韓国と日本の史資料オープンデータの現況と展望
－韓国史データベースと国立公文書館デジタルアーカイブの比較を通じて－

扈素妍（奈良文化財研究所）

The Present State and Prospects of Opening up Historical Datasets in Korea and Japan

Ho Soyeon（Nara National Research Institute for Cultural Properties）

・オープンデータ／Open Data・ユーザーインターフェース／User Interface
・韓国史データベース／Korean History Database
・国立公文書館／National Archives of Japan

はじめに

近年、全世界においてデジタル人文学が話題になっている。西洋では、1980年代から人文学研究においてコンピュータ技術の導入が試されてきたが（이재연・송인재 외 2019）、近年までは主に「デジタル道具を学問と知識の再構成のために動員する」（김동윤 2014）ものに過ぎなかった。ところが、2010年代からはデジタル人文学とは「情報記述の補助で、新しい方式で遂行する人文学研究と教育、そしてこれと関わった創造的な著作活動」（김현 2015）へ拡張して来た。ここでいう新しい方式とはテキストマイニング、GISなどの方法がある。ところが、人文学者にとって以上の新しい方式を利用するためには、まずこれまでの文献データのデジタル化する作業が必要であろう。

そのため、デジタル人文学を論じる際には、文献の史資料がどのくらいデジタル化されて、データベースとして提供されているのかが重要な論点になる。世界各国においてはアメリカのミネルヴァデジタルアーカイブのような、重要図書館を中心に1990年代中半からインターネット知的資源を収集し、保存するプロジェクトが進行されてきた（이남희 2017）。

韓国では国家施策として、国史編纂委員会を中心に、公的著作物のデジタル化事業が1990年代末から推進され、今や朝鮮王朝時代の国家編纂記録類や近現代の新聞記事などのあらゆる記録物がデジタルコンテンツ化されている（이남희 2017／김현 2015）。特に『韓国史データベース』は古代から現代にまでわたる広大な文字資料を原文と現代語訳で確認できる、韓国史を研究する上で必須的なデータベースである。

一方、日本では、2005年に運営をはじめ、2010年にリニュアルした国立公文書館デジタルアーカイブが歴史資料を提供する代表的なウェブサイトである。国立公文書館デジタルアーカイブは国の行政機関等から移管をうけて、国立公文書館が所蔵する、歴史資料として重要な公文書などの目録情報の検索と資料原本のデジタル画像が閲覧できる情報提供サービス（風間吉之2012）である。

以上の二つの史資料提供ウェブサイトは、韓国側は「データベース」と、日本側は「デジタルアーカイブ」と名付けられており、アーカイブとデータベースが同じものか、また、そもそもアーカイブというのは何かなどの質問が生じえる。しかし、本稿は両サイトの意義や本質について問いをかけるものではない。本稿では両サイトが場所・時間とは関係なく、誰にでもデジタル化した史資料を提供するオープンデータとして、社会において共通の役割を果たすものと位置付けて、両サイトのユーザーインターフェースに着目して論じる。

筆者はデータやアーカイブ学の専門家ではない。そのため、ここでは両国のデータベースを使用する

図１　韓国史データベーストップページ

図２　韓国史データベースの文字入力機、「日本語音読入力」で「AN」をクリックした画面

利用者として、また、人文学研究者として、両方のデータベースの仕組みとユーザーインターフェースを紹介及び比較し、これからのオープンデータの展望を提示したい。

韓国史データベースの特徴

韓国の国史編纂委員会では1989年6月に「電算化推進総合発展計画」を立てて、情報化時代を準備しはじめ（이남희 2017）、1997年からは韓国史データベースデジタル化事業を開始し、2000年から2009年までの国家による知識情報支援管理事業及び国家DB事業の支援によって拡大・発展してきた。

2019年3月を基準に、韓国史データベースでは合わせて101種、約848万件、15億1千万字の資料を提供している。データベースの趣旨は、「韓国史の重要資料を利用者が探しやすく、本文まで閲覧できるように精製されたデータを資料利用者の便宜にかなうように提供」することである。

まず、トップページは図１のようであるが、真上にキーワード検索欄があって、その横に人気検索用語が昇順で一個ずつ表示され、その横に「詳細検索」「シソーラス検索」がある。また、その下には各資料群を「時代別一覧」「形態別一覧」「ガナダ順一覧」「国史編纂委員会資料」で、所蔵資料群の分類を変更して見ることが出来るようになっている。このような一覧変更は、各資料の基本的な性格を理解しやすくする。

「詳細検索」では「類義語／異体字（‘ㅇ，ㄹ，ㅿ’などのこと）」の適用有無を選択でき、「全体検索」「記事題目」「著者／編者／発・受信者」「記事内容」「索引語」で検索でき、「対象期間」「（形態別・時代別・国史編纂委員会資料）分類」「資料名」によっても検索できるようになっている。また、検索語入力の際に、「文字入力機」というポップアップウィンドウで「音価」「コード」「部首」「総画数」「新出漢字」「漢語拼音」「日本語音読」の入力ができるようになっている。この文字入力機はトップページでも利用でき、キーワード検索欄の右にあるキーボード模様のボタンがそれである。

資料一覧から、もしくは検索した資料ページに入って、左上の資料名をクリックすると、その資料全体の一覧を確認できる。そこで右上に「資料紹介」「下位資料検索」「資料詳細検索」「ダウンロード」「印刷」「問題申告」のボタンがあり、「資料紹介」は各資料の概要と利用のツボなどが述べられていて、資料解題のPDFファイルのダウンロードもでき、資

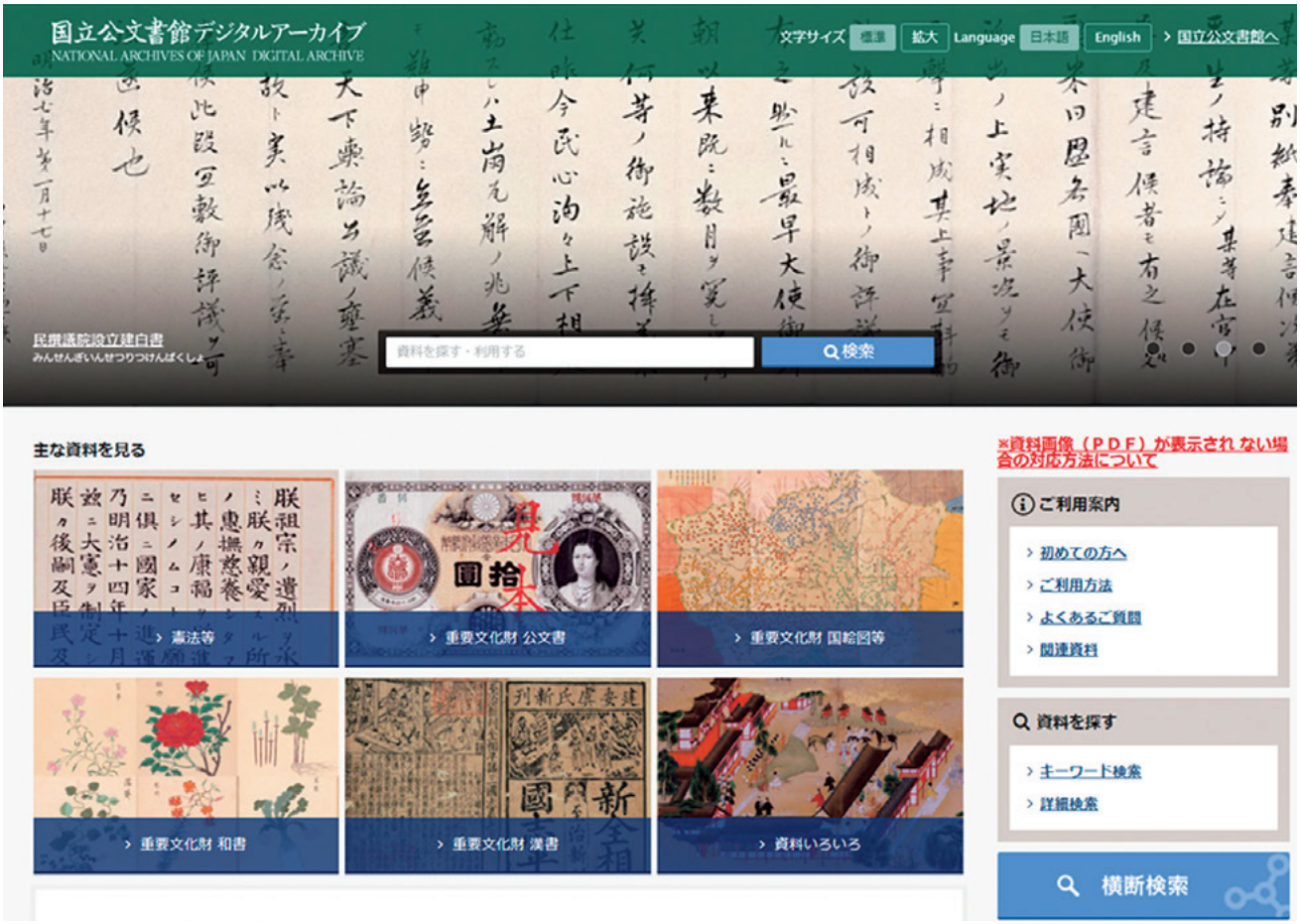

図３　国立公文書館デジタルアーカイブのトップページ

料について詳しい情報を取得できる。一方、「問題申告」を通じて、誤字・脱字、画像エラーなど、間違えた情報や利用中不便な事項について申告することができる。

また、すべての資料は原文画像の確認ができる上、『朝鮮王朝実録』や「韓国近現代雑誌」など、多くの資料はその原文のみ、もしくは現代韓国語訳文までを全部起こして提供しているため、原文の内容からも検索ができる。しかし、近代の新聞記事などはまだ題目のみの検索にとどまっていて、『東亜日報』や『朝鮮日報』の場合、NAVER で提供するニュースライブラリー、また、『毎日申報』は国立中央図書館のデジタルコレクションを検索した方がより豊かな資料に触れることができる。

国立公文書館デジタルアーカイブの特徴

国立公文書館デジタルアーカイブは「デジタルアーカイブの概要」からうかがえるように、「インターネットを通じて、「いつでも、どこでも、だれでも、自由に、無料で」、館所蔵の特定歴史公文書等の目録情報の検索、公文書や重要文化財等のデジタル画像等の閲覧、印刷、ダウンロードが可能な」サービスの提供を目的としている。

検索は韓国史データベースと同じく、基本的にはキーワード検索であるが、トップページの右下の方の「キーワード検索」をクリックすると、「資料群」「検索対象」「辞書の使用有無」「表示件数」「一覧の表示内容の量」を選択して検索ができる。それから、同じ位置の一つ下の「詳細検索」には、前述の「キーワード検索」に検索オプションとして、「作成年月日」「作成年」「資料群指定」「公開／非公開の利用制限区分」の選択、「本館／分館」の保存場所の選択が提供されている。また、その下に「行政文書」と「内閣文庫」に分けてその請求番号などの情

報からも検索できるようになっている。

実際にキーワードで検索すると、その表示順を「標名／件名／書名」「資料群／簿冊件名順」「作成部局／人名」「(作成) 年月日」「請求番号」に基づいて、昇順もしくは降順に、また表示される文書の分量も選択して切り替えることができるようになっている。また、各資料は「概要情報」にその資料の階層が表記されていて、その上、表示形式を「階層表示」に変更することもでき、省庁部局の文書のというその資料の位置や性質をより理解しやすくしている。

以上のように、国立公文書館の資料は国家機関の文書という特性のため、文書の性質を理解するためには、その文書の発行機関についての情報をも必要である。国立公文書館デジタルアーカイブでは、そのために、「省庁組織変遷図」と「太政類典の構成」のページを提供している。

「省庁組織変遷図」は、万延元年から平成 22 年に至るまで、省庁組織の名前が如何に変わってきたのかを把握できる表で、公文書の発行機関の歴史を理解することに役立つ。また、各省庁及び部局名をクリックすると、その設置時期・廃止時期・英語表記がポップアップウィンドウで表示されて、そのポップアップウィンドウの「部局レベルで表記」をクリックすると、その局が所属した組織図を確認できる。また、「検索する」をクリックすると、その部局と関連のある文書が新しいタブで表示される。

また、「太政類典の構成」では、太政類典の簡略な紹介と、構成表の第1表と第2表、そして、目録まで提供している。構成表は太政類典の構成を把握しやすくするため、分類項目という簿冊との関係を表にしたもので、第 2 表は第 1 表をより詳しい分類で表したものである。表の内容をクリックすると、デジタルアーカイブ検索結果へつながる。

以上のページに加えて、「概要情報」の資料群の名前をクリックすると、「資料群情報」という頁に移

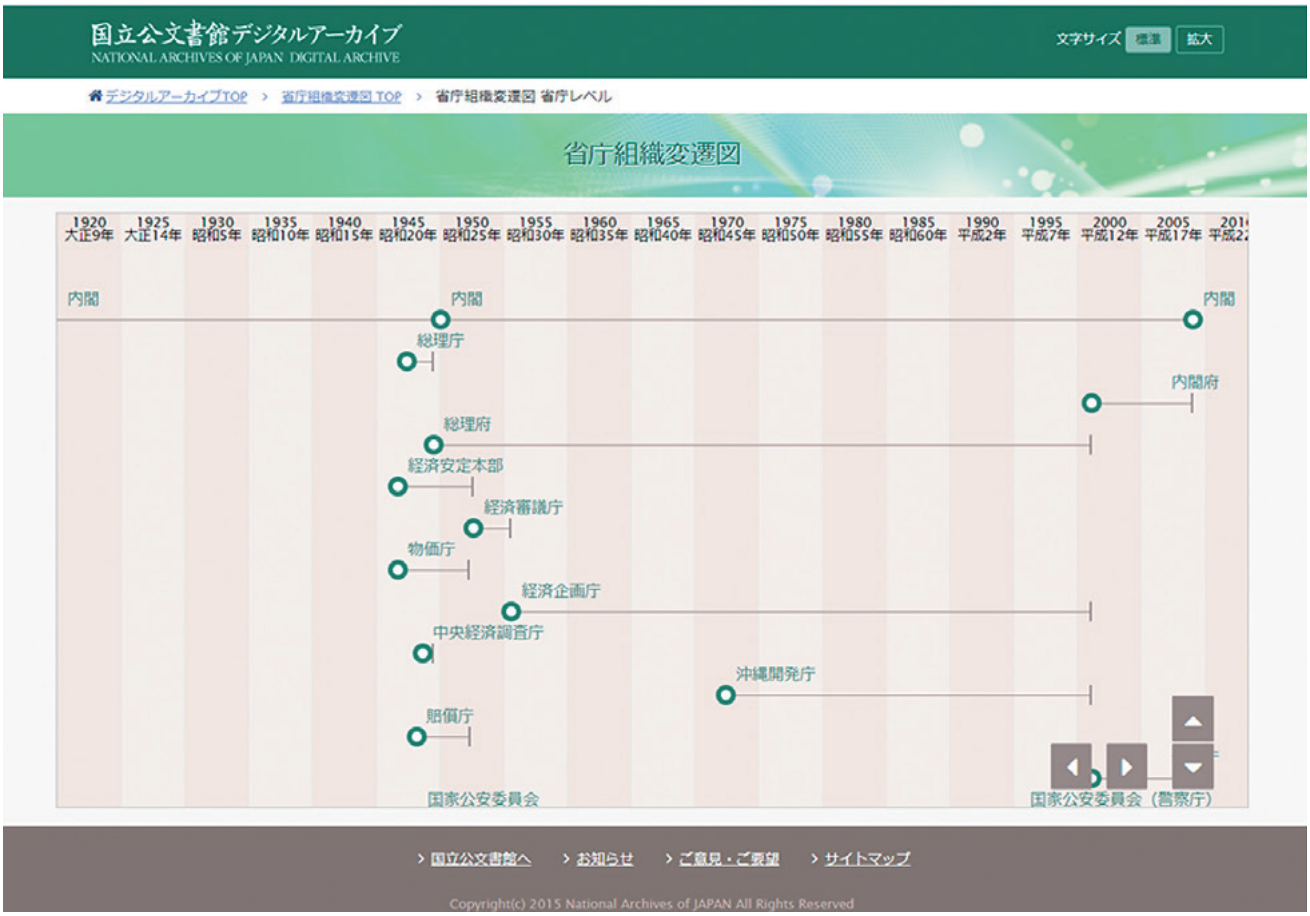

図 4　国立公文書館デジタルアーカイブの省庁組織変遷図

り、そのページの「資料群詳細を表示」に入ると、資料の「階層」「作成期間」「簿冊件数」「組織歴」「資料履歴」「移管元省庁等」「内容」「参考文献」までの情報が提示され、各資料群の性格がより明確に理解できる。

これからの展望

以上のように、両ウェブサイトはそれぞれ提供するデータの特徴に合わせて、利用者がより便利で、より豊かな情報を得ることができるように検索方法や資料紹介などのサービスを提供している。これは、歴史研究者のみならず、歴史に興味のある人々にとっても歴史を理解し、楽しめるようにしたという点で、意義のある作業である。ところが、これからより多くの人の利用を促すためには、いくつか改善を要する点がある。

まず、モバイル環境でのインターフェースを考慮することである。両サイトは、モバイル環境で訪問すると、ウェブサイトのトップページがそのまま出てくる。韓国史データベースの場合、本当にウェブサイトそのままで、国立公文書館デジタルアーカイブの場合はウェブサイトの項目がモバイル環境でより見やすく表示されるものの、ウェブサイトのトップページにあった簡単な検索欄は表示されない。今はデスクトップやノートパソコンより、スマートフォンの利用者が多くなっていることを考えると、両サイトともモバイル環境でのインターフェースにつき工夫が必要だと考えられる。

また、全ての資料の原文検索ができるようにする必要がある。もちろん、著作権法や公文書管理法などによって、全面公開できない資料はあると考えられる。しかし、少なくとも画像で全文を公開している資料ならば、全文をOCR化して、検索可能にすることは、キーワードと関連のある資料をより多く探し、歴史をより多層的に理解できる糸口になるに違いない。

そして、ユニバーサルな利用者を想定することも改善点になるだろう。例えば、利用者には障害などによって文字を読めない人もいることを想定し、音声案内などができるようにすることや、外国人の利用により積極的に備えることが必要と考えられる。韓国史データベースは韓国語以外の言語では表示されない。一方、検索は文字入力機を利用してコード・中国語・日本語などでも検索できるようにしているが、ハングル・漢字・英語で検索した時、それぞれ表示される資料件数は大きく違う。例えば、同じ意味である「산파」「産婆」「midwife」の三か国語で検索した時、連続刊行物の表示件数は「1535」「611」「0」件になる。一方、国立公文書デジタルアーカイブは英語のサイトも提供しているが、英語のサイトでは、「省庁組織変遷図」と「太政類典の構成」は表示されない上、検索においては韓国史データベースと同じ問題がある。このような状態は、研究者ならともかく、両国の歴史に興味を持つ外国人を戸惑わせる。原文を全てOCR化し、検索システムにおいて多言語間の連結をより緊密にして提供すると、原文をすぐには理解できなくても、機械翻訳機を利用して、その内容が大まかにでもわかるようにすれば、閲覧者の理解に役立ち、両サイトの利用もより活発になるだろう。

オープンデータは資料をどこでも、誰でも確認できるという便利さや、歴史資料を破損の恐れなく保存及び提供できるという点で、これからその利用は活発になっていくと考えられる。よって、その提供データの量を増やすことも重要であるが、それ以上にユーザーインターフェースをより便利で、分かりやすくする工夫も必要であろう。

【参考資料】

한국사 데이터베이스（http://db.history.go.kr/）。

国立公文書館デジタルアーカイブ（https://www.digital.archives.go.jp/）。

橋本雄太 「人文学資料オープンデータの可能性と現状」『情報の科学と技術』64-12、2015。

風間吉之 「国立公文書館のデジタルアーカイブ」『アーカイブズ』48、2012。

杉本重雄 「デジタルアーカイブへの期待と課題－コミュニティの違いを超えた知的資源の保存に向けて－」『アーカイブズ』45、2011。
平野健一郎 「デジタルアーカイブズと歴史理解および歴史研究」『アーカイブズ』48、2012。
이재연・송인재 외『세계 디지털 인문학의 현황과 전망』(커뮤니케이션북스、2019)。
이남희「한국사 문헌자료 디지털화의 현황과 과제」『열린정신 인문학 연구』18-2、2017。
김현「디지털 인문학과 선비문화 콘텐츠」『유학연구』33、2015。
송인재「동아시아 개념사와 디지털인문학의 만남」『개념과 소통』18-0、2016。
김동윤「프랑스‘디지털 인문학’의 인문학 맥락과 동향」『인문콘텐츠』34、2014。
하은아・이성숙「공공도서관 OPAC 인터페이스의 발전방안 연구」『한국 비블리아학회지』24-3、2013。
윤정옥「고서목록 데이터베이스의 검색 인터페이스에 관한 연구 －검색 기능을 중심으로－」『한국도서관・정보학회지』42-2、2011。