



平成25年度技術委員会

オープンデータ流通推進コンソーシアム オープンデータ化の評価指標案

2014.02.14

オープンデータ流通推進コンソーシアム 事務局

背景

■ オープンデータ化への取組の拡大

- ▶ 政府データポータルサイト (data.go.jp)
- ▶ Open Data METI (経済産業省)
- ▶ 情報通信白書のオープンデータ化 (総務省)
- ▶ 自治体によるオープンデータ化への取組 (鯖江市・横浜市・流山市・静岡県など)
◇さまざまな規模の組織がデータを公開している。

■ オープンデータの取組状況を評価する指標が必要

- ▶ 目標を立て、それに基づいてオープンデータ化の実践を推進するため。
- ▶ 電子行政オープンデータ実務者会議においても、検討事項の1つとして挙げられている。
◇「第5回電子行政オープンデータ実務者会議」資料5による

既存のオープンデータ指標

1. 取組全体を評価する指標

- ▶ Open Data Index (World Wide Web Foundation)
- ▶ Open Data Barometer (Open Data Institute)

2. データを評価する指標

- ▶ Open Data Certificate (Open Data Institute)

…など

- 本技術委員会での検討事項は、「2. データを評価する指標」との関係性が高い。

Open Data Index

■ World Wide Web Foundationによる指標のうち、オープンデータに関する14の指標。下記の項目について10段階で評価したもの。(*1)

1. ウェブに情報を公開している政府機関が、どの程度オープンライセンスを使用しているか
2. 国際貿易に関する政府データが、Webにどの程度公開されているか
3. 各部門の予算と実支出に関する詳細な政府データが、Webにどの程度公開されているか
4. 保険医療セクター(病院、医者など)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
5. 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
6. 交通機関の状況やスケジュールに関する政府データが、Webにどの程度公開されているか
7. 統計(年齢、収入、投票、移住など)に関する政府データが、Webにどの程度公開されているか
8. 政府の地図データが、Webにどの程度公開されているか
9. 納税申告書のデータとその提出方法に関する政府データが、Webにどの程度公開されているか
10. 政府の各サービス提供者(例えば、地方の警察所、図書館など)の連絡先に関する政府データが、Webにどの程度公開されているか
11. 国内の犯罪に関する政府データが、Webにどの程度公開されているか
12. Web上の政府データ(例えば上記質問に示したデータ)に対して、オープンでマシンリーダブルな形式(.csvや.xls、XML、RDFなど)で、どの程度容易にアクセスできるか
13. 政府は特別にオープンデータイニシアティブを実施しているか
14. 保険医療、教育、安全、予算などの分野において、政府データを活用したWebアプリケーションやサービスがどの程度存在しているか(つまり、それらの分野において、公開されている政府データから導き出された新しく価値のある情報やサービスが存在しているかどうかということ)

(*1) <http://www.webfoundation.org/2012/09/introducing-the-open-data-index/>
およびこれに関する Open Knowledge Foundation Japan の解説記事による

Open Data Barometer

- Open Data Instituteが提唱する、各国政府のオープンデータ度を測定する指標。
- 以下の3種類の指標からなる。
 1. 準備度 (Readiness)
 - ◇政府・起業家／ビジネス・市民社会に対するオープンデータの認知度をはかる。
 2. 実施度 (Implementation)
 - ◇説明責任・社会政策・イノベーションに必要なオープンデータが公開されているかをはかる。
 3. 影響 (Emerging impacts)
 - ◇政治的・経済的・社会的な影響力をはかる。

Open Data Certificate

- Open Data Instituteによって運用されている、そのデータが「どれくらい利用しやすいか」を示す指標。
- データ提供者が質問に答えることにより、該当するデータに対して以下の4つのいずれかの証明書が発行される。
 1. Raw: A great start at the basics of publishing open data
(オープンデータ公開の基本に基づく第一歩)
 2. Pilot: Data users receive extra support from, and can provide feedback to the publisher.
(データ利用者はデータ提供者に対してサポートを受ける、またはフィードバックを与えることができる)
 3. Standard: Regularly published open data with robust support that people can rely on.
(信頼できるしっかりしたサポートが受けられる、標準的なオープンデータ)
 4. Expert: An exceptional example of information infrastructure.
(情報基盤の例となり得る、非常に優れたオープンデータ)

Open Data Certificateにおける質問

■ 基本的事項

- ▶ データの名称
- ▶ データが掲載されているURL
- ▶ データの提供者
- ▶ データ提供者のURL
- ▶ データのリリース頻度と種類
 - ◇ データ形式は単一か複数か
 - ◇ 1回限り／定期的リリース／API提供

■ 法的関係（ライセンス・プライバシー）

- ▶ オープンデータとして公開できるか／権利は明確か
 - ◇ Yes／No／不明／複雑すぎて説明できない。
- ▶ rights statementを宣言しているURL
- ▶ 人々がこのデータをオープンデータとして利用できるライセンスが存在するか
- ▶ 著作権を誰が持っているか
 - ◇ なし／単一の個人・団体／複数の個人・団体
- ▶ このデータから個人が識別できるか

■ 実用性

- ▶ そのデータに対してホームページ（主たるwebサイト）からリンクされているか
- ▶ （データカタログ等の）目録に掲載されているか
- ▶ 出版物等からそのデータが参照されているか
- ▶ そのデータは最新か古くなるか
- ▶ データの信頼性について記載したURL
- ▶ データの品質管理プロセスについて記載されたURL
- ▶ いつでもアクセスできるか
 - ◇ オフサイトバックアップを取っているか
- ▶ そのデータはいつまで利用可能か？

※ <https://certificates.theodi.org> に掲載されている、自己診断の質問文から抜粋

Open Data Certificateにおける質問

■ 技術的情報（データ形式・信頼性）

- ▶ そのデータにどこでどのようにアクセスするか。そのアクセス方法
- ▶ そのデータは機械可読か
 - ◇画像や表形式ソフト形式のデータでなく、CSV等で提供しているか
- ▶ そのデータは標準的・オープンなフォーマットで提供されているか
 - ◇たとえばXML、CSV、JSONなど
- ▶ 公開しているデータの種類
 - ◇人間が読む文書／統計データ／地理データ／その他構造化データ
- ▶ 永続的なIDがあるか？
- ▶ 機械可読なデータの取得先（どのようなデータをどこで取得して、どのような処理をしたかという由来、起源のデータ）を提供しているか
- ▶ データの提供元（取得したデータが本当にそこで提供されたものか／途中で改ざんや偽データで置き換えられていないか）を検証する手段を提供しているURL

■ 社会的な情報

- ▶ そのデータに関する技術的解説文書のURL
- ▶ そのデータはスキーマやボキャブラリを利用しているか。
- ▶ そのデータにはコード（分類・識別用の番号等）が使われているか
- ▶ そのデータに関する質問をする方法（問い合わせ先）
- ▶ そのデータに対するフィードバックをするためのURL
- ▶ そのデータに関するプライバシーに関する質問をするための方法（問い合わせ先）
- ▶ そのデータを利用する人々と交流するためのソーシャルメディアを利用しているか
- ▶ そのデータについて議論するフォーラムやメーリングリストがあるか
- ▶ そのデータについて積極的な議論するコミュニティを育成する担当者がいるか
- ▶ データを利用するためのツールを示したURL

評価指標に関する案

■ Open Data Certificateに記載されている評価指標に下記を加えてはどうか？

1. 日本独自の指標

1. データを提供する言語

- ◇国際的なオープンデータ化を目指すならば、多言語（少なくとも英語）のデータが必要がある。
- ◇従って、「多言語（少なくとも英語）のデータがあるか？」という指標が必要ではないか。

2. 文字コードの問題

- ◇日本語を記述する文字コードには、ISO-2022-JP(JIS)、Shift-JIS、EUC、UTF-8など複数存在する。
 - ◆これらが明記されていなければ、機械はデータを解読できない。
 - ◆JSONやSimple Data Formatなど他の規格との整合性を考慮すると、データはUTF-8で提供されることが望ましい。
- ◇従って、「データを提供する文字コードが明記されているか」「UTF-8を利用しているか」という指標が必要ではないか。

2. 機械可読性に関して

- ▶ 機械可読に関する指標は、「オープンデータ化のための技術ガイド」の方が詳しい。
- ▶ 「オープンデータ化のための技術ガイド」を満たしている項目数は、機械可読性をはかる指標にならないか。



OPEN DATA

オープンデータ流通推進コンソーシアム