



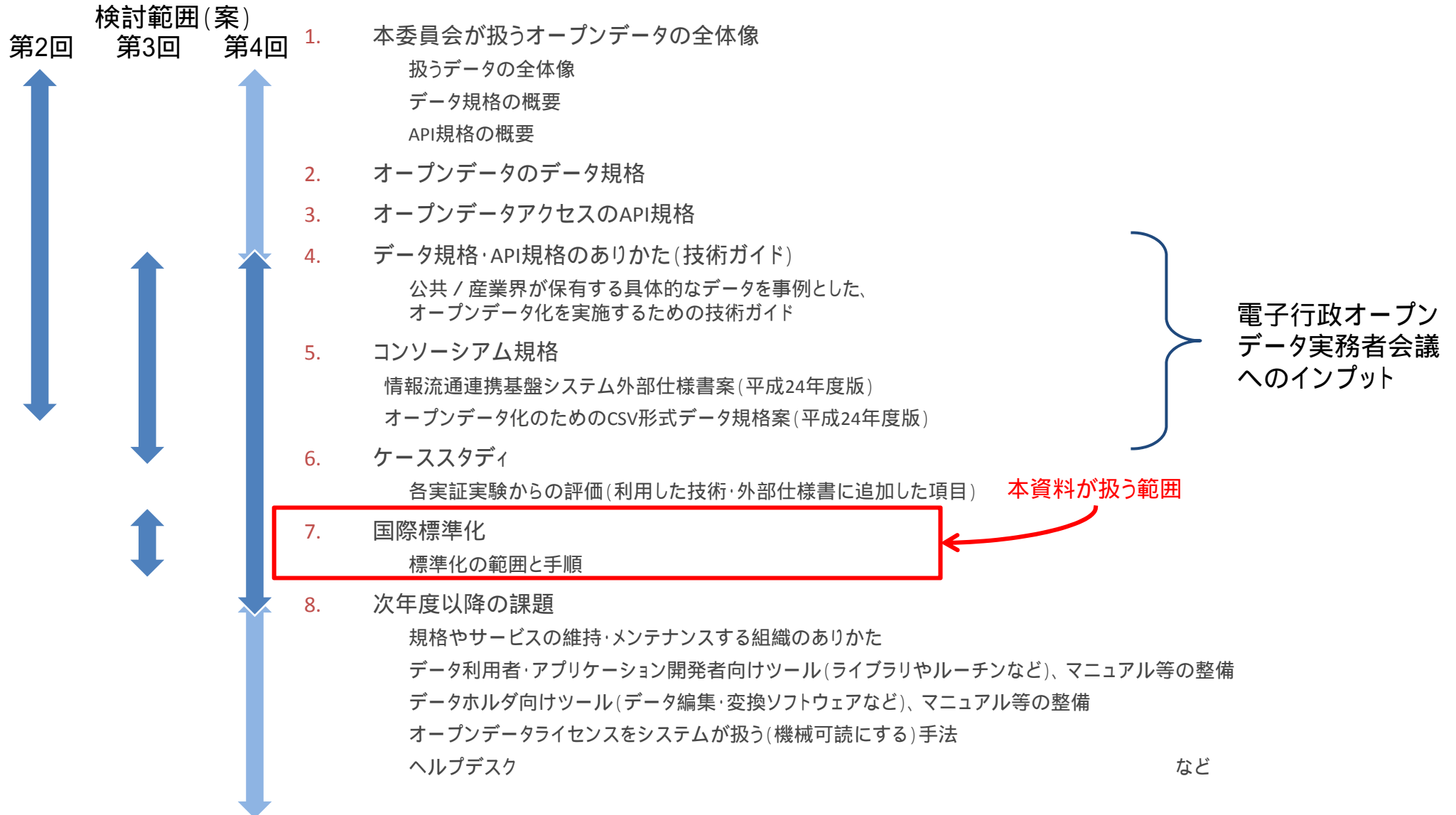
第三回 技術委員会資料

オープンデータ流通推進コンソーシアム 国際標準動向調査結果

2013.2.26

オープンデータ流通推進コンソーシアム 事務局

技術委員会の論点(再掲)



概要

目的

本コンソーシアムで提案するガイドや規格案を、広い範囲で利用してもらうための手法を考察する。

◇特に、標準化により採用を促す場合に、どのような標準活動に提案したらよいか、現状のオープンデータプロジェクトの動向、実際に使われているデータフォーマット、関係する標準団体をサーベイした結果をもとに論ずる。

主な用語定義

公共機関

◇公共的なサービスを提供する機関をいい、行政機関だけではなく報道機関・交通機関・ライフライン・医療機関などを含む。

公共データ (Public Sector Information : PSI)

◇公共機関が保持する、または公共サービスに資する公開可能なデータ。

◇OECDによるPSIの定義は、以下の通り^(*1)。

Information, including information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the government or public institution.

構成

1. オープンデータと標準化に関する概説
2. PSIに関する事例調査
3. データポータルの実態
4. まとめ

付録

◇調査に使用した標準やPSIの一覧

(*1) Organization of Economic Co-operation and Development (OECD) Council, April 2008, Recommendation of the OECD Council for enhanced access and more effective use of public sector information, <http://www.oecd.org/dataoecd/0/27/40826024.pdf>, p. 4



1. オープンデータと標準化に関する概説

1. オープンデータと標準化に関する概説

標準化の方法

- a. de jureの国際標準化団体に提案して標準規格とする。
- b. de factoの標準にするために任意団体をつくって、そこで規格し、広める努力をする。
de factoだが、広く認知されており、de jureに劣らず広く認知され影響力のある規格設定団体もある。

標準案を提案するメリット

標準化団体のトラックや標準化団体が運営するラウンドテーブルに載せることにより、提案仕様を広く共通して利用するものにできる。

透明でアカウントビリティのある手法で標準化された規格であるといえる。

◇標準化団体に提出せず、自分たちで団体を作りde facto標準になるように努力する方法もある。

この場合も、利用者の安心を得るには、透明でアカウントビリティのある手法で、誰もが自由に使える規格を制定し、将来改変される場合も同じ手続きをとることをその団体が保証することが必要である。

標準案を提案するデメリット

規格を自由に改変できない。

決定に時間がかかる。

規格仕様書に関する一般的な考え方

標準化する / しないに関わらず、規格文書自身とその技術が自由に無料で入手可能でないと、普及は進まない。

標準化提案先の選定方針

PSIを公開 / 利用している主要な団体を調査し、そこで使われているデータや規格、それを交付している標準化団体を調査する。



2. PSIに関する事例調査

「2. PSIに関する事例調査」の構成

以下の5つの文献・事例を調査する。

1. Engage, Getting on with Government 2.0, Report of the Government 2.0 Taskforce
2. Government Data and the Invisible Hand
3. ベストプラクティス / ガイドラインの事例
4. オープンデータフォーマット仕様を提案している団体
5. メタデータ記述の事例

1. Engage, Getting on with Government 2.0, Report of the Government 2.0 Taskforce (*2)

オーストラリア政府に2009年末に提出された委員会レポート。主な論点は以下の通り。

ライセンスの重要性

◇ライセンスも機械可読な標準として扱うことが重要である。

メタデータの重要性

◇検索発見できないデータは、存在しないに等しい。(If it can't be spidered or indexed, it doesn't exist.)

◇オーストラリアでは、政府情報の検索に使うシソーラスに使う目的でAGLS (Australian Government Locator Service)メタデータを規定している。これは、Dublin Coreを拡張したボキャブラリ規格である。

2. Government Data and the Invisible Hand (*3)

Yale Journal of Law and Technology (vol. 160,2009) に掲載された論文。論点は以下の通り。

Obama大統領のもとで、政府が Internet を利用したアカウントビリティのある政府をめざすとしても、政府がデータをきれいにみせるように自らウェブを運営するのではなく、再利用できるデータを公開すること、そしてそれを可能にする技術インフラに主眼をおくべき。

変化の速いウェブ技術においつくのを目的とはせず、簡単に使える信頼性の高い政府のもつデータを容易にだれもがアクセスできるインフラ整備することが重要である。

個々のエンドユーザの必要とするような表示、対話機能をもつウェブは政府ではなく民間が自発的につくるのがベストである。

3. ベストプラクティス / ガイドラインの事例

アメリカでの事例

Requirements and Best Practices Checklist(*4)

◇具体的な Web 公開についてのベストプラクティスをまとめたチェックリスト。

Draft Concept of Operation(*5)

◇data.govの運用に対する方針をまとめた文書。

オーストラリアの事例

Enhancing the discoverability and accessibility of government information(*6)

◇Accessibilityを向上すること、政府全体で共通のメタデータをつかうなど、検索を容易にするなどに関する提言。

Early leadership in semantic web(*7)

◇政府機関のウェブやデータセットに“semantic tag”をつけるためのガイドライン。

Whole of government information publication scheme(*8)

◇政府によるPSI公開にあたってのベストプラクティスの方針の議論。

◇集約した検索サイトの必要性を言及している。これが元となり<http://www.data.gov.au/> が誕生したと考えられる。

ニュージーランドの事例

Government ICT Direction and Priorities(*9)

◇政府関連の PSI データについての情報をあつめるときに誰もが参照するページ。W3CのOpen Data five-star Modelが解説されており、PSIはレベル3以上(non-proprietaryで、マシン可読、ウェブ経由でアクセスできて、NZGOAL に応じてライセンスされていること)であるべきと規定している。

◇実際のPSI データ公開サイトの運用の際のチェックリスト(*10)も掲載されている。

(*4) <http://www.howto.gov/web-content/requirements-and-best-practices/checklists/long>

(*5) <http://www.ideascale.com/userimages/sub-1/736312/ConOpsFinal.doc>

(*6) <http://gov2.net.au/projects/project-1/>

(*7) <http://gov2.net.au/projects/project-5/>

(*8) <http://gov2.net.au/projects/project-7/>

(*9) <http://ict.govt.nz/>

(*10) <http://ict.govt.nz/programme/opening-government-data-and-information/toolkit-agencies>

3. ベストプラクティス / ガイドラインの事例

イギリスでの事例

Public sector information Directive and Regulations^(*11)

- ◇イギリスがPSI をオープンデータとして公開する政策や、ベストプラクティスをまとめたサイト。
- ◇以下のようなレポートが掲載されている。

- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2010
- ◆ Guide to the Regulations and Best Practice
- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2009
- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2008 UK implementation of the EU Directive on the Re-use of Public Sector Information - the first two years

オープンデータ政策が主流になる前に、特定の政府機関などに独占的にデータを利用、販売させる契約がなされた事例も紹介されている。

IATIの事例

IATI: International Aid Transparency Initiative^(*12)

- ◇国外から経済支援をうけるような支援開発途上国が、その必要性を自ら評価し、実際に手に入りそうな経済支援を探すのを助け、さらに支援をする側の国の市民が、税金が有効に使われているかを助けるためのNPOである。
- ◇その活動の中で、世界中の経済支援の資金の流れ(目的、実際の使い道など)を比較できるように、共通のオープンデータフォーマット形式を提案し、利用を促している。
 - ◆ 世界中の団体が標準形式で公開したデータは、IATI レジストリへリンクを登録することで公開、周知、共有ができる。
 - ◆ IATI レジストリはデータそのものもホストするが、リンクを収集することも主要な目的である。
- ◇データ提供者への「入力補助、データの検証、データビューワー」の3点セットを提供することで、その利用促進を促している。あるルールに従ってCSV データを用意すれば、CSV - XML 変換ツールにより、自動的に IATI 標準のXML データが作成できる^(*13)。

(*11) <http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm>

(*12) <http://www.aidtransparency.net/>

(*13) <http://iatistandard.org/getting-started/tools-to-help/>

3. ベストプラクティス / ガイドラインの事例

XBRL (eXtensible Business Reporting Language)

XBRLは、財務諸表などのビジネスレポートを電子化して、ビジネスレポートの作成の効率化、比較・分析などの二次利用が行えることを目的として、XMLの規格をベースに作られた言語である。

これは、データの主なユーザである会計士の団体が提案のきっかけの活動を担い、さらにその利用を政府機関が促したことから標準になるという、標準化に対する成功事例である。

◇webの初期の段階でアメリカのSEC (Securities and Exchange Commissions, www.sec.gov/) は、自らが監督する株式市場、証券会社のデータのWEBでの公表に躊躇した。

◇しかし個人の努力で一部のデータがInternetに流れる、その情報源へのアクセスが大量に殺到した。

◇その人気をみた結果SECも公開に前向きになった。

◇この過程で、XMLを財務データ向きにした表現を使って財務情報を交換するアイデアが民間から提案され、検討の末XBPLという標準規格になった。

(*11) <http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm>

(*12) <http://www.aidtransparency.net/>

(*13) <http://iatistandard.org/getting-started/tools-to-help/>

4. オープンデータフォーマット仕様を提案している団体

FGDC (Federal Geographical Data Committee)

地理データを国家機関でのユーザ、一般市民に提供する地理情報のインフラ提供を行なうための調整を行う機関。

◇FGDCの活動は、National Spatial Data Infrastructure (NSDI) として知られている。

FGDA はNSDI向けに自分たちでアメリカ国内の規格を提案制定すると同時に、NSDIにおいて利用すべき外部で制定された規格 (de jure, de facto の両方の標準) の推薦リストを作成している。

◇内部標準 (27件): <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/fgdc-endorsed-standards>

◇外部標準 (65件): http://www.fgdc.gov/standards/fgdc-endorsed-external-standards/index_html

FGDAはさらに、地理データを扱える既存のソフトウェア製品の推奨リストを作成している。(割引値段の交渉済み)

Big Data Initiative

アメリカでは、ホワイトハウスが2012年3月に、2億ドルの予算をつけた Big Data Initiative という R&D プロジェクトをアナウンスした。目的は以下の通り。

◇Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data.

◇Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning;

◇and Expand the workforce needed to develop and use Big Data technology

第1の目的から、ファイルフォーマットのより厳密な定義、拡張、ポータビリティをあげていろいろなプラットフォームで動作するようにするなどの標準活動がなされると考えられる。

5. メタデータ記述

Australia AGLS


政府情報の検索に使うシソーラスに使う目的で規定されたメタデータ。
Dublin Coreを拡張したボキャブラリ規格である。

米国NIEM

米国政府が後援する、公共および民間部門の組織間での情報共有を促進するためのイニシアチブ。
もともとは司法関係のデータ交換記述フォーマットであったが、他の分野（ドメイン）にも拡大され続けている。

EU ISA

EU政府間のデータ交換のための技術仕様。
W3Cやfoafのボキャブラリをベースに定義されている。



3. データポータル調査

「3.データポータル」の調査」の構成

すでに公開・運用されている公共データのポータルを調査する。

調査の目的

利用されている標準データフォーマット

利用されている標準を規定している団体

メタデータ記述の例

調査の対象

政府が公開するデータポータル

1. アメリカ: www.data.gov
2. イギリス: www.data.gov.uk
3. オーストラリア: www.data.gov.au
4. ニュージーランド: data.govt.nz
5. カナダ: data.gc.ca
6. インド: data.gov.in

市のレベルでコンテストを開きアプリケーションを集めたことで知られるデータポータル

7. Washington DC: data.octo.dc.gov/
8. New York City: nyc.gov/data

1. data.gov

データセット: 38万件

ほとんどは地理空間情報であり、それは外部のサイトから取得できるようになっている。

特徴

提供しているデータの一覧表をCSV / XLS / XML等の各種データ形式でダウンロードできる。

表形式データは、その1行目がタイトル、残りがデータセルと解釈される。

◇タイトルに関する規約はなく、このためタイトルの意味が分からないというユーザの不満もある。

現在は、Socrata社のソフトウェアを利用して作成されている。

◇将来は、Open Government Platform (OGPL: Drupal + CPAN) に移行する方針であることを、2013年1月に表明。

data.gov内のデータが準拠する標準規格

地理空間情報関連の標準規格は、OGC (Open Geospatial Consortium) が定めたものが圧倒的に多い。

◇その標準規格の多くは、ISO標準としても採択されている。

ISO/IECやITU-Tが関与する標準もあるが、これらはJPEGやMPEGなど画像データの圧縮基準である。

| 標準化団体 | 標準規格数 |
|---------------|-------|
| OGC | 21 |
| INCITS/ISO | 16 |
| INCITS | 7 |
| ISO/TS | 4 |
| ISO | 7 |
| ISO/IEC ITU-T | 2 |
| ISO/IEC | 2 |

2. data.gov.uk

データセット数: 9,024

CKANシステムを利用している。

オープンな標準規格に基づかないデータ、機械可読でないデータも多く登録されている。

| オープンで機械可読なフォーマット | | オープンでない / 機械可読でないフォーマット | |
|------------------|-------|-------------------------|-------|
| データ形式 | 登録件数 | データ形式 | 登録件数 |
| CSV | 2,346 | XLS | 1,595 |
| HTML | 415 | PDF | 562 |
| RDF | 191 | DOC | 57 |
| XML | 183 | PPT | 41 |
| JSON | 21 | Zip | 51 |
| TXT | 14 | CSV + ZIP | 26 |
| RDFa | 14 | ... | |
| KML | 13 | | |
| ... | | | |

3. data.gov.au

データセット数: 1,126

人口あたりのデータセット数に換算すると、アメリカやイギリスと比べて遜色がない

特徴

PSI を公開する際のガイドラインがトップレベルからリンクされている。

内容に関するユーザのフィードバックポストの最近の抜粋が最初のページにある。

利用しているデータフォーマット

◇ CSV/XLS / TXT / XML / RDF / KML/KMZ / Shapefile / Catalog

PSI を公開する際のガイドライン: データを公開する際に、下記に留意すべきと記されている。

Accessibility

◇ 機械可読であるべき。

Non-discrimination

◇ いつでも誰でも、登録とか、個人情報の入力を必要せず、何の制限もなくアクセスできるべき。

Open Standards

◇ proprietary なソフトウェアに依存するデータフォーマットを避けるべき。

◆ proprietary データ形式とそれを置き換えることができるフォーマットの具体例も示されている。例: Access データベース → XML, CSV, RDF

◇ 選択する機会があれば、つねにプラットフォームに依存しない、機械可読なデータ標準を選択すべき。

◇ Open Standards に関する具体的な事例も紹介されている。

◆ 人が使う文書 (具体例: Microsoft Word, RTF and PDF files) は、ダウンロードして利用するデータには向かない。

◆ 大きなデータセットは、ダウンロード時間を短縮するために圧縮するべき (例 zip 形式)。

◆ 可能なところではリンクを使ったり、API を公開することでアクセスがより容易になる。

Metadata Standards and Documentation

◇ データセットの特徴づけをメタデータを利用して記述すること。

◇ 具体的なメタデータの記述例:

◆ Title / Description / Abstract / Date Published / Authoring Agency / Subject / category / Licence / Temporal Coverage (いつのデータか) / Spatial Coverage (どの地域のデータか)

◇ データセットの利用価値を上げるメタデータの例:

◆ Low-level keywords / tags / Granularity / Update frequency / Date of last update / Agency program / Agency jurisdiction / Collection mode

4. data.govt.nz

データセット数: 2,274

特徴

他の政府機関が保持するデータを参照するポータルとしての機能のみを提供し、自らはデータをホスティングしない。

データ公開の方針として、「機械可読な、しっかりしたフォーマットかつオープンなデータセットを中心とするが、webページのようなデータも役立つことがある」と表明している。

☆実際、データセットの約50%である1,226件は、CSV形式のデータである。

2次利用に関しては関連機関に問い合わせよ、という表記が多く見られる。

その他のポータル調査結果

5. カナダ: data.gc.ca

ダウンロード可能なデータのフォーマット

◇CSV/Text / KML/KMZ / RDF / RSS/GeoRSS / Shapefile / XML

2013年に、オープンソースのOpen Government Platform (OGPL) に移行

6. インド: data.gov.in

オープンソースのOpen Government Platform (OGPL)を利用して構築されている

7. ワシントンDC: data.octo.dc.gov

ダウンロード可能なデータのフォーマット

◇Text/CSV / Atom feed / XML / Shapefile / KML

8. ニューヨーク市: nyc.gov

data.govと同じSocrata社のソフトウェアを利用しているが、データセットカタログのフォーマットはdata.govとは異なっている。

◇アメリカ政府全体での、検索用のメタデータ記述の統一がすすんでいないか、またはそれを具体的なデータフォーマットに埋め込むルールのベストプラクティスがないか。

◇メタデータ記述に限らず、CSVのようなデータ記述においても、タイトルの説明を丁寧に行うなどのベストプラクティスを提示するだけでも、PSIのオープンデータポータルを将来作成して運用するにあたって、ユーザの利益になることは間違いない。



4. まとめ

4. まとめ

現状分析の結果

データを検索・発見するためのメタデータ記述の標準は進んでいる。

しかし、データそのもののメタデータ基準は進んでいない。

データのカラム名をわかり易いものにする、具体的な意味の厳密な説明をする、等により、データを容易にアクセスできるための具体的な指針、ベストプラクティスがないのが、現状の大きな問題である。

当面目指す標準の範囲

PSI公開のワークフローの中で、データを準備する部分に対するベストプラクティスを、標準として提案する。

