員にとって、メタデータは頭の痛い問題ではないだろうか。従来図書館が扱ってきた出版物は、紙にせよ電子にせよ一定の形式で出版されたものであった。多種多様な研究データを登録する際に、メタデータをどうすればよいのか。

筆者は、メタデータはエディンバラ大学のような基本的な要素にとどめ、研究データの作成者による「データ記述 (data description)」を充実させることを提案したい。この点について、データジャーナルの例から解説する。

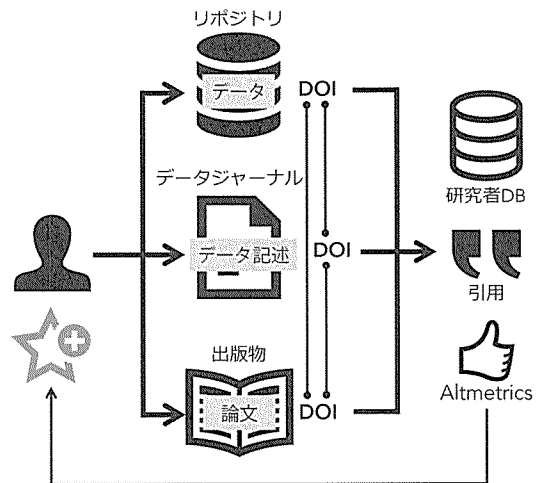
公開した研究データの登録先や、収集方法などをまとめたデータ記述を掲載するオープンアクセスのデータジャーナルが相次いで刊行されている³⁶⁾ (図3)。このデータ記述は、研究データの詳細な内容を表しているという意味でメタデータの役割を果たしていると言えるだろう。事実、Ubiquity Press は、自社が発行するデータジャーナルを「メタジャーナル」と称している³⁷⁾。

さて、図2のDataShareの検索結果画面にも、データセット記述(抄録)が表示されている(点線部分)。この部分をデータジャーナルのデータ記述のように充実させることで詳細なメタデータに代えられないだろうか。

さらにDOIで論文や研究者データベースへのリンクを確保すれば、論文の本文や著者情報が、より詳細なメタデータの役割を果たすだろう。完璧なメタデータを単一のリポジトリ内に作成しようとするのではなく、外部の情報で補完することによって、効率的にデータの発見可能性を高め、内容の理解を深めることができる。と考える。

筆者は元図書館員として、メタデータをシンプルにしようと提案することに不安も覚える。かつてダブリンコア (DC) の15エレメントを策定する過程では、より詳細に作り込むよう主張した情報専門家もいたそうである。しかし、エレメントは「コア」なレベルにとどめ、かつ詳細な記述も可能とした結果として、DCが長期にわたって広く使われる語彙となったのは歴史が示すとおりである。RDMサービスの障壁とならないよう、メ

図3 データ出版と学術情報流通³⁸⁾



タデータの必須項目は基本的なものにとどめつつ、DCに倣って拡張性を備えておくことが重要だと思われる。

いずれにせよ、利用者にとって究極の理想は世界中のデータをワンストップで検索できることだろう。そのためには相互運用性が高いメタデータスキーマを検討する必要があるが、RDAの「メタデータ標準ディレクトリWG」³⁹⁾やJiscによる「研究データレジストリとディスカバリーサービス」⁴⁰⁾の活動が参考になるだろう。

おわりに

研究者や図書館員の方から「研究データを公開しても、果たして役に立つのか?」と尋ねられることも多い。確かに現状では、公開データの再分析や追試には限界がある。また、図3に示したデータ引用や評価の取り組みもはじまったばかりである。

ルイスと訪れた東京大学附属図書館では、杉田玄白らの『解体約図』が特別展示されていた⁴¹⁾。当時の最新医学を伝えるこの書は、関東大震災当時は貴重資料とみなされておらず、全壊した図書館とは別の倉庫に入れられていたため難をのがれたという。作者の死後、数十年経ってから評価された著作物や芸術作品は枚挙にいとまがない。か

つての『解体約図』のように、未整備のまま放置されている「ダークデータ」⁴²⁾の中には、数十年後に価値が見いだされるものが含まれているかもしれない。

『図書館学序説』でバトラーは、図書館は“人類の記憶を保存する一種の社会的メカニズムである”と表現している⁴³⁾。研究データも人類の記憶の一つとして、図書館が長期にわたって保存し、共有の財産として公開する必要があるのではないだろうか。

謝辞

スチュワート・ルイス氏には、2013年8月の訪問調査をはじめ、本稿にも全面的にご協力いただきました。また、機関リポジトリ推進委員会とデジタルリポジトリ連合(DRF)の皆さま、筑波大学の逸村裕教授、杉本重雄教授、国立国会図書館の福山樹里氏から多くのご助力と示唆をいただきました。心よりお礼申し上げます。

<参考文献・注>

- 1) Vines, Timothy H : Albert, Arianne Y. K : Andrew, Rose L : Debarre, Florence : Bock, Dan G : Franklin, Michelle T : Gilbert, Kimberly J : Moore, Jean-Sebastien : Renaut, Sebastien : Rennison, Diana J. The availability of research data declines rapidly with article age. *Current Biology*. Vol. 24, No. 1, p.94-97. doi : 10.1016/j.cub.2013.11.014
- 2) Vines Timothy H : et. al. 2013. Data from : The availability of research data declines rapidly with article age. Dryad Digital Repository. doi : 10.5061/dryad.q3g37
- 3) Thomson Reuters. Unlocking the Value of Research Data : A Report from the Thomson Reuters Industry Forum July 2013. 2013, p.3.
<http://researchanalytics.thomsonreuters.com/m/pdfs/1003903-1.pdf>, (accessed 2014-12-4)
- 4) 下記文献によれば、トップ40カ国の研究開発費の対GDP比は平均2.0%、日本は3.4%である。2014 Global R&D Funding Forecast. 2013, 35p.
http://www.rdmag.com/sites/rdmag.com/files/gff-2014-5_7%20875x10_0.pdf, (accessed 2014-12-4)
- 5) ジャック・アンドレイク : 有望な膵臓がん検査—なんとティーンエイジャーが開発。TED. 2013-2.
http://www.ted.com/talks/jack_andraka_a_promising_test_for_pancreatic_cancer_from_a_teenager?language=ja, (accessed 2014-12-4)
- 6) 日本でも、総務省が「データサイエンス・スクール：統計力向上サイト」を開設したり、日本版MOOCのgaccoで「統計学I：データ分析の基礎」を開講するなど、データサイエンティストの育成が盛んである。
<http://www.stat.go.jp/dss/>,
https://lms.gacco.org/courses/gacco/ga014/2014_11/about, (accessed-2014-12-4)
- 7) 宮入暢子. オープンサイエンスと科学データの可能性. 情報管理. Vol. 57, No. 2, 2014, p. 80-89. doi : 10.1241/johokanri.57.80
- 8) Research Data Alliance (RDA)
<https://rd-alliance.org>
- 9) DataCite
<https://www.datacite.org/node>
- 10) Digital Curation Centre (DCC)
<http://www.dcc.ac.uk>
- 11) Australian National Data Service (ANDS)
<http://www.ands.org.au>
- 12) Ikeuchi, Ui. Research Data Sharing and Open Science. figshare. 2015.
doi : /10.6084/m9.figshare.1270371
- 13) New York University Health Science Library. Data Sharing and Management Snafu in 3 Short Acts. YouTube. 2012-12-19.
<http://www.youtube.com/watch?v=N2zK3sAtr-4>, (accessed 2014-12-4)
- 14) Establishing Incentives and hanging Cultures to Support Data Access. 2014. 28p.
http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf, (accessed 2014-12-4)
- 15) Ferreira, Filipe. et al. Data governance in e-science. LIBER conference 2014.
http://liber2014.1nb.lv/wp-content/uploads/sites/9/2014/04/4_July_11_1_Raquel_Bairrao.pdf, (accessed 2014-12-4)
- 16) 高祖歩美. 生命科学分野におけるデータの共有の現状と課題. 情報管理. Vol. 56, No. 5, 2013, p.294-301. doi : 10.1241/johokanri.56.294
- 17) “「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」の決定について”. 文部科学省. 2014-8-26.
http://www.mext.go.jp/b_menu/houdou/26/08/1351568.htm, (accessed 2014-12-4)
- 18) オープンデータに関する権利と義務：本格的なデータジャーナルに向けて. 日本学術会議情報学委員会国際サイエンスデータ分科会. 2014, 15p.
<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h140930-3.pdf>, (accessed 2014-12-4)
- 19) Piwowar, Heather A : Vision, Todd J. Data reuse and the open data citation advantage. PeerJ. 2013, 1 : e175. doi : 10.7717/peerj.175 ほか
- 20) “World University Rankings 2014-2015”. Times Higher Education.
<http://www.timeshighereducation.co.uk/world-university-rankings/2014-15/world-ranking/methodology>, (accessed

- 2014-12-4)
- 21) Lewis, Stuart. Research Data Management : Edinburgh University Library Experience. 第16回国書館総合展「大学の知の発信システムの構築に向けて—機関リポジトリの新たな可能性を探る, 2. 研究データへのアプローチ: エジンバラ大学図書館の実践事例を中心に」. 神奈川, パシフィコ横浜. 2014-11-6. [機関リポジトリ推進委員会コンテンツ WG による日本語訳つき]
<http://www.research.ed.ac.uk/portal/en/publications/research-data-management-edinburgh-university-library-experience> (59147c46-1c47-4e70-8bd5-b1da24c9218f) .html , (accessed 2014-12-4)
- 22) Research Data Management (RDM) Roadmap : August 2012 - May 2015. Information Services RDM Policy Implementation Committee, University of Edinburgh. January 2014 : Ver. 1.2.
http://www.ed.ac.uk/polopoly_fs/1.1168731/fileManager/UoE-RDM-Roadmap-140106.pdf, (accessed 2014-12-4)
- 23) Lewis, Stuart. "Open data repository - file size analysis" . Edinburgh Research Data Blog. 2014-6-14.
<http://datablog.is.ed.ac.uk/2014/06/19/open-data-repository-file-size-analysis/>, (accessed 2014-12-4)
- 24) Joint Academic Coding System (JACS) Version 3.0. Higher Education Statistics Agency : HESA.
<https://www.hesa.ac.uk/content/view/1776/649/>, (accessed 2014-12-4)
- 25) 最新版は, DataCite Metadata Schema Repository (<http://schema.datacite.org>) から入手できる。
- 26) MANTRA : Research Data Management Training. University of Edinburgh.
<http://datalib.edina.ac.uk/mantra/>, (accessed 2014-12-4)
- 27) Rice, Robin. "New release of Research Data MANTRA (Management Training) online course".
<http://datablog.is.ed.ac.uk/2014/09/05/new-release-of-research-data-mantra-management-training-online-course/>, (accessed 2014-12-4)
- 28) Information Services. The University of Edinburgh. Edinburgh Research Data Blog. 2014-9-5.
<http://www.ed.ac.uk/schools-departments/information-services>, (accessed 2014-12-4)
- 29) QS World University Rankings 2014/15. QS : Top Universities. エディンバラ大学は17位。
<http://www.topuniversities.com/qs-world-university-rankings>, (accessed 2014-12-4)
- 30) "機関リポジトリ一覧". 学術機関リポジトリ構築連携支援事業.
<http://www.nii.ac.jp/irp/list/>, (accessed 2014-12-4)
- 31) ルイスによれば, 歴史的経緯から IR とは別に DataShare を構築したが, やり直すならば IR を活用するだろうとのことである。なお, 英国の大学の8割が IR を利用しており, 残る2割が ckan などを使ってデータ専用のリポジトリを構築しているという。
- 32) 資源タイプ別コンテンツ数内訳 (割合) [2014年11月分]. IRDB コンテンツ分析.
<http://irdb.nii.ac.jp/analysis/index.php>, (accessed 2014-12-4)
- 33) "Members". DataCite.
<https://www.datacite.org/about-datacite/members>, (accessed 2014-12-4)
- 34) ジャパンリンクセンター運営委員会. "研究データへの登録実験プロジェクト計画書". ジャパンリンクセンター (JaLC). 2014-9-10.
http://japanlinkcenter.org/top/doc/JaLC_kouboH26-1-0.pdf, (accessed 2014-12-4)
- 35) 教育研究の革新的な機能強化とイノベーション創出のための学術情報基盤整備について—クラウド時代の学術情報ネットワークの在り方— (審議まとめ). 文部科学省科学技術・学術審議会学術分科学術情報委員会, 2014-8.
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/031/houkoku/1351113.htm, (accessed 2014-12-4)
- 36) Hrynaskiewicz, Ian ; Shintani, Yoko. Scientific Data : データの再利用を促進するオープンアクセス・オープンデータジャーナル. 情報管理. Vol. 57, No. 9, 2014, p. 629-640. doi : 10.1241/johokanri.57.629
- 37) "Publishing with Ubiquity Press". Ubiquity Press.
<http://www.ubiquitypress.com/site/publish/>, (accessed 2014-12-4)
- 38) Ikeuchi, Ui. Data publishing and scholarly communications. figshare. 2015.
 doi : /10.6084/m9.figshare.1270372
- 39) Metadata Standards Directory Working Group. Research Data Alliance.
<https://rd-alliance.org/groups/metadata-standards-directory-working-group.html>, (accessed 2014-12-4)
- 40) Jisc Research Data Registry and Discovery Service. Digital Curation Centre.
<http://www.dcc.ac.uk/projects/research-data-registry-pilot>, (accessed 2014-12-4)
- 41) 平成26年度東京大学附属図書館特別展示 : 展示資料リスト. 東京大学附属図書館.
<http://www.lib.u-tokyo.ac.jp/tenjikai/tenjikai2014/list.pdf>, (accessed 2014-12-4)
- 42) Heidorn, P. Bryan. Shedding light on the dark data in the long tail of science. *Library Trends*. Vol. 57, No. 2, 2008, p.280-299. doi : 10.1353/lib.0.0036
- 43) Butler, Pierce. 図書館学序説. 藤野幸雄訳. 日本図書館協会. 1978, p.23. (2014.12.11 受理)