

Title	自然科学と社会科学のデータ連携に向けたオープンデータ政策の課題分析
Author(s)	佐野, 仁美; 和泉, 潔; 林, 隆之
Citation	年次学術大会講演要旨集, 37: 482-485
Issue Date	2022-10-29
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18512
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

1 F 2 2

自然科学と社会科学のデータ連携に向けたオープンデータ政策の課題分析

○佐野仁美（政策研究大学院大学），和泉潔（東京大学），林隆之（政策研究大学院大学）
hitomi.sano.jp@gmail.com

1. はじめに

AI 技術の進展により、特定分野や組織内のデータだけでなく、異なる領域のデータを連携させて新しい価値を発見し、問題解決を行う重要性が高まっている。特に、世界的な課題である「気候変動に伴う社会経済への影響」の分析には、自然科学分野と社会科学分野の連携が不可欠であり、両分野の連携による総合知は、課題解決に向けた客観的エビデンスとして有用な知見になると考えられる。そこで、本研究では、自然科学分野と社会科学分野の連携事例の一つとして「気候変動に伴う社会経済への影響」にフォーカスし、同連携における代表的なデータである気象関連データと金融関連データについて、現状のデータ項目や公開状況等の検証を行った。この結果、気象関連データの多くがオープン化されている一方、金融関連データはオープン化が進んでいないほか、データ項目にも課題を内包し、本分野の研究進化の妨げになっていることが判明した。本稿では、これらの検証結果をもとに、研究進化に向けた我が国のオープンデータ政策および分野間連携における課題を議論する。

2. 先行研究

国内のオープンデータ政策は、先行研究[1][2][3][4][5]やオープンデータ政策を推進する内閣府やデジタル庁の資料等[6][7][8][9][10]に、オープンデータのあり方や、国内外の動向およびオープンデータの活用事例等が整理されている。しかし、本研究のテーマである「自然科学分野と社会科学分野のデータ連携」は、一般的に必要性が認識されつつも、第5期科学技術基本計画「超スマート社会」の実現（Society 5.0）の中で部分的に記載されるに留まり、具体的な連携に向けた検討は十分ではない。一方、2022年6月に閣議決定された2022年版「環境・循環型社会・生物多様性白書（環境白書）」では、世界中で気候関連の被害が増加し被害額が急増していることを踏まえ、「気候変動問題は経済・金融のリスク」と強調されており、同分野の分析の重要性が高まっている。国内における関連研究としては、水害が実体経済に与える影響を定量分析した研究[11]や、局地的大雨が預金者のATM引出額に与える影響を分析した研究[12]等がある。前者は、過去、わが国において発生した水害が、GDPなどの実体経済に与えた間接効果を推計した研究であり、データはいずれも社会科学分野である国土交通省の水害統計調査[13]と県民経済計算を利用している。気象状況との相関が主眼ではないため、分野連携はないが、水害統計調査は市町村単位の詳細区分で表されており、疑似的な位置情報付き金融関連データの利用事例として挙げた。次に、先行研究[12]は、局所的大雨の降雨強度とATM引出動向を時間別・地点別に解析し、キャッシュレス決済が進む中での現金ニーズにおける客観的エビデンス取得を目指した。データは気象関連データとして国土交通省の高解像度降雨情報（XRAIN）、金融関連データとして大手都市銀行のATM引出統計情報を利用している。後述のとおり、一般的には、金融関連データの多くに位置情報が付されず、気象関連データと金融関連データの連携は困難であるが、本研究ではATM設置地点の位置情報が付された引出統計情報を利用したため、気象関連データとの連携を可能にした。

3. 異分野間データ連携における全般的課題

令和3年4月から施行された科学技術・イノベーション基本法では、科学技術・イノベーション政策が、自然科学や人文・社会科学を含む「総合知」により、人間や社会の総合的理解と課題解決に資する政策となる必要性が提示されている。特に、自然科学分野と社会科学分野の連携の必要性は、科学技術イノベーション関係者間でも課題として挙げられているが、現状の具体策は乏しく、組織的・人的連携

に重点を置く提案に留まっている。しかし、実際の研究現場では、根本的理由として、両分野のデータ連携が技術的に困難化する要素が多いことが課題となっている。

まず、「オープンデータ」の前提を整理する。国及び地方公共団体は、官民データ活用推進基本法（平成 28 年法律第 103 号）により、オープンデータへの取り組みが義務付けられている。平成 23 年 3 月 11 日の東日本大震災以降のデータの公開・活用への意識の高まりから、政府、地方公共団体や事業者等が保有する公共データは「国民共有の財産」とされ[14]、現在では、データ公開のみでなくデータ活用を前提とした「課題解決型のオープンデータの推進」として、デジタル庁（2021 年 9 月 1 日発足）を中心に取り組みが推進されている。デジタル庁の資料「オープンデータ基本指針」（令和 3 年 6 月 15 日改正）[15]によると、「オープンデータ」は下記のとおり定義されている。

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工、編集、再配布等）できるよう、次のいずれの項目にも該当する形で公開されたデータをオープンデータと定義する。

- ① 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- ② 機械判読に適したもの
- ③ 無償で利用できるもの

上記定義は、一般的なデータ連携において重要な前提条件であるが、これらの対象は公共データであり、民間企業には適用義務がない。このため、企業内で生成された多くのデータは上述の項目①～③を満たさず、結果として、多くの民間企業のデータはオープン化が困難になっている。

4. 気象関連データと金融関連データのオープン化動向

気象関連データは、国内最大規模の地球環境情報プラットフォームである「データ統合・解析システム」（Data Integration and Analysis System：通称「DIAS」）の「データ・アプリケーション」[16]、気象庁の「各種資料・データ」[17]および国土交通省の「防災情報提供センター」[18]等を中心にオープン化がされている。つまり、国内で扱われる主要な気象関連データのほぼ全てについて、オープン化がなされている状況と言える。特に、降雨観測情報は、上述の DIAS 上の公開データ「XRRAIN」が高解像度（250m メッシュ）の観測情報を観測からほぼリアルタイム（1 分毎）で配信する等、国内でオープン化されている気象関連データは、量および質とも世界でトップレベルの有益な情報となっている。

一方、金融関連データは、本稿では気象状況で影響を受ける業態として、預金関連の「銀行」と物損関連の「損害保険」に特化して言及するが、両業態とも個別企業のオープンデータ化は進んでいない。ただし、業態内の統計情報として一部のオープン化はされており、銀行では日本銀行による統計情報[19]で金融機関の取引総計が公開され、損害保険では一般社団法人日本損害保険協会により「自然災害での支払額」[20]が公開されている。また、国土交通省も「水害統計調査」[13]として、1 年間に発生した水害被害を対象とした個人・法人が所有する各種資産、河川・道路等の公共土木施設及び運輸・通信等の公益事業施設等に発生した被害の統計を公開している。ただし、いずれのデータも統計情報であるため、公表までに一定期間を要し、タイムリーに取得可能な情報ではない。

5. 気象関連データと金融関連データの連携に向けた課題

前述のとおり、現状、気象関連データと金融関連データの連携は進んでいない。この要因を検証した結果、主に金融関連データ側に、次の（1）～（3）に挙げる課題を検出した。これらについて、以降、改善提案も含めて述べる。

- （1）個別企業におけるデータ占有
- （2）位置情報の設定
- （3）時空間解像度の設定

5.1 個別企業におけるデータ占有

個別企業のオープンデータ化を阻む最大の要因としては、個人情報保護法の影響が挙げられる。金融機関のデータには、個人または法人の取引明細データが多く含まれるが、個人情報保護法遵守の観点か

ら、これらの取引明細データを未加工では開示できない。従って、対外開示の際は、取引明細における個人情報のマスキングや、明細データを累計して個別明細を特定不可能にする等、秘匿性を確保するための加工処理が必要となる。これらは当該機関内で人件費等のコストを要するため、対価回収を目的に、加工データはビジネスとして利用される傾向にある。ビジネスでのデータ利用は、特定契約先への有償提供か、特定機関との共同利用に大別される。有償での提供は、有力な販売先（高額かつ継続的なデータ販売先）の販路確立に力点が置かれ、データ形式も提供先の要望に沿うよう加工される傾向がある。これによって、オープンデータの条件である「汎用的なデータ形式」は事実上無効化されている。また、特定機関と共同利用する場合（主に共同研究等）、稀少データの特異性による成果の期待や知名度向上等のインセンティブが高まることから、データが占有され、第三者の参入機会が消失する可能性もある。このように、個別企業では、データ生成にコストが生じるため、利益と照らした利用にならざるを得ず、結果として、独自のデータ形式やデータの占有を招き、研究進化のためのオープンデータ化を妨げる要因になっている。当該機関のオープンデータ化におけるインセンティブ向上は現状困難と考えられることから、解決策としては、現在の統計情報[19][20]等の業態別全企業の統計情報を細分化することにより、疑似的な個別企業のオープンデータ化を図ることが一案として考えられる。

5.2 位置情報の設定

気候関連の研究では、時間（Time）、位置情報（Location）、観測値（Value）を基本項目とする。更に、分析手法としては、キー項目（Time および Location）毎に Value を整理・分析してモデル化し、将来予測を行う手法が一般的である。データ連携の際には、両データのキー項目を同期させる必要があるが、大半の金融関連データには位置情報（Location）が含まれず、時間（Time）と金額等（Value）に留まる事例が多い。これは、これまでのデータの収集や管理が組織内各拠点に一任される運用であり、「データの発生場所」を考慮する必要性がなかった故と推測される。気象関連データは「空間情報」であり、位置情報を削除すると意味をなさない。このため、キー項目が同期できない金融関連データとの連携は、事実上不可能となっている。解決策として、今後の金融関連データには、統計情報として最適解を取得し得る位置情報（地点、市町村、都道府県等）をデータに付加する改善が望まれる。これにより、気象関連データとのデータ連携が可能になるだけでなく、中央管理部署が各拠点のデータを俯瞰して管理するなど、当該企業の全般的状況を詳細かつタイムリーに把握可能になる。

5.3 時空間解像度の設定

前述 5.2 のとおり、金融関連データには位置情報が付されないことが多いため、現状は、「全国」「県別」等の地域単位での統計を地理的情報として扱わざるを得ない。しかし、気象関連データがリアルタイムかつ局地的情報が得られている一方、金融関連データの統計情報は地理的な解像度が粗い故に、相互の局地的な特徴の検出が困難となっている。加えて、時間幅も「年次」や「月次」等の総計が多く、公表タイミングも当該集計期間最終日を待たなければならない。このため、現在のオープンデータからは局地的情報を取得できないほか、タイムリーな検証・分析も行えない。従って、空間情報としては「地点」または「市町村」のレベル、時間幅は「日」「時間」「分」レベルで解析可能な統計情報として整備し、可能な限りタイムリーに公表可能な情報に改善されることが望まれる。

6. まとめ

本研究では、昨今の世界的な課題である「気候変動に伴う社会経済への影響」を事例の一つとし、自然科学分野と社会科学分野のデータ連携に向けた課題の検証を行った。この過程において、気象関連データ（自然科学分野）と金融関連データ（社会科学分野）のオープン化の動向および連携する上での課題を検証した。この結果、気象関連データは国内の主要データのほぼ全てについてオープン化がされている一方、金融関連データは個別企業データの大半がオープン化されていないことが判明した。加えて、金融関連データには、個別企業におけるデータの占有、位置情報が付されていない、時空間解像度が粗い等の課題があり、これらの改善が気象関連データとの連携に向けた前提条件になることを示した。

最後に全般的な課題を述べる。現在、民間の各企業は、これまでの「日々のデータの収集・蓄積」から「自社を飛躍させる財産」としてデータを積極的に利活用する方針に転じており、幅広いデータを収集し利活用するニーズが高まっている。このため、各企業のデータ生成のコストを勘案しながらも、多くの民間企業にオープンデータの情報基盤への参画を促す対策は、互いに有益な情報を共有する観点か

らも有用と考える。また、個人情報保護法遵守により、公表データには統計処理が施されることが一般的であるが、より正確に状況を可視化するには、統計の括りを可能な限り細分化することが望ましく、その適正範囲も今後の課題である。オープンデータは、多くの研究成果や分野を超えた総合知を生み出す情報基盤である。公共データのみでなく、多様な業態のオープンデータ化により、分野間データ連携の場が生み出され、新たな研究進化の一助となることを期待する。

謝辞

本研究は、文部科学省、科学技術イノベーション政策における「政策のための科学」推進事業の支援を受けたものである。また、論文の中で示された内容や意見は執筆者個人に属し、本学の公式見解を示すものではない。

参考文献

- [1] 庄司昌彦, オープンデータの動向とこれから(<特集>オープンデータ), 情報の科学と技術, 65(12), 496-502(2015)
- [2] 林和弘, オープンサイエンス政策の広がり第 6 期科学技術・イノベーション基本計画, 2021.
- [3] 早田吉伸, 前野隆司, 保井俊之, オープンデータ推進に向けた国内先進地域の特徴分析, 地域活性化研究, 6, 61-70 (2015)
- [4] 高木聡一郎, 欧州におけるオープンデータ政策の最新動向, 情報管理, 55(10), 746-753 (2012)
- [5] 庄司昌彦, オープンデータの定義・目的・最新の課題, オープンデータ』, 国際大学 CLOCOM, 4-15 (2014)
- [6] デジタル庁「政府 CIO ポータル」<https://cio.go.jp/policy-opendata>
- [7] 内閣府 科学技術・イノベーション推進事務局 「総合知」の基本的考え方及び戦略的に推進する方策中間とりまとめ(令和4年3月17日)<https://www8.cao.go.jp/cstp/sogochi/index.html>
- [8] 総務省「第2節オープンデータの活用の推進」
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/pdf/n3200000.pdf>
- [9] 自治体オープンデータ <https://www.open-governmentdata.org/>
- [10] データカタログサイト「DATA GO.JP」<https://www.data.go.jp/>
- [11] 芹沢拓郎, 須藤直, 山本弘樹, 水害が実態経済に与える影響に対する定量分析, 日本銀行ワーキングペーパーシリーズ 2022 年 (2022)
- [12] 佐野仁美, 皆川直人, 坂地泰紀, 和泉潔, 位置情報を利用したデータ融合による局地的大雨発生時の現金引出動向分析, 人工知能学会全国大会論文集, JSAI2022 (0), 4N1GS302, 1-4(2022)
- [13] 国土交通省「水害統計調査」
- [14] 「電子行政オープンデータ戦略」(平成 24 年 7 月 4 日 高度情報通信ネットワーク社会推進戦略本部決定) (2012)
- [15] デジタル庁「オープンデータ基本指針」(令和 3 年 6 月 15 日改正)
- [16] 「データ統合・解析システム」(Data Integration and Analysis System)「データ・アプリケーション」https://diasjp.net/apps_search/
- [17] 気象庁「各種データ・資料」<https://www.jma.go.jp/jma/menu/menureport.html>
- [18] 国土交通省「防災情報提供センター」<https://www.mlit.go.jp/saigai/bosaijoho/>
- [19] 日本銀行「統計」<https://www.boj.or.jp/statistics/index.htm/>
- [20] 一般社団法人日本損害保険協会「統計」<https://www.sonpo.or.jp/report/statistics/index.html>