

## CA1750

### 英国とオランダの国立図書館にみる 新聞資料デジタル化プロジェクト

#### はじめに

新聞は発行当時の情報を調査するのに有効なツールであるが、新聞紙は劣化しやすく、長期保存が難しい。また、記事を探すにも掲載日等の詳細な情報が無い場合、情報へのアクセスが難しい資料である。しかし近年、古い時代の新聞資料についてもデジタル化が進み、キーワードによる記事の検索や、該当部分の紙面画像閲覧も可能となり始めている。新聞記事へのアクセスを容易にするツールとして利用者のニーズは高まっており、新聞資料のデジタル化を進める国が増えてきている (CA1577 参照)。

本稿では、筆者が2011年3月中旬に訪問し、見学とインタビューを行った、英国図書館 (BL) とオランダ王立図書館 (KB) における新聞のデジタル化プロジェクトについて紹介する。

#### 1. BL の新聞資料デジタル化プロジェクト

##### 1.1. 概要

BL は、歴史的に重要な新聞をウェブ上で提供し、学術コミュニティが記事の全文検索を行えるようにすることを目的として、「英国新聞 1800-1900 年」<sup>(1)</sup> (British Newspapers 1800-1900 : BN) プロジェクト (CA1577 参照) を 2004 年から 2007 年にかけて実施した<sup>(2)</sup>。BL が所蔵する、英国で発行された 1800 年から 1900 年までの新聞 49 タイトル、およそ 200 万ページがデジタル化され、記事全文検索および紙面画像閲覧が可能なデータベースとして公開されている。BL 館内のデータベース用端末および英国内の高等教育機関等では無料で利用ができる (ただし、高等教育機関の実際の利用に際しては、Gale 社から許諾を得る必要がある)。なお、この中には、既に新聞社等によって



図1 BL新聞図書館入口

商業目的でデジタル化がなされたもの、例えば Times 等は含まれていない<sup>(3)</sup>。

また、BL は 2010 年 5 月に、brightsolid 社と協力関係を結び、10 年間で 4,000 万ページ以上の新聞をデジタル化するという非常に大規模なプロジェクトを発表した<sup>(4)</sup>。現時点で “British Newspaper Archive” (BNA) という名称でウェブサイトが開設されており<sup>(5)</sup>、一部紙面画像も公開されている。2011 年秋には、このプロジェクトの成果として 19 世紀までの新聞 100 万ページが公開予定である<sup>(6)</sup>。

#### 1.2. デジタル化の実作業の流れ

1.1. の BN プロジェクトは、BL が Gale 社と提携して行ったプロジェクトである。デジタル化にかかる 200 万ポンド (当時のレートで約 4 億円) の費用は、英国情報システム合同委員会 (JISC)<sup>(7)</sup> のデジタル化プログラムの予算 2,200 万ポンドの中から拠出された。データのアップロードに関する費用は Gale 社が全額負担した<sup>(8)</sup>。オンラインユーザーは一部のタイトルを除き、紙面画像のダウンロードには有料のパスが必要である。

デジタル化には全てマイクロフィルムが使用され、原紙のみの資料や、劣化の進んでいたフィルムは、デジタル化のために新たにフィルムが作製された<sup>(9)</sup>。当時は原紙を使用したデジタル化は品質が悪く、マイクロフィルムからデジタル化を行わざるを得なかったからだという<sup>(10)</sup>。

BNA プロジェクトは、BL が brightsolid 社に協力するという位置づけで行われている。最初の 2 年間で 400 万ページ以上のデジタル化が予定されているこのプロジェクトでは、マイクロフィルムと原紙の両方をデジタル化の対象としている<sup>(11)</sup>。ページ毎の検査等のスキャニング準備作業、スキャニング、データのアップロードは brightsolid 社が行うこととなっており、費用は全て同社が負担している<sup>(12)</sup>。デジタル化の実作業は、ロンドン北部コリンデールにある、対象資料を保管している BL 新聞図書館<sup>(13)</sup> 内で行われており、製本済みの新聞原紙は解体せずに、1 日あたり 8,000 ページがスキャニングされている<sup>(14)</sup>。スキャニングによって作成されたデータに欠号や欠ページ、破損等の情報を付与し、画像調整等を経て、提供される流れとなっている。

#### 1.3. 著作権上の問題への取り組みと今後

BNA プロジェクトでは著作権保護期間内の資料もデジタル化の対象になっているが、そのための著作権者との協議は brightsolid 社が行っている<sup>(15)</sup>。このプロジェクトではできるだけ多くのタイトルのデジタル

化を目指しており、19世紀までの新聞に次いで20世紀の新聞のデジタル化に着手し、2011年からの2年間でさらに数百タイトルのデジタル化が行なわれる予定である。BNAによるデジタル化資料はBL館内では紙面画像の閲覧を含めて無料で利用できるが<sup>(16)</sup>、オンラインユーザーは記事検索のみ無料で、紙面画像の閲覧は有料となる。また、オンラインユーザーがウェブ上でアクセスできるデータには著作権保護期間内の資料も含まれる予定であるという<sup>(17)</sup>。

BLは、所蔵資料の中でも特に新聞資料の劣化が著しいという理由から<sup>(18)</sup>、2008年から2011年にかけてデジタル化を進める資料の中で新聞資料を最優先順位に挙げている(E832参照)。また、2011年から2015年の戦略的優先事項<sup>(19)</sup>のひとつ「研究を望む者は誰でもアクセスできるようにする」(Enable access to everyone who wants to do research)に、brightsolid社との協力による新聞2,000万ページのデジタル化を挙げている(E1163参照)。

## 2. KBの新聞資料デジタル化プロジェクト

### 2.1. 概要

KBは、重要な資料である新聞をデジタル化によってウェブサイト上で提供し、誰でもアクセス可能にすること、原資料を長期保存すること<sup>(20)</sup>を目的として、「日刊紙デジタルデータバンク」(Databank Digitale Dagbladen : DDD)プロジェクト<sup>(21)</sup>を2007年に開始した。

DDDプロジェクトは1618年から1995年の期間にオランダと旧植民地で発行された新聞の中から選別されたタイトル、およそ800万ページを対象としたプロジェクト<sup>(22)</sup>である。1618年から1945年までの期間だけで1,400タイトルが対象となっており<sup>(23)</sup>、最終的には、1,736タイトル、約918万ページがデジタル化されるという<sup>(24)</sup>。2010年6月に約100万ページがウェブサイト上で公開され<sup>(25)</sup>、デジタル化された資料のうち、2011年7月現在、1618年から1945年までの期間の記事全文検索および紙面画像の閲覧が可能である。今後、残りのページが順次公開されるほか、発行地や記事の種類等による絞り込み検索機能の実装も予定されている。

### 2.2. デジタル化の実作業の流れ

DDDプロジェクトは1999年から2004年に約35万ページ(対象期間は1910年から1945年まで)の新聞デジタル化を試行的に実施した<sup>(26)</sup>のちに、2007年から5年計画で開始<sup>(27)</sup>された。プロジェクトの予算は5年間で1,250万ユーロ(約15億円)である。

まず、歴史家、ジャーナリズム研究者等の専門家



図2 KB概観

による諮問委員会<sup>(28)</sup>が基準を定めてタイトルを選別し、デジタル化候補のリストを作成する。それを元に、KBが著作権の確認作業、デジタルデータの有無、フィルムや原紙の所蔵の確認、そして他機関の所蔵の確認を行ったのち、デジタル化が行われる<sup>(29)</sup>。

著作権保護期間内の資料は著作権保持者の許諾を取得しなければデジタル化はできない。そのため、許諾を取得できなかった資料については対象から外される。

また、デジタル化は、KBの所蔵資料だけでなく、オランダ国内の機関のほか、ヴァチカン図書館やBL、スリナム国立公文書館等の国外も含め30以上の機関<sup>(30)</sup>の所蔵資料を借りて、既存のデジタルデータ、マイクロフィルム、原紙から行われている。既存のデジタルデータはそのまま流用するのではなく、KBの規格に修正のうえ、デジタル化作業に利用された。ページ毎の検査や補修作業、号(issue)のメタデータ入力等の品質管理が行われたのち、資料とメタデータは委託業者へ送られる。業者によってスキャニングや光学式文字認識(Optical Character Recognition : OCR)処理等が施され、画像の調整やページの順序等の検査が行われると、KBへと送られ、利用者に提供される流れとなっている。

### 2.3. 著作権上の問題への取り組みと今後

DDDプロジェクトでは著作権保護期間内の資料も対象となっている。そのため、KBは新聞団体や著作権保持者の代表団体等、各団体と協議を行い、2011年3月時点で、15の新聞社から、102タイトルのデジタル化を無償で行う許諾を得ており、現在も交渉が進められているとのことであった<sup>(31)</sup>。

また、KBは2010年から2013年までの戦略計画<sup>(32)</sup>で、1470年以降にオランダで出版された全ての図書、雑誌、そして新聞をデジタル化することを挙げている(E1017参照)。KBの担当者によると、2012年には新



たな新聞デジタル化プロジェクトを予定しており、準備を進めているという。

### おわりに

今回紹介した2館のプロジェクトでは、BLが民間企業との提携により非常に大規模なデジタル化を進めている点や、KBが外部機関と協力して所蔵資料以外にもデジタル化している点等の相違点もあるが、資料へのアクセスを向上させ、保存と両立させることを目的としている点と、著作権保護期間内の資料もデジタル化の対象としている点で共通している。

国立国会図書館(NDL)の新聞デジタル化は、技術的な問題や著作権法上の問題等もあり、現時点では実施には至っていない。もちろん両館の事例は単純にNDLの問題解決につながるものではない。例えば、アルファベットが使用されている資料はOCRでの可読性が高いが、日本の古い新聞には漢字と仮名に加えて旧字体や略字等が交じっており、同程度の成果を期待するのは難しい。しかし、著作権上の問題への取り組みには、今回紹介した2館から学べる点は多いと思われる。どちらの図書館もデジタル化後の原資料は保存し続けるとのことだが、劣化状態を考えると、原紙からのデジタル化に残された時間は少ない。日本も対応を急ぐべきであろう。

(主題情報部新聞課：佐々木美穂)

- (1) "British Newspapers 1800-1900". British Library. <http://newspapers.bl.uk/blcs/>, (accessed 2011-08-03).
- (2) "19th century newspapers". JISC. <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/bln.aspx>, (accessed 2011-08-03).
- (3) "Frequently Asked Questions". British Newspapers 1800-1900. [http://newspapers.bl.uk/blcs/blcs\\_25.htm](http://newspapers.bl.uk/blcs/blcs_25.htm), (accessed 2011-08-03).
- (4) "British Library and brightsolid partnership to digitise up to 40 million pages of historic newspapers". British Library. <http://pressandpolicy.bl.uk/Press-Releases/British-Library-and-brightsolid-partnership-to-digitise-up-to-40-million-pages-of-historic-newspapers-271.aspx>, (accessed 2011-08-03).
- (5) British Newspaper Archive. <http://www.britishnewspaperarchive.co.uk/>, (accessed 2011-08-03).
- (6) British Newspaper Archive. <http://www.britishnewspaperarchive.co.uk/>, (accessed 2011-08-06).
- (7) "Who we are". JISC. <http://www.jisc.ac.uk/aboutus/whoweare.aspx>, (accessed 2011-08-03).
- (8) BNプロジェクトと同様に、JISCの助成を受けGale社と提携して実施したBLによる新聞デジタル化事業には、「17-18世紀パーニーコレクションデータベース」(17th and 18th Century Burney Collection Database)と「英国新聞1620-1900年」(British Newspapers 1620-1900)がある。前者は17~18世紀に発行された新聞とニュースパンフレットを集めたBL所蔵のCharles Burneyコレクションをデジタル化したフルテキストデータベースである。後者は前者とBNプロジェクトの拡充プロジェクトであり、およそ100万ページがデジタル化された。BNプロジェクトと同様に、BL館内のデータベース用端末および英国内の高等教育機関等では無償で利用できる(ただし、高等教育機関の実際の利用に際しては、Gale社から許諾を得る必要がある)。「17th and 18th Century Burney Collection Database». British Library. <http://www.bl.uk/reshelp/findhelprestype/news/newspdigproj/burney/index.html>, (accessed 2011-08-03).
- (9) "British newspapers 1620-1900". JISC. <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/newspapers2.aspx>, (accessed 2011-08-03).
- (10) "Frequently Asked Questions". British Newspapers 1800-1900. [http://newspapers.bl.uk/blcs/blcs\\_25.htm](http://newspapers.bl.uk/blcs/blcs_25.htm), (accessed 2011-08-03).
- (11) 2011年3月にBLのデジタル化プロジェクト担当者に行ったインタビューに基づく。
- (12) "British Library and brightsolid partnership to digitise up to 40 million pages of historic newspapers". British Newspaper Archive. <http://www.britishnewspaperarchive.co.uk/archive-media.php>, (accessed 2011-08-03).
- (13) "News from the Crimea arrives at the digital age via British Library". Times. 2010-05-20. [http://technology.timesonline.co.uk/tol/news/tech\\_and\\_web/the\\_web/article7131113.ece](http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article7131113.ece), (accessed 2011-08-03).
- (14) 2011年3月にBLの新聞担当者に行ったインタビューによると、計画に遅れが生じているとの話であったが、以下に挙げた報道によると、BL新聞図書館は2013年に閉館を予定しているとのことである。Kynaston, David. "British Newspaper Library: Tough decisions to be made on hard copy". History Today, 2011, 61(8). <http://www.historytoday.com/david-kynaston/british-newspaper-library-tough-decisions-be-made-hard-copy>, (accessed 2011-08-03).
- (15) "Latest scanning news". British Newspaper Archive. <http://www.britishnewspaperarchive.co.uk/news-item.php?id=9>, (accessed 2011-08-03).
- (16) "News from the Crimea arrives at the digital age via British Library". Times. 2010-05-20. [http://technology.timesonline.co.uk/tol/news/tech\\_and\\_web/the\\_web/article7131113.ece](http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article7131113.ece), (accessed 2011-08-03).
- (17) "British Library creates a 'national memory' with digital newspaper archive". Guardian. 2011-05-30. <http://www.guardian.co.uk/media/2011/may/30/british-library-digital-newspaper-archive>, (accessed 2011-08-03).
- (18) 2011年7月にBLの新聞デジタル化プロジェクト担当者に行ったインタビューに基づく。
- (19) BLの資料保存部長であるノボトニー(Deborah Novotny)氏の資料保存環境に関する講演会が2011年2月17日にNDLで開催された。講演では、他の資料と比較して新聞資料の劣化の割合が最も高いことが指摘された。また、資料の劣化に関連して、講演会配布資料の図35-36にあるように、従来の新聞書庫では製本済み新聞が立てて排架されていることや、自重による損傷の激しい点が紹介された。今後の保管方法として、ウェストヨークシャーのポストンスバに建設予定のBL新書庫では、講演会配布資料の図43-44にあるように、資料を板で挟んでベルトで固定し、平積み状態でコンテナに保管する予定であることが紹介された。Novotny, Deborah. "Lecture 2: Stack Management and storage environment at the British Library". 国立国会図書館. 2011-02-17. [http://www.ndl.go.jp/jp/aboutus/pdf/forum2\\_no25.pdf](http://www.ndl.go.jp/jp/aboutus/pdf/forum2_no25.pdf), (accessed 2011-08-03).
- (20) "Growing Knowledge: British Library's Strategy 2011-2015". British Library. <http://portico.bl.uk/aboutus/stratpolprog/strategy1115/strategy1115.pdf>, (accessed 2011-08-03).
- (21) "Policy digitization". Koninklijke Bibliotheek. <http://www.kb.nl/hrd/digitalisering/beleid-en.html>, (accessed 2011-08-03).
- (22) "Databank Digitale Dagbladen". Koninklijke Bibliotheek. <http://www.kb.nl/hrd/digi/ddd/>, (accessed 2011-08-03).
- (23) "KB presenteert 400 jaar kranten online". Koninklijke Bibliotheek. 2010-5-27. [http://www.kb.nl/nieuws/2010/historische\\_kranten.html](http://www.kb.nl/nieuws/2010/historische_kranten.html), (accessed 2011-08-03).
- (24) "Selected titles and selection procedure: Selection process". Koninklijke Bibliotheek. <http://www.kb.nl/hrd/digi/ddd/selectie-en.html>, (accessed 2011-08-03).
- (25) 2011年7月にKBのDDDプロジェクト担当者に行ったインタビューに基づく。
- (26) "One million pages of Dutch historical newspapers

- online". Koninklijke Bibliotheek.  
[http://www.kb.nl/nieuws/2010/historische\\_kranten-en.html](http://www.kb.nl/nieuws/2010/historische_kranten-en.html), (accessed 2011-08-03).
- (26) "Project". Koninklijke Bibliotheek.  
<http://www.kb.nl/hrd/digi/ddd/aanpak-en.html>, (accessed 2011-08-03).
- (27) "Project: Planning". Koninklijke Bibliotheek.  
<http://www.kb.nl/hrd/digi/ddd/planning-en.html>, (accessed 2011-08-03).
- (28) "Selected titles and selection procedure". Koninklijke Bibliotheek.  
<http://www.kb.nl/hrd/digi/ddd/selectie-en.html>, (accessed 2011-08-03).
- (29) "Selected titles and selection procedure: Selection process". Koninklijke Bibliotheek.  
<http://www.kb.nl/hrd/digi/ddd/sproces-en.html>, (accessed 2011-08-03).
- (30) Faase, Jasper et al. "Quantity meets Quality: Towards a digital library". IFLA International Newspaper Conference 2011: Newspaper in Multiple Scripts and Multiple Languages: Issues and Challenges for National Heritage. Kuala Lumpur, Malaysia, 2011-04-25/27, IFLA. 2011.  
<http://www.ifla.org/files/newspapers/documents/IFLA%20International%20Newspaper%20Conference%202011-%20Papers.pdf>, (accessed 2011-08-03).
- (31) 2011年3月にKBのDDDプロジェクト担当者に行ったインタビューに基づく。
- (32) "Strategic Plan 2010-2013". Koninklijke Bibliotheek.  
<http://expo.kb.nl/blader/beleidsplan-2010-2013-en/pageflip.html>, (accessed 2011-08-03).