

CA1825

動向レビュー

オープンデータと図書館

おおむかいっ き
大向一輝*

1. はじめに

ネットワークの高速化やサーバ・ストレージの低廉化に伴い、ウェブにおける情報公開のコストは低減し続けている。その中で、いわば完成品の情報である文書・文献だけを公開するのではなく、それらを作成するための基礎資料やデータを同時に共有する事例が増加している。これらの情報を参照することで、元の文書・文献の信頼性を確認することが可能になるとともに、さまざまな情報源からのデータを組み合わせる新たな知見を引き出すことや、新規ビジネスの基盤になることが期待される。近年、このようなデータの公開・共有の取り組みがオープンデータと呼ばれ、主に政府・自治体や学術コミュニティにおける活動が注目されている。本稿ではオープンデータの現状について概説し、図書館による支援あるいは貢献の可能性について述べる。

2. オープンデータの基礎

オープンデータは字義通りに解釈すれば「データの公開」にすぎないが、より詳細には「ウェブ上での再利用性の高いデータの公開」と定義できる。再利用性には制度と技術の2つの観点がある。制度面での再利用性とは、データの2次利用や再配布に制約が設けられていないことを指す。英国 Open Knowledge Foundation (OKFN) は、商用・非商用を問わず誰もが利用・再利用・再配布できることがオープンデータの定義であり、その他の制約としてはクレジットの明記と再配布に対して同様の利用条件を課すことのみが可能であるとしている⁽¹⁾。技術面では、公開されたデータがコンピュータで容易に取り扱えることが重要である。例えば紙の文書をスキャンした画像データなどは再利用性が低いと、オープンであるとは言い難い。

これらの諸条件をわかりやすく整理したのがウェブの開発者であるティム・バーナーズ＝リーの5 star Open Dataである⁽²⁾。5 star Open Dataは5つ星でオープン化のステップを示す枠組みである。1つ星に認められるためにはオープンライセンスが必要である。オープンライセンスは、制度面でのオープンさを満たすために、データ提供元が提示する利用者への事前許諾である。ライセンスは提供元が独自に定めることが

できるが、他サイトとの互換性を重視して Creative Commons⁽³⁾ や Open Data Commons (ODC)⁽⁴⁾ が採用される例が多い。なお、米国では政府・行政機関が提供するデータは原則としてパブリックドメインに置かれており、ライセンスを明示する必要はない。

2つ星ならびに3つ星は技術面での再利用性に関する指標である。プログラムで処理できるようにデジタル化されたデータには2つ星が、データの形式が特定の商用プログラムに依存していないものには3つ星が与えられる。3つ星の例としてはカンマ区切りテキスト (CSV) や XML が挙げられる。4つ星、5つ星は Linked Open Data に関する指標であるが、本稿では詳述せず、別稿⁽⁵⁾を参照されたい。

オープンデータ自体は分野を限定するものではないが、現状では政府・自治体を中心とした公共セクターと、大学・研究機関等を含む学術コミュニティの動きが目立っている。前者においては政府・自治体の透明性を高め、市民との対話・協働による行政の実現を目指すオープンガバメント運動の実現手段のひとつとして認識されており、全世界的にデータ公開の動きが広まった。政府レベルでは米国の Data.gov⁽⁶⁾ や英国の data.gov.uk⁽⁷⁾ など、先進諸国はいずれもポータルサイトを開設し、オープンデータに関する取り組みの紹介、データの検索機能を提供している。日本でも電子行政オープンデータ戦略⁽⁸⁾に基づき各府省のデータのオープン化が順次進められており、これらを一元的に検索・アクセスするためのデータカタログサイト試行版が開設された⁽⁹⁾。2013年のG8サミットではオープンデータ憲章が採択され、行動計画の策定や進捗の確認が義務づけられた⁽¹⁰⁾。

自治体の取り組みは国内・国外を問わず年々増加している。この勢いを象徴するイベントとして、2014年2月22日に行われた International Open Data Day (IODD) を挙げる⁽¹¹⁾。IODDはオープンデータ関連の催しを都市ごとに同時多発的に行うもので、2013年度は102都市の参加であったものが、2014年度は194都市にまで増加している。日本からはそれぞれ8都市から32都市に拡大しており、伸び率が最も高い。

学術コミュニティでは、研究活動の過程で得られるデータを学術論文と同様に共有し、参照・引用による評価のサイクルを形成する動きが活発化している。その意味では研究データを対象としたオープンアクセスと捉えることも可能である。研究データ共有においてはデータ管理のためのポリシー策定や技術的な支援、人材育成などの課題について図書館の参画が期待されている (CA1818 参照)。

オープンデータが利活用されるためには、単に公開するだけでなく市民や開発者のコミュニティを巻き込

*国立情報学研究所

むことが重要である。データ活用アイデアを募るアイデアソン、アプリケーションを短期間に開発するハッカソンといった形式のワークショップが各地で行われている。また、オープンデータに特化したアプリケーションコンテストも多数開催され、活況を呈している⁽¹²⁾。

3. オープンデータと図書館の関係

3.1 情報源としての図書館

図書館コミュニティにおいては、書誌データ・典拠データのオープン化が積極的に進められている。米国議会図書館をはじめとする各国の国立図書館、OCLC、米国デジタル公共図書館(DPLA: E1429参照)といった大規模なデータベースにおいてCreative CommonsあるいはODCのライセンスが付与され、自由な利活用が認められた⁽¹³⁾⁽¹⁴⁾。データの公開と並行して、各国国立図書館やOCLCの連携によるバーチャル国際典拠ファイル(VIAF)⁽¹⁵⁾や、EU参加国の図書館・美術館・博物館によるデジタルアーカイブEuropeana⁽¹⁶⁾といった国際的なプロジェクトが立ち上がり、図書館データのプレゼンスを高めている。

国内では2010年に国立国会図書館が典拠データを公開し⁽¹⁷⁾、2012年にはVIAFに参加した。「国立国会図書館の書誌データ作成・提供の新展開(2013)」⁽¹⁸⁾では書誌データの開放性を主要テーマのひとつとして掲げている。また国立情報学研究所が運営し、大学図書館が参加する総合目録データベースNACSIS-CATでは、書誌データのオープン化の方針が決定され、ライセンスや公開方法に関する詳細が検討されている⁽¹⁹⁾。

デジタルアーカイブの分野でもオープンライセンスの適用が進んでいる。京都府立総合資料館の東寺百合文書WEBは書誌だけでなく画像データに対してもCreative CommonsのCC BY(クレジットの明記のみを求める)が適用されており、自由な利用・改変が認められている⁽²⁰⁾。

3.2 利用者支援としての図書館

以上のように、オープンデータの情報源としての図書館の役割は拡大していくと思われる。またこれらのデータを利用したアプリケーションも増加傾向にある。一方、データの利用者として、あるいは利用者の支援機関としての図書館の役割については十分な議論がなされていない。文献あるいは視聴覚資料以外の情報源として入手可能になったオープンデータをどのように捉えるべきであろうか。

例えば、白書などの政府刊行物はこれまでもオンライン、オフラインを問わず入手・閲覧することが可能であったが、その後の利用方法を含めて考慮されてい

たとはいない。オープンデータとして公開されている総務省の情報通信白書では、本文がテキストデータとしてコピーできるだけでなく、文中の図表については画像の他に数値データをCSV形式で入手することができる⁽²¹⁾。さらに詳細な情報が必要な場合には政府のデータカタログサイトを通じて元データを得ることもできる。これらはすべてオープンライセンスが付与されており、利用者はクレジットを明示するのみでそのまま再利用、再配布が可能である。

数値データが入手できれば、表計算ソフト等を使って複数のデータを組み合わせ、付属の可視化ツールでわかりやすく表現することも容易である。例として、筆者が作成した人口ピラミッドの時間的推移の可視化を示す⁽²²⁾。元データは実績値が統計センターが公開する国勢調査のデータから「年齢(5歳階級)、男女別人口及び人口性比-全国(大正9年~平成22年)」⁽²³⁾、将来予測が国立社会保障・人口問題研究所の将来推計人口から「男女年齢各歳別人口:出生中位(死亡中位)推計」⁽²⁴⁾である。双方のデータの形式が異なるため多少の整形が必要であるものの、表計算ソフトの基本的な操作を理解していれば簡単にインタラクティブなグラフを作成することができる。またデータ中の時間情報や位置情報を用いて時間軸あるいは地図にマッピングするTimeMapperはデータの時空間的分布を確認するのに有用である⁽²⁵⁾。このように、オープンデータを処理して利用者の課題や疑問を解決するようなスキルが今後のレファレンス業務において求められるのではないだろうか。

こういった一連の作業を行うためには、まずデータの所在を知ることが必要である。これまで公共セクターのデータの多くは組織の縦割り構造を反映してウェブサイトの中に散在していることが多かったが、データの管理に特化したポータルサイト(データカタログサイト)を立ち上げ、一括管理する事例が増えている。データカタログサイトでは個々のデータにメタデータが付与されており、名称や作成年月日、作成部署といった項目で検索することができる。経済産業省ではデータカタログサイトで掲載しているデータ以外にも、省全体で保有しているすべてのデータの棚卸しを行い、一覧を公開している⁽²⁶⁾。これらのデータは情報公開制度に基づく開示請求を行うことで入手できる可能性がある。

得られたデータの成り立ちや作成プロセスを理解することも利用者の支援につながる。ある統計情報が悉皆調査に基づいているのかサンプリングによるものなのかといった由来に関する情報や、各項目の分類の方法について知ることは、利活用のためにデータを加工するうえで重要な情報となる。オープンデータを用い

たレファレンス業務を行う際にはこのような背景情報の調査を含めて提供する必要があると思われる。

市民自らがオープンデータを作成する事例も増加している。図書館によるデータ作成の支援という観点では、Wikipedia タウンの試みが注目されている⁽²⁷⁾。Wikipedia タウンは、ボランティアが地域の観光施設や文化財を取材し、その結果を Wikipedia に掲載して共有することを目的とするワークショップである。Wikipedia では掲載される情報については何らかの裏付けが求められる。そこで、Wikipedia タウンでは現地取材の前後に図書館で関連資料を探し、その資料に基づいて記事を執筆するというプロセスが設けられている。そこでは図書館員の資料に対する知識やファシリテーション能力が求められており、今後も需要が高まるものと思われる。

2013年10月に起こったアメリカ政府機関の停止では、Data.gov のサイトもアクセス不能になり、データの入手が困難になった。また2014年4月には日本政府のデータカタログサイトが予算の関係から一時的に休止したことも記憶に新しい。このような事態は望ましいことではないが、それぞれのサイトに格納されているデータはオープンライセンスが付与されていることから、事前に第三者がバックアップしておき、代替サイトを立ち上げることも原理的には可能である。実際に日本政府の例では筆者が関わる Data for Japan コミュニティにおいてミラーサイトを立ち上げた⁽²⁸⁾。学術情報流通の分野では電子ジャーナルの長期保存を目的として国際的にアーカイブを持ち合う CLOCKSS⁽²⁹⁾ が存在するが、それに類する枠組みによってデータの永続性を担保することが図書館コミュニティにも求められるようになるのではないか。

いまや多くの文献がデジタル化され、単なるキーワード検索で探せるような情報はユーザ自身が発見できるようになった。今後は複数の情報源を組み合わせる新たな知識を作り出すことや、利用者の情報発信を支援することが図書館のミッションになるだろう。その変化の中でオープンデータの重要性はますます高まるものと期待される。

- http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryoutu.pdf, (参照 2014-05-07).
- (9) データカタログサイト (試行版).
http://data.go.jp, (参照 2014-05-16).
 - (10) "オープンデータ憲章 (概要)". 外務省.
http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html, (参照 2014-05-07).
 - (11) International Open Data Day.
http://opendataday.org, (accessed 2014-05-07).
 - (12) Linked Open Data チャレンジ Japan.
http://lod.sfc.keio.ac.jp, (accessed 2014-05-07).
 - (13) "Data licenses & attribution". OCLC.
http://www.oclc.org/data/attribution.en.html, (accessed 2014-05-07).
 - (14) "Policies". Digital Public Library of America.
http://dp.la/info/about/policies/, (accessed 2014-05-07).
 - (15) VIAF.
http://viaf.org, (accessed 2014-05-07).
 - (16) Europeana.
http://www.europeana.eu, (accessed 2014-05-07).
 - (17) Web NDL Authorities. 国立国会図書館.
http://id.ndl.go.jp/auth/ndla, (参照 2014-05-07).
 - (18) "国立国会図書館の書誌データ作成・提供の新展開 (2013)". 国立国会図書館.
http://www.ndl.go.jp/jp/library/data/shintenkai2013.pdf, (参照 2014-05-07).
 - (19) "総合目録データベースのデータ公開方針". 国立情報学研究所.
http://www.nii.ac.jp/CAT-ILL/about/infocat/od/, (参照 2014-05-07).
 - (20) 東寺百合文書 WEB. 京都府立総合資料館.
http://hyakugo.kyoto.jp, (参照 2014-05-07).
 - (21) 情報通信白書. 総務省.
http://www.soumu.go.jp/johotsusintokei/whitepaper/, (参照 2014-05-07).
 - (22) "人口ピラミッド (国勢調査・将来人口推計)". 大向一輝.
http://bit.ly/japan-population, (参照 2014-05-07).
 - (23) "年齢 (5歳階級), 男女別人口及び人口性比 - 全国 (大正9年~平成22年)". 統計センター.
http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000001085926, (参照 2014-05-07).
 - (24) "男女年齢各歳別人口: 出生中位 (死亡中位) 推計". 国立社会保障・人口問題研究所.
http://www.ipss.go.jp/pp-newest/j/newest03/02_syosai/01/Mm1-9.xls, (参照 2014-05-07).
 - (25) TimeMapper. Open Knowledge Foundation Labs.
http://timemapper.okfnlabs.org, (accessed 2014-05-07).
 - (26) "オープンデータに関する調査研究 (2012年度)". 経済産業省.
http://datameti.go.jp/data/ja/dataset/report-001-2012, (参照 2014-05-07).
 - (27) "YOKOHAMA International Open Data Day 2014 [Wikipedia Town]レポート". 横浜オープンデータソリューション発展委員会.
http://yokohamaopendata.jp/?p=239, (参照 2014-05-07).
 - (28) datago.jp.
http://datago.jp, (accessed 2014-05-07).
 - (29) CLOCKSS.
http://www.clockss.org, (accessed 2014-05-07).

[受理: 2014-05-19]

Ohmukai Ikki.
Open Data in Libraries.

- (1) "Open Definition". Open Knowledge Foundation.
http://opendefinition.org, (accessed 2014-05-07).
- (2) "5 star Open Data".
http://5stardata.info, (accessed 2014-05-07).
- (3) Creative Commons.
http://creativecommons.org, (accessed 2014-05-07).
- (4) Open Data Commons.
http://opendatacommons.org, (accessed 2014-05-07).
- (5) 大向一輝. オープンデータと Linked Open Data. 情報処理. 2013, Vol. 54, No. 12, p. 1204-1210.
- (6) Data.gov.
http://www.data.gov, (accessed 2014-05-07).
- (7) data.gov.uk.
http://data.gov.uk, (accessed 2014-05-07).
- (8) "電子行政オープンデータ戦略".