

歴史データをひらくこと ーオープンデーター

橋本雄太（国立歴史民俗博物館）

1. はじめに

インターネットを通じて誰でも入手でき、自由に改変や再配布が認められているデータを一般に「オープンデータ」と呼びます。インターネットの世界において、オープンデータはいわば公共財としての役割を果たします。オープンデータの利用者は、著作物の二次利用にともなう複雑な権利関係に頭を悩ますことなく、そのデータを活用した自由な研究開発や創作活動、あるいは商業的活動に取り組むことができます。

データのオープン化によってもたらされる利益は広く認識されており、行政や学術など公共的性格の強い分野において、オープンデータの公開は着実に根付きつつあります。歴史学を含む人文学研究の領域もその例外ではありません。本章では、オープンデータとその関連概念について説明し、人文学特に歴史学においてオープンデータの公開がもたらすさまざまな可能性について議論します。

2. オープンデータ普及の背景

オープンデータをめぐる今日の状況はさまざまな歴史的経緯の上に成り立っていますが、「データへのアクセスを自由化する」考え方が最初に確立したのは自然科学分野であるといわれます。天文学や地球惑星科学の分野では、1957-58年の国際地球観測年にあたって世界資料センター（World Data Center/WDC）が設立されました^[01]。この機関は世界中の研究機関や観測施設から観測データを受け入れるとともに、研究者の求めに応じて（必要経費を除き）無償でこれらのデータを提供するものです。その目的は、予見できない災害などによるデータの逸失を防ぐこと、また特定の機関や国家によって重要な科学研究データが専有されることを防ぐことにありました。

科学研究のオープンデータ公開は、インターネットの発達と普及を通じてさらに加速しました。データの公開とアクセスにかかるコストが劇的に削減されたためです。1995年には、全米研究評議会（National Research Council）の委員会により『科学データの完全かつオープンな交換について』^[02]と題されたレポートが刊行されました。このレポートは環境モニタリングの結果など、観測範囲や影響範囲が地球全体にわたるデータの全世界的な共有を提唱するもので、「オープンデータ」という用語の初出ともされます。2004年には、OECD全加盟国の科学技術大臣が、公的助成を受けた科学研究による研究データを原則オープン化する宣言に署名しました。

オープンデータの公開は政府や地方自治体など行政分野においても活発です。行政のオープンデータ公開は「オープンガバメント」運動に強く影響されたものです。オープンガバメントとは、市民にとって透明かつオープンな政府を目指す政策指針であり、2009年に当時のオバマ米国大統領が公表した大統領メモによってその方針が明らかにされました^[03]。このメモでは（1）透明性、（2）市民参加、（3）官民連携の3つを基本原則として、政府のオープン化の推進が提唱されています。その実現にあたっては行政データの積極的公開が必要不可欠でした。オバマ大統領のメモを受け、早くも2009年5

月には米国政府のオープンデータを公開する Data.gov^[04] が設立されました。米国を除く各国政府もオープンガバメントの運動に追随し、日本では2012年にIT総合戦略本部が「電子行政オープンデータ戦略」^[05]を策定しています。

自然科学や行政分野のオープン化運動の強い影響を受け、文化学術機関においてもオープンデータの公開は急速に普及しつつあります。とりわけ文化資料を公開するデジタルアーカイブでは、特別な法的制約が存在しない限り資料情報をオープンデータとして公開することが一般化しつつあります。

3. ライセンス

オープンデータは何らかの利用条件を明示して公開されますが、データごとに逐一利用条件を設定してはデータの提供者と利用者の双方に負担になります。そこで一般には政府やNPOなどの団体が策定した標準ライセンスが使用されます。以下では、デジタルアーカイブやデータベースで利用されることの多い代表的なライセンスについて説明します。

3.1. クリエイティブ・コモンズ・ライセンス

インターネット法の専門家ローレンス・レッシグらによって創設された非営利組織クリエイティブ・コモンズ (Creative Commons/CC) によって提供されるライセンスです^[06]。電子データを含む著作物一般に適用することができます。

著作物の利用条件には、著作者のクレジット表示を求める「表示」(BY)、営利目的での利用を禁じる「非営利」(NC)、改変を禁じる「改変禁止」(ND)、派生著作物に同じCCライセンスの付与を求める「継承」(SA)など4種類の条件が設定されています。これらの利用条件の組み合わせによって、6種類のCCライセンスが定義されています。例えば「CC BY-SA ライセンス」は、著作者のクレジットを表示し、また著作物を改変した場合には同じCCライセンスで公開することを条件に、営利目的での二次利用も認めるものです。CCライセンスはあくまで著作物の利用条件を明示化するもので、著作権の

放棄とは異なることに注意が必要です。著作物に関するあらゆる権利を放棄する際には、CC0 という表示によってその意志を示す方法が用意されています^[07]。なお、著作権を含む知的財産権が消滅し存在しない著作物の状態を「パブリックドメイン (PD)」と呼びます。CC0 は著作物をパブリックドメインに置くことを宣言するものです。

CC ライセンスは Wikipedia などインターネット上のさまざまなプロジェクトで利用されており、一部の歴史資料デジタルアーカイブの利用規約にも採用されています。ただし CC ライセンスは**著作者自身**が著作物の利用条件を明示化するためのライセンスです。文化学術施設が所蔵する（著作権が失効した）歴史資料に CC ライセンスを付与することについては、法的な意味を疑問視する声もあります^[08]。

3.2. 政府標準利用規約

2014 年に作成された、日本政府の府省の Web サイトの利用ルールを定める規約です^[09]。出典の記載や第三者の権利を侵害しないこと、個別法令による利用制約を遵守することなどを条件に、コンテンツを自由に利用、複製、加工、再頒布することを認めています。第 1.0 版では「公序良俗に反する利用や国家・国民の安全に脅威を与える利用」の形態が禁止されていましたが、条件が不明確でありデータ利用の萎縮を招くとして、2015 年に改定された第 2.0 版ではこの条件が削除されました。第 2.0 版で定められた利用規約は、クリエイティブ・コモンズの表示ライセンス (CC BY) と互換性があることが明記されています。

3.3. Rights Statements

CC ライセンスは著作者自身が著作物のライセンス表示をするための仕組みです。これに対して Rights Statements は、インターネット上で公開される資料の**権利の所在**のみを表記するための仕組みです^[10]。文化学術施設が収蔵する資料は、通常は当該施設ではなく第三者によって作成されたものだから、資料の権利表示のための仕組みとしては CC ライセンスよりも Rights

Statements のほうが適している場合があります。

Rights Statements では、資料の著作権が残存しているケースと失効しているケースの2グループについて、合わせて12種類の権利表記が定義されています。例えばNO COPYRIGHT - CONTRACTUAL RESTRICTIONS^[11]は、「資料に著作権は存在しないものの、利用にあたってはその他何らかの法的制約が課されている」ことを明示するものです。

Rights Statements は欧州の統合デジタルアーカイブである Europeana^[12] や米国デジタル公共図書館 (Digital Public Library of America/DPLA)^[13] などですすでに採用されています。まだ日本国内での採用例は少数ですが、今後デジタルアーカイブにおける権利表示の国際的標準として広く普及する可能性があります。

4. 機械可読性

ライセンスと並んでオープンデータの利用可能性を左右するのが機械可読性です。データの価値を最大化するためには、人間が見て理解できるだけでなく、「コンピューターによって容易に処理できる (機械可読性の高い)」形式で公開することが重要です。例えば文書データを公開する際には、紙の文書をスキャンした画像データを公開するよりも、全文検索が可能なPDFファイルとして公開したほうが利用は容易です。さらにテキストデータそのものも公開すれば、機械処理はさらに容易になります。

こうした考え方にに基づき、World Wide Web の発明者であるティム・バーナーズ＝リーは、オープンデータの公開形式を5段階で評価する指標である「5つ星スキーム」を提案しました (図1)^[14]。各段階が推奨する公開形式は次のようなものです。

★ (どんな形式でも良いので) あなたのデータをオープンライセンスで Web 上に公開しましょう

★★ データを構造化データとして公開しましょう (例: 表のスキャン画像)

よりも Excel)

★★★ 非独占の形式を使いましょう (例: Excel よりも CSV)

★★★★ 物事を示すのに URI を使いましょう。そうすることで他の人々があなたのデータにリンクすることができます

★★★★★ あなたのデータのコンテキストを提供するために他のデータへリンクしましょう

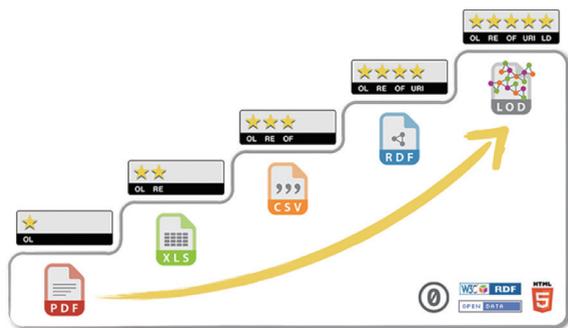


図 1 5つ星オープンデータの各段階

「5つ星スキーム」は、CSVのように特定のプラットフォームに依存せず、また機械処理が容易なフォーマットでデータを公開することで、データの利用可能性を最大化することを推奨しています。さらに URI (Universal Resource Identifier) を用いてほかのデータとの関係やコンテキストを記述することで、データを Linked Data として公開することを勧めています。

Linked Data は Web 上で構造化データを公開・共有するための技術群であり、URI (Universal Resource Identifier) や RDF (Resource Description Framework) などの Web 標準技術を駆使して、Web 上に公開されるさまざまなデータを相互に「リンク」する仕組みです^[15]。オープンライセンスのもとに公開される Linked Data は Linked Open Data (LOD) と呼ばれます。

5. 人文学資料のオープンデータ化がもたらす可能性

すでに多くの文化学術機関が、自館が所蔵する資料をデジタル化し、インターネット上でデジタルアーカイブやデータベースとして公開しています。一部機関では、デジタル化した資料をオープンデータとして公開する取り組みも進められています。公開されるデータには、古文書や絵図、写真などをスキャンした画像データ、文学作品や歴史文献資料を翻刻したテキストデータ、博物館資料や美術作品の目録情報などが含まれます。以下では、こうした人文学資料をオープンデータとして公開する意義について考えてみましょう。

5.1. 二次利用の促進

人文学資料をオープン化する最大の意義のひとつは、資料の二次利用にかかるさまざまな制約を解消することでしょう。1990年代から2000年代にかけて、各地の図書館や博物館、美術館において所蔵資料のデジタルアーカイブ化が急ピッチで進められました。しかしながら公開資料を研究論文に引用したり Web サイトに転載したりするためには、別途申請が必要となる場合が少なくありませんでした。

デジタル化した資料を明確なライセンス表記のもとオープンデータとして公開することで、資料の利用にかかる煩雑な手続きを解消することができます。これはデータ提供者側の作業負荷を低減することにもつながります。

5.2. マッシュアップによる価値創造

オープンデータとして公開されたデータは、利用者が自由に加工し再配布することができます。あるデータをほかのデータと組み合わせたり、利用者が開発したソフトウェアにデータを組み込んだりすることも可能です。データの組み合わせによって新しい価値を生み出す手法は「マッシュアップ」と呼ばれ、オープンデータの活用手段として盛んに実践されています。

歴史資料のオープンデータにもマッシュアップは有効です。歴史資料の目録データやテキストデータベースには、歴史的な人物や団体、著作、事件な

ど、さまざまな事項に関する情報が記述されています。これらの事項に関するオープンデータを組み合わせることで、歴史資料を立体的に捉える視覚化が可能になるかもしれません。例えば人間文化研究機構は、歴史地名辞典や明治期の地図をもとに作成した30万件以上の座標情報付き地名データからなる「歴史地名データ」をCC BYライセンスのもとで公開しています^[16]。このデータセットを古文書や古記録など文献資料のテキストデータベースとマッシュアップし、座標情報を地図上にプロットすることで、歴史記録の空間的把握を容易にする可視化アプリケーションを開発するといった活用法が考えられます。

5.3. データ駆動型の人文学研究の促進

従来型の人文学資料のデジタルアーカイブやデジタルライブラリーは、公開資料の用途を「資料の閲覧・鑑賞」に限定するかたちで設計されてきました。機械処理が容易なフォーマットでデータを提供することで、図書館や博物館などのデータ提供者がこれまで想定してこなかった、新しい人文学資料の活用法が生まれる可能性があります。

人文学資料の新しい活用法として期待される手法のひとつが、統計的手法や機械学習技術を駆使した「データ駆動型」の人文学研究です。例えば文学研究の分野では、特定の作家／時代／地域／言語の文学作品群に対して自然言語処理や統計処理を適用することで、個別作品の熟読からは導出が困難な新しい知見が得られる可能性があります。このような手法は「遠読 (distant reading)」と呼ばれ、文学研究の新しい手法として高い注目を受けています^[17]。遠読を駆使した研究には大量のテキストが必要となるため、過去の文学作品のオープンデータ化は分野全体の発展に大きく影響します。

6. 歴史学・人文学分野のオープンデータ

以下では、実際に日本国内で公開されている人文学資料のオープンデータをいくつか紹介します。

▶ 国立国会図書館デジタルコレクション^[18]

国立国会図書館が所蔵しデジタル化した資料を公開する Web 上のコレクションです。文献や絵画をデジタルスキャンした画像資料のほかに、レコードなどをデジタル化した音源資料も公開されています。著作権が失効したパブリックドメイン資料は原則としてインターネット公開され、国立国会図書館に申請することなく複製や再頒布も含めて自由に利用することができます。さらに、国立国会図書館デジタルコレクション全体の書誌データなどのデータセットもオープンデータとして公開されています。

▶ 国立文化財機構 ColBase^[19]

2017年に公開された国立文化財機構の4つの博物館（東京国立博物館、京都国立博物館、奈良国立博物館、九州国立博物館）の所蔵品を横断的に検索することのできるシステムです（図2）。資料の目録情報に加えてデジタル画像も閲覧することができます。ColBaseの利用規約は政府標準利用規約第2.0版に

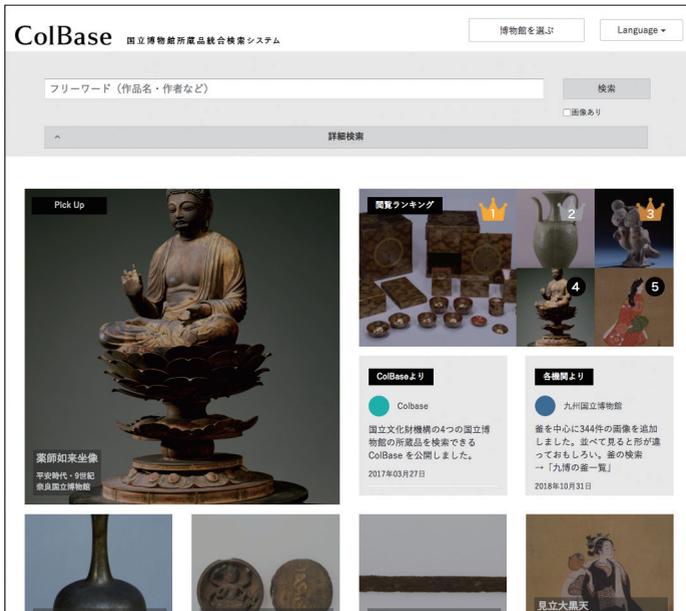


図2 ColBase

準拠しており、出典を記載する限りにおいて商用目的も含めて自由にコンテンツの複製や翻案が可能です。

▶ 歴史的典籍 NW 事業オープンデータセット^[20]

2014年度より、国文学研究資料館は「日本語の歴史的典籍の国際共同研究ネットワーク構築計画（歴史的典籍 NW 事業）」と題された大規模な研究事業を実施しています。この事業では国文学研究資料館とその連携機関が所蔵する約30万点の古典籍資料のデジタル化が進められています。このデジタル化資料の活用を図るため、国文研では古典籍に関わるデータをオープンデータとして公開しています。2016年11月からは、情報・システム研究機構に設立された人文学オープンデータ共同利用センター（CODH）と連携し各種データセットの公開を進めています。

2018年12月時点では、古典籍1,767点の画像データと書誌データなどを公開する「日本古典籍データセット」、古典籍から採取した3,999文字種の字形データ403,242文字からなる「日本古典籍くずし字データセット」、江戸の料理本に書かれたレシピを現代語訳も含めてまとめた「江戸料理レシピデータセット」などが公開されています。データセットのライセンスには多くの場合クリエイティブ・コモンズ表示 - 継承 4.0 国際ライセンス（CC BY-SA）が用いられています。

▶ 国立歴史民俗博物館 khirin^[21]

国立歴史民俗博物館が公開する歴史・民俗・考古学資料のデータベースです（図3）。同館が推し進める「総合資料学の創成と日本歴史資料の共同利用基盤構築」事業の一環として開発・運営されています。2018年12月時点では、国立歴史民俗博物館の館蔵資料のほか、歴史民俗調査カード（1970年代に文化庁が作成した文化財の調査カード）、千葉大学所蔵町野家文書^{まちのけもんじょ}などが公開されています。khirinの利用規約は政府標準利用規約第2.0版に準拠しており、特に断りがない場合は出典を明記するだけで自由に利用することが可能です。一部資料については、目録情報だけでなく画像データも IIF (International

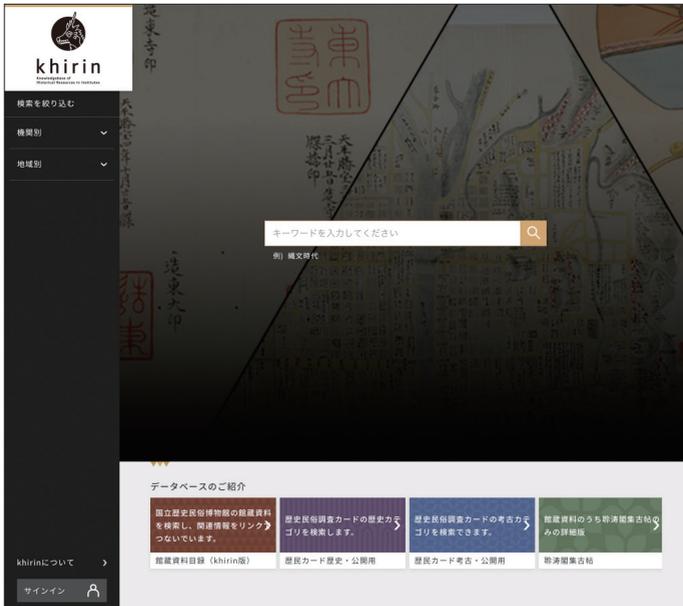


図3 国立歴史民俗博物館 khirin

Image Interoperability Framework) のフォーマットで提供されています。

▶ 人文学オープンデータ共同利用センター^[22]

2017年に設立された情報・システム研究機構データサイエンス共同利用基盤施設人文学オープンデータセンター (Center for Open Data in the Humanities / CODH) は、情報学・統計学の技術を駆使した人文学データのアクセス改善、データ駆動型の人文学研究、超学際的な人文学研究の実現などを目的とした組織です。

CODHは前述の歴史的典籍NW事業にあたって国文学研究資料館と連携しており、日本古典籍データセットや日本古典籍くずし字データセット(図4)はCODHから公開されています。2018年12月には、日本古典籍くずし字データセットを加工し、機械学習分野で著名なMNISTデータセット互換のくずし字データセットKMNIST^[23]を公開しました。KMNISTは機械学習研究者から高い注目を集めており、くずし字の自動認識の実現につながる事が期



図4 日本古典籍くずし字データセット

待されています。また CODH では『武鑑全集』や『近代雑誌データセット』など、さまざまな人文学資料を機械処理が容易な形式で公開しています。

▶ 京都府立京都学・歴史館 東寺百合文書 WEB ^[24]

京都府立京都学・歴史館（旧京都府立総合資料館）は、同館の Web サービス「東寺百合文書 WEB」にて東寺百合文書のデジタル画像をオンライン配信しています。東寺百合文書は京都・東寺に伝えられた約 2.5 万点の古文書群であり、東寺と関係のあった有力な荘園の経営に関する文書や寺内会議の議事録などを大量に含むことから、中世日本史研究における第一級史料とみなされています。東寺百合文書 WEB ではこれらの文書のデジタルスキャン画像に加えて書誌データや翻刻データ（一部資料のみ）を公開しています。公開データは CC-BY ライセンスにて利用することができます。東寺百合文書ほどの規模と重要性を有した日本史史料がオープンデータとして公開された事例はほかになく、2014 年の公開時には大きな反響がありました。

▶ 愛知県美術館コレクション ^[25]

愛知県美術館は2018年11月に同館が所蔵する1,200件超のコレクション画像をパブリックドメインのコンテンツとして公開しました。日本国外ではメトロポリタン美術館やシカゴ美術館といった大規模美術館がすでにパブリックドメイン化した所蔵作品を公開していますが、日本では非常に先進的な事例といえます。パブリックドメイン化した作品はCCライセンスで公開される著作物と異なり、転載にあたって所蔵元のクレジット表記が必要ありません。愛知県美術館では「公開画像利用時のお願い」として、同館のクレジット表記に協力を求めています。

7. オープンデータの普及に向けた諸課題

日本国内でも、文化学術機関による人文学資料のオープンデータ化は一般化しつつあります。しかし研究者や研究グループによるオープンデータの公開事例はほとんどありません。歴史学を含む人文学の研究では、資料目録や翻刻テキストなど多種多様なデータが生まれます。そうしたデータが研究者個人や研究室のPCに「死蔵」されているケースも少なくありません。以下では、研究者レベルでのオープンデータ公開の障害となる要因について考えてみます。

7.1. 評価の問題

人文学における研究成果の発表方法の主流は、今もって学会あるいは出版社を通じた紙媒体での出版です。研究上のデータを電子化しWeb上で公開しても、それが業績として評価の対象になることは少なく、そのために研究者が労を取って研究データのオープン化に取り組むモチベーションが生じにくいという構造があります。他方で紙媒体での出版は紙数の制限を受けるとともに、機械的な情報の加工や探索に適しておらず、Webの利用が浸透した今日においては必ずしも最適な研究成果の公開方法とはいえません。

人文学研究者が研究の過程で生み出した公共性・再利用性の高い研究データのオープン化を促すためには、データ公開を研究者の業績として適切に評

価する仕組みが必要です。自然科学分野では、「データジャーナル」がこうしたデータを評価する仕組みとして期待されています。データジャーナルとは、研究成果発表の場である通常の論文誌とは異なり、重要度の高い基礎的な研究データの共有と再利用を目的として主にオンラインで刊行される学術雑誌です。投稿される研究データは、各ジャーナルの求めるデータの記述様式（データディスクリプター）を満たす品質の高いデータであるかどうか査読による評価を受け、適切なメタデータを付与された上でCC BYなどのライセンスで公開されます。2014年には、ネイチャー・パブリッシング・グループによるデータジャーナル“Scientific Data”が創刊されました。

人文学分野では、2015年に創刊された日本デジタル・ヒューマニティーズ学会の学会誌 Journal of the Japanese Association for Digital Humanities が、通常の論文と並んで人文学データセットの投稿を受け付けることを表明しています^[26]。今後、有用な人文学データのオープン化事例が増加していくことが期待されます。

7.2. 出版文化との衝突

人文学の基礎資料の編纂と出版は、これまで出版社の力に依拠してきました。例えば日本史・国文学分野の最重要基礎資料である『国書総目録』は、昭和期に岩波書店によって発行された目録です。このため出版社の意向が人文学資料のオープン化の障害になる場合があります。

2013年に、かねて「近代デジタルライブラリー」（国立国会図書館が公開していた近代資料のデジタルコレクション。2014年に国立国会図書館デジタルコレクションに統合）上で公開されていた『大正新脩大藏経』および『南伝大藏経』が、一般社団法人日本出版者協会（出版協）の抗議を受けて一時的に公開が停止されたことがありました。この抗議は、上記の書籍が仏教出版社の大藏出版により現在も刊行、販売されており、国立国会図書館によるデジタル公開によって同社の事業が損害を被っているという内容のものでした。両書の著作権はすでに消滅していましたが、最終的に国立国会図書館は出版

社側に配慮し、『南伝大蔵経』のインターネット公開を当面の間停止するという措置を下しています^[27]。

研究書や史料集の刊行を通じて、出版社は日本の人文学研究を長年にわたり支えてきました。しかしながら出版社が人文学の基礎資料を抱え込んでいるために、そのオープンデータ化が遅々として進まないという皮肉な構図があります。この課題については研究者側と出版社側が協議を重ねていく必要があるでしょう。

8. おわりに

本章では、人文学領域におけるオープンデータ普及の背景、意義、現状、今後の課題について概説的に論じました。自然科学や情報学の分野と比較すると、人文学領域におけるオープンデータ公開の取り組みは活発であるとはいえません。しかし同時に今後の可能性を秘めた分野であるともいえます。

歴史を振り返っても、人文学は時代に合わせて研究データの共有・公開手法を進歩させてきました。例えば江戸時代の国学者・^{はなわ ほ きいち}埴保己一は、貴重な古典籍の散逸と国学の衰退を恐れ、古代から近世に書かれた重要文献を集書し、一大叢書として刊行する事業に取り組みました。埴の事業は三十余年をかけて『^{ぐんしよるいじゆう}群書類従』として結実し、その後の国文学および史学研究の発展に大きく貢献したことは広く知られています。『群書類従』の編纂は、^{しゃほん}写本というかたちで閉じられていた古典の世界を、出版によって「オープン化」する企てであったといえるでしょう。今後、人文学の諸分野においてオープンデータについての活発な議論が起こることが期待されます。

——注（Web ページはいずれも 2018-12-1 参照）

[01] Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council. Earth Observations from Space: The First 50 Years of Scientific Achievements. *The National Academies Press*. pp. 6, 2008.

[02] National Research Council. On the Full and Open Exchange of Scientific Data. *National Academies*, 1995.

[03] Barack Obama. “Memorandum on Transparency and Open Government”. 2009. <https://www.archives.gov>

- gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf.
- [04] Data.gov. <https://www.data.gov/>.
 - [05] 電子行政オープンデータ戦略, https://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf.
 - [06] クリエイティブ・コモンズ・ジャパン, <https://creativecommons.jp>.
 - [07] CC0 について, <https://creativecommons.jp/sciencecommons/aboutcc0/>.
 - [08] 生貝直人「デジタルアーカイブと利用条件」、『カレントアウェアネス・ポータル』322, 2014年.
 - [09] 政府標準利用規約 (第 2.0 版), https://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl2_betten_1.pdf.
 - [10] Rights Statements, <https://rightsstatements.org/>.
 - [11] COPYRIGHT - CONTRACTUAL RESTRICTIONS, <https://rightsstatements.org/page/NoC-CR/1.0/>.
 - [12] Europeana, <https://www.europeana.eu/>.
 - [13] Digital Public Library of America, <https://dp.la/>.
 - [14] 5-star opendata, <https://5stardata.info/>.
 - [15] Linked Data とその周辺技術については次を参照してください。
Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. S Morgan & Claypool Publishers, 2011 (武田英明監訳『Linked Data : Web をグローバルなデータ空間にする仕組み』近代科学社、2013年)。
 - [16] 歴史地名データ, https://www.nihu.jp/ja/publication/source_map/.
 - [17] Franco Moretti. Distant reading. Verso Books, 2013 (秋草俊一郎訳『遠読 - 〈世界文学システム〉への挑戦』みすず書房、2016年)。
 - [18] 国立国会図書館デジタルコレクション, <http://dl.ndl.go.jp/>.
 - [19] ColBase, <https://colbase.nich.go.jp/>.
 - [20] 歴史的典籍 NW 事業オープンデータセット, https://www.nijl.ac.jp/pages/cijproject/data_set_list.html.
 - [21] khirin, <https://khirin-ld.rekihaku.ac.jp/>.
 - [22] 人文学オープンデータ共同利用センター, <http://codh.rois.ac.jp/>.
 - [23] KMNIST データセット, <http://codh.rois.ac.jp/kmnist/>.
 - [24] 東寺百合文書 WEB, <https://hyakugo.kyoto.jp/>.
 - [25] 愛知県美術館コレクション, <https://www-art.aac.pref.aichi.jp/collection/index.html>.
 - [26] Journal of the Japanese Association for Digital Humanities, <https://www.jstage.jst.go.jp/browse/jjadh>.
 - [27] カレントアウェアネス・ポータル「国立国会図書館、インターネット提供に対する出版社の申出への対応についての報告を公開」2014年1月8日, <http://current.ndl.go.jp/node/25212/>.