

教育分野におけるオープンデータ活用の実態

豊福晋平^{†1}

概要：Open Data Index 等オープンデータに関する国際ランキングには教育領域も含まれている。しかしながら日本の教育分野のオープンデータ活用に対する評価は必ずしも高くない。本論では、各種国際ランキングでの評価状況、オープンデータの整備状況、活用方法の課題と解決策についてまとめる。

キーワード：アカウントビリティ、国際比較、学力調査結果データ、オンラインデータベース

1. はじめに

2012 年高度情報通信ネットワーク社会推進戦略本部は公共データを社会全体で効率的に活用促進する基本戦略として「電子行政オープンデータ戦略」[1]を決定した。国、地方公共団体等のオープンデータへの取組みは活発化しており、2016 年 12 月現在、政府データカタログサイト「DATA.GO.JP」の掲載データセット数は17,000件を超え、参加自治体数は約 230 団体にのぼる。

一方、World Wide Web Foundation の「Open Data Index 2012 年」[2]によると日本は 19 位であり、主な減点評価としてオープンデータのライセンス使用度、機械可読データの公開等と並んで、教育関連データの公開が十分でないことが指摘されている。

本調査研究では、教育分野のオープンデータに関して、主に基礎的なデータ提供と活用に関して、我が国及び主要国の実態を明らかにするとともに、その利活用に向けた課題等の整理とその解決策の分析等を行った。

2. 教育分野オープンデータの定義

電子行政オープンデータ戦略等で、我が国のオープンデータの目的・意義として挙げられているのは次の 4 点である。

- 透明性・信頼性の向上
- 国民参加・官民協働の推進
- 経済の活性化・行政の効率化
- 組織・企業としての社会貢献

これに対し、教育分野のオープンデータに関しては体系的な記述がみられないので、海外文献[3][4][5]をもとに意義・データセットカテゴリ等を以下の通り整理した(なお、本稿ではオープンデータに関わるステークホルダとオープンデータ集計レベルは省略した)。

2.1 教育分野のオープンデータの意義

教育分野のオープンデータの意義は、次の 3 点にまとめられる。

- ① アカウントビリティ：透明性と信頼性の向上・参加と協働

教育機関や研究機関における各種情報、教育・研究活動により生成される情報がオープンデータとして提供されることで透明性が高まり、ステークホルダからの信頼を得るとともに、教育・研究活動への参加と協働を促す。

- ② 効率化・高度化・価値向上

主に教育機関における各種情報、教育・研究活動により生成される情報がオープンデータとして提供されることで、国・地方公共団体の限られた部署・関係者にのみで共有されていた分析考察が一般に公開される。考察可能機会・比較範囲の拡大によって効率化と知見の高度化が図られ、学習者・保護者のみならず、学校経営、政策検討に至るまで幅広くメリットを与える。

- ③ イノベーション

教育機関や研究機関における各種情報、教育・研究活動により生成される情報がオープンデータとして提供されることで、市場における編集、加工、分析等の各段階を通じた新ビジネス、あるいは、研究・開発等への参入障壁の低減を通じたイノベーションの創出が促される。

2.2 オープンデータセットのカテゴリ

オープンデータセットのカテゴリは表 1 の通りである。

表 1 オープンデータセットの分類

番号	データカテゴリ	内容
1	政策データ	予算、支出、KPI、ROI、集約統計等
2	機関データ	所在地、施設・設備、在籍数、学力、保健、出欠率、特別支援、監査報告、論文リポジトリ等
3	コースデータ	カリキュラム、シラバス、教材リソース等

^{†1} 国際大学
International University of Japan

4	個人データ	個人プロフィール, 出欠, コース選択, 成績, 試験結果, 研究者情報, 研究業績等
5	履歴データ	学習履歴ログ, パフォーマンスデータ, e-ポートフォリオ, 研究バックデータ等

3. 日本の国際的位置付け

教育分野のオープンデータ推進とその利活用に関する、世界での我が国の位置付けと、ベンチマークとすべき国の抽出について、主要な3種類のオープンデータ国際ランキングをもとに検討を行った。

3.1 Open Data Index

先に述べた WWF の Open Data Index は不定期で実施され、2012 年を最後に実施されていない。2012 年の日本の順位は 19 位であった。

教育指標は「教育データの Web への公開教育の業績に関する政府データが、Web にどの程度公開されているか(1=全くなく、10=広範囲に)」。調査ロードデータ[6]を参照すると日本の評点は 6 点であった。

3.2 Open Data Barometer

WWF による Open Data Barometer [7]は 2013 年から調査実施されており、2017 年 3 月時点の最新版は 2015 である。2015 年の結果で日本は総合 13 位、2014 年は 19 位、2013 年は 14 位であった。

教育分野の基準は「初等中等教育のパフォーマンスデータ」であり、具体的には国家統一試験の生徒テストスコア、学校出席率、教員出勤率であり、単純な学校一覧情報は含まれないとしている。評価観点としては、データの存在以外に、機械読み取り可能な形式か、一括入手可能か、ライセンスがオープンか、などで 0~95 点の評点レンジで日本の評点は 15 点であった[8]。

3.3 Global Open Data Index

OKF (Open Knowledge Foundation) の Global Open Data Index[9]は 2013 年から調査実施されており、2017 年 3 月時点の最新版は 2015 である。

2015 年の結果で日本は総合 31 位 (46%オープン)、2014 年は 19 位 (61%オープン)、2013 年は 32 位 (43%オープン)であった。評価対象となるサイトの URL と評価結果が公開されており評価透明性は高いが、教育に関する評価項目は含まれない。

表 2 教育分野オープンデータの各国位置付け

対象	Open Data Index 教育	Open Data Barometer 教育	Global Open Data Index 総合
英国	10	95	76%
豪州	10	95	67%
米国	9	80	64%
日本	6	15	46%

3.4 ベンチマーク国との比較

以上 3 つの国際指標をみる限りでは、日本の教育分野におけるオープンデータの位置付けはトップラインにはないことが明らかである。ベンチマーク対象国としては、英国、オーストラリア、米国の 3 国が考えられる。表 2 には各国の点数比較を示した。

4. オープンデータの公開状況

教育分野におけるオープンデータの公開活用として、基本的なオープンデータセットと、アクセスや活用が容易なデータベース整備について整理した。

4.1 学校名称・所在地リストの入手

単純な学校名称・所在地リストは Open Data Barometer の評価基準から除外されているが、米国では Common Core Data の 1 つとみなされ、最も基本的かつ重要なオープンデータセットの一つである。政府関連機関のみならず、NPO・NGO が各自治体や各学校のパフォーマンスや進捗(情報機器整備状況や校務労働実態など)の比較データを作成するには学校名称・所在地の基本情報は欠かせない。

表 3 国別教育機関名称・所在地データの提供状況

提供方法	英国	豪州	米国	日本
データベース	Edubase2	My School	CCD	なし
一括取得	○	△	○	×

○:一括取得可 △:各州から取得可 ×:取得不可

英国・オーストラリア・米国はいずれも国レベルでの教育情報オンラインデータベース(後述)を提供しており、米英は一括ダウンロードも可能だが、日本には国レベルのデータベース・サービスがない。

日本では、都道府県・市区町村レベルでデータ提供されているが、その対応にはばらつきがあり、機械可読性の低い書式が多く含まれている。いずれにしてもオンラインで完全なデータを取得することは困難である。

表 4 日本の教育機関名称・所在地データの提供状況

提供情報	オープンデータカタログサイト	学校基本調査統計	学校一覧
対応 都道府県数	23	45	38

4.2 教育情報のオンライン・データベース・サービス

データセット単位でファイル提供されたオープンデータはユーザーによる加工・整形・分析を必要とするが、集約的なオンライン・データベース・サービスを用いれば、情報入力・集約・アップデート、膨大なデータセットの蓄積・提供、分析・要約生成といった作業を極めて効率的に行うことが出来る。英米豪いずれも教育分野の大規模なオンライン・データベース・サービスを展開している。

4.2.1 英国

英国 School performance tables[10]は、全国的な学力調査 National Tests, General Certificate of Secondary Education (GCSE) に加え、各種アセスメント以外のデータも含む全学校のデータを提供している。サイト上で公表されたデータは一切アクセス制限がなく、自由に閲覧可能であり、データもダウンロード可能である。

利用者は特定の学校を検索し、学校情報(所在地・地図・児童生徒数・学校サイトリンク)、各種アセスメント調査結果、出欠率、教員数、財政状況、教育水準局による学校監査レポートを参照することが出来る。また、複数学校を比較する機能も付いている。

4.2.2 オーストラリア

オーストラリアの My School サイト[11] は、保護者や教育関係者向けに学校情報をオープンに提供するため 2010 年開設され、ACARA (Australian Curriculum, Assessment and Reporting Authority) が全学校のデータ (NAPLAN: National Assessment Program-Literacy and Numeracy 以外のデータも含む) を集計・分析した結果を公表している。2018 年には改訂版が提供される予定である。

My School では、特定学校を検索して、学校情報(児童生徒数や教員数、所在地、社会経済的な環境等)、財務状況、NAPLAN の結果(グラフ、得点、分布状況、経年変化、類似学校との比較)、近隣の学校情報等を閲覧できる。

4.2.3 米国

米国の教育省 NCES が運用する NAEP Data Explorer[12] サイトは、NAEP(全米学力調査)のデータを公表しており、利用者は科目、学年別にデータを閲覧することができる。また、生徒、教員、学校から収集したデータに基づき、

性別、人種、学校設置者(公立/私立)、教員の経験等、多数の指標から学力データを確認することができる。さらに、これらのデータを使い、統計表やグラフの作成や、地図上への表示を行うことができる。

4.3 学力調査結果のデータベース提供

公教育のアウトカムと政策決定のために各国では統一学力調査の実施が強化されてきた。我が国でも 2007 年から「全国学力・学習状況調査」が実施されている。各国では、得られた膨大な調査結果をデータベース化して提供している。データベースには調査個票データも含まれる。児童生徒のプライバシーに関する情報が含まれるため、その運用と利用対象には制限が設けられるのが一般的である。

例えば、先に述べた米国 NAEP の Data Explorer に含まれる項目には次のようなものが含まれる。

表 5 米国 NDE に含まれる項目(制限項目を含む)

	分類	詳細	項目
児童生徒	コア背景変数	家庭環境、保護者教育レベル、欠席日数、母語、人種、学校種、在校生の人種比率等	40
	数学関連質問	学習方法、テストに対する評価態度、数学学習に対する態度等	50
	読解関連質問	学習方法、テストに対する評価態度、読解学習に対する態度等	47
教師	背景・教育/研修	最終学歴・勤務年数・学位/履修単位・教科主任・研修経験・人種等	152
	教授法・組織	学級人数、宿題、教育方針、教育方法、ICT 活用等	128
学校		欠席率、教師欠席率、在籍完了(中退)率、クラス人数、特別支援比率、カリキュラム、研修、学校運用形態、常勤教員数等	272

制限事項が含まれるデータへの第三者のアクセスは各国により条件が異なる。オーストラリアの場合、アクセスの許可が得られるのは、連邦政府・各州・特別地域の職員、国内大学及び専門学校の研究者、大学院生、その他 管理機関が認めた研究・行政機関の職員等である。また、利用目的によっては、国際機関や海外の機関の利用者もデータ利用可としている。

表 6 には各国のデータ提供状況を示した。

表 6 各国の学力調査結果のデータ提供状況

	英国	豪州	米国	日本
アクセス 申請制度	○	○	○	事業委託 者のみ
オンライン データベース	○	○	○	×

三菱総合研究所[13]が関係者に行ったヒアリングによると、我が国の全国学力・学習状況調査のデータは、文部科学省委託調査「学力調査を活用した専門的な課題分析に関する調査研究」を通じ、文部科学省よりデータの貸与を受け課題分析が実施されており、委託調査は、複数大学にまたがる研究チームにより実施することもある（その他の研究・企業に対してはオープンにされていない）。

5. 教育分野のオープンデータ利用の課題

以上述べた通り、我が国のオープンデータ利用は英国・オーストラリア・米国と比較すると遅れが目立つ。本稿の最後に各領域にまたがる複合的な課題についてまとめる。

5.1 「Excel 方眼紙問題」

奥村晴彦[14]は「Excel に代表される表計算ソフトは、簡便なデータ入力・解析・可視化ツールであるが、これを柔軟な罫線の引ける DTP ソフトとして用い、データとしての再利用の困難な複雑な帳票を作成してしまうこと」と述べている。オープンデータとして提供されるデータは罫線やセル結合を多用した紙の帳票作成を最終目的とするものが多く、機械可読性を下げている。

5.2 教育関係者のデータリテラシー課題

教育分野にオープンデータのコンセプトを普及させるにあたっては、第一の当事者である関係者のデータリテラシー(data literacy)習得が必要とされている。

データリテラシーは「意思決定を導くための効果的なデータ利用能力」であり、教育者がデータから情報・最終的な実用的知識へと変換することを可能にする特定知識スキルで構成されている(Mandinach *et al.* 2008)。

例えば、我が国の学校教職員に関して言えば、学習者の膨大な学習履歴・出欠や生活実態調査、保護者や地域から収集する学校評価アンケートなど、データを扱う機会は教育情報化に伴って確実に増えてきているが、現職教員向け研修や教職養成課程では体系的なデータリテラシーを学習する機会が十分用意されている訳ではない。

5.3 データベース・ビッグデータ活用への批判

Glyn Moody[15]は過去 4 年間に NPD (National Pupil Database) から子どもの詳細情報が警察に 21 回、内務省に

18 回照会されたことを明らかにしている。また、Guardian 誌[16]によると、教育省は 800 万人分の子どもの出生と国籍データを収集し NPD に投入する事を計画していたが、これはかねてから英国内でくすぶっていた移民問題や人種プロファイリングの差別に繋がるとの批判を引き起こすことになった。

英国 NPD は全国の児童生徒個票データを収容するデータベースだが、児童生徒を特定可能な機微性の高い情報も扱っていることから、これに対する批判が生じていると考えられる。社会的な不安や不信を放置したまま、一部のステークホルダの意見のみでビッグデータの集約や選択的で不透明なオープン化を進めれば、大きな反対が起こる事が予想される。

参考文献

- [1] 首相官邸. 電子行政オープンデータ戦略,2012, <http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html>, (参照 2017-03-30).
- [2] WORLD WIDE WEB FOUNDATION Introducing the Open Data Index, 2012. <http://webfoundation.org/2012/09/introducing-the-open-data-index/>, (参照 2017-03-30).
- [3] Marieke Guy. Open data in Education. 2014, <https://www.slideshare.net/MariekeGuy/edtalk2>, (参照 2017-03-30).
- [4] Louis Coiffait. Changing education: Open Education Data #opendata 2013, <https://www.slideshare.net/louisrusoe/open-education-data>, (参照 2017-03-30).
- [5] Open Knowledge. Open Education Handbook 2014, https://en.wikibooks.org/wiki/Open_Education_Handbook/Print_version#Types_of_Open_Data, (参照 2017-03-30).
- [6] <http://thewebindex.org/wp-content/uploads/2012/10/Open-Data-Index/Open-Data-Index.xlsx>, (参照 2017-11-07).
- [7] Open Data Barometer, <http://opendatabarometer.org/>, (参照 2017-03-30).
- [8] <http://opendatabarometer.org/3rdEdition/data/>, (参照 2017-03-30).
- [9] Global Open Data Index, <http://index.okfn.org/>, (参照 2017-03-30).
- [10] <https://www.gov.uk/school-performance-tables>, (参照 2017-11-07).
- [11] <https://www.myschool.edu.au/>, (参照 2017-11-07).
- [12] <https://nces.ed.gov/nationsreportcard/naepdata/>, (参照 2017-03-30).
- [13] 三菱総合研究所. 学力調査を活用した専門的な課題分析に関する調査研究〔全国学力・学習状況調査の結果データベースの構築・運用に関する調査研究〕,2016, http://www.mext.go.jp/b_menu/shingi/chousa/shotou/112/shiryo/_icsFiles/afldfile/2016/06/06/1371753_2.pdf.
- [14] 奥村晴彦. 「ネ申 Excel」問題,情報処理学会・情報教育シンポジウム論文集, Vol. 2013, No. 2, pp. 93-98. <http://oku.edu.mie-u.ac.jp/~okumura/SSS2013.pdf>.
- [15] Glyn Moody. National Pupil Database: Schoolkids' info passed to cops and home office. 2016, <https://arstechnica.co.uk/tech-policy/2016/10/national-pupil-database-info-passed-cops-home-office/>, (参照 2017-03-30).
- [16] The Guardian. The national pupil database puts children at risk of racial profiling. <https://www.theguardian.com/commentisfree/2016/oct/10/national-pupil-database-children-risk-racial-profiling>, (参照 2017-11-07).