

オープンガバメントを実現するシステムの可能性等についての調査研究
統計情報提供基盤(通称:データボックス)の機能評価・検証について

平成22年11月

経済産業省

委託先 株式会社サンビジネス

目次

第1章 調査研究の概要	1
1.1 背景	1
1.2 目的	2
第2章 統計情報提供基盤（プロトタイプ）実証実験の概要	5
2.1 システムの概要（開発・提供した機能）	5
2.1.1 OpenIDによるユーザ登録・ログイン	8
2.1.2 検索機能	9
2.1.3 検索結果の表示	10
2.1.4 個別データの詳細表示	11
2.1.5 広告エリア掲載機能	18
2.2 調査対象の概要	19
2.3 アンケートの内容	21
2.4 有識者ヒアリングの内容	22
2.5 実施スケジュール	24
第3章 統計情報基盤に向けられている期待/要望	25
3.1 統計情報基盤のニーズ	25
3.2 統計情報基盤の機能に関する要望（プロトタイプ構築にあたって）	27
3.3 統計情報提供形式（XML/SDMX）の利用イメージ	29
3.4 統計情報基盤開発の留意点	31
3.5 中長期の課題・要望	32
第4章 実証実験の結果に基づく評価・考察	34
4.1 利用者観点での考察	34
4.1.1 実証実験に参加した利用者層	34
4.1.2 統計情報の利用目的の分析	38
4.1.3 GUIなどシステム操作性の評価	41
4.1.4 検索機能の評価	43
4.1.5 提供されたデータ表現形式（XML/SDMX）の評価	45
4.1.6 提供されたデータ構造の評価	49
4.1.7 グラフ化機能の評価	51
4.1.8 詳細分析機能の評価	52
4.1.9 OpenIDの評価	54
4.1.10 その他の個別機能の評価	56
4.1.11 利用者の期待度の考察	60
4.2 開発者・運用者の観点での考察	65

4.2.1 開発における工夫・課題	65
4.2.2 運用／保守時における課題.....	68
4.2.3 XML/SDMX活用に関する課題.....	69
第5章 全体評価.....	70
5.1. 有用性・利便性の評価.....	70
5.1.1 一般利用者の観点での評価.....	70
5.1.2 ヘビーユーザ（統計専門家）の観点での評価	71
5.1.3 サービス提供の観点での評価	72
5.2 統計情報データの新たな活用形態に基づく市場活性化への期待	73
5.3 利用者の認知度について	75
第6章 今後の展望・提言.....	76
6.1 今後の課題等について.....	76
6.2 今後の取り組みについて	78
6.3 最後に.....	83
用語集.....	84
参照資料	87
附録A. 「データボックス オンラインアンケート」の結果一覧.....	89
附録B. オープンガバメント系サイトのツールやマッシュアップ事例.....	102
附録C. SDMX関連情報の概要	126

第1章 調査研究の概要

今回の調査研究の概要として、以下に調査の背景と目的を記す。

1.1 背景

2010年5月11日に高度情報通信ネットワーク社会推進戦略本部（IT戦略本部）がまとめた「新たな情報通信技術戦略」においては「国民本位の電子行政の実現」の記述が見られ、「オープンガバメント等の確立」が重点施策として位置づけられている。これは、双方向性を有するインターネットの機能を活用することで、行政情報の公開や行政への市民参加を促進する世界各国政府の「オープンガバメント」化の流れに沿った提言である。

特に米国においては、オープンガバメントの方針として「透明性の向上 (transparency)」、「市民参加 (participation)」、「政府間及び官民協力 (collaboration)」の3原則の下、国民に向けた様々なサイトの構築が進められている。

□透明性の向上 : *transparency*

－国民・行政間の情報格差を解消すること

□市民参加 : *participation*

－インターネットを活用し、行政に対し直接声を届けられるようにすること

□政府間及び官民協力 : *collaboration*

－官庁間の連携、地方行政と中央官庁との連携、政府と民間との協力を行うこと

例えば、「市民参加 (participation)」については、国民が意見を投稿し、賛同する意見に投票し、得票数が多い意見が上位に上がってくるクラウドソーシング型ツールを、大統領による市民集会の開催や、各政府機関のオープンガバメント計画の作成等に用いている。

日本においても、昨年度2009年10月、2010年2月の2度にわたり、同様のクラウドソーシング型ツールを用いて、政策アイデアを募集する「アイデアボックス」を実施したところ、多数の参加を得、様々な意見をいただいた。

特に2010年2月に実施された経済産業省アイデアボックスでは、3,799名の登録があり、モデルとなった米国での参加者（アメリカでは2010年2月～4月までのユーザ登録者数が約1,500名）を上回る反響があり、日本でのオープンガバメントへの関心の高さを物語る結果となった。また、そのアイデアボックスにおいて、統計情報を含む行政情報全般に関し、「日本版 DATA.gov」の開設、行政情報 API の公開など、多くの要望が寄せられた。

これによって、今回は、今までの「市民参加 (participation)」の試みを更に高度化し、活用していくことが強く求められているという認識に立ち、オープンガバメント3原則の他の原則である「透明性の向上 (transparency)」、「政府間及び官民協力 (collaboration)」にも目を向け、これまでのオープンガバメントの仕組みを補完する実証実験を行うことと

した。

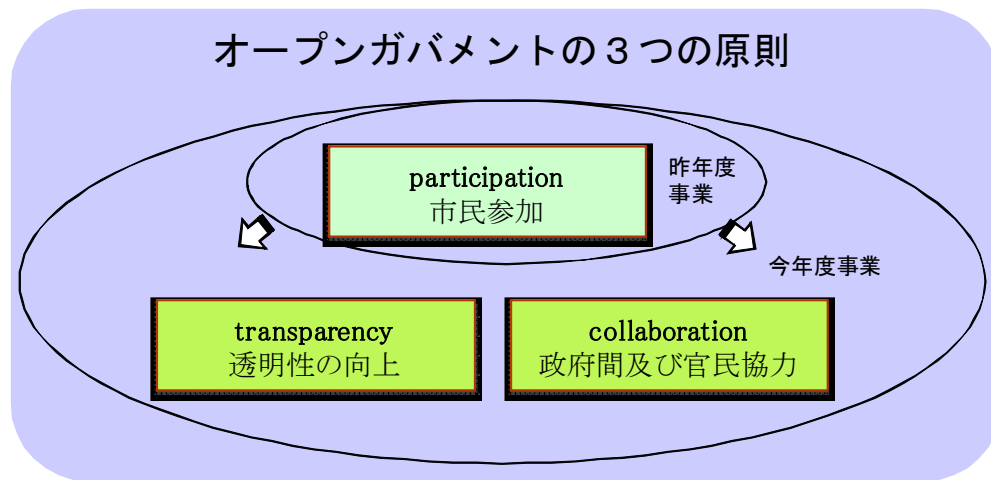


図 1.1 オープンガバメントの3つの原則

1.2 目的

本事業では、昨年度までの現状を踏まえ、積極的な行政情報の公開や行政への国民参加を実現する IT システムの可能性等について、モデルシステムを用いた具体的な実証実験の実施・検証を通じ、日本での「オープンガバメント」における IT システムの可能性の調査・研究を行うことを目的とする。

前述のIT戦略本部「新たな情報通信技術戦略」（2010年5月11日決定）においては、重点施策として「オープンガバメント等の確立」が取り上げられ、下記の方針が打ち出されている。（<http://www.kantei.go.jp/jp/singi/it2/100511honbun.pdf>）

-
- 行政が保有する情報を二次利用可能な形で公開して、原則としてすべてインターネットで容易に入手できるようにするなど、行政が保有する情報の公開を積極的に推進する。
 - 行政が保有する統計・調査などの情報について、回答者の個人情報保護の観点から、個人が特定できない形に情報の集約化・匿名化を行い、それらを原則としてすべて二次利用可能な形でインターネットで容易に入手し、活用できるようにすることにより、新事業の創出を促進する。
-

この方針を実現するため、本事業中において、特に公開する行政情報を「統計情報」としたモデルシステムを構築した。

統計情報を対象とした一つの理由には、昨年来全面施行されている「新統計法」の存在

が挙げられる。60 年振りに改正されたこの統計法では、公的統計は「国民にとって合理的な意志決定を行うための基盤となる重要な情報である」と位置づけられている。また、新統計法の施行に合わせて閣議決定された「公的統計の整備に関する基本的な計画」では、統計情報の有効活用を推進し、多様化・高度化する利用者のニーズに的確に対応することが目標として掲げられている。

この点で、今回開発した統計情報基盤は、インターネットを用いた統計データの公表や検索機能の充実を図るものであり、統計情報を公表するだけでなく、統計データの集計・分析機能や提供ファイル形式の多様化にも対応した統計情報の提供手段とすることを目指している。

このモデルシステムの検討・構築に当たっては、平成 21 年度委託調査研究の成果により以下の点が重要との指摘があり、これを参考にした。

「統計情報の XML 利用環境の整備に関する調査研究に係る報告書」

委託先：伊藤忠テクノソリューションズ株式会社、

URL：http://www.meti.go.jp/statistics/toppage/topics/kenkyuShiryo/hokokusho/pdf/xml_houkokusyo.pdf

- ①多様なユーザに対応した検索機能の強化
- ②ライブラリ整備等統計情報管理基盤の強化
- ③調査対象者等へのフィードバック機能の提供
- ④都道府県等へのパターンオーダー型提供基盤の整備
- ⑤統計 XML フォーマット策定の検討
- ⑥クラウドを活用した統計情報提供基盤の機能等検証

ここに挙げられた要件に従い、国民が統計情報に容易にアクセスし活用できるようにするためのシステムのあり方（モデル仕様書・設計書）及びデータフォーマット等についてモデルシステムを構築し、それを使用した実証実験を通じて調査研究を行った。また、この中で、行政職員が、行政情報に書誌情報を付してインターネット上に設置したリポジトリにアップロードしたデータを、国民が簡単に検索して入手でき、当該行政情報に対してコメントや評価を付すことができるシステムについても検討を行った。

また、統計情報等の行政情報の二次的利用を促進するためのデータ形式や提供方法についても、SDMX（※）など XML 書式の適用検証等を通じて検討を行った。

※SDMX：OECD の統計情報提供基盤（SIS）にて情報交換のために標準化された XML をベースとした言語。

更に、統計情報基盤を構築するにあたっては、クラウドコンピューティング、OSS（オープンソースソフトウェア）という 2 つの技術的潮流に沿って開発を行うこととした。

クラウドコンピューティングとは、コンピュータネットワークを運営する組織から、ハードウェア、ソフトウェア、サービス等の提供をインターネットを介して受け、利用した

分だけの費用を支払うという新しいコンピュータ処理の概念である。これまでは、各組織が自前でハードウェア、ソフトウェアを調達し運用しなければならなかったため、管理者の人件費も含め多大な IT 投資が必要とされたが、期限付きサービスとしてハードウェア、ソフトウェアを利用できることによって、システム導入期間の短縮、初期・運用コストの削減、サービスの迅速な開始が実現できる。更に、システムリソースの拡張にも柔軟に対応できるなどのメリットもある。

従って、オープンガバメントを含め、ニーズに対する迅速な対応やコスト軽減が眼目となる情報系システムでは、クラウドコンピューティングはこの要件を満たす技術のひとつとして官民合わせその動向が注目されている。

一方、OSS（オープンソースソフトウェア）とは、コンピュータプログラムのソースコードが無償で公開され、その改変を自由に行えるソフトウェアのことであり、旧来の有料パッケージソフトウェアの考え方とは大きく異なるソフトウェアの開発・利用の形態である。

OSS は、独自の目的に沿った機能追加など、旧来のベンダ依存のパッケージソフトウェアには求めることのできなかった柔軟性が得られ、更には特定ベンダへのロックインを防げるというメリットもある。また、政府機関では、オープンソースソフトウェアを調達要件に入れることによって、入札参加者の参入障壁を低くし、公平な競争を実現できるという点も挙げられ、オープンガバメントの「透明性」にも寄与するものとなる。

そこで、今回は民間のクラウドサービスの上にいくつかのオープンソースソフトウェアを活用し、システム開発を行うことにより、これらの政府系システムへの適用可能性についても実証・検討することとした。

第2章 統計情報提供基盤（プロトタイプ）実証実験の概要

前章の目的に資するシステムとして本事業では、「A. 実証実験共通基盤」「B. 国民・行政職員等との共同作業基盤」「C. 行政情報提供基盤」「D. 統計情報提供基盤」の4つの基盤から構成されるモデルシステムを構築し、特に「D. 統計情報提供基盤」について評価・検証を行った。この章では、システムの機能、それに対する評価の方法など、実証実験の概要を述べる。

2.1 システムの概要（開発・提供した機能）

次の図は、モデルシステムの4つの基盤とそこで提供される機能を示したものである。

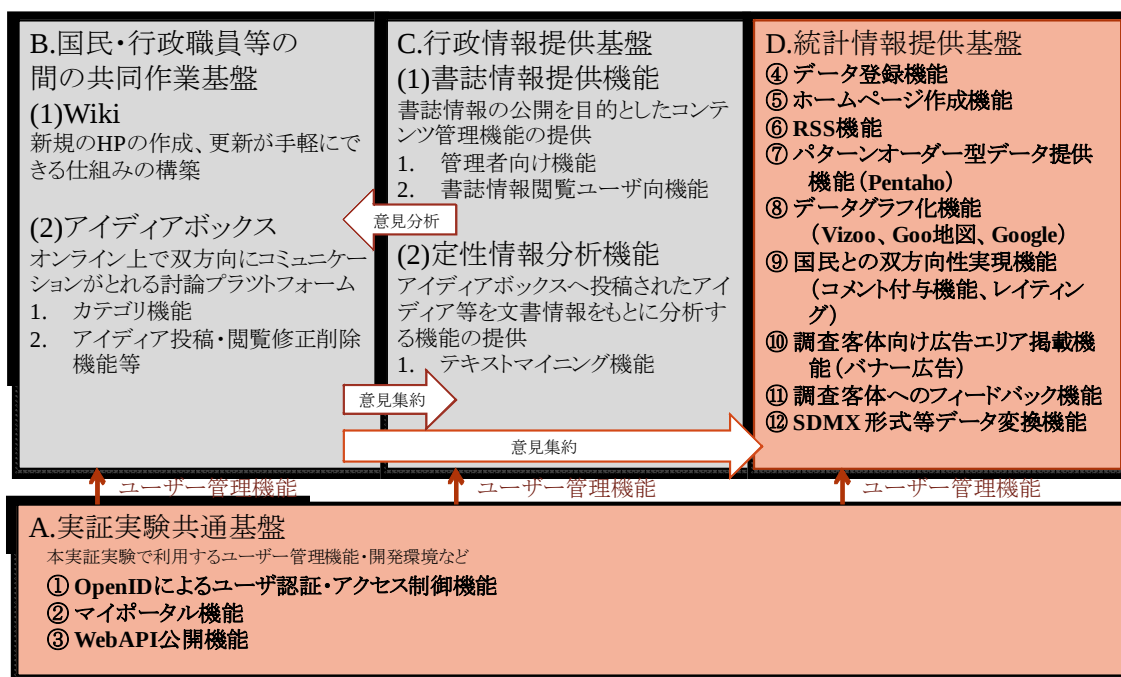


図 2.1 モデルシステム全体における統計情報提供基盤（プロトタイプ）の位置付け

今回の評価の対象となる機能は主に「データボックス」として実現されている「D. 統計情報提供基盤」であるが、ログインに関連する「ユーザ認証・アクセス制御機能」などいくつかの機能は「A. 実証実験共通基盤」の中で実現している。これらの機能を以下に列挙する。

- ① OpenID によるユーザ認証・アクセス制御機能
 - ーOpenID を活用したユーザ認証・アクセス制御機能を提供
- ② マイポータル機能
 - ーユーザごとに画面カスタマイズが可能な機能を提供
- ③ Web API 公開機能
 - ー本事業において提供する Web API を一般開発者向けに公開する機能を提供
- ④ データ登録機能
 - ー経済産業省の担当職員からアップロードしたデータを登録する機能を提供
- ⑤ ホームページ作成機能
 - ー登録した統計データを公開、検索、ダウンロードできるホームページを作成するための機能を提供
- ⑥ RSS 機能
 - ー統計情報の公表、修正情報の提供等のための RSS 機能を提供
- ⑦ パターンオーダー型データ提供機能
 - ー非定形分析機能で、ユーザが自由に表頭・表側の選択や、ドリルダウンやドロップダウンを実現できる機能を提供。『Pentaho』を利用してサービスを提供
- ⑧データグラフ化機能
 - ーユーザの操作に応じて、容易にグラフ化できる機能を提供する。『Vizoo Custom』『goo 地図』『Google モーションチャート』を利用してサービスを提供
- ⑨国民との双方向性実現機能
 - ー提供された統計情報に関するレイティング機能やコメント付与機能を提供
- ⑩ 調査対象者向け広告エリア掲載機能
 - ー調査対象者向けにバナー広告エリアを掲載する機能を提供
- ⑪ 調査対象者へのフィードバック機能
 - ー調査対象者から提供されたデータを含む統計データを検索・ダウンロードできる機能を提供
- ⑫ SDMX 形式等データ変換機能
 - ーアップロードされた統計データを SDMX 形式に変換する機能を提供

これらの機能を使って実現したデータボックスの GUI および操作方法を以下に記す。

まず、データボックスのメイン画面は以下のような構成となっており、左の区画で検索機能が提供され、右の区画で検索結果の表示が行われる。

The screenshot shows the Databox web application. On the left is a sidebar with search filters. The main area on the right displays a list of data items with columns for title, evaluation, responsible ministry, format, and number of downloads.

タイトル	評価	担当府省	形式	DL件数
全産業活動指数、全産業供給指数・月次開示・2010年9月分・平成22年11月19日公表	☆☆☆☆☆	経済産業省	PDF	13
全産業活動指数、全産業供給指数(全産業活動指数・原指数・月次 9月分)・平成22年11月19日公表	☆☆☆☆☆	経済産業省	Excel	10
全産業活動指数、全産業供給指数(全産業活動指数・季節調整済指数・月次 9月分)・平成22年11月19日公表	☆☆☆☆☆	経済産業省	Excel	0
全産業活動指数、全産業供給指数(全産業供給指数・季節調整済指数・月次 9月分)・平成22年11月19日公表	☆☆☆☆☆	経済産業省	Excel	0
全産業活動指数、全産業供給指数(全産業供給指数・季節調整済指数・四半期、原指数・年 9月分)・平成22年11月19日公表	☆☆☆☆☆	経済産業省	Excel	5
経済産業省特定業種石油等消費統計 平成22年09月確報・平成22年11月10日公表	☆☆☆☆☆	経済産業省	Excel	2
特定サービス産業動態統計調査 2010年9月分(確報)対事業所サービス業 その1・平成22年11月16日公表	☆☆☆☆☆	経済産業省	Excel	34
特定サービス産業動態統計調査 2010年9月分(確報)対事業所サービス業 その2・平成22年11月16日公表	☆☆☆☆☆	経済産業省	Excel	12
特定サービス産業動態統計調査 2010年9月分(確報)対個人サービス業・趣味・娯楽関連・平成22年11月16日公表	☆☆☆☆☆	経済産業省	Excel	14
特定サービス産業動態統計調査 2010年9月分(確報)対個人サービス業・娯楽・健康関連・平成22年11月16日公表	☆☆☆☆☆	経済産業省	Excel	26

図 2.2 データボックスのメイン画面

2.1.1 OpenID によるユーザ登録・ログイン

データボックスでは以下の 2 つの方法で OpenID を利用できる。

1. Yahoo! JAPAN、Mixi、livedoor、Google の ID でログイン。
2. 経済産業省が発行する OpenID でログイン（経済産業省が発行する OpenID はアイディアボックス等と共通）。



図 2.3 ユーザ登録画面



図 2.4 ログイン画面

2.1.2 検索機能

統計情報を様々な条件を指定して検索することができるが、その操作は以下の検索エリアで行う。

図 2.5 検索エリア

ここで行える操作は以下の通り。

■新着ダウンロード

新着のデータ一覧を表示。

■データ検索

以下の条件指定が可能（各項目間の関係は AND 条件）。

- ☐ キーワード： 自由なキーワードで統計データを検索
- ☐ 書誌情報・統計情報： チェックした情報を対象に検索
- ☐ 評価： 利用者が付けた評価を条件に検索
- ☐ ファイル形式： PDF、Excel、XML（SDMX データを含む）の形式でファイルが提供されている統計データを検索（各ファイル形式の関係は OR 条件）
- ☐ 属性情報：
 - ー政策分野から探す
(例：「製造産業」、「情報政策」、「流通・商務」、「技術革新」..)

ー統計分野から探す

(例:「鉱工業」、「商業・サービス業」、「人口・世帯」、「労働・賃金」..)

ー作成機関から探す ※今回は実証実験であり、収録データは「経済産業省」のみ

(例:「経済産業省」、「総務省」、「特許庁」、「中小企業庁」..)

ー統計種類から探す

(例:「基幹統計」、「一般統計」、「業務統計」、「加工統計」..)

ー分析種別から探す

- ・ 共有資料：一般的なデータ
- ・ 簡易グラフ：折れ線グラフ等のグラフが提供されているデータ
- ・ 都道府県グラフ：都道府県の地図グラフが提供されているデータ
- ・ 時系列グラフ：時系列グラフが提供されているデータ
- ・ 詳細分析：パターンオーダー型分析機能が提供されているデータ
- ・ SDMX 形式：SDMX 形式のデータが提供されているデータ

2.1.3 検索結果の表示

これらの条件を設定し検索することにより、結果が下記のように表示される。

ゲストさん、ようこそ。 [ログイン](#) [アンケートは](#)

Labs オープンガバメントラボ Govter がはったー IdeaBox アイディアボックス IdeaBox Local 地方自治体アイディアボックス Wiki オープンガバメント ウィキ DataBox データボックス

1 2 3 4 5 6 7

(1 - 50 / 316)

[この一覧のRSSを取得](#)

タイトル	評価	担当府省	形式	DL件数
特定サービス産業動態統計調査 2010年9月分(速報)対事業所サービス業 その1-平成22年11月9日公表-	☆☆☆☆☆	経済産業省	Excel	24
特定サービス産業動態統計調査 2010年9月分(速報)対事業所サービス業 その2-平成22年11月9日公表-	☆☆☆☆☆	経済産業省	Excel	11
特定サービス産業動態統計調査 2010年9月分(速報)対個人サービス業・趣味・娯楽関連 -平成22年11月9日公表-	☆☆☆☆☆	経済産業省	Excel	9
特定サービス産業動態統計調査 2010年9月分(速報)対個人サービス業・教養・生活関連 -平成22年11月9日公表-	☆☆☆☆☆	経済産業省	Excel	21
鉱工業出荷内訳表、鉱工業総供給表 2010年9月分速報概要(PDF)-平成22年11月4日公表-	☆☆☆☆☆	経済産業省	PDF	42
鉱工業出荷内訳表 時系列データ 2003年1月～2010年9月分速報 -平成22年11月4日公表-	☆☆☆☆☆	経済産業省	Excel	29
鉱工業総供給表 時系列データ 2003年1月～2010年9月分(速報)-平成22年11月4日公表-	☆☆☆☆☆	経済産業省	Excel	21
工業統計調査 平成20年確報 企業統計編・利用上の注意 -平成22年10月29日公表-	☆☆☆☆☆	経済産業省	PDF	1

図 2.6 検索結果表示エリア

リスト上部の[タイトル]、[評価]、[担当府省]、[形式]、[DL 件数]をクリックすると、そ

の項目毎の並べ替えが行える（「タイトル」「担当府省」「形式」は名称によるソート、「評価」「DL 件数」は数の高い順に並べるソート）。

【RSS 機能】

[この一覧の RSS を取得]をクリックすると、表示されている画面の RSS を取得できる。



図 2.7 取得した RSS の表示

2.1.4 個別データの詳細表示

検索結果として表示されたものの中からいずれか 1 件のデータを選んでクリックすると、そのデータの詳細表示エリアへ遷移する。



図 2.8 詳細表示エリア

【データのダウンロード】

ユーザは、公開されているデータを自分のコンピュータにダウンロードし活用することができる。図 2.8 の場合、エクセル形式のデータが提供されており、「Excel 形式 ダウンロード」をクリックするとデータがダウンロードされる。

【コメント付与】

国民との双方向性を実現するための機能の一つとして、データに対してコメントを投稿することができる。図 2.8 の「コメント送信」ボタンの左のコメント欄にテキストを入力し「コメント送信」をクリックする。

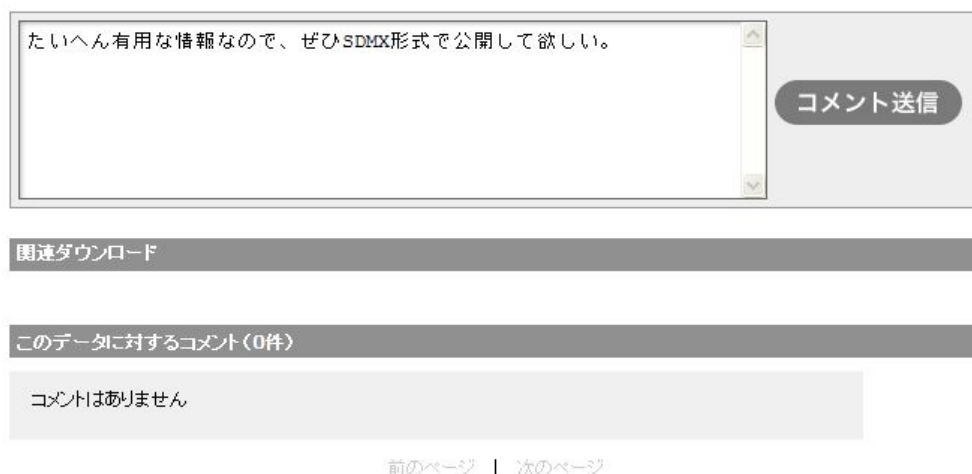


図 2.9 コメント付与機能

【このコメントに返信】をクリックすることによって他の利用者のコメントに返信することもできる。これによってユーザ間の一種の意見交換を行うことも実現できる。



図 2.10 他者のコメントへの返信による意見交換

【レイティング】

国民との双方向性を実現するためのもう一つの機能として、ログインした状態であれば、マウスカースルを☆の上に移動し、評価点の上で左クリックすることにより、データの評価を行うことができる。



図 2.11 レイティング機能

【簡易グラフ】

選択したデータが「簡易グラフ」に対応していると下記のようにグラフ表示される。(検索条件の[分析種別から探す]で[簡易グラフ]をチェックし検索すると、簡易グラフが提供されているデータを検索できる)

簡易グラフ機能は「vizoo」(<http://www.visualzoo.com/>)を利用している。

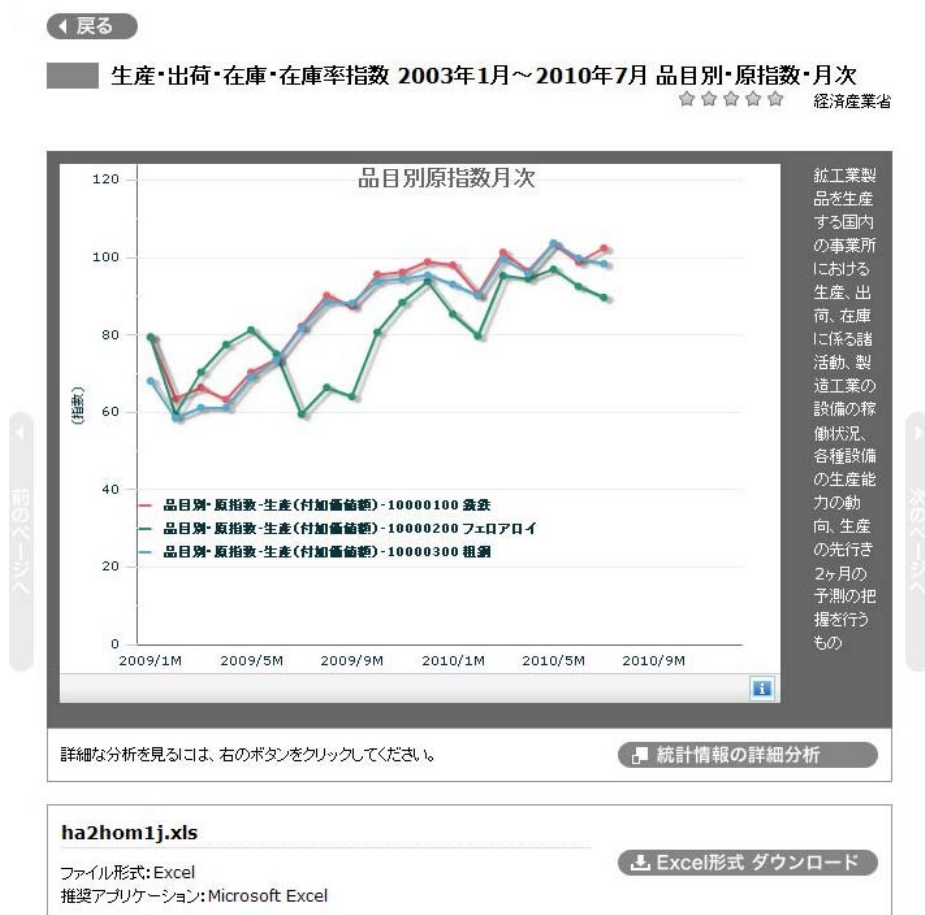


図 2.12 簡易グラフ

【都道府県グラフ】

選択したデータが「都道府県グラフ機能」に対応していると下記のようにグラフ表示される。(検索条件の[分析種別から探す]で[都道府県グラフ]をチェックし検索すると、都道府県グラフが提供されているデータを検索できる)

都道府県グラフ機能は、「goo統計地図」(<http://map.goo.ne.jp/datamap/>)を利用している。



図 2.13 都道府県グラフ

【時系列グラフ】

選択したデータが「時系列グラフ機能」に対応していると下記のようにグラフ表示される。(検索条件の[分析種別から探す]で[時系列グラフ]をチェックし検索すると、時系列グラフが提供されているデータを検索できる)

時系列グラフ機能は、Google モーションチャートを利用している。

戻る

商業統計 平成19年確報・第1巻 産業編(総括表)



経済産業省

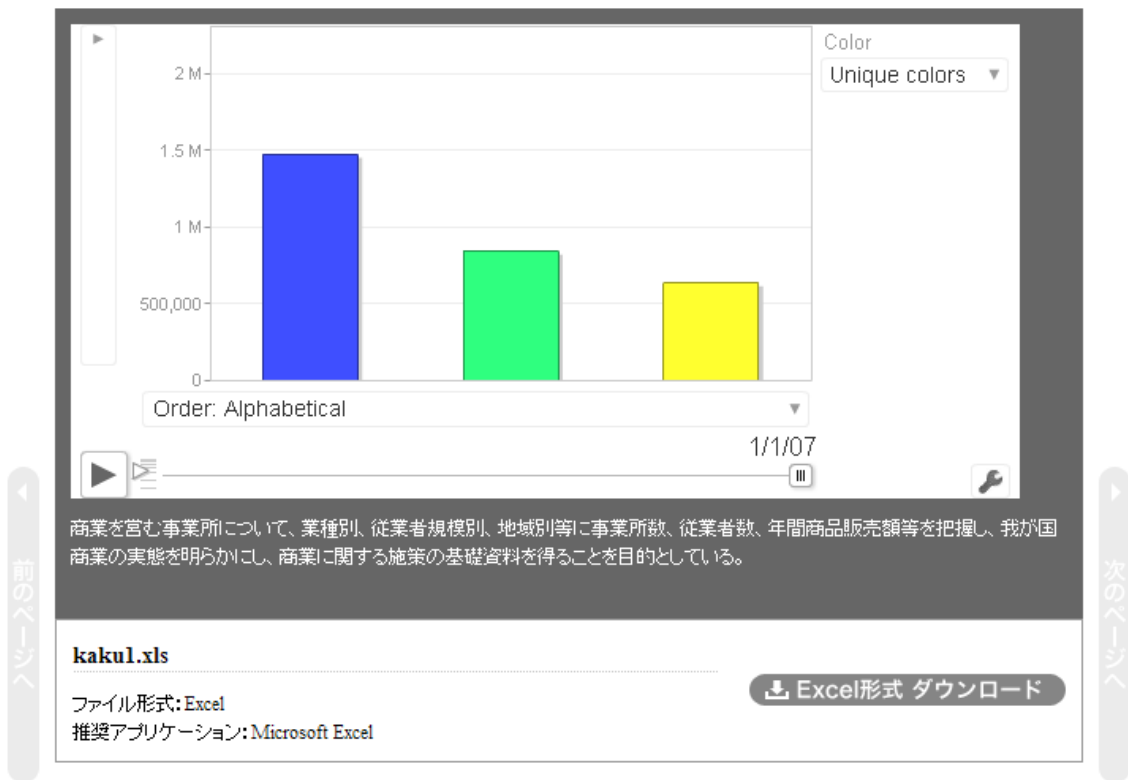


図 2.14 時系列グラフ

【詳細分析機能】

選択したデータが「詳細分析機能」に対応していると下記のようにグラフ表示される。(検索条件の[分析種別から探す]で[詳細分析]をチェックし検索すると、詳細分析が提供されているデータを検索できる)

詳細分析機能は、『Pentaho』を利用している。

原指数-出荷				
品目分類	時点			
	201005		201004	
	Measures		Measures	
	● ウェイト	● 指数値(加重平均)	● ウェイト	● 指数値(加重平均)
■ 鉱工業	10,000	86.5	10,000	91.4
■ 製造工業	9,990.1	86.5	9,990.1	91.4
+ 鉄鋼業	547.4	94.6	547.4	87.9
+ 非鉄金属工業	262.8	84.2	262.8	89.4
+ 金属製品工業	479.9	74.6	479.9	80.5
+ 一般機械工業	1,159.9	69.9	1,159.9	75.3
+ 電気機械工業	641.3	83.7	641.3	84.2
+ 情報通信機械工業	461.2	92.3	461.2	98.3
+ 電子部品・デバイス工業	763.7	121.5	763.7	121.6
+ 輸送機械工業	2,014.7	82.5	2,014.7	88.9
+ 精密機械工業	98.1	92.2	98.1	98.7
+ 窯業・土石製品工業	205.6	75.9	205.6	80.8
+ 化学工業	953.8	90.2	953.8	93.5
+ 石油・石炭製品工業	520.8	81.6	520.8	88.1
+ プラスチック製品工業	337.9	84.1	337.9	91.9
+ パルプ・紙・紙加工品工業	238.8	85.5	238.8	94.3
+ 繊維工業	159.6	75.6	159.6	84.5
+ 食料品・たばこ工業	779.4	97.1	779.4	108.4
+ その他工業	365.2	78.5	365.2	85.2
+ 鉱業	9.9	92.2	9.9	98.8

図 2.15 詳細分析機能

【SDMX 形式データ提供機能】

選択したデータが「SDMX形式」に対応しているとSDMXデータをダウンロードできる。
 (検索条件の[分析種別から探す]で[SDMX形式]をチェックし検索すると、SDMX形式のデータファイルが提供されているデータを検索できる)



図 2.16 SDMX 形式のデータ提供

2.1.5 広告エリア掲載機能

データボックスでは、調査対象者による調査票情報の提出を促すため、バナーを貼る機能を設けている。

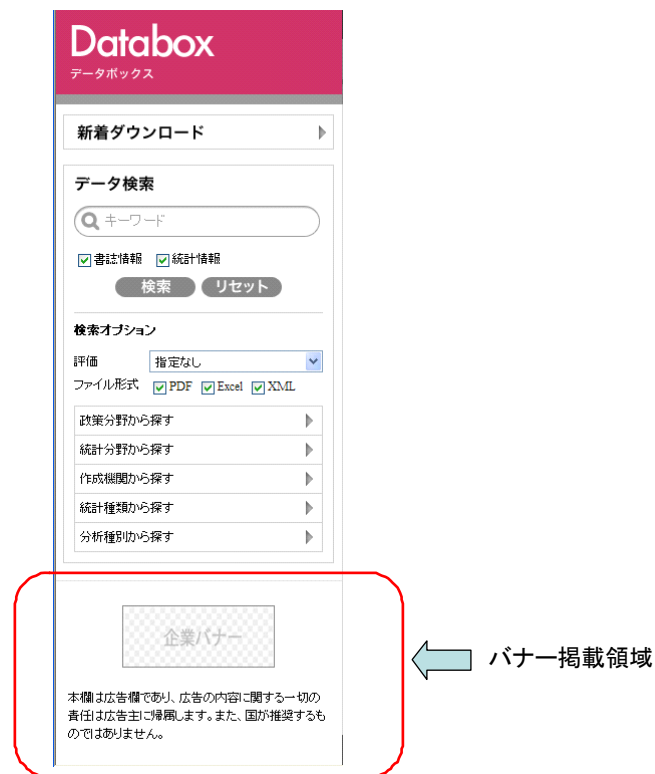


図 2.17 バナー掲載エリア

統計情報提供基盤は、2010 年 9 月 28 日にオープンガバメントラボ (※) 内で「データボックス」として機能提供を開始した。これに関連するモデルシステム全体を取り巻く関係者を図示すると以下のようになる。

Figure 1: Model System Development Process

The diagram illustrates the development and evaluation of a Model System. The central component is the **Model System**, which includes:

- A. 実証実験共通基盤 (Empirical Experiment Common Base)**
- B. 国民・行政職員等の間の共同作業基盤 (National/Local Government Joint Work Base)**
- C. 行政情報提供基盤 (Administrative Information Provision Base)**
- D. 統計情報提供基盤 (Statistical Information Provision Base)**

The process flow is as follows:

- データ提供 (Data Provision):** Provided by the **経済産業省 (Ministry of Economy, Trade and Industry)** to the **B. 国民・行政職員等の間の共同作業基盤**.
- 開発 (Development):** Conducted by the **開発者・運用者チーム (Developer/Operator Team)**.
- 利用/評価 (Utilization/Evaluation):** Conducted by **統計データ利用者 (Statistical Data Users)** and **有識者 (Experts)**.
- 評価と検証 (Evaluation and Verification):** Includes **有識者ヒアリング (事前・事後) (Stakeholder Hearings (Before/After))** and **ユーザアンケート調査 (User Survey)**. This phase feeds back into the **有識者** and **開発者・運用者チーム** via **提案要望 (Proposal/Requirement)**.

統計データ利用者：

有識者： データボックス公開前には事前ヒアリングを実施し、開発すべき機能への提案・要望を伺うことによってデータボックスの機能評価への指

導・助言をお願いした。また、公開後には事後ヒアリングを実施し、事業への評価と提言をいただいた。

開発者・運用者チーム：

有識者への事前ヒアリングで挙げられた提案・要望も念頭に、経済産業省と協力してモデルシステムの要件調整を行いつつシステムを構築した。また、今回のシステムや開発手法に関して、開発側のスタンスから評価を行うため、開発者・運用者チームへのヒアリングも実施した。

2.3 アンケートの内容

「データボックス」サイト内のアンケート機能を使った一般利用者へのアンケートは以下の設問により実施した。(実施期間：2010年9月28日～10月24日；有効回答数：55)

表 2.1 オンラインアンケートの設問一覧

設問	内容
1	あなたの業種を選択してください。
2	あなたの職務を選択してください。
3	経済産業省を含む公的組織から公開されている統計データを利用していますか。
4	ご回答頂いた、利用したことのある統計データについてお聞きします。：利用頻度を左の選択肢の中から選んでください
5	ご回答頂いた、利用したことのある統計データについてお聞きします。：最も多く利用する取得方法を左の選択肢の中から選んでください
6	ご回答頂いた、利用したことのある統計データについてお聞きします。：最も多く利用するデータ形式を左の選択肢の中から選んでください
7	ご回答頂いた、利用したことのある統計データについてお聞きします。：利用目的を左の選択肢の中から選んでください
8	本システムを何で知りましたか
9	本システムの全般的な使い勝手や操作性はいかがでしたか
10	本システムの全般的な画面デザインはいかがでしたか
11	本システムで関心を持った(便利と思った)機能は何ですか。(最大3つまで)
12	検索機能の充実は、統計データの利用を促進する要因になると思いますか。
13	グラフ化機能は、統計データの利用を促進する要因になると思いますか。
14	詳細分析機能は、統計データの利用を促進する要因になると思いますか。
15	XML 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。
16	Google や Yahoo の ID でログインできる機能(OpenID)が利用を促進する要因になると思いますか。
17	SDMX 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。
18	今後も本システムを利用したいと思いますか
19	使いたい理由について、あてはまるものをお選びください。(最大3つまで)
20	使いたくない理由について、あてはまるものをお選びください。(最大3つまで)
21	本システムに期待することは何ですか。(最大3つまで)
22	統計データの利用を促進する施策について、ご意見がありましたらお願い致します(※自由記述)

2.4 有識者ヒアリングの内容

本事業で利用された要素技術は、統計、アイデンティティ管理、クラウド、XML/SDMX など多岐に渡るが、それらに対する包括的な意見をお聞きするために、各界の専門分野で著名な 9 名の有識者にヒアリングを実施し、システム機能・運用に関する提案・要望を伺い、更に今回の実証実験への評価も頂いた。なおヒアリングは、実証実験の実施前と実施後の 2 回行った。

表 2.2 有識者一覧（50 音順、敬称略）

	氏名	所属	役職など
1	大山 敬三	国立情報学研究所 コンテンツ科学研究系	教授・研究主幹
2	小澤 英昭	NTT レゾナント株式会社 サーチ事業部	ビジネス推進部門長 工学博士
3	葛西 重雄	株式会社トリエス	代表取締役パートナー
4	崎村 夏彦	株式会社野村総合研究所 DI ソリューション事業部	上級研究員
		OpenID ファウンデーション・ジャパン	発起人代表
		The OpenID Foundation	副理事長
		Kantara Initiative	理事
5	高橋 陸春	財団法人経済産業調査会 経済統計 情報センター	理事 経済統計情報センター所長
6	椿 広計	統計数理研究所	副所長 リスク解析戦略研究センタ ー教授
7	廣松 毅	情報セキュリティ大学院大学 情報セキュリティ研究科	教授
8	安田 聖	一橋大学経済研究所 附属社会科学統計情報研究センター	教授
9	渡辺 弘美	国際大学 GLOCOM	客員研究員

【事前ヒアリング】

「データボックス」の開発が始まった 2010 年 7 月に、実現予定の機能、提供するデータの内容、データ形式などについて事務局より概要を説明し、有識者の各氏から、それぞれの専門分野での知見を基に、機能に関する要望、開発の留意点、中長期の課題・要望などに関するコメントを頂いた。これらのコメントの中で開発中の機能として反映できるものについては実証期間の中で実現している。

【事後ヒアリング】

「データボックス」の公開後、2010 年 10 月に有識者へのヒアリングを再び実施した。その際にご意見を伺った中心となる論点は以下の通りである。

1. 本プロトタイプ of 機能全般、使い勝手、操作性、画面デザイン等について
2. 本プロトタイプで特に興味がある機能や改善が必要な機能について
 - (1) 検索機能
 - (2) グラフ機能
 - (3) ファイルダウンロード機能
 - (4) 詳細分析機能（パターンオーダー機能）
 - (5) SDMX 形式 of データ提供
 - (6) コメント付与機能
 - (7) レイティング機能（データ評価機能）
 - (8) マイポータル機能
 - (9) Web API 公開機能
 - (10) RSS 機能
 - (11) OpenID によるユーザ認証・アクセス制御機能
3. データ構造の有効性について（統計データの形式、検索属性等）
4. XML/SDMX 書式適用について
5. 本プロトタイプ of 今後の期待／意見／要望について
6. 統計データの利用を促進する施策について

2.5 実施スケジュール

オープンガバメントラボのサイトは、2010 年 7 月 29 日にオープンし、まず「アイデアボックス」、「オープンガバメント Wiki」のサービスを提供した後、「がばったー」、そして本書で報告を行っている「データボックス」が順次提供された。

更に、このシステムを利用した他府省・地方公共団体の取り組みとして、「北海道陸別町」、「京都府伊根町」、「埼玉県鶴ヶ島市」の地域別アイデアボックス、「国民の声アイデアボックス」といったサービスも展開した。

これらのサービスの提供スケジュールを以下に記す。

2010年	6月	7月	8月	9月	10月	11月
【オープンガバメントラボ自体の機能／サービスとして】						
☐ アイデアボックス			7/29～			
☐ オープンガバメントWiki			7/29～			
☐ がばったー			8/1～			
☐ データボックス				9/28～		
【他府省や地方公共団体との連携】※経済産業省はシステム提供のみ						
☐ 地域別アイデアボックス						
ー 北海道陸別町				8/21～		
ー 京都府伊根町				9/2～		
ー 埼玉県鶴ヶ島市					10/1～	
☐ 国民の声アイデアボックス				9/24～10/14		

図 2.19 オープンガバメントラボのサービス提供スケジュール

第3章 統計情報基盤に向けられている期待/要望

本章では、主に有識者への事前ヒアリングで得られた情報を基に、統計情報基盤に対する期待や要望をまとめる。これは今後のオープンガバメントの取り組みを充実させていく方向性を定める上で重要な情報となる（これらの期待や要望の中で、一部のシステム機能を実装している）。

※以下『』で囲まれた内容は、有識者ヒアリングで得られた主なコメントである（一字一句の発言ではなく、その内容や意図を表現したもの）。

3.1 統計情報基盤のニーズ

■ オープンガバメントの意義とそこでの統計情報基盤の位置づけ

まずは、オープンガバメントという大きな枠組みの中で『オープンガバメントは政府のプラットフォーム化であり、プラットフォームを作ることによって、その上で様々なサービス提供が行われていくようにする必要がある』との認識がある。今回の統計情報基盤は、「統計情報」という一つのデータ分野におけるプラットフォームであり、本実証実験はこの期待に応えるための第一歩と位置付けることができる。

ただし、『米国のオープンガバメントでも DATA.gov が必ずしもメインである訳ではなく、色々な取り組みがある』と言った指摘はもちろんのことであり、その範囲を統計以外の分野にも広げて行く取り組みが今後必要である。

■ 統計情報基盤のユーザモデル／ビジネスモデル

求められる機能については『利用者モデルがまずあり、その上で必要な機能の検討が出てくる。ターゲットとするユーザモデルを明確にするべき』、『政府のデータを受け取る側がどのようなビジネスモデルでデータを活用しているか調べるべき』との指摘があったが、この点で一つの利用シーンとして『中学校・高等学校の統計の授業』が挙げられた。これについては総務省においてグラフコンクールを実施しているとのことである。

■ 統計情報基盤で公開すべきデータ

『ツールの配付は利活用を促進する一つの方法』であり、『基本的にはデータソースを如何に公開するかを中心に考えるべき』ということで、RawData の提供を望む声が聞かれた。

（※「RawData」とは、ここでは利用者による加工を行う前の統計データで、匿名性が担保された公開データのことを表す。）

RawData の提供が基本であるというこの考え方は『まずはデータソースを公開することと、そのリファレンスを作ることを先行させるべきである』、『RSS 機能までが提供側の説

明責任の範囲であり、少なくともデータソース公開の場所が指定されていれば良い』というコメントにも表れている。

もちろん、ユーザ層のニーズに応える基盤とするには、『経済成長戦略に則ってセグメント分析を行い、製造業、情報産業、商社などの各業種、各企業が求めるデータセットを提供すべき』でもあることは言うまでもない。

■提供されるサービス

RawData が使えるようになると『政府提供の統計データに自分の持っているデータを加えて相関を調べるなど、データを融合して使える環境が望まれる』が、これを広く統計利用の促進材料として使うというアイディアもある。たとえば、『マッシュアップしたアプリケーションのコンテストなどにより、データとアプリケーションの利用促進を図ることも重要である』との指摘もなされ、オープンガバメントが提供する情報が利用者と提供者の間でうまく還元されるような仕組みも必要である。

また、『利用者の分析ニーズにあったデータセットを提供すべきであり、米国でも行われていない特定業種向けのポータルというのも一つのアイディアである』という提案もあった。これは、調査票情報の提出のインセンティブに関し、『政府の統計調査に協力した方がより充実したフィードバックを受けられる仕組みがあれば、統計環境の改善にもつながる』というコメントの一つとして捉えることができる。

本節のポイントをまとめると以下のようなになる。

【ポイント】

- ☐ まずユーザモデル／ビジネスモデルを明確することが重要
- ☐ オープンガバメントのプラットフォームでは利用者の分析に資する RawData を出すべき
- ☐ 公開されたデータのマッシュアップなど、二次加工が幅広く行われるようになることが理想

3.2 統計情報基盤の機能に関する要望（プロトタイプ構築にあたって）

プロトタイプ構築にあたっては、有識者から機能、GUI、データなどのシステム面から、更には普及周知活動などの運用面に至るまで様々な要望が挙げられた。

■機能全般に対する要望

機能については、最新技術動向に合わせて『統計データの表現方法として、2次元だけでなく地図情報との連携やモーション技術の活用など視認性を高めるような機能が望まれる』、『iPad など様々なデバイスへ対応可能であると良い』という要望が出てきたが、これについては今回、goo 統計地図による都道府県地図表示機能を実装し、またデバイスとして iPad にも対応している。

■GUI・使い勝手への要望

多数の統計情報を扱う今回のようなシステムでは、ユーザにとっての使いやすさを追求するには検索機能が重要な鍵を握る。

この点で『統計情報は、政策分野から入る人もいればデータの実態から入る人もいる。タグ付けをうまく行う必要がある。ユーザが必要な統計データへ容易にたどり着けるようなインターフェイスが必要』との指摘があり、今回のシステムでは、2.1.2 項にあるように、政策分野以外にも「統計分野」、「統計種類」、「分析種別」など複数の観点での検索が可能となっている。

■公開データへの要望

公開されるデータは、国内のみならず国際展開も視野に入れる必要があるが、『文字コードは UTF-8 を採用すべき』という意見にも表れているように、この点でインターネット環境での国際文字標準に対応した文字コード体系の採用は必須である。今回のシステムは UTF-8 に対応している。

また、コンテンツの国際化・標準化の対象としては、データ規模の小さいメタデータに注目し、『メタデータの表現形式を決める必要がある（例えば RDF）』との提案も出されたが、この点では今回 SDMX 形式のメタデータの提供を行っている。

■データ提供のインセンティブ

オープンガバメントのプラットフォームは、できるだけ多くのユーザに使ってもらうことによって認知度が高まると同時に、調査票情報の提出のインセンティブともなる。この点からサイト内での広告機能として実現したバナー広告については、『バナー広告によって、統計調査に協力するというスタンスが得られる』との指摘があった。

更に『公開データの良い利用事例についての周知が大事である。リアルメディアとの連

携も進めるべきである。ビジネスで経済産業省の統計が使われ効果があった事例を見せることが重要』との指摘にもあるように、ベストプラクティスとしての事例を紹介するなど、様々なチャネルを通じて普及広報に努め、一層の統計情報利用を呼び掛ける必要もあるだろう。

本節のポイントをまとめると以下のようなになる。

【ポイント】

- ☐ 一部の専門家ではなく様々な利用者が検索できるような仕組みを提供すべきである
- ☐ 国際展開を意識したデータやコンテンツを整備する必要がある
- ☐ 今回のバナー広告を始め、調査票情報の提出のインセンティブとなるような仕掛け、取り組みが必要

3.3 統計情報提供形式（XML/SDMX）の利用イメージ

統計情報を SDMX 書式で提供する目的は、組織間もしくは統計手法間における類似データの比較検討能力を向上させ統計利用者（民間）の利用促進を図ると共に、政府・自治体間での情報交換を円滑に行うことにある。

今回の実証実験では、「鉱工業指数」の一部と「商業統計」の一部を対象として SDMX 書式の検討を行ったが、これに対する以下のような期待が寄せられている。

■データ標準化による互換性確保

政府だけではなく、地方自治体や民間企業での統計情報の利用を考えると、何らかの統一フォーマットが要件となるが、これは『都道府県のユーザは自前のシステムを持っており、そうしたユーザに対しては、規格化された XML が適している』、『XML のタグの組み合わせを国際標準にしておく、政府だけでなく個別企業が分析に使う標準にもなり、色々と分析もし易い』、『各国との比較を念頭に入れると、XML 形式にしておくことは重要』と言ったコメントにも表れている。

XML は、1998 年制定以来 10 数年を経てあらゆるデータ交換の標準フォーマットとして各分野で採用が進んでおり、これは統計データにも当てはめることができる。

■CSV との関係

一方、統計情報でも利用されるエクセル等の表形式データは、カンマ区切りのテキスト表現である CSV 形式が簡易なフォーマットとして利用され、対応アプリケーションも多い。

このような観点で『提供するデータセット形式としては、国際的なデータモデルを意識するならば SDMX は必須であろう。一方、国内では CSV が一般的である。従って SDMX と CSV の二本立てということも考えられる』というコメントや、『SDMX のコンバータをヘビーユーザ向けに用意することが望まれる。エクセルでは大変なので、CSV などにコンバートするツールがあれば、利用者がそれぞれ自分用に書き換えるはず』というコメントに見られるように、CSV との何らかの併用を意識したシステム作りが必要であることが分かる。

『諸外国より抜きん出ることを目指すなら、SDMX でのみ公開するのも一案である』という先進的な意見もあったが、その場合でも前述のような CSV へのコンバータツールの提供など既存 CSV ユーザへの配慮は最低限求められるであろう。

■書誌情報の XML 化

統計情報だけではなく広く行政情報全般の課題に関連するが、XML 化の一般的な留意点として『書誌情報を検索し易い形とするには構造化する必要がある。例えば、議事録のフォーマットひとつとっても厳密に書式を確定するというのは難しい面がある。一方で、内

容が何なのかが分かるようなレベルの話であれば、タグを2,3個付けるという作業であり、それほど難しいものではない。それによって検索性を高めることはメリットも大きい』というコメントもあった。

更に踏み込んで、文書全体の XML 化についても、『文書の構造化が望まれる。それによって長大な白書と1枚の審議会の議事録が混在していても適切な検索を行うことができる。文書を作る時から XML 化を意識していれば統計情報のマイニングを行う際にも関連部分が容易に絞り込める』という意見があった。具体的な対象文書として『例えば、総務省や経済産業省で作る白書のスキーマを決めることもそれほど無理はないはず』との提案もいただいた。今後のプラットフォーム拡充の際、行政情報公開にも範囲を広げた場合に参考となる知見である。

■SDMXのAPI化

この他、『ユーザ向け機能のAPIとしてSDMXを利用できる。SDMXの形式でどれほどのAPI部品を提供できるかということが課題となる』のように、SDMXによるAPI公開の可能性も示唆されている。これは今後の課題である。

本節のポイントをまとめると以下ようになる。

【ポイント】

- ☐ XMLによる標準化は、地方自治体や民間企業のシステムとの連携を行うためには必須
- ☐ SDMXの普及においては、当面は既存ユーザのことも考慮し、CSV形式やExcel形式との併用を考慮すべき
- ☐ 将来的にはSDMXを使ったAPI公開も検討すべき
- ☐ 統計情報に限らず行政情報全体に渡ってXML化を行えば検索性が向上する

3.4 統計情報基盤開発の留意点

今回の統計情報基盤の開発に関連して、セキュリティ面での考慮点、クラウド化に伴う課題なども明らかになった。

■セキュリティの2つの観点

情報公開時の一般的な要件として『情報セキュリティについては、2つの側面で考える必要がある。公開が前提となっている統計データの場合は、法制度上秘匿が十分であるということが求められる。一方、一般的な情報セキュリティの観点では、対象データの機密性や完全性、可用性を担保することにある』という観点での検討が求められている。今後、統計情報を含めた様々な行政情報の提供を行う場合に留意すべき点である。

また、3.1節で述べたような統計情報の二次加工サービスに関しては、その期待がある一方で、『他のデータと突き合わせていくと匿名性を担保できなくなるので、統計データの二次的利用を行うのであれば、事業所・企業関係の匿名データの提供は困難な面が多い点にも留意すべき』という指摘があった。

『クロス集計を一定以上重ねると警告を出すような機能』などの具体的な機能作りなど、データの秘匿性を担保するための枠組みをどのように構築すべきか、どの様なデータセットなら公開できるか等の議論が必要であることをこれらは示唆している。

■クラウド利用の観点

クラウド利用については、アプリケーション利用の観点で『データをダウンロードし処理を行うには道具立てが必要だが、今回はクラウド型のサービスなので、道具立てを持たない学生には大変ありがたいシステムである』との期待が寄せられる一方で、各国の法制度の相違に絡んで『国際法上は確定していないが、米国では国内のデータセンターのデータは米国の持ち物であるとしている。海外のクラウドサービスを利用した場合、押収される可能性もないとは言えない』との課題が指摘されている。

本節のポイントをまとめると以下ようになる。

【ポイント】

- ☐ 統計情報の秘匿性をどのように担保するか、二次的利用などの際に検討すべき課題が残っている
- ☐ クラウドの利用にあたっては、データの所有権など各国の法制度の相違を念頭に置く必要がある

3.5 中長期の課題・要望

統計情報提供基盤の今後の課題として様々な観点での提案がなされている。それらを以下に記す。

■統計データへの要望

統計データ自体に対しては、『省庁では時系列でデータを繋げて公表している所は少ないが、グラフで様々な処理を行い始めると時系列のデータが必要。大分類のレベルだけでも時系列で繋げて貰いたい』とのコメントがあるように、時系列データの提供など、統計データの充実が望まれている。

■他システムとの連携

統計情報については『STATS と e-Stat^{注1} との間での連携が更に必要』などのコメントに表れているように、特に総務省との連携重視が求められる。更にオープンガバメントのプラットフォーム全体としては、『総務省と経済産業省で連携しながらある程度のモデルを示し、他省庁も含めデータソースのセットとデータ形式、メッセージ形式等について議論を早期に進めるべき』という提案もあり、連携するには標準化作業が必要であるとの認識が提示されている。

(注1)

e-Stat :

政府が作成、公表する統計に関する情報提供をワンストップで提供する総務省のポータルサイト。学習、ビジネス、研究などに統計を活用する機能が提供されている。

STATS :

経済産業省が構築した「調査統計システム」の通称であり、統計調査等業務における基幹業務（統計調査の「企画」、「実施準備」、「実施」、「審査・集計」、「分析」、「公表・提供」）、及びこれを支援する名簿管理業務、並びにこれらの業務を処理するシステム。

■統計情報基盤普及のための方策

統計情報は、まず利用者の使いやすさの観点から『統計リテラシーがあれば自分の組織、企業の状態をチェックできる機能になっていると良い』との要望があり、利用者にとってメリットの大きい形での情報提供が求められている

この点で『将来的に国の中に企業の実態・生産性・活力を「見える化」する研究チームのようなものができ、その成果が民間で活用できるようになれば、企業からも喜ばれ、情報提供が加速するはず』との提案にもあるように、データの可視化は重要なポイントであり、今後の体制強化が求められている。

本節のポイントをまとめると以下ようになる。

【ポイント】

- ☐ 時系列データなど統計データを充実して貰いたい
- ☐ 総務省の e-Stat や経済産業省の STATS との連携統合が望まれる
- ☐ 互換性確保のためには省庁横断的なデータ形式等の標準化作業が必要
- ☐ 可視化アプリケーションを充実されることによりユーザの満足度を高める必要あり

第4章 実証実験の結果に基づく評価・考察

4.1 利用者観点での考察

本節では、「データボックス」サイト内に設けられたオンラインアンケート結果および有識者への事後ヒアリングで得られた評価を主たる参考情報として、今回の統計情報基盤に対する評価を検証する。(実施期間：2010年9月28日～10月24日；有効回答数：55)。
オンラインアンケートの単純集計結果は「附録A」にも掲載)

※独立した各設問において、回答数は必ずしも上記の有効回答数とは一致していない。

4.1.1 実証実験に参加した利用者層

ここでは、本節で回答内容を分析するにあたって利用者層を念頭におくため、今回オンラインアンケートへの回答を頂いた利用者の業種、職務、統計データ利用実態、利用媒体について「設問1」から「設問5」の結果を基に整理する。

【設問1】 あなたの業種を選択してください。

[n=55]

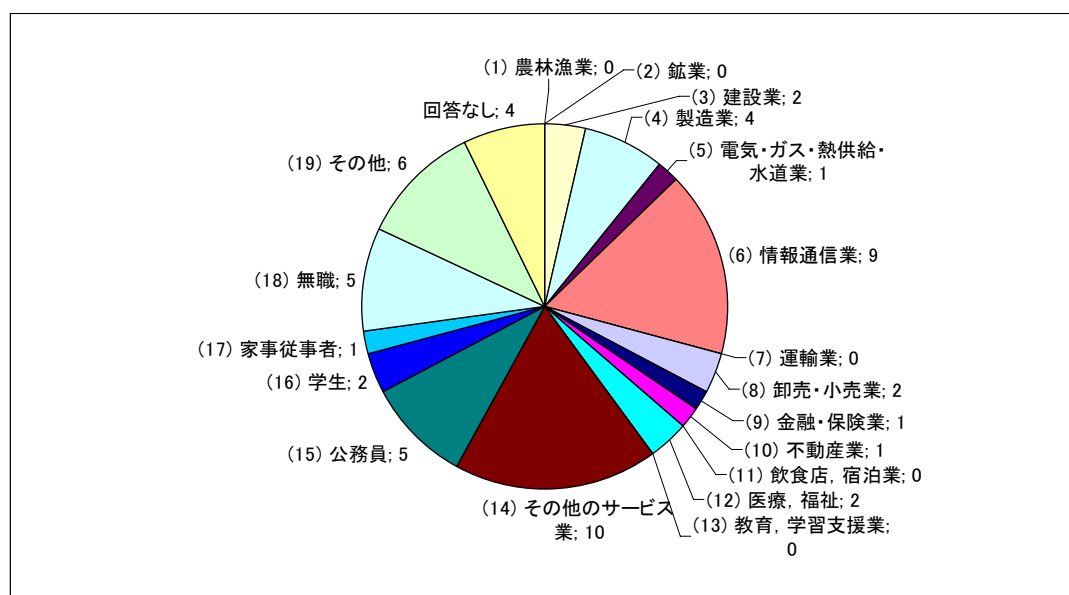


図 4.1 アンケート結果—実証実験参加者の業種

設問1の集計結果を見ると、今回アンケートに回答をいただいた方は、「その他のサービス業(10)」が筆頭であり、その他や無職などを除くと、「情報通信業(9)」、「公務員(5)」、「製造業(4)」などが続いている。これらは統計情報に関連のある業種と見ることができ、今回の実証実験にはこれら統計に関心のある人が参加したことが分かる。

【設問 2】 あなたの職務を選択してください。

[n=50]

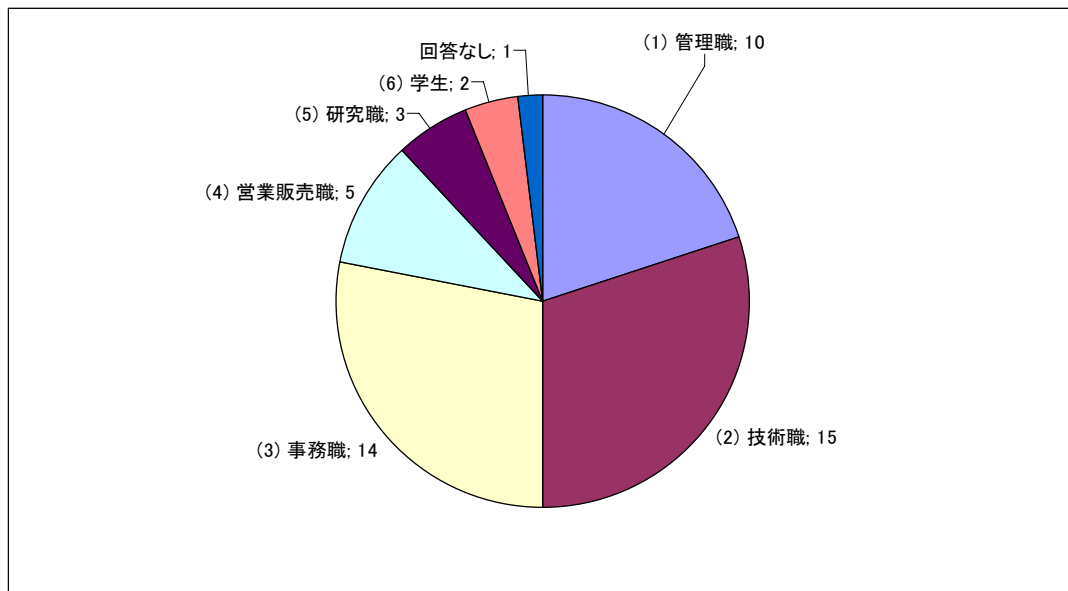


図 4.2 アンケート結果—実証実験参加者の職務

設問 2 の集計結果を見ると、職務として「技術職 (15)」、「事務職 (14)」、「管理職 (10)」の順に多く参加している。実際にデータを作ったり利用したりする人が主に参加したと考えられる。

【設問 3】 経済産業省を含む公的組織から公開されている統計データを利用していますか。

[n=54]

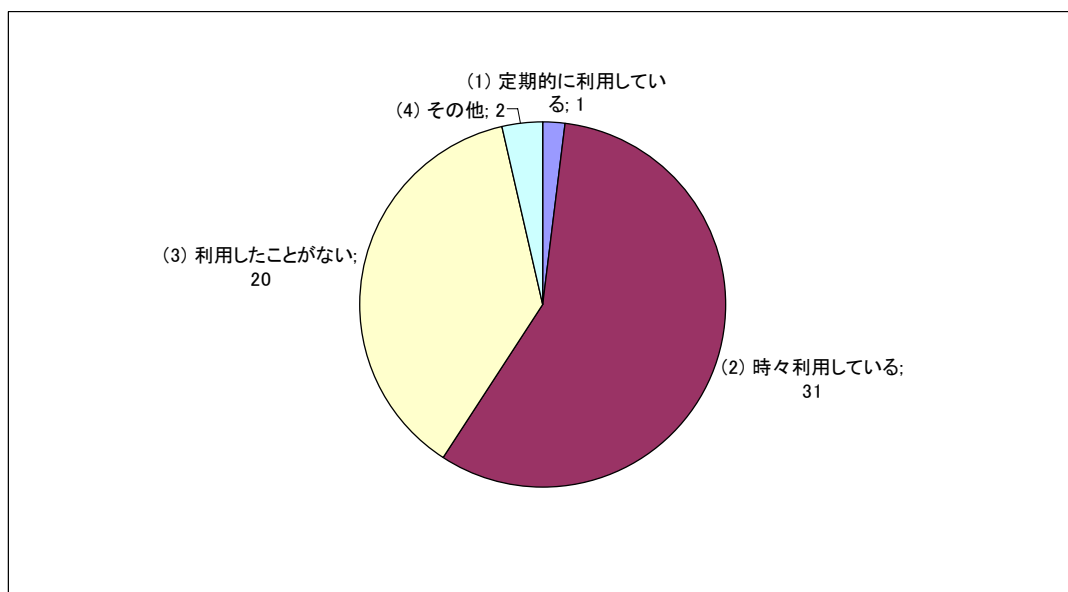


図 4.3 アンケート結果—公的統計データの利用状況

設問 3 の集計結果を見ると、「時々利用している (31)」が最も多いことが分かる。統計というデータの特性を考えると現実的な利用状況と考えられる。一方で「利用したことがない (20)」も数が多いことに気付く。これは、これまで統計情報とは関わりの無かった人が今回の統計データのデータボックスに関心を持って参加して頂いたことを表しており、統計データ活用の今後に期待が持てる結果となっている。

【設問 4】 ご回答頂いた、利用したことのある統計データについてお聞きます。：利用頻度を左の選択肢の中から選んでください

[n=45]

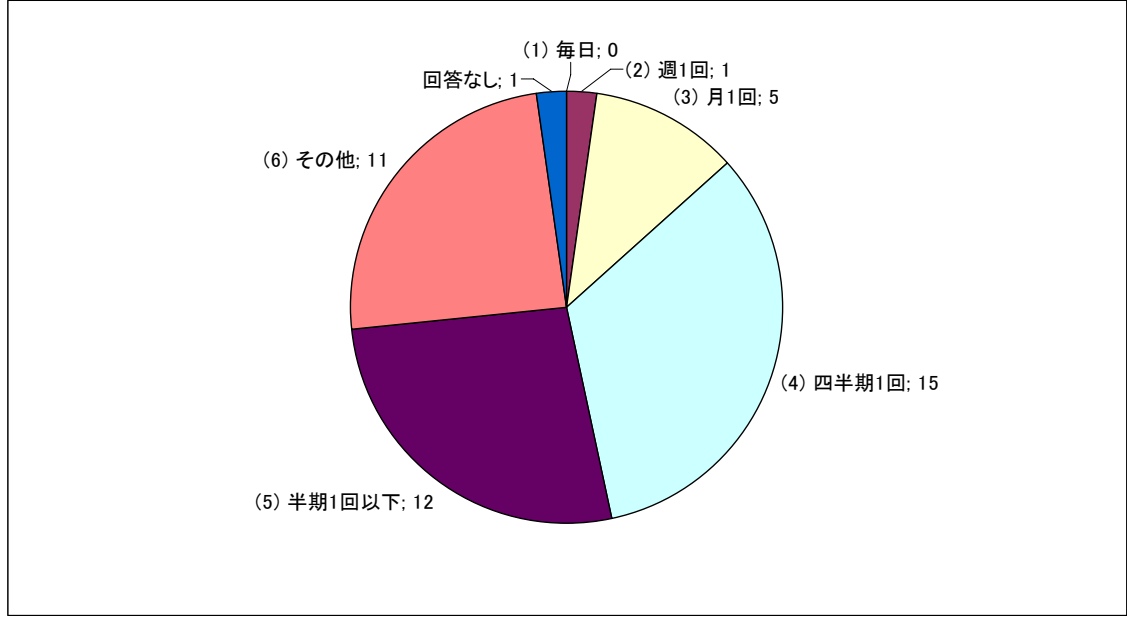


図 4.4 アンケート結果—統計データの利用頻度

設問 4 の集計結果を見ると、「週 1 回 (1)」、「月 1 回 (5)」、「四半期 1 回 (15)」、「半期 1 回以下 (12)」となっているが、統計の公表時期は統計の種類によって月毎、四半期毎、年毎になっていることを考えると、主に公表時期に合わせて統計データが利用されていることが窺える。

【設問 5】 ご回答頂いた、利用したことのある統計データについてお聞きます。：最も多く利用する取得方法を左の選択肢の中から選んでください

[n=44]

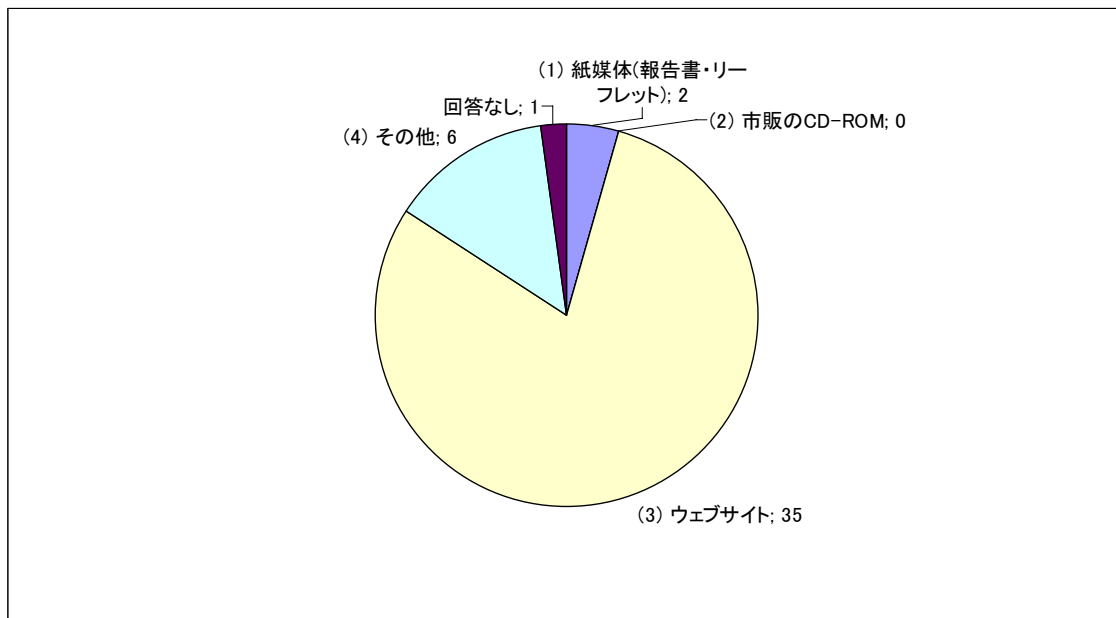


図 4.5 アンケート結果—統計データ利用時の取得媒体

設問 5 の集計結果を見ると、「ウェブサイト (35)」が「紙媒体 (2)」に対して圧倒的な割合を占めている。更に「市販の CD-ROM」は 0 (ゼロ) であったことは注目に値する。情報取得元としてのインターネットが完全に定着していることを如実に示す結果であろう。

【実証実験参加利用者層のあらまし】

- ☐ 統計に関心が高い層が参加しているが、これまで公的な統計データを利用したことのない人も大勢参加している
- ☐ 統計データのヘビーユーザは少数であり、利用頻度は必ずしも高いとは言えない
- ☐ 統計データの取得源は、圧倒的にインターネットである

4.1.2 統計情報の利用目的の分析

次に、統計情報基盤がどのような形で利用されているかの分析を行う。オンラインアンケートでは、「設問 7」が利用目的に関連する質問となっている。

【設問 7】 ご回答頂いた、利用したことのある統計データについてお聞きます。：利用目的を左の選択肢の中から選んでください

[n=42]

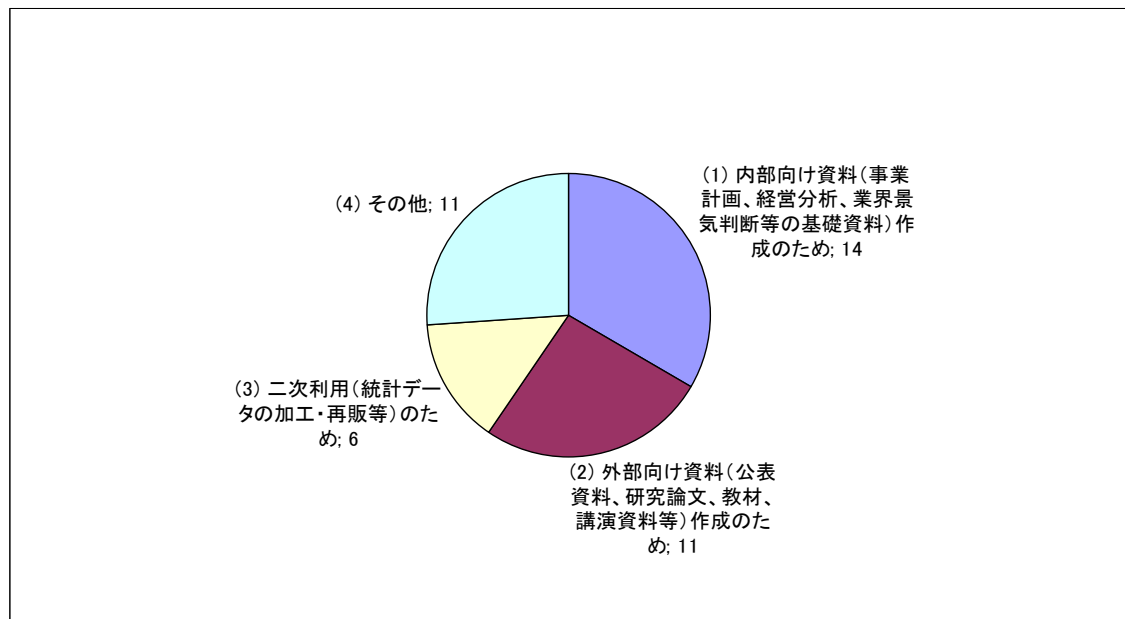


図 4.6 アンケート結果—統計データの利用目的

設問 7 の集計結果を見ると、「内部向け資料作成のため (14)」と「外部向け資料作成のため (11)」が主要な用途を二分している。いずれにしても様々な組織で資料作成の根拠となる統計情報を得る目的で統計情報を取得したいというニーズがあることが分かる。単なる閲覧による情報収集ではなく、他者への説明のために統計情報が活用されている様子が窺える。

更に「二次利用のため (6)」という利用者も一定数存在することは、今後の統計情報整備の促進要因と捉えることができ、RawData の提供など二次的利用の際のデータ提供の在り方を検討する必要性を示すものでもある。

■統計情報二次加工を意識したデータ提供の検討

この二次加工に関しては、有識者ヒアリングでもコメントがあり、例えば、RawData の提供を前提として『データの切り出し方に取り組んで、グラフは Excel で書いてもらう』

という用途があることが指摘されている。これは『本当にユーザにとって必要なものかどうか分類し、優先順位を付けて考えるべき。更に分析をしたいという特定ユーザについては、その人自身の努力や責任に任せても良いのではないか』との意見を踏まえた今後のデータ提供形式の一つの解決策であろう。

その際のデータ処理については、『マクロを組むことになるはずであり、そのようなプログラムを公開することもできる』との提案もあり、データだけではなく、それとペアになる形で、処理プログラムもオープンな形で残るような仕組みの検討の余地もあることが示されている。

この点では、Excelのような簡易ツールを超えた先進的な取り組みを期待する指摘もあった。例えば、具体的な利用形態として『県庁レベルでは Web API を使ってプログラムを組むことが考えられるが、そのようなプログラムが公開されれば都道府県コードを変えて使うことも可能』とのアイディアも出され、国が提供するデータが地方自治体で活用されるという国・地方連携のイメージも示されている。その際の留意点としては、データ形式の標準化の必要性が、『公表データの書式が同じという前提で、新しい年次や四半期データが出た時に、そのデータを使って以前と同じ処理をしたいという要望が出るはず』とのコメントから浮かび上がって来る。

■学校教育での統計情報活用の検討

前述の考察は、本格的な統計分析を指向するユーザに関する利用イメージであるが、この他、学校での統計教育での活用も考えられる。この点、『今年から学校で統計教育が始まるので、教科書でどのような地図や分析データを掲載しているのかを調べ、そこに出ている分析までを対象とするのも一つの方法である』という背景が存在するとの指摘がある。

学校教育における統計情報利用については、『小中高校生を対象とするならば、生徒が自分でデータを選択しグラフを表示できるようにする仕組みがあると面白い』というアイディアも提案されている。

ただし、統計自体でどこまでのユーザ層を対象とするかを検討するにあたっては、『統計自体が広いユーザ向けであるのか否かは疑問である。データを比較的容易に電子媒体として出せるのは良いが、それ以上の機能を含める必要があるかどうかは議論の余地がある』というコメントも参考にし、対象ユーザとその利用イメージを明らかにすることによって、データ形式やサービスアプリケーション等を検討する必要がある。

本項で明らかになった統計情報提供基盤の利用目的と、それに合わせたデータの提供、機能の提供のポイントを整理すると以下ようになる。

[利用目的の現状と今後の展望]

- ☐ 本格的な統計分析を行う利用者の存在が確認された
- ☐ 小中学校での統計教育での活用も視野に入れるべき
- ☐ 二次加工及び個別の要件へ対応するため、**RawData** を提供し、処理はユーザに任せるという選択もある

4.1.3 GUI などシステム操作性の評価

オンラインアンケートの「設問 9」での全般的な使い勝手や操作性に関する質問と、「設問 10」の全般的な画面デザインに関する質問から、GUI と操作性に関する利用者の評価を考察できる。

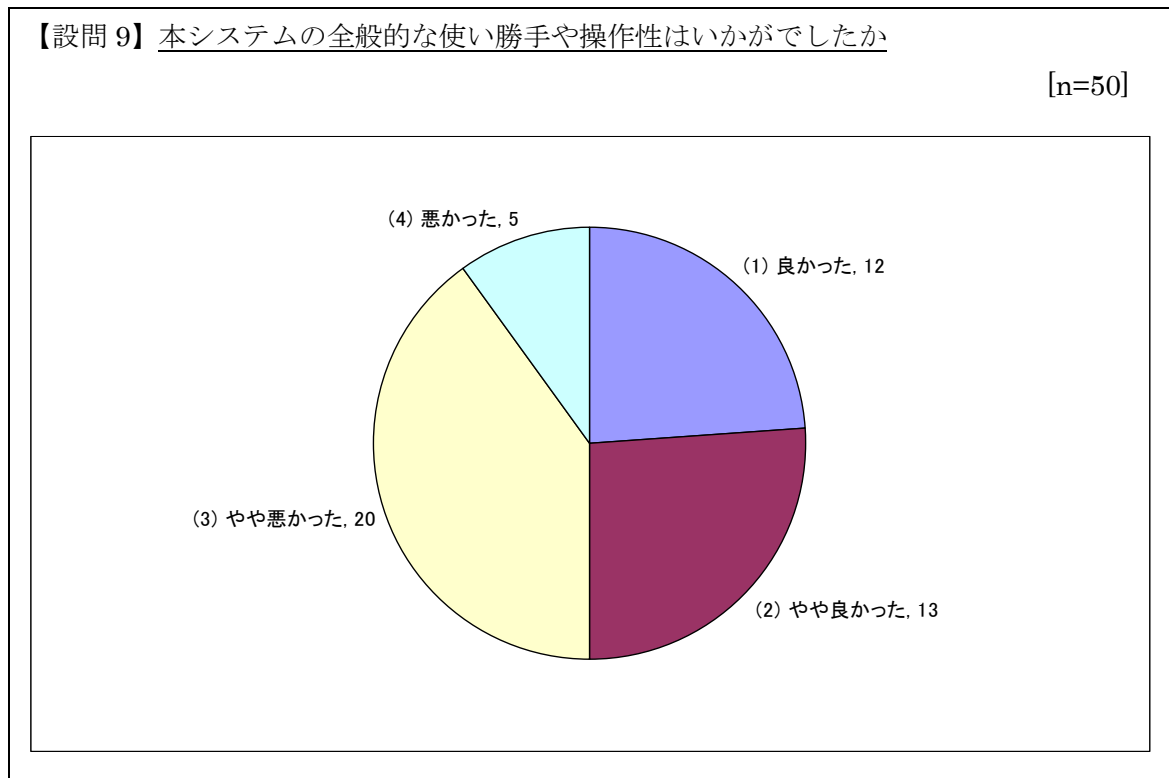


図 4.7 アンケート結果—システムの操作性に対する評価

設問 9 の集計結果を見ると、「良かった (12)」と「やや良かった (13)」でちょうど半数の評価となり、操作性（使い勝手）については評価が分かれている。アンケートの自由記述コメントとして寄せられた『サイトが重く、表示されるまでに時間がかかりすぎ、ストレスがある』や『選択欄や記入欄が画面枠から外れていて、見るのを飛ばしてしまう事が多々ある』などの指摘を見ると、パフォーマンスや情報表示の面で改善の余地があることが分かる。設問 9 以外の自由記述コメントでも、パフォーマンスの改善を指摘する声が散見され、統計情報提供基盤の普及・本格運用に向けては、性能向上が一つのポイントとなることが分かる。

【設問 10】 本システムの全般的な画面デザインはいかがでしたか

[n=50]

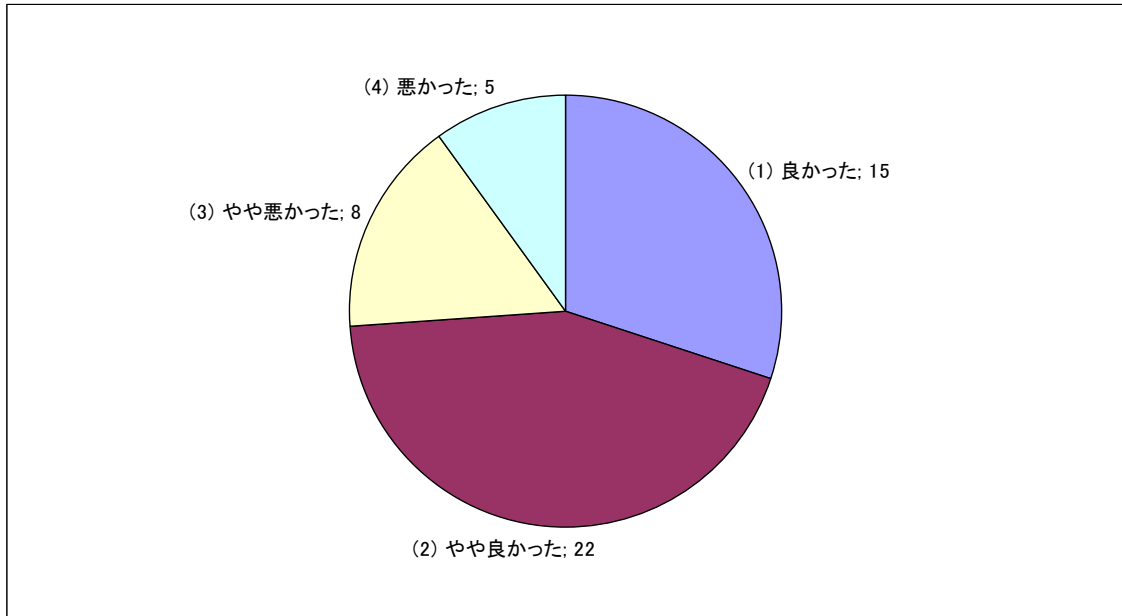


図 4.8 アンケート結果—画面デザインに対する評価

設問 10 の集計結果を見ると、GUI としての画面デザインは好評であり、「良かった (15)」と「やや良かった (22)」で全体の 4 分の 3 に近い評価を受けている。自由記述のコメントを見ると『初めて観るサイトでも、どこに情報が配置されているか、分かりやすい』という評価を得ており、“配置の良さ”がポイントとして挙げられている。

ただし、設問 9 のコメントと同様、『ノート PC 等の小さい画面では表示される部分が少なくなり、分かり辛くなる』というような指摘が挙げられている。有識者ヒアリングからも『画面の構成は、もう少し小さいほうが良い』とのコメントを得ており、様々な条件指定や、結果一覧、グラフ描画などをどのようにユーザに使いやすく表示するかが課題として残る。

【システムの操作性に関する評価と課題】

- ☐ 一般的な使い勝手や GUI は好評
- ☐ 画面の大きさに合わせた利用し易い情報配置が必要
- ☐ 更にパフォーマンスの向上の検討が必要

4.1.4 検索機能の評価

前項の使い勝手の評価につながる基本的な操作として、今回「検索機能」にポイントをあて分析を行う。検索機能が統計情報基盤の利用正否の鍵を握ることはオンラインアンケートの「設問 12」の結果にも表れている。

【設問 12】 検索機能の充実は、統計データの利用を促進する要因になると思いますか。

[n=51]

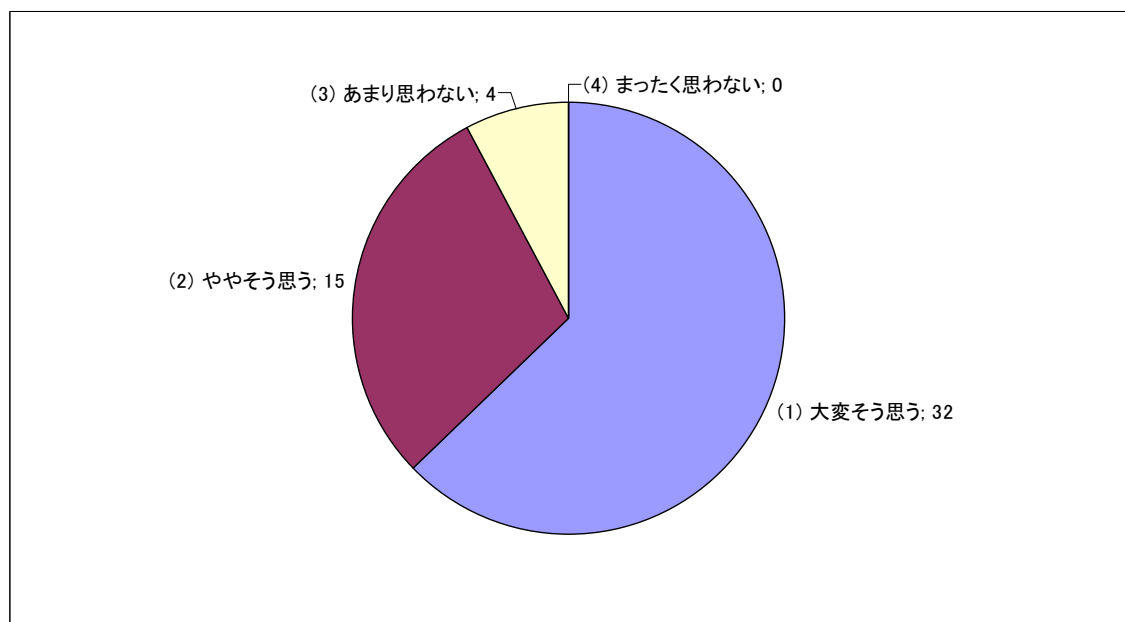


図 4.9 アンケート結果—検索機能充実に対する期待

設問 12 の集計結果を見ると、「大変そう思う (32)」「ややそう思う (15)」でほとんどの回答者が統計データ利用促進の鍵として検索機能の充実を挙げている。今回の収録データだけでも数百ある統計データから自分の知りたい情報を探し出すには検索機能は不可欠であり、今後も検索機能の充実が望まれる。

この検索機能に関しては、事後の有識者ヒアリングでも以下のような様々な提案がなされた。

- 『キーワード検索をした時にどこでヒットしたのかが分からない。一般的なサーチエンジンでは色を変えて表示していることなどを参考に、統計データの中身のヒットした所を表示するなどの工夫があると良い』
- 『ファイルの中身も検索対象とすることについては、特に PDF データがキーワード検索や文字検索できるようになると良い』

更に事前の有識者ヒアリングでは具体的な手法の提案もあり、これらの先進的技術も検討の余地がある。

- 『データには色々な属性があり、属性の全体を把握しながら検索するのが難しい。これがデータにたどり着くのが難しい要因となる。解決の一つの方法として「ファセット検索」というのがある。属性を沢山持つものに対する検索インターフェイスとしては有効と思われる。政府系のサイトでは年度毎にデータが公開され、経年で見るとは大変である。データ提供にあたっては、ある程度時間を追って見ることが容易になると良い。その一助がファセット検索になると思う』

[検索機能への期待と提案]

- ☐ 検索機能は今後の統計情報基盤の利用促進に不可欠な要素
- ☐ キーワード検索で絞り込んだ結果の表示は、サーチエンジンなどを参考に工夫すべき
- ☐ 新規技術導入（ファセット検索など）も検索機能向上の一助として検討すべき

4.1.5 提供されたデータ表現形式（XML/SDMX）の評価

提供するデータ形式としての XML 及び SDMX については、オンラインアンケートの「設問 15」と「設問 17」が関連した質問となっている。

【設問 15】 XML 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。

[n=50]

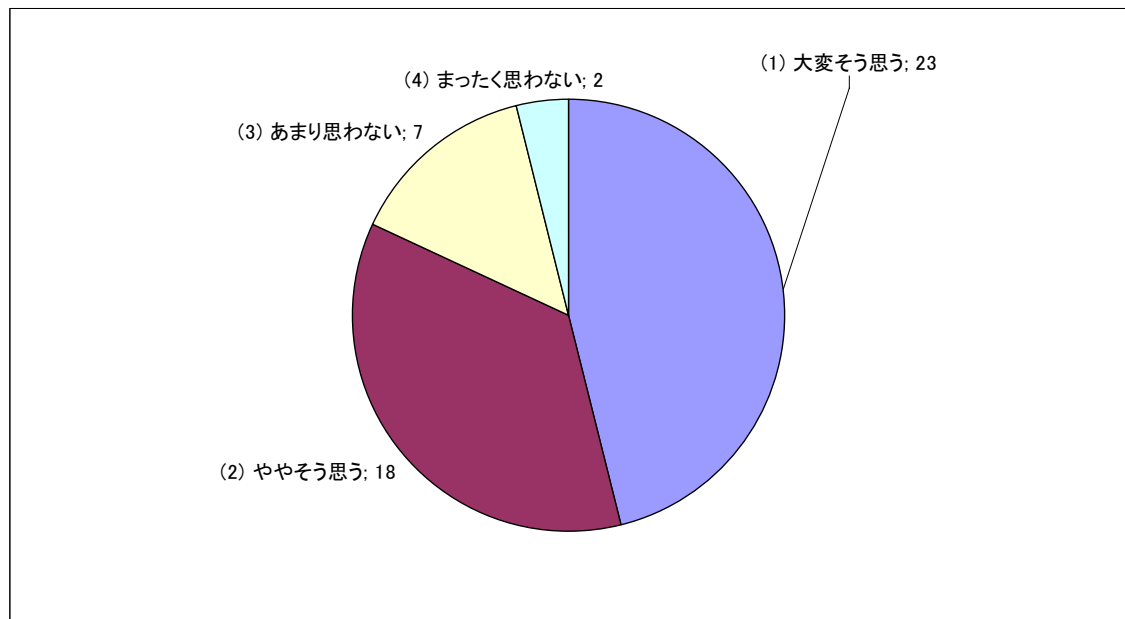


図 4.10 アンケート結果—XML に対する期待

設問 15 の集計結果を見ると、「大変そう思う（23）」「ややそう思う（18）」という回答があり、ほとんどの回答者が XML への理解を示していることが分かる。XML 形式の場合、具体的なタグセットの理解が必要であり、それに合わせたアプリケーションも必要となるが、現時点では『汎用的で一般的なデータは利用促進には必要』との利用者コメントにもあるように、一般的な標準規格準拠への期待の段階と受け取ることができる。そのような認識はそれ自体でも重要なことであり、この意識を利用者が持っていることを念頭に、XML および後述の具体的なタグセットである SDMX による統計データの標準化を促進することが求められている。

有識者ヒアリングでも『ここ 5 年間ほどの間に各省庁も Excel データや CSV データで情報提供をするようになった。これは、印刷物や PDF データが主であった以前と比べると革命的なことである。ただし、Excel データでは、表頭、表側など余分なデータが付いていることが多いので機械的な処理が出来ず、その点で XML データでの公開には期待をしてい

る』とのコメントがあり、XML への期待が高い。

【設問 17】SDMX 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。

[n=46]

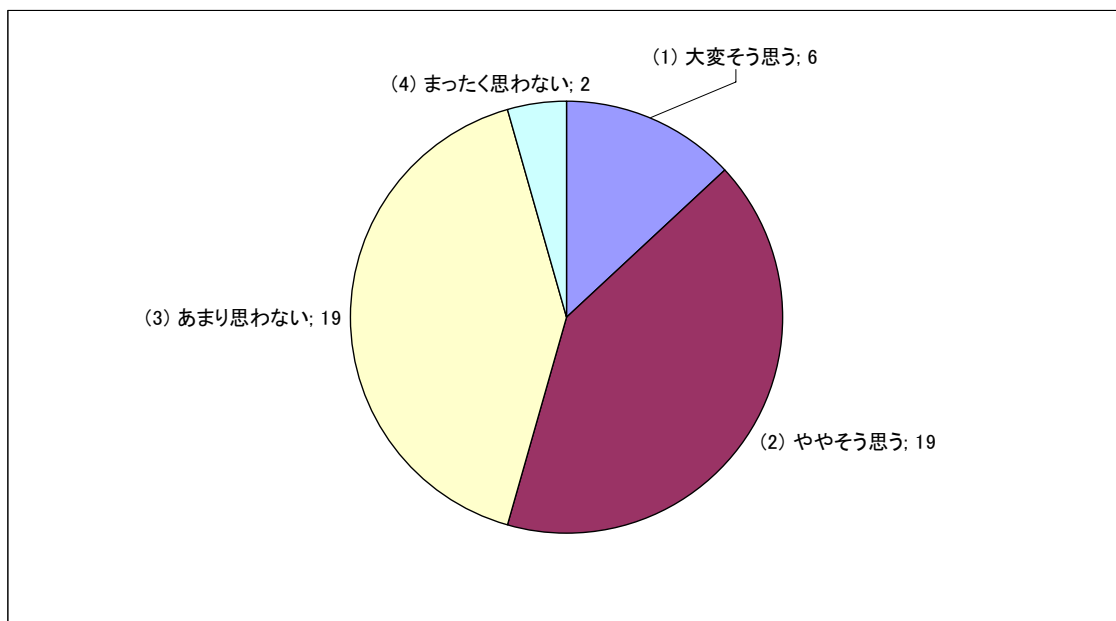


図 4.11 アンケート結果—SDMX に対する期待

設問 17 の集計結果を見ると、「大変そう思う (6)」「ややそう思う (19)」が回答総数の半数を超えてはいるものの、「あまり思わない (19)」「まったく思わない (2)」も合わせると半数近くに迫っている。XML への理解が高かったことを考えると、必ずしも SDMX 形式の意義や有用性が理解されていない状況が浮かびがる。

この点、『対応しているアプリケーションの数に左右されてくる』との利用者コメントや、有識者からの『SDMX を使えるツールがない』『SDMX は、アプリケーションとして対応するものが無いので、現時点で特段 SDMX である必要性は感じていない』という指摘にも見られるように、今回は処理アプリケーションが提供されなかったためメリットが見えないところにその第一の原因があるのではないかと同時に、統計情報標準化の一般的なメリットを啓蒙する必要性も課題として残る。

XML (設問 15) と SDMX (設問 17) で個別に評価を尋ねた 2 つの結果を考察するならば、相互運用性のための標準フォーマットへの漠然とした期待はあるものの、統計情報のための標準 XML タグセットが SDMX かどうかについては必ずしも認知されていないことが分かる。データ標準は決まっても処理アプリケーションが充実してこなければユーザは

付いて来ないということであろう。

SDMX 自体についても、有識者から『SDMX 自体、一部では普及を目指していく動きもあるものの、広く受け入れられているかどうか不明である。SDMX にこだわらずに、一般的なデータ形式に対して柔軟性を持つておくのが良い』というコメントがあり、ここにも、XML には賛成だが SDMX には懐疑的である根本的な理由が見て取れる。

■既存のデータ形式との関係

ところで、XML/SDMX の検討に関しては、既存のデータ形式との関連も考慮に入れる必要がある。これについては「設問 6」が関連する質問となっている。

【設問 6】 ご回答頂いた、利用したことのある統計データについてお聞きます。：最も多く利用するデータ形式を左の選択肢の中から選んでください

[n=43]

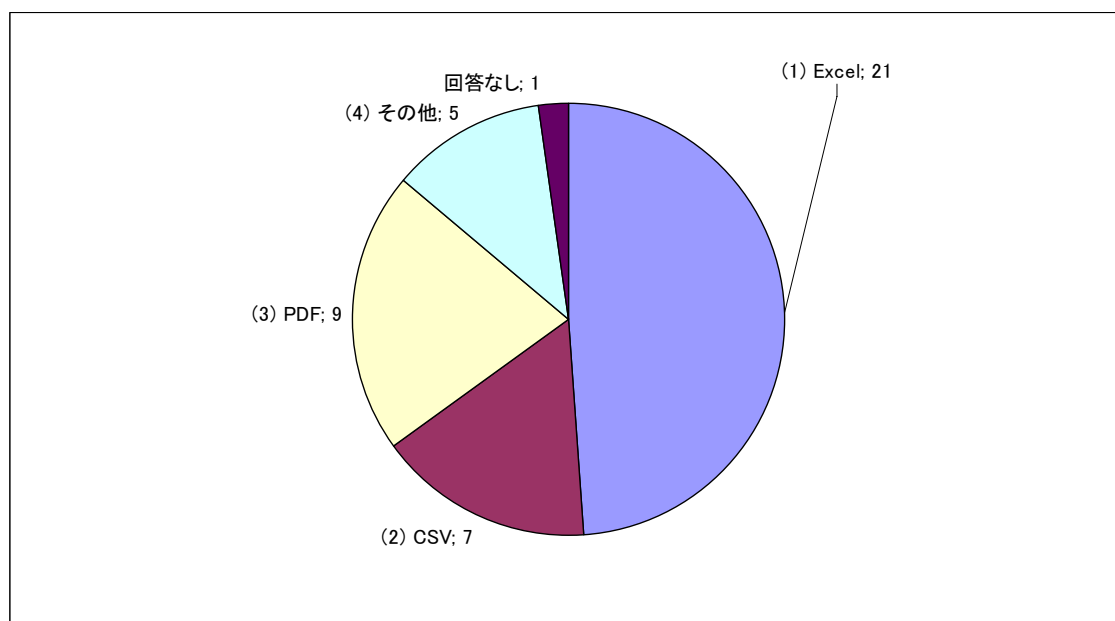


図 4.12 アンケート結果—利用する統計データの形式

設問 6 の集計結果を見ると、「Excel (21)」が他を圧倒している。CSV も Excel で処理できる形式であることを考えると「CSV(7)」を加えて全体の 3 分の 2 程度に上る。これらから、統計情報の処理アプリケーションとして Excel などの表形式ソフトが定番化していることが分かる。

一方で、既存表計算ソフトと XML/SDMX との関連性を考慮する際に、XML/SDMX と Excel (CSV) を対極の存在として捉える必要は必ずしもない。有識者ヒアリングでも『Office 2007 を使って、XML でデータを利用できるかどうかテストをしている』という先進的な

取り組みが紹介されているように、マイクロソフト社の Office 2007 では、Excel2007 を含め、XML 処理機能が標準で組み込まれており、そのような既存表計算ソフトとの連携も検討する余地がある。

その他の回答として得られた「PDF (9)」は紙媒体の延長としての Web 表現形式と考えられる。この場合、書誌情報として PDF の中に記載された統計データを読むということになるが、PDF は数値を取り出して加工することができない。SDMX/XML などを利用し、何らかの形での“数値の電子化”が求められるであろう。

[XML/SDMX への評価および課題]

- ☐ XML の意義については概ね理解されている
- ☐ SDMX は、処理アプリケーションが無いことがネックとなり評価は低い
- ☐ 統計情報を扱う定番アプリケーションとしての Excel (CSV) との関連性を念頭に置く必要あり

4.1.6 提供されたデータ構造の評価

データ形式としては前項で XML/SDMX に関する評価を行ったので、ここでは検索時の属性を用いた条件指定について論じる。これはオンラインアンケートでは直接触れられていないので、主に有識者ヒアリングでのコメントを参考に考察する。

■属性情報による絞り込み

今回のデータ検索においては、属性情報として、「政策分野」、「統計分野」、「作成機関」、「統計種類」、「分析種別」の各条件を指定できるようになっている。

この属性についての全般的な評価として以下の指摘があった。

- 『「分析種別から探す」という用語が良く分からなかった』
- 『政策分野や作成機関をユーザが知っていれば目的のデータを探することができるのだが、実際知らない方が多い』

これらは、仕組みとして用意されたものをユーザに的確に説明するための方策が求められていることを表している。

検索の操作性については以下のコメントがあった。

- 『自分の必要なデータを絞り込んでいく時にもう少し便利になればと思う。データや統計の構造を分かっている人はピンポイントで絞り込むことができるが、知らない方にとっては絞り込みが面倒であろう』
- 『統計知識を持ったユーザでないと入り口から分からない。今後、経済産業省以外のデータも投入されると、更に入り口が分からなくなり、なかなかデータに到達しない。対象範囲を広げれば広げるほどこの問題が出てくる』

ここから、「政策分野」、「統計分野」、「作成機関」、「統計種類」、「分析種別」という複数ある条件をどのように組み合わせて絞り込みを行うかは、専門知識がないユーザを念頭においた更なる検討の必要性が浮かび上がる。

■タイトルの付け方

統計データの看板とも言えるタイトルは、その内容を類推するのに重要なメタ情報の一つと捉えられる。これについて、以下の有識者コメントが挙がっている。

- 『統計の名前だけでは一般にはその内容が分かり難い。何かの意図を持って政府情報を探す方にとっては、検索オプションが分野別、省庁別といった条件ではなく、「～

について調べたい」といった目的別のところから入っていけるようにすることが必要』

- 『今回のタイトルは統計の種類と中身が一緒に入っており、統計を知っている人しか分からない。タイトルの付け方は極力簡単にして、その先で内容を説明する階層的な形にしないと統計の素人には分からない』
- 『誰を対象とするかという点でもあるが、タイトルの付け方は工夫したほうが良い。素人の方へは中身の説明を厚くして対応すべき』

これらのコメントから、今回のタイトルは統計データの内容を的確に表しているかどうか再検討が必要であるという課題が見える。タイトルだけで難しい場合は「説明情報」のような他の情報付与の検討も必要であろう。

また、今回は統計情報のみの提供であったが、汎用的な行政情報公開サイトとしては、『文書情報と統計情報は入り口から分けたほうがユーザは使い易い』というコメントにもあるように、公開データの種類別に検索させることも提案されている。統計情報はその利用形態を含め特別なユーザ層、活用法があるものと考えたと今後の開発の参考になる意見である。

[データ構造と関連した検索機能の課題]

- ☐ 統計の専門知識のない人でも絞り込みが容易になる条件設定など工夫が必要
- ☐ タイトルが内容を的確に表すような工夫が必要

4.1.7 グラフ化機能の評価

オンラインアンケートでは、「設問 13」がグラフ化機能に関連した質問となっている。

【設問 13】 グラフ化機能は、統計データの利用を促進する要因になると思いますか。

[n=53]

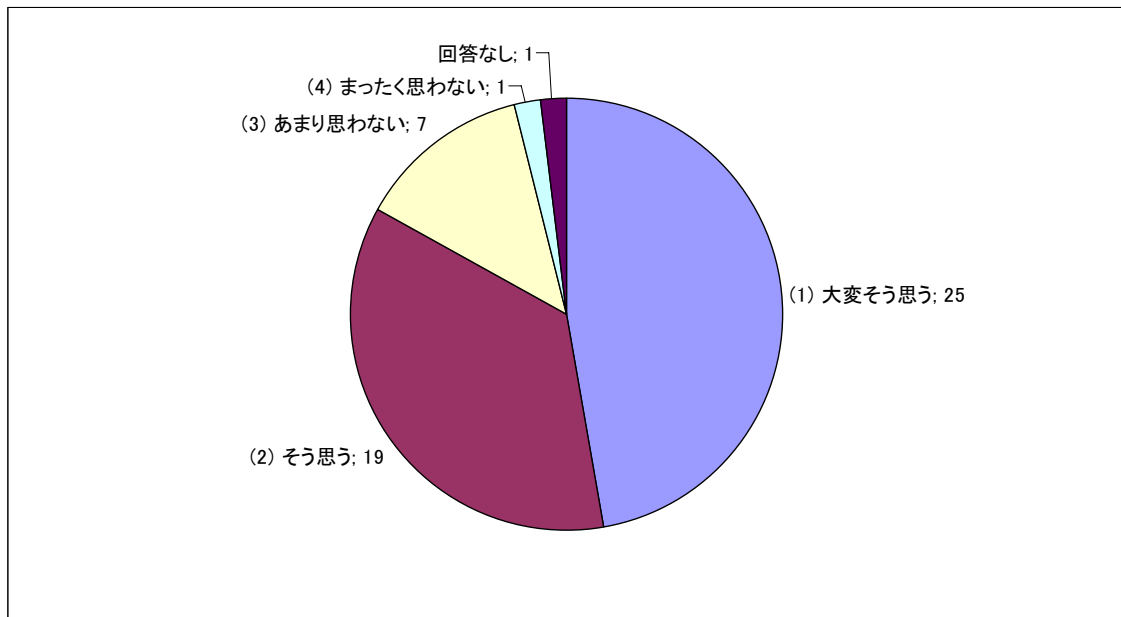


図 4.13 アンケート結果—グラフ化機能への期待

設問 13 の集計結果を見ると、「大変そう思う (25)」、「そう思う (19)」となっており、8 割を超える回答者がグラフ機能への期待を表明している。数値だけの情報から内容を読み取るのも一つの使い方であるが、サンプルとしての可視化ツールの提供がサイトに好印象を与えていることが分かる。二次的利用のための **RawData** の提供への期待も有識者から挙がっていたが、統計情報基盤としてサンプルとなるグラフ化ツールを今回用意したことが利用者の満足に繋がっていることが分かる。

更にグラフ化ツールの機能については『教材として使用するために、グラフがそのまま電子媒体にできるとか、パワーポイント用データにできるといったことがあると便利である』という要望もあり、グラフ化機能の拡充も検討する必要がある。

また、機能改善の余地として『グラフ表示ができるデータの詳細分析に入っていくと、最初に表示したグラフを表示できない』という指摘も参考になる。

【グラフ化機能への期待】

☐ グラフ化機能への期待は高く、機能拡充の検討が必要

4.1.8 詳細分析機能の評価

オンラインアンケートでは、設問 14 が詳細分析機能に関連した質問となっている。

【設問 14】 詳細分析機能は、統計データの利用を促進する要因になると思いますか。

[n=49]

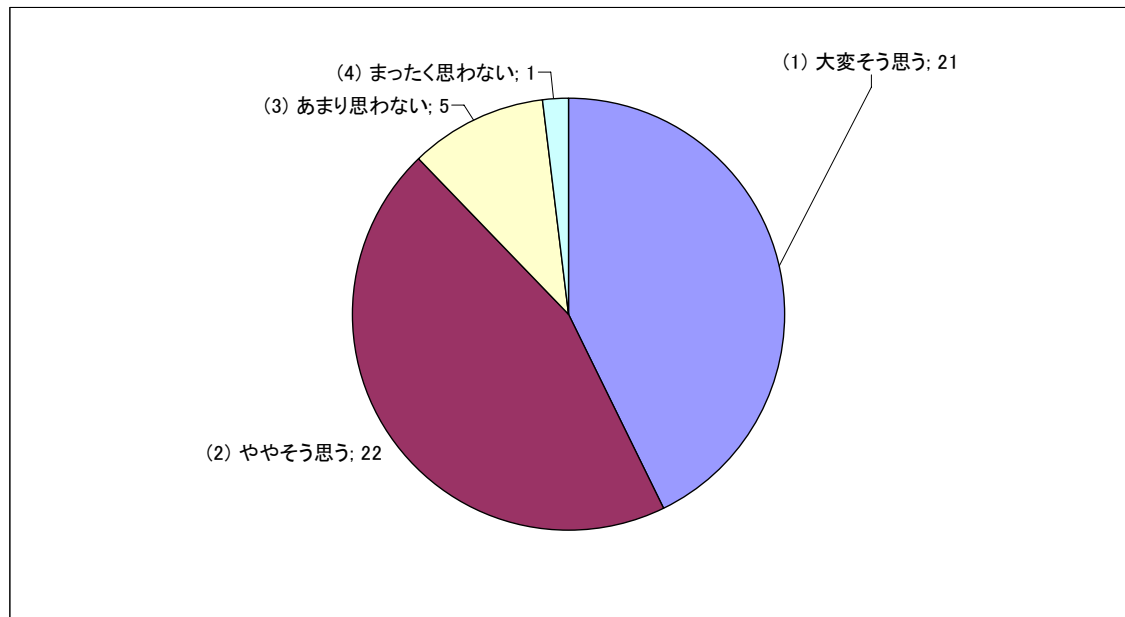


図 4.14 アンケート結果—詳細分析機能への期待

設問 14 の集計結果を見ると、前述のグラフ化機能と同様 8 割を超える利用者が「大変そう思う (21)」、「ややそう思う (22)」と回答しており、詳細分析機能への期待を表している。

詳細分析機能は、グラフなどの可視化ツールほどの有用性は一般利用者に与えないかもしれないが、統計情報を様々な角度から分析したいユーザにとっては有用なツールであり、このアンケート結果は、グラフ化機能だけではなく、詳細分析機能などの道具立てもサイト側で用意しておくことによって、サイト活用の活性化を期待できることを示している。

また、この結果は、統計データを詳細分析するニーズの存在を裏付けするものであり、二次活用のための RawData の提供への期待と繋がるものである。ただし、RawData の二次加工は誰もが出来ることではないので、今回の詳細分析機能としての「Pentaho」のようなサンプルツールを用意することも必要であろう。

ただし機能改善の余地として、有識者から寄せられた『Pentaho は時間が掛かり、ボタンの意味や使い方が良く分からなかった』や『詳細表示では時系列で表示されるが、詳細分析へ行くと同じ時系列表示ができなかった。一般利用者を利用者として想定すると、も

う少し丁寧な説明があると良い』などの指摘も参考にして今後の機能改善を図る必要がある。

【詳細分析機能への期待】

- ☐ 詳細分析機能への期待は高く、機能拡充の検討が必要
- ☐ パフォーマンスの改善や丁寧な説明などの余地あり

4.1.9 OpenID の評価

オンラインアンケートでは、「設問 16」が OpenID に関連した質問となっている。

【設問 16】 Google や Yahoo の ID でログインできる機能(OpenID)が利用を促進する要因になると思いますか。

[n=51]

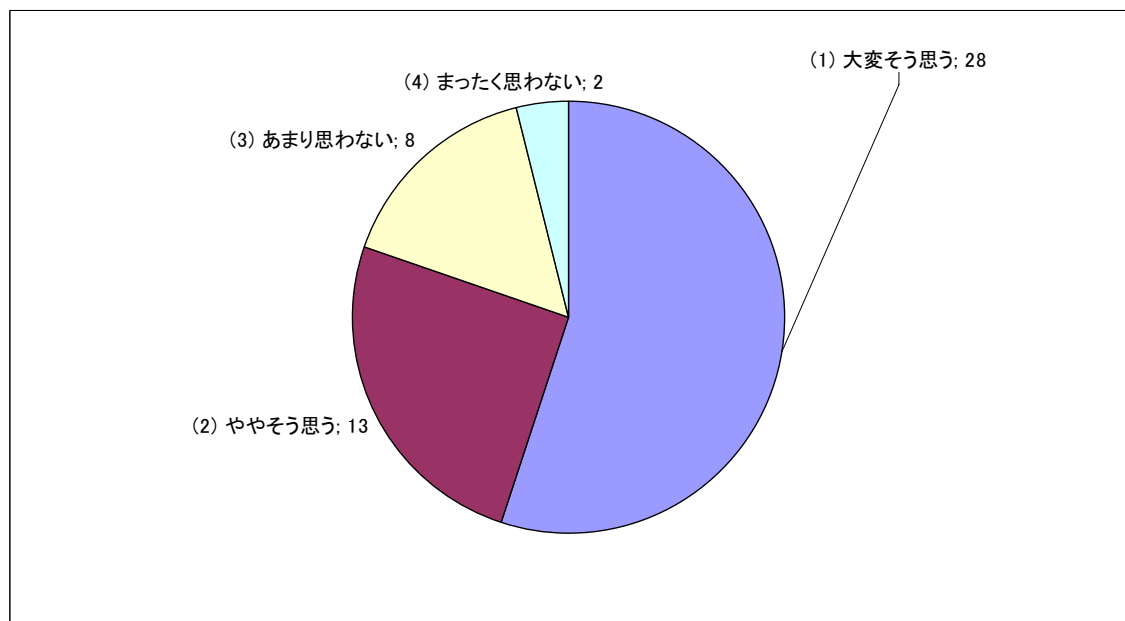


図 4.15 アンケート結果—OpenID への期待

設問 16 の集計結果を見ると、「大変そう思う (28)」、「ややそう思う (13)」となっており、4 分の 3 を超える回答者が OpenID 利用への理解を示している。更に「大変そう思う」自体で半数以上を占めていることにも OpenID の受容度および期待の高さが窺える。また、有識者ヒアリングからも、その意義について『OpenID だけでも今回は十分に成果と言える。政府システムで使うことに関する議論はあるものの、政府系システムで実証した意義は大きい』との評価が得られている。

一方で、この設問での自由記述コメントとして『そもそもなぜ統計データをダウンロードするのに ID 登録が必要なのか理解できない』という指摘もあるように、現時点の内容で ID 登録をする必要性についての疑問も出ている。今回はプロトタイプであり、今後統計情報に限らず行政情報一般について広く情報公開し、様々な機能拡充が行われることによってこれらの疑問は解決され则认为られる。

この理由を有識者ヒアリングから探ると、『別のシステムの ID を入れたにもかかわらず再入力が必要であったことは、それらの情報が関連付けられてしまうのではないかという

プライバシー上の危惧を持たせるものであり、利用を躊躇する方もいるのではないか』というコメントを一つの理由として考えることができる。これは『既に ID を持っている方はオーソライズされているので、属性を取得して、そのまま利用するというのが本来の流れで、認証用の文字列入力などは不要』との別のコメントにも表れており、OpenID の運用方法に更に検討の余地があることを示している。

[OpenID に対する評価と課題]

- ☐ OpenID のへの理解は幅広く浸透している
- ☐ 情報の再入力必要性など OpenID の運用には検討の余地がある

4.1.10 その他の個別機能の評価

ここではオンラインアンケートの「設問 11」の回答を参照しつつ、今回提供された機能に対する評価について、これまでに触れなかった機能を中心に考察する。

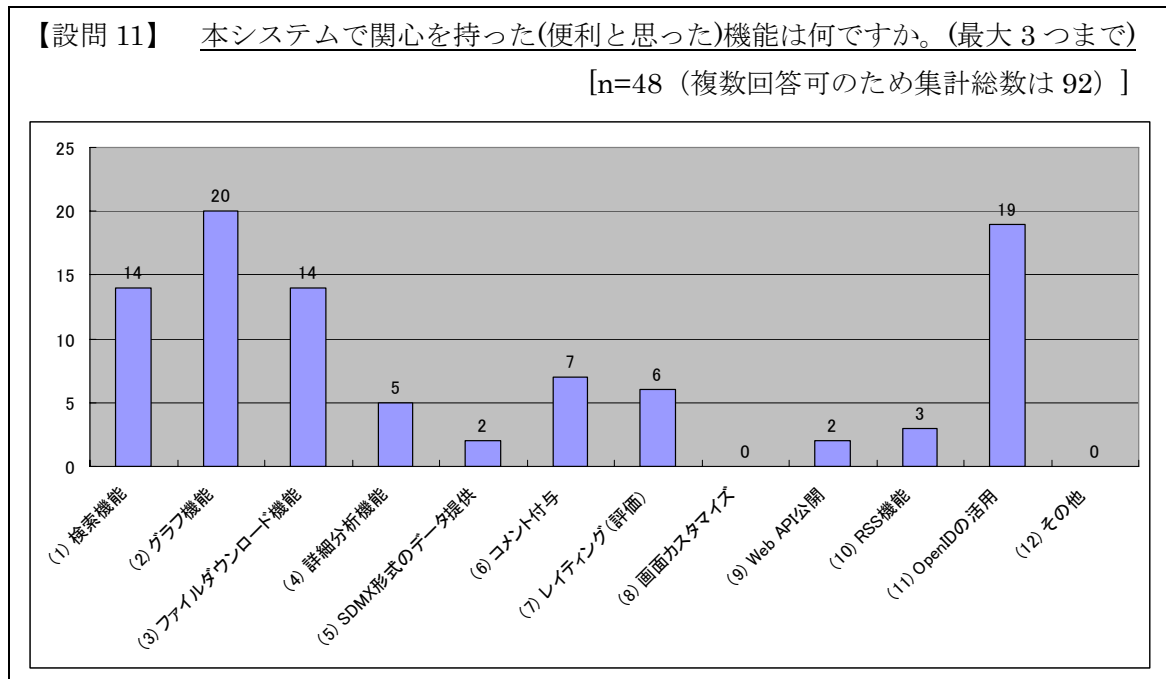


図 4.16 アンケート結果—評価の高かった機能

設問 11 の集計結果を見ると、最も多かったのが「グラフ機能 (20)」であることが分かる。これは統計情報サイト上で可視化された統計情報を見たいというニーズの表れと見ることができる。つまり、閲覧ユーザが多数存在するということである。一方で「詳細分析機能 (5)」が少ないのは、可視化の方が直感的な興味を引くことの表れであろう。

これに対して「ファイルダウンロード機能 (14)」も多くの支持を集めた。これはその場での閲覧ではなく、自分なりにデータ処理を行うユーザも存在することを類推させるものである。

「SDMX 形式のデータ提供 (2)」の支持が少なかった点を見ると、新しい統計情報標準を持ち込むには普及周知を含め、処理アプリケーションの充実なども必要であることを示している。

機能面では、「コメント付与 (7)」、「レイティング (評価) (6)」もある程度の評価を得ており、双方向コミュニケーションの機能についても、潜在的なニーズがあることが分かる。

以下、「ファイルダウンロード機能」、「コメント付与機能」、「レイティング機能」、「RSS 機能」、「WebAPI 公開」について、主に有識者ヒアリングでのコメントや提案をまとめる。

■ファイルダウンロード機能

ダウンロード機能への関心は 14 票ということで、非常に高い水準にある。

- 『PDF や EXCEL ファイルをそのまま開く流れもあるだろう。現行では、ファイルを開くアプリケーションが設定されていない。勿論、一回は必ずダウンロードするという主旨ならば良い』
- 『ダウンロード機能については、環境にも依存することでもあるが、時間的にも問題なく不満は無かった』

ダウンロード機能への評価が高いだけに、前者のコメントからデータの提供方式のさらなる検討が必要である。

■コメント付与機能

コメント付与への関心は 7 票ということで一定の評価を得ている。有識者ヒアリングでも『コメント付与機能については、記入欄も目に付く場所に用意されていて書き易く良かった』との好意的なコメントを得ている。

コメント付与機能は、情報提供側とユーザとの双方向のコミュニケーションを図る手段として直接的で効果のあるものであり、活用の広がりが期待されると同時に機能の拡充にも取り組む必要があるものである。この点で、有識者ヒアリングでは以下の提案がなされている。

- 『コメントを付けることができるようになっていたが、怪しいコメントや誹謗中傷をする人が出てくる可能性もあるので、報告ボタン等を付けて他の利用者から報告を受け、マニュアルに沿って削除するようにするのが良い』
- 『レコメンデーションがあると面白い。こういった統計を見ている人はこういう統計も見たいというような仕組みがあると、関連情報を探し易い』

同じく有識者ヒアリングで挙げた今回の機能への改善要求も念頭に置きつつ機能を充実させたいところである。

- 『コメントを入力し他ページに遷移すると、自分のコメントを編集／削除することができなかった。少なくとも自分の書いたコメントの一覧を簡単に閲覧し編集できる機能が欲しい』

■レイティング機能

レイティング機能への関心は 6 票ということで一定の評価を得ている。ただし以下の有

識者コメントも寄せられており、その活用法についてユーザに分かりやすい説明を行うことや機能拡充などの検討の余地がある。

- 『どういう目的でレイティングを使うのか、何に使うのかが見えなかった。明確な評価軸を決めたほうが良い』
- 『レイティングについては、1つの系列を良い悪いという評価をすることは考えていないと思う。データそのものに対しての評価なのか、画面に対しての評価なのか、系列毎にレイティング機能が付いているべきなのかが良く分からない』
- 『海外サイトとも比較してみたが、ソーシャル機能を付けても良いかと思った。統計情報のページから **Twitter** や **Facebook** へリンクを飛ばすようなことをすれば、レイティングだけではなく更に情報共有を図ることができる』

■RSS 機能

RSS 機能への関心は 3 票ということで少ないが、専門的見地からは有識者ヒアリングで好意的な反応を得ている。

- 『RSS が普通に出せるようになったので、STATS 側からのリアルタイムの RSS の取得を最終的には考える必要がある。RSS は必須と思う。メールでの新着情報も持っているなのでそこでの連携も必要。STATS 側に API がないので API は作っていくしかない』

これは STATS との連携を視野に入れたコメントであり、将来的に今回の RSS 機能が意味を持つようになることを表している。この点で更なる検討を進める意義は大きい。その際には、今回の提供方法に関しての以下のコメントを念頭におく必要がある。

- 『検索結果が上限 50 件となっている関係から、RSS もこれと連動し最大 50 件となっている。これは少ない。RSS は検索結果の全件表示としても良い』
- 『検索結果について、上限 50 件は少ない。どのように絞り込めば良いのか分からない人は、画面上でサーチしたり検索結果を一つずつ見ている。論文検索でも通常 200～300 件の一覧を見ている』

■WebAPI 公開

WebAPI 公開への関心は 2 票と少ないが、今回はプロトタイプの実証実験であり、実際にそれを使うまでは至らなかったことが原因であろう。WebAPI 公開の意義が小さいということでは決して無く、それは次の有識者コメントにも表れている。

- 『メニューを省略して Web API で直接入る形にすると良い。日本地図に合わせて書いたプログラムソースだけを公開すれば、市レベルでも行えるようにする人が出てくるかも知れない。全部を提供するのも一つの方法であり、ソースを公開するのも一つの方法である』

ここから、統計情報提供基盤が本格的に運用されると、WebAPI によってユーザに使いやすいサイトとして使う試みが様々出てくることの期待が読み取れる。WebAPI 公開は個別ユーザのカスタマイズの道具としてさらなる拡充が求められる。

【その他の機能対する評価と課題】

- ☐ データのダウンロード機能への評価は高く、データの提供方式の更なる検討が必要
- ☐ コメント付与機能も一定の評価を得ており、機能改善の検討も必要
- ☐ レイティング機能については、その意義の説明が必要
- ☐ RSS 機能は、将来的な STATS との連携に重なる機能である
- ☐ WebAPI 公開は、個別ユーザのカスタマイズの道具として重要な要素である

4.1.11 利用者の期待度の考察

ここでは、今後も統計情報基盤が使われるようにするための参考情報として「設問 18」から「設問 21」までの結果を基に整理する。

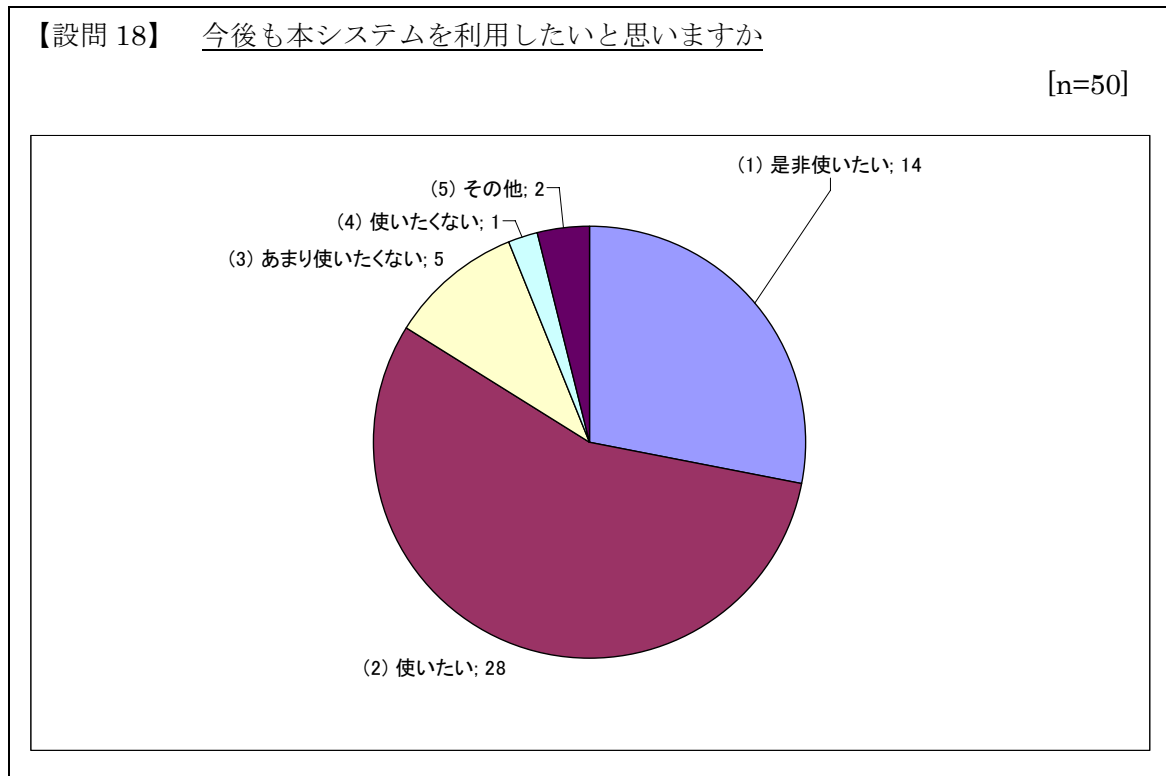


図 4.17 アンケート結果—今後の利用意向

設問 18 の集計結果を見ると、「是非使いたい (14)」、「使いたい (28)」が回答総数の 8 割を超えており、今後のユーザの支援を期待できる結果となっている。その際、鍵を握るのは、『データ更新の頻度と内容による』、『新データが更新されていけば面白いと思う』などの利用者コメントにもあるように、統計情報が最新の状態で提供されているか否かが重要である。実運用においては、公開データをどのタイミングでアップロードしメンテナンスするかが課題となる。

【設問 19】 使いたい理由について、あてはまるものをお選びください。(最大 3 つまで)

[n=42 (複数回答可のため集計総数は 78)]

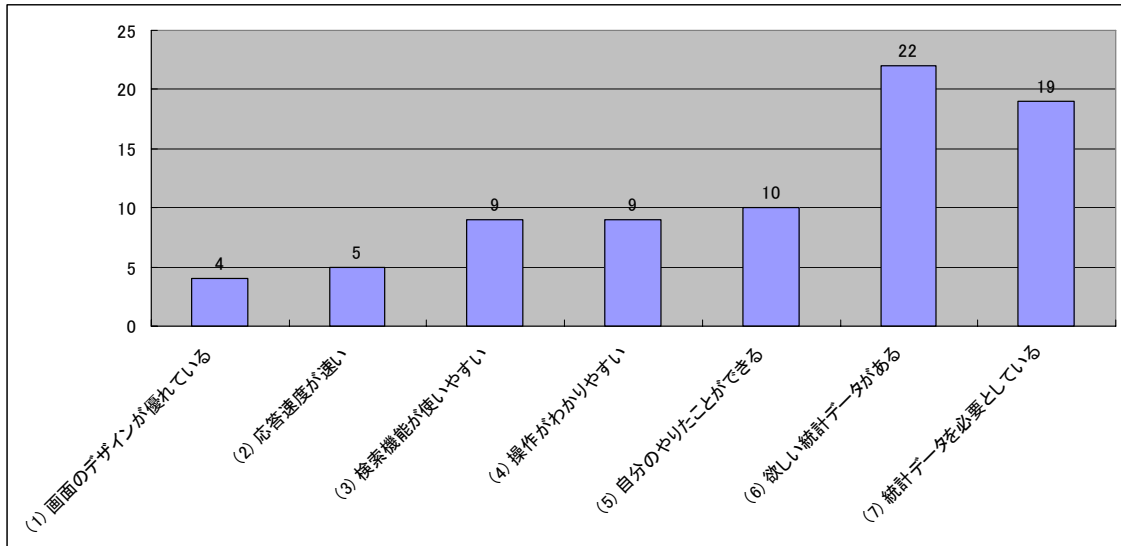


図 4.18 アンケート結果—今後も使いたい理由

設問 19 の集計結果を見ると、回答として最も多かったのが「欲しい統計データがある (22)」、「統計データを必要としている (19)」というものである。これは統計情報への実際のニーズが統計情報基盤の後押しをする大きな要因となることを示すものである。その際重要なのは「自分のやりたことができる (10)」機能があり、「検索機能が使いやすい (9)」というメリットをユーザに見出してもらえかどうかのポイントであろう。それに付随して「画面のデザインが優れている (4)」、「応答速度が速い (5)」、「操作がわかりやすい (9)」などのシステム上の考慮も必要となる。

【設問 20】使いたくない理由について、あてはまるものをお選びください。(最大 3 つまで)

[n=34 (複数回答可のため集計総数は 66)]

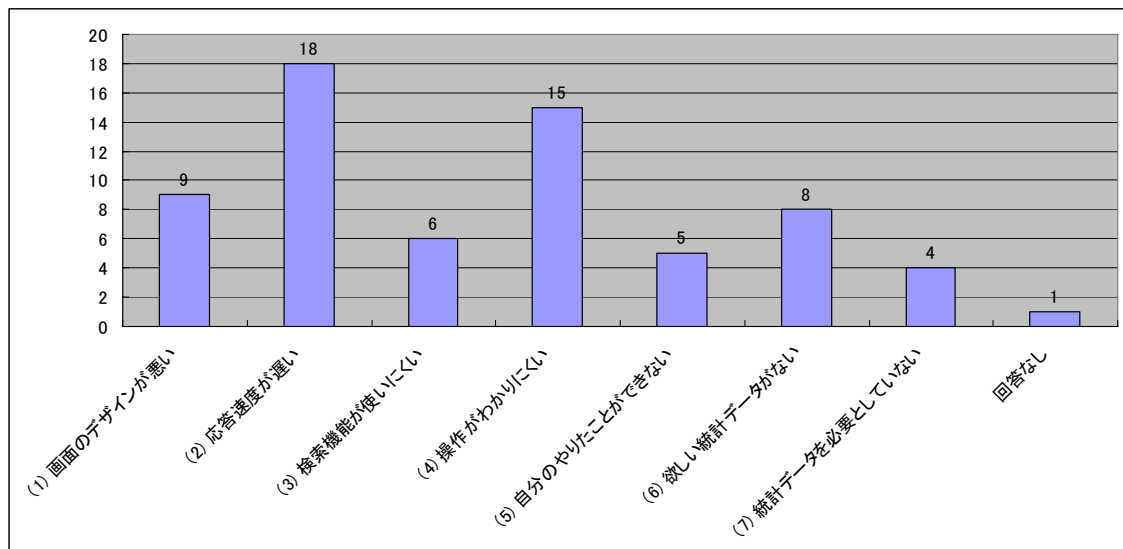


図 4.19 アンケート結果—今後使いたくない理由

設問 20 の集計結果を見ると、最も多かったのが「応答速度が遅い (18)」である。「操作がわかりにくい (15)」も改善の余地を示唆している。この設問は前述の設問 19 と表裏の関係にあり、その意味で「自分のやりたことができない (5)」、「欲しい統計データがない (8)」、「統計データを必要としていない (4)」が少なかったことは、設問 19 の結果の補強証拠とも言える。その意味では設問 19 ではっきりとは現れなかった「画面のデザインが悪い (9)」、「検索機能が使いにくい (6)」という評価が一定数挙がっていることは、今後の課題として捉えるべきである。

【設問 21】 本システムに期待することは何ですか。(最大 3 つまで)

[n=50 (複数回答可のため集計総数は 112)]

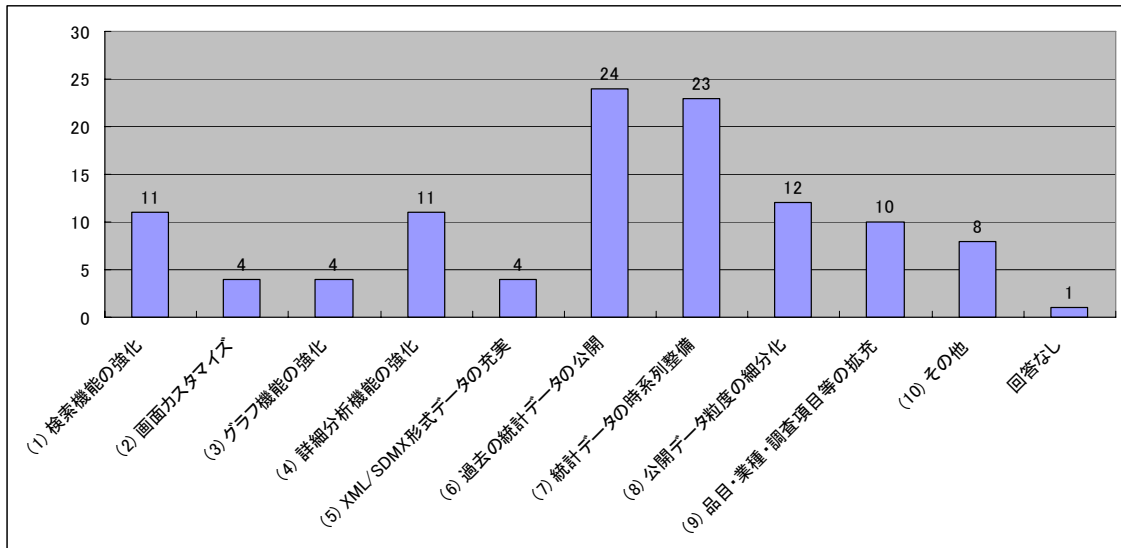


図 4.20 アンケート結果—システムへの期待

設問 21 の集計結果を見ると、最も高かった期待は、「過去の統計データの公開 (24)」、「統計データの時系列整備 (23)」である。自由記述コメントでも『過去データの整備、時系列整備、時系列でのグラフ化機能の強化が進めばとても使いやすいサイトになる』との要望が寄せられており、今後の統計情報提供基盤が持つべき第一の視点として、統計データの中身の充実が浮かび上がった。

これは、それに続く「公開データ粒度の細分化 (12)」、「品目・業種・調査項目等の拡充 (10)」にも表れている。機能面では「検索機能の強化 (11)」、「詳細分析機能の強化 (11)」、「画面カスタマイズ (4)」、「グラフ機能の強化 (4)」となっているが、データの内容の充実への要求に比べると少ない。「XML/SDMX 形式データの充実 (4)」は、やはり統計情報標準化のメリットを十分説明する必要性も浮き彫りにしている。

【設問 8】 本システムを何で知りましたか

[n=51]

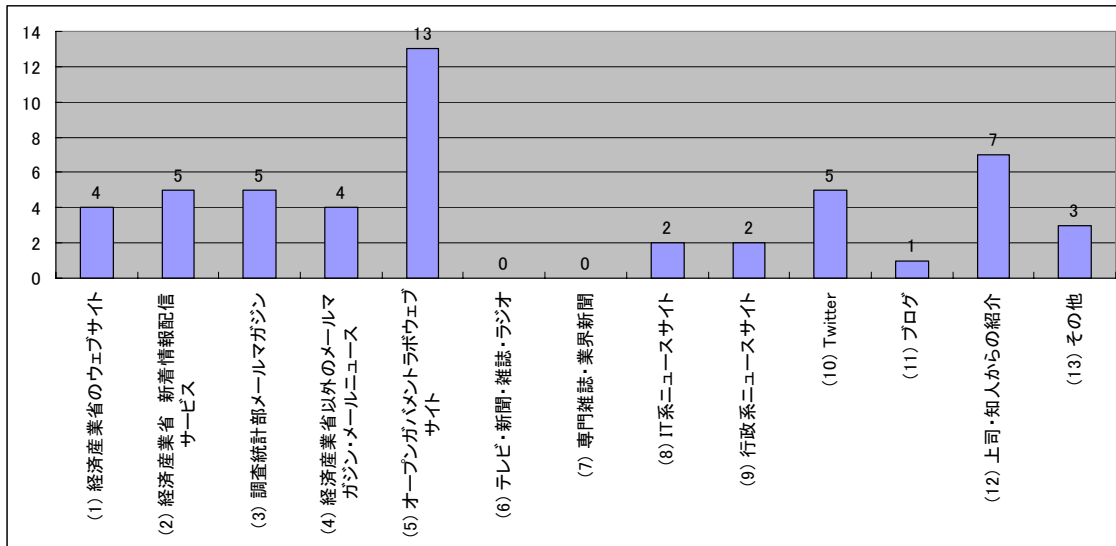


図 4.21 アンケート結果—システム認知の手段

設問 8 からは、今回の実証実験サイトを知るきっかけが分かるが、これは今後の統計情報基盤の認知・周知の参考となる情報である。ここでの集計結果を見ると、最も多いのが「オープンガバメントラボウェブサイト (13)」だが、これは既存ユーザの継続と考えることができる。

これに対し、「経済産業省のウェブサイト (4)」、「経済産業省 新着情報配信サービス (5)」、「調査統計部メールマガジン (5)」、「経済産業省以外のメールマガジン・メールニュース (4)」、「Twitter (5)」などは、様々な Web 上の情報から今回のシステムのユーザとなった人々であり、今後の普及周知活動の一つの参考となる。一方で、既存メディアである「テレビ・新聞・雑誌・ラジオ (0)」、「専門雑誌・業界新聞 (0)」は全くなかったが、これらの“リアルメディア”の情報周知度は高く、今後の課題であろう。

なお、「上司・知人からの紹介 (7)」も、社内外での“口コミ”効果と考えるならば歓迎すべき経路である。

【今後のシステム利用を見据えた方策など】

- ☐ 統計情報提供基盤への期待は高く、今後も使い続けられることが予想される
- ☐ 情報の更新頻度を検討し、最新の情報が提供されるような工夫が求められる
- ☐ 満足度の向上のためパフォーマンス改善が課題
- ☐ 時系列データや過去のデータ公開など、情報のバリエーション拡充の検討が必要
- ☐ サービス周知のため、新聞雑誌などの既存メディア活用の可能性も要検討

4.2 開発者・運用者の観点での考察

ここでは、情報統計基盤の開発者・運用者チームへのヒアリングから得られたコメントを中心に、有識者ヒアリングでのコメントも加え今回のシステム開発と運用を振り返り、開発手法への評価や今後の課題などをまとめる。

4.2.1 開発における工夫・課題

今回のシステムは、クラウド上での構築となり、これに関する開発者・運用者チームの評価を以下にまとめる。

■クラウド利用

まず、クラウド利用に関しては、以下のようなメリットに関する意見があった。

- 『急激にアクセスが増えた事象でも、一般利用者に大きな影響のない形でスケールアップし対応できた』
- 『例えば「がばったー」を行いたいという意向を伝えてから数時間でハードウェアを用意できるなど、サービス開始までの時間を短縮できるのは大きなメリットである。これはコスト削減にも繋がるメリットである』
- 『クラウドサービスは、購入手続きを踏まなくても利用できる。ハードウェアのリソースの変更が容易で、集約率を高めることでその中にサービスを入れることができる』
- 『パブリック・クラウドは仮想レイヤなので、裏で何が起こっているのか分からない場合があり、それを公開する会社と公開しない会社がある。非常に大きなリスクではないものの業者依存となるリスクではある。一方、プライベート・クラウドでは、仮想レイヤの管理も一緒に行える。レスポンスが遅くなった時なども迅速に対応が可能であった』

これらは開発におけるスケーラビリティ向上と開発期間の短縮にクラウドが大いに貢献することを示すものである。一方でクラウドならではの課題の指摘もあった。

- 『プロジェクト当初、必要なリソースの量が明確ではなかったため、想定が難しかった。クラウドでは従量課金となるが、この点は、官公庁が自らクラウドを活用することによって、リソースのやり繰りができるようになるのではないかとと思われる』
- 『省庁の予算計画が、目標値や上限を決める枠組みとすると、上限が無いというクラウドのメリットは相いれない部分があるかも知れない』

今後、官公庁におけるクラウド利用を普及するには、官公庁ならではのプロジェクト形態とクラウドの考え方との調整をする必要があることが分かる。

■アジャイル開発

今回のシステム開発はアジャイルの手法を用いた。これに関する開発者・運用者の評価は、最終的な要件・仕様が明確に共有されていない状況で、短期間に 200 件以上の修正要望に対して迅速に対応できたことは、アジャイル開発のメリットであり、通常の開発手法ではかなり困難なことであった。また、以下のような意見もあった。

- 『開発スピードが早まり、短期のスケジュールに対応できた』
- 『今回はアジャイル開発を用いたが、取り急ぎ形を作ってから品質を向上させるという通常とは逆のアプローチであった。この方法が馴染むケースと馴染まないケースがあるので、使い分けが必要となるだろう』
- 『開発者側も遠隔で作業を行う開発体制モデルは、マネジメントの観点においても新たなモデルとしての可能性がある』
- 『全員が接続できる VPN サーバを介し、そこで付与するグローバル IP を持って本番環境へアクセスするという形態が一つの提案できる形ではないか』
- 『業務要件の整理を行うに当たって、ユーザ側に具体的なシステムイメージがない場合、実際に動くシステムを開発し、見てもらいながら要件整理をすることで、要件ずれの発生を抑えつつ、迅速に開発を実施することができた』

これらのコメントから、アジャイル開発はシステム開発のスピードアップに貢献するという大きなメリットがあり、開発管理などのマネジメント上の課題はあるものの、今後新たな開発手法のモデルケースとなる可能性があると言える。

■オープンソースソフトウェア（OSS）の活用

今回は、前述の詳細分析機能に用いた Pentaho を含め、SugarCRM、Wordpress、MediaWiki といった多種多様の OSS を組み合わせ開発を行った。これに対する開発者・運用者の評価は、今回のプロジェクトにて既製ソースコードだけでは達成できなかったとのコメントを含め、提出されたコメントを以下に挙げる。

- 『OSS は、短期間でシステムを構築できるというメリットがある。今回のプロジェクトも、急なリクエストに対して短期間で構築ができた』
- 『システム開発要件は、業務運用でシステムを使用してもらい、修正を行っていくサイクルが必要である。今回は品質を後から高めていく手法で行っている。PDCA サイクルを回しながら、システム構築をしていけるのは大きなメリットではないか。』

この手法は、OSS などソースコードが公開されていないと取り難い手法ではないか』

- 『OSS を使う時の難しさとして、品質基準がないので品質担保のための方法論が必要となる点が上げられる』
- 『ベンダとしては、OSS との付き合い方は難しい面があり、例えばライセンスは無料であっても技術者の報酬をゼロとはできない』

これらのコメントから、OSS の利用は短期間の開発に威力を発揮することは分かるが、品質保証の面で課題があり、目標設定を明確にした上で達成評価を行う必要があることが分かる。更に、OSS の場合、ベースとなるソースコードの品質にばらつきがあること、参入ベンダがビジネス上のメリットにより対応度合いが異なることなどの問題があることも分かった。

[今回の開発手法の利点や課題など]

- ☐ クラウドは、スケーラビリティ向上と準備期間の短縮に貢献する
- ☐ アジャイル開発は、開発のスピードアップに貢献するが、品質保証や目標設定・達成管理の面で課題がある
- ☐ OSS の利用は、短期間での開発に威力を発揮する

4.2.2 運用／保守時における課題

ここでは、データの更新に関して有識者ヒアリングで得られたコメントを検討する。

- 『データボックスを今後メンテナンスしていく時に、データの更新や過去データにどこまで遡及させるかという点で、データ管理が大変ではないかと思われる。研究者でも科学研究費が出ている間は専門分野のデータベースを作って研究することがあるが、研究費がなくなるとその後はメンテナンスをやめてしまう例が多い。データの更新や管理の手間からシステムを継続していくのは難しい面もある』
- 『次の段階の作業となるかと思うが、自動的にデータのアップデートができるシステムになれば、より発展した形態となるであろう』

これらのコメントからは、今後の統計情報基盤の拡充への期待と課題が分かり、更に自動アップデートなどの機能追加もそのために必要であることが分かる。

【統計情報の更新に関する課題など】

- ☐ 新規データ、過去データのメンテナンスは作業の負担が大きい
- ☐ 自動アップデートなどの機能追加も必要

4.2.3 XML/SDMX 活用に関する課題

今回のプロトタイプでは、SDMX 書式の検討のポイントをメタデータの定義と互換性確保に置き、「鉱工業指数」の一部と「商業統計」の一部を対象として SDMX データを作成した。

今回は、諸外国において上記検討対象と同様のデータが SDMX 書式で提供されている場合、メタデータを含む書式データに類似性を持たせるため、諸外国の例を調査しつつ書式を定めた。

今回の SDMX 化について開発者・運用者からの評価は、以下のようなものであった。

- 『SDMX については、より具体的なメリットなどの議論が必要ではないかと思われる』
- 『SDMX の仕様書が膨大であり、調査が大変だった』

今回の SDMX 化は試行段階にあり、更なる調査及び活用のための議論の必要性が浮かび上がる。また、運用側で今後の課題として意識している点を以下に挙げる。

- メタデータに外国語による説明情報を入れるべきか要検討（海外の事例は、米国作成のデータは **description** が英語のみ、EUROSTAT のデータは **description** が英仏独語となっている）
- SDMX 書式を直接読み込み・処理を行うツールが現時点で確認できておらず、ツールの充実を図る必要がある。
- EUROSTAT の例を見てもすべてのデータが SDMX 書式で公開されている訳ではなく、一部のみの対応となっている。これはどのデータをどの書式で公開するか、妥当性の検討が必要であることを示唆している。
- メタデータにより、統計情報に説明情報が付与されていたとしても、その意味論の解釈と調整はデータ処理担当者に委ねられており、フォーマットの統合により得られるメリットに関して更に議論が必要。

[XML/SDMX の活用に関する課題]

- ☐ SDMX を扱うツールの充実が求められている
- ☐ メタデータへの多国語対応は要検討
- ☐ SDMX の適用範囲に関する議論が必要
- ☐ SDMX 仕様の調査と解釈に注力する必要あり

第5章 全体評価

この章では、前章までに個別機能について整理した統計情報基盤に関する分析を基に、今回の実証実験に関する全体的な評価を行う。

5.1. 有用性・利便性の評価

有用性・利便性を評価するにあたっては、本システムを利用する利用者層を大きく分けるとすれば、一般利用者とヘビーユーザ（統計専門家）という背景知識が異なる層に分かれるということを念頭に置かなければならない。

ヘビーユーザ（統計専門家）はデータが公開されていれば、自力で所望のデータを探し出し、それを処理・加工する作業も容易かも知れないが、後者の一般利用者の場合は、情報を探し出すところから、何らかの支援の手立てを考える必要があるだろう。

この点で、有識者ヒアリングにおいて『今回のサイトでは、初心者ではないが本格的な分析の専門家でもない人（例えば学生など）には丁度良いだろう』というコメントがあり、これら利用者 2 層の中間向けの内容となっているとの受け止め方がある。このため、今後の本格的な利用イメージを見据え、一般利用者とヘビーユーザ（統計専門家）というそれぞれの層のニーズを的確に捉え、それぞれの層に適合した機能やデータを用意することが重要である。そこで、それぞれの観点での検討を以下を行う。

5.1.1 一般利用者の観点での評価

一般利用者にとっては、まず検索機能を用いてデータを探し出すところが最も重要なポイントとなる。これに関連する利用者コメントとして[設問 22]（「統計データの利用を促進する施策について、ご意見がありましたらお願い致します」）に対する回答（自由記述）の中に次のような意見が見られた。

- 『サイトの使いやすさ、検索機能を充実させて欲しい』
- 『資料間の関係（引用や参照関係）や、この資料を見た人はこんな資料も見ているというようなレコメンデーション機能があると便利』
- 『検索ワードに関連する統計情報を探し出すことができるよう、検索機能を強化して欲しい』

ここから、まずは検索機能を一般利用者向けにどのように簡易で利用し易いものにするかが一つのポイントとして浮かび上がる。更に有識者ヒアリング（事後）でも以下の点が指摘されている。

- 『検索機能については、もう少し全体を見通せるような工夫が欲しい』

- 『利用者がこういった観点やアプローチで検索するかを把握することが重要。一般利用者をスコープに含めるならば、その部分をシステムでカバーするか、画面の中の分かり易い説明でカバーするのかなどを検討することが必要であろう』
- 『絞り込みの方法が分かり難いところがある。基本がキーワード検索ならば、フリーキーワードで検索した内容を絞り込む過程が容易に分かるインターフェイスにする必要がある。今までのツリー状の表示形式に則っていると、そのツリー構造を知っている利用者しかその検索方法が使えなくなってしまう』

これらから、現在の検索機能への一層の工夫が求められていることが分かる。4章のアンケート結果でも触れたように、画面デザインなどの GUI 設計は概ね好評であったので、残る課題は、今後公開する統計情報が増えて行くに従って、膨大な資料の山から利用者が必要とする情報を容易に取り出しやすくする工夫、つまり検索機能の操作性向上ということになる。

5.1.2 ヘビーユーザ（統計専門家）の観点での評価

ヘビーユーザ（統計専門家）は別のニーズがあるであろう。例えば、取得した統計情報からどのような分析を行い、それを加工・編集したり、論文や報告書などにまとめるかというポイントである。この点での対応の一つは、統計情報処理ツールの充実である。

有識者コメントでは、『他システム等と比べ、地図上にデータ表示できるといった点が特徴と感じた。統計データが自分の目的に合ったものかどうかを確認するには非常に良い』などのように、今回のシステムの特徴を評価している。

これは、統計情報のような数値データを公開する際には、旧来の表形式での提示だけではなく、地図やグラフなどを使った可視化ツールがユーザにその有効性を理解してもらうための重要な手段であり、今回のいくつかの可視化ツールが好意的に評価されていることを表している。更にツール提供自体についても、有識者コメントから以下の提案が出されている。

- 『ユーザのノウハウを蓄積し共有できると良い。また、ソースコードを公開できるような環境を用意しておくが良い』
- 『学会誌などでは、データと使ったソフトをペアにして提示することが多い』

これらのコメントは、利用者間（これに行政側が加わってももちろん良い）でのノウハウ交換の場として、あるいは開発された統計情報ツールの公開及び評価の場として、統計情報基盤を活用することの可能性を示唆しているものと受け取れる。これは、今後検討すべき課題であろう。

5.1.3 サービス提供の観点での評価

これについては、有識者ヒアリングで『今後もトライアルを続けるならば、サービスモデルをしっかりと決める必要がある』と指摘されているように、まずはサービスモデルの構築がこのようなプロジェクトの成否の鍵を握る。

これはシステムや機能の観点というより、まずは最初にオープンガバメントという大きな概念を整理し、その枠内で提供するサービスや機能の定義を行うべきであるという考え方でもある。総務省を始めとする各省庁、地方自治体との連携を視野に入れた上での適用・展開が容易なサービスモデルの構築が望まれる。これは下記の有識者コメントにも表れている。

- 『オープンガバメントの定義だが、これまで出て来なかった情報をオープンにしていくことなのか、自治体や各省庁が同じような情報アーキテクチャやメニューの定義とすることで情報が探しやすくなることなのか等、まず定義を決めて、次いでモデルを作ると良くなるのではないか』
- 『本プロトタイプの方角性を整理する必要がある。全省庁の統計データを公開している総務省サイトと本プロトタイプ、及び経済産業省の統計サイトとの関係を整理しなければならない』

ここでは、政府全体のプラットフォームからの統計情報提供という観点で各省庁、地方自治体との連携も視野に入れたサービス提供モデルを構築する必要性が課題として提起されている。

【ポイント】

- ☐ グラフ、地図などを活用した統計情報の可視化は有効
- ☐ 利用者が作成したツールを共有できる場とすることも統計情報提供基盤の有効活用につながる
- ☐ オープンガバメントの枠組みでの全体的な定義、サービスやモデルの定義などを明確にする必要がある
- ☐ 総務省をはじめとする他省庁との連携統合を進めることにより、政府統計共同利用システムの充実を図ることが必要である

5.2 統計情報データの新たな活用形態に基づく市場活性化への期待

今回は可視化ツールの一つとして「都道府県グラフ」を提供したが、統計データの新たな利用形態として、地図情報との連携を更に進めた様々なマッシュアップによる統計情報コンテンツの充実を一つの方向性として捉えたい。これについては有識者ヒアリングから以下のようなコメントや提案が出ている。

- 『米国の例で、ウルフラム・アルファ（[Wolfram Alpha](http://www.wolframalpha.com/)、<http://www.wolframalpha.com/>）というサービスがある。統計的な情報等を集め、例えば、ロサンゼルスで地震が起きた時にマグニチュードを地図上に表現したり、平易な問いに対して答えを返すことなどが可能である』
- 『紙の地図帳の教科書を、例えば地図とデータを可変に扱い、自由に組み合わせることができれば、教育や教材の質の向上が図れるのではないか』

統計情報は、RawData または Excel など含む表形式だけではなく、地図情報とのマッシュアップなど様々な活用形態が考えられるが、今回のグラフ化機能への高い評価やこれらの事例紹介、提案を考えると、地図情報に統計情報をオーバーレイさせるマッシュアップ技術が今後の重要なアプリケーション開発の対象となると予想される。

更に、有識者ヒアリングで『簡単な分析結果を統計データと一緒に載せると良い。それを見て内容を把握し、利用者が更にデータを深堀するかどうかの判断ができるようになる』と語られているように、分析結果を何らかの方法で公表するのも利用者の利便性向上につながる場合がある。更に、『文書内容のキーワード検索ができると良いのではないか。号数、キーワード、内容等の切り口を工夫する必要がある』との有識者コメントにあるように、書誌情報の拡充を求める声も聞かれる。

それと同時に、『統計を知らない人は何を対象に調査しているのか中身が分からない。基礎的な情報とリンクし、ある対象が調査されているのかが見えるようになる』との有識者コメントにもあるように、統計情報と書誌情報のリンクなどもハイパーテキストの仕組みを使って実現することも必要であろう。

このような新たな技術を用いたより分かりやすく使いやすい情報提供を行うことにより、ユーザの利便性向上につながれば、統計情報提供基盤によるデータ提供がユーザの間で自律的・発展的に活用され、新たな市場の創出に結びつくことも期待できるであろう。

【ポイント】

- ☐ 地図情報とのマッシュアップは今後期待が持てる分野である
- ☐ 分析結果や分析手法などを統計データと共に提供することも要検討
- ☐ 統計情報と書誌情報をリンクさせることにより統計の内容を理解して貰うための仕組みを実現することが重要

5.3 利用者の認知度について

今回は短期間での実証実験ということもあり、利用者はある程度限られたものにならざるを得なかったが、この点で、今後どのような方法で統計情報提供基盤の存在を周知するかという課題が残されている。

これについては 3 章でも提案されていたように、新聞などのリアルメディアを活用した周知なども検討する必要もあるだろう。Web の中でも何らかの手段を用いてその存在を知らなければ、このサイトに行きつくことはできない。

更に、サイトの存在を知った上で、サイトを魅力的なものとする工夫も必要である。有識者ヒアリングでは、以下のような取り組みが紹介された。

- 『論文検索では CiNii (※Citation Information by NII : 国立情報学研究所が運営する論文情報ナビゲータ) というシステムがあり、API コンテストが行われている。ID 登録者は数万人、IP 認証で利用している人は数十万人いる。募集期間は 4・5 カ月であり、周知期間を十分に取ることが重要である。学会誌といった古くからあるメディアが有効である。内容に興味を持っている人が集まる所でイベントをするのも良いであろう』

ここでも学会誌というリアルメディアを使った周知について触れられているので参考にしたい。

更にここでは、API コンテストという興味深い取り組みが紹介されている。これを統計情報に適用した取り組みとして、統計情報と地図情報のマッシュアップコンテストなどを実施することも一つの方策である。この点での具体例を含む提案については 6 章で後述する。

【ポイント】

- ☐ 新聞、雑誌、学会誌などのリアルメディアによる周知も要検討
- ☐ コンテスト開催は利用拡大に有効な手段となりうる

第6章 今後の展望・提言

本章では、統計情報提供基盤に関するここまでの評価・分析に基づき、今後のサービス提供に関する展望を行い、提言としてまとめる。

6.1 今後の課題等について

まず、今回の実証実験では扱うことのできなかった課題についていくつかの観点で整理する。

■総務省との連携によるシステム強化

今回の実証実験では経済産業省のデータを使った公開となったが、統計情報基盤自体は他省庁との連携を念頭においたものであり、特に統計情報について e-Stat を運用する総務省との連携統合が求められることは言うまでもない。総務省は、膨大な統計情報を作成・管理・提供しており、統計データの二次的利用を考える際に、連携統合を検討すべき対象である。3章で触れた内容の再掲を含め、事前および事後の有識者ヒアリングでも総務省との連携強化や統合の重要性が以下のように語られている。

- 『STATS と e-Stat との間での連携が更に必要』
- 『総務省と経済産業省で連携しながらある程度のモデルを示し、各省庁も含めデータソースのセットとデータ形式、メッセージ形式等について議論を早期に進めるべき』
- 『本プロトタイプの方角性を整理する必要がある。全省庁の統計データを公開している総務省サイトと本プロトタイプ、及び経済産業省の統計サイトとの関係を整理しなければならない』

これらのコメントの背景には、有識者ヒアリングで次のように語られたように、政府から公表される統計データが総務省、経済産業省の2本立てになってしまうことの懸念が挙げられる。

- 『政府としてデータ提供の利便性を向上させるという考え方が重要。将来的に e-Stat と別立てで2つのシステムが同時に稼動することになれば、片方だけが使い易いとか、データの整備レベルが違うといった様ないろいろ不都合が生じる』

これは、統計情報ユーザの観点から政府統計の利用普及を考えた場合に非常に重要なポイントである。

統計情報提供基盤を統計情報のワンストップポータルとして機能させるには、現在、調査協力者と総務省 e-stat、総務省 e-stat と経済省 STATS との間での連携にとどまらず、更に連携・強化を図り、迅速な情報提供を図るとともにデータやその形式の充実を図り利便性の向上を図っていくべきである。

総務省 e-stat と経済産業省 STATS のシステム連携統合の検討への期待と共に、今回の統計情報提供基盤の意義を評価するコメントが有識者ヒアリングから得られていることは、今後の統計情報提供基盤の更なる検討を期待する現れでもあろう。

なお、システム連携によって、調査協力者などへ総務省経由で過去データのフィードバックを可能とすれば、調査協力者にとっても有効なインセンティブとなるであろう。

- 『政府の公的統計の拡充に向けた取組みという点においては、業務・システムの最適化計画の下で、政府統計共同利用システムが統計行政を横に繋ぐ方向で動いている。現在は、政府の統計行政として 1 つの理想である政府統計共同利用システムを目標にしながら、各省庁もシステム構築に取り組んでいる。その中で、今回の実証実験は、総務省の取組みと競合するものではなく、政府統計共同利用システムの理想へと更に進めていく過程の中での一つの取組みとして位置付けることもできるであろう』

■統計データの秘匿性担保

今回の実証実験の公開データは、秘匿性について特に議論の対象となる内容のものではなかったが、この点で、有識者ヒアリングで下記のような指摘があった。

- 『匿名性については、絶対的な匿名性と事実上の匿名性に分けて考えるべきであろう。絶対的匿名性を担保すると、情報として役に立つ内容とはなり得ない。現在は新しい統計法が制定され、統計の二次的利用を促進するために個票に近いデータを公開すべきであるという方向性が見られる。データ公表時に事前に特定化されないことを検証し、事実上の匿名性を担保する必要があるが、個票に近いデータになればなるほど秘匿性についてのリスクは高くなる』

今後、統計データの二次的利用を推進するときに、秘匿性が重要な検討課題となることが予想される。特に、新統計法との関係で匿名性を担保するための統計データ提供の仕方について更なる検討が重要である。

6.2 今後の取り組みについて

ここでは、新たな活動を中心に、今後取り組むべき分野について整理し提案を行う。

■インセンティブとしてのマッシュアップ技術の活用

5 章では、論文検索のシステムに関連し API コンテストの開催が有識者ヒアリングで紹介されていたが、このようなコンテスト形式を統計情報に当てはめたマッシュアップコンテストなども既に海外では実施されている。

これは様々な統計情報を地図情報など他のアプリケーションとマッシュアップすることによって情報価値を高める仕組みを作るものであり、コンテスト形式にすることによって、アプリケーション構築を推進しようというものである。これは周知普及の一環ともなり、認知度の向上にもつながる有効な方策として注目できる。有識者ヒアリングでもその点が『米国や英国のように、マッシュアップに対して賞金を出すとかインセンティブを付ける」と認知度が上がると思う』として指摘されている。

例として、ここでは米国ニューヨーク市での「NYC BigApps」コンテストを紹介する(<http://nycbigapps.com/>)。

これは 2009 年に第一回のコンテストが実施され（2009 年 10 月 6 日から 12 月 8 日まで提出受付）、総額 20,000 ドルの賞金を目指し様々なアプリケーションが登録された。このときの結果は <http://nycbigapps.s3.amazonaws.com/index.html> に掲載されているが、例えば、その中で「NYC Health and Demographics」というアプリケーションがある。これは、米国勢調査局（US Census Bureau）からの人口学的情報と、NYC Community Health Surveyからの結果を地図上にオーバーレイし、比較することができるものである。Google Mapに色分けして表示され、健康レイヤ（高血圧、糖尿病などの比率）や人口学的レイヤ（例：平均収入、人口密度など）などの間で切り替えが可能である。

NYC Health and Demographics

by [Webfoot Maps](#)

On this map, you can compare or overlay information from the [NYC Community Health Survey](#) with demographic information from the [US Census Bureau](#).

Use the controls below to switch [Health](#), [Demographic](#), or [Dots](#) layers, and then to modify colours and opacity.

▶ How the area colour controls work

Health Layers:

% of population with high blood pressure ever

▼ Hide controls

Min

fffff

17

Mapping:

linear

<-- transparent | opaque -->

▼ Hide data information

Percent of population who have ever had high blood pressure; data from New York City Department of Health and Mental Hygiene (2007). See [here](#).

Demographic Layers:

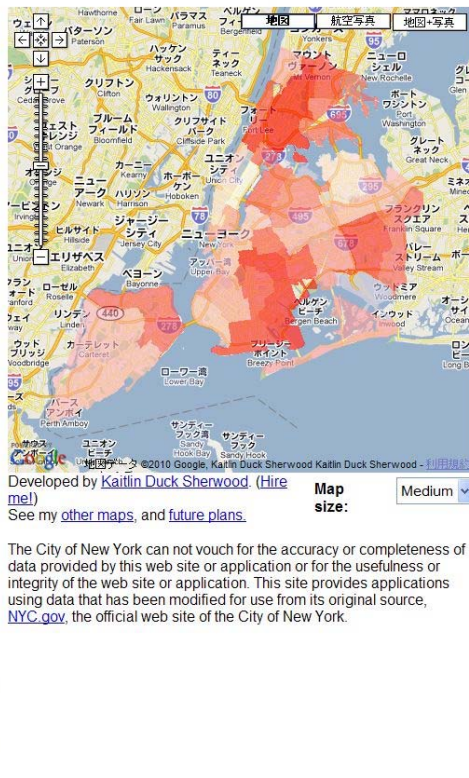


図 6.1 統計と地図のマッシュアップアプリケーション例「NYC Health and Demographics」
(<http://maps.webfoot.com/demos/nyhealth/NYHealth.html>)

今年（2010 年 10 月 12 日から 2011 年 1 月 12 日まで提出受付）も「NYC BigApps 2.0」というタイトルで 2 回目のコンテストが実施されている。

■ 地方自治体との連携

オープンガバメントとして、中央省庁でのプラットフォーム活用が考えられる一方、地方自治体も含めた形のより大きな枠組みも視野に入れると、一層のサービス向上につながるはずである。

これは単に一般利用者から見たメリットだけではない。例えば、統計情報提供基盤にとっても、地方自治体の様々な情報を取り込むことができるのは基盤運営側にとってはデータ拡充の上で重要なことである。一方で、データを提供する地方自治体にとっても、システム開発の負担軽減にもつながり、更には提供されたデータが様々な形で活用されることによって情報に対する価値が付加され、その恩恵を受けることにもつながるはずである。この点は有識者ヒアリングでも指摘されているところである。

- 『自治体にも同じような取組みをして貰い、同じプラットフォームを使って貰うようにした方が良い。日本では各自治体で個別にシステムを作っているが、オープン

ソースでインフラをシェアするということが必要となってくるのではないか。例えば、2010年9月から米国では「Civic Commons」というサイトがオープンし、テクノロジーをシェアすることを始めている』

- 『現在の分散システムの考え方で地方自治体がシステムを持つ場合には、前述のようにメッセージ形式などの標準を作る必要もあり、財政的にも負荷が大きいなど課題が多い。そこで、クラウドを使って国も情報を還元するという仕組みになれば歓迎されるかもしれない。その上で、地方自治体の統計、経済産業省の統計、総務省の統計という区別をするのではなく「ガバメント STATS」のようなものが構築されれば、それが本当のオープンガバメント的なものとなるのではないか』

これらのコメントを念頭に、今後は地方自治体からの協力をも得て、双方がメリットを見いだせるプラットフォームの姿を検討する必要もあるであろう。

■SDMX 標準化への取り組み

今回、統計情報の一つの形式として SDMX を使用したことは、統計情報提供基盤の大きな特徴であった。この基盤上でのデータ集約を図る意味においても、更には今後、国内の地方自治体との連携、国外データとの比較などを行うためには何らかの標準形式に合わせることが重要であり、その一つの方法として SDMX が考えられる。

ただし、SDMX の標準化は欧州を中心に国外で行われていることもあり、この標準規格をより使いやすいものとするためには、標準化プロセスに関与することにより、日本独自のニーズなどを取り込むような活動も重要である。

有識者ヒアリングでも、これらの課題に関する指摘を含めた次のような提案が行われている。

- 『ツールの作成や統計情報の SDMX 化等、経済産業省として SDMX を普及させるために、どこまで推進役となるかが検討の一つのポイントである。もう一つ重要なポイントとして、SDMX の標準化にどこまでコミットするかという課題もある。国際標準として存在するものを日本へローカライズすることと、日本で決めた内容を国際標準としてコミットしていくことの2面において、これらの動きがうまく進まないと、データ形式として SDMX が使える姿とはならないのではないか』
- 『国際標準の相互接続性を持つ規格として SDMX を経済産業省が普及させ、世界中の色々な情報を SDMX 化することに日本として貢献したいのであれば、ツール作成や国際会議への影響力を持つよう努力することが必要となる。そうすることで、色々なソフトウェアのライブラリに日本で決めた仕様が入ってくるようになり、そうするとライブラリを作るための知的財産権の確立やソフトウェアとしての市場価値が出てくる。ただ単に利用しているだけの立場では、そういうデータ形式があるとい

うことから脱却しにくい』

■明確なユーザモデル／ビジネスモデルに基づく経済・産業活性化のインフラ整備

統計情報基盤としては、提供されるデータが報告書での引用や経済アナリストの分析という特定領域内で止まるのではなく、一歩進んで産業の活性化・地域経済の発展に結びつく形で活用されるなら、それは単なる情報提供を超えた、実際の経済活動へのフィードバックを実現できる原動力となるということであり、非常に望ましい姿である。

そのためには、本報告で述べた地方自治体との連携、SDMX などのデータ標準化、そしてマッシュアップ技術の活用などのインセンティブ付与によるツール開発の促進など、総合的な取り組みが求められる。それに加え、個々の取り組みの前に、3 章でも触れたユーザモデル／ビジネスモデルの明確化も重要である。

例えば、ある地域で店舗開設を計画している企業や個人は、その地域でのニーズ、既存類似店舗の展開状況、現在のサービス提供の地理的範囲など様々な情報を活用することにより、どの場所にどのようなサービスを提供する店舗を配置すれば、ビジネスとして成り立つかなどの判断が下せるはずである。この様な情報及び活用ツールが統計情報基盤から提供されるとすれば、統計情報基盤が、地域および企業／個人に対する経済最適化のための有効な道具立てにもなりうる。

サービス機能の前提となる統計データとしては、地域情報として地方自治体が持つ商業データ、住民データ、都市計画データなどが不可欠であることは言うまでもなく、そのようなサービス提供基盤は、もはや単なる統計情報提供サイトではなく、「地域経済活性化・最適化のためのインフラストラクチャ」として位置付けるべきものとなる。これは現在の統計情報提供の一つの望ましい将来的な発展形である。

このような総合サービス提供基盤を構築することによって、地域の産業・経済の発展への寄与というメリットが明確になってくれば、情報提供のインセンティブとしても働き、国と地方の Win-Win 関係も構築される。

そのような枠組みの中では、国は総合情報データベース構築・整備の役割を担うことになる。そして、その上に地方自治体、民間企業、情報の実利用者などの様々なプレーヤが参加できるインフラストラクチャを構築することができれば、経済活性化という形を通じた地方自治体／企業への利益還元ともなり、オープンガバメントが標榜する「政府間及び官民協力」の一つの実現形としてのプロトタイプシステムにもなることが期待される。

6.1 と 6.2 をまとめると、次のような今後の課題や取り組みが浮かび上がる。

【ポイント】

- ☐ 統計情報提供の仕組みを充実させるためには総務省との連携統合が不可欠
- ☐ インセンティブとしてのマッシュアップ技術の活用などは利用拡大に有効な手段
- ☐ 地方自治体を含めたオープンガバメントの構築は、利用者のみならず、中央省庁や地方自治体それぞれにメリットを見出すことができる

6.3 最後に

今回の統計情報基盤による実証実験においては、データ提供の仕組みとしてのシステム、提供されるデータの整備、提供方法など様々な角度での検討が行われ、その中から今後対応すべき課題や取り組むべき新たな分野が明らかになった。

有識者からも『一年限りの実証実験で終わらずに、継続できれば良いと思う。経済産業省だけでなく、政府全体へ広げていくのが望ましい』、『非常に良い取組みであったので、他省庁へ広げていって頂きたい。また、例えば政府のレポートや科学技術系のデータ等一元的になっていないものが沢山あるので、統計法上の統計データ以外にも広げて貰いたい』などの声が寄せられており、今回の実証実験が、オープンガバメントにおける行政情報提供の一つのひな型となり、各省庁や地方自治体を含めた横断的統一プラットフォームの端緒として意義あるものであったと評価されたことは意義深い。その中でも統計情報に関し言えば、特に総務省との連携を更に強化することが重要な課題であることが明確になっている。これに加え、利用者にとって必要な統計データは何かを特定し、どのようなサービスを提供するかを定義するサービスモデルの確立、対象利用者や利用形態の明確化、データ形式標準化のプロセスへの寄与、統計利用者へのインセンティブ策の一つとなるマッシュアップ技術の活用など、与えられた課題は多い。

今後、これらの課題に積極的に取り組むことにより、「透明性の向上（transparency）」、「市民参加（participation）」、「政府間及び官民協力（collaboration）」の3原則を真に実現するオープンガバメントのプラットフォーム構築が望まれている。

用語集

CSV

Comma Separated Values の略称。表計算ソフトなどのデータを項目別にカンマを区切って出力、保存、読み込みを行うためのデータ形式。

e-Stat

政府が作成、公表する統計に関する情報提供をワンストップで提供する総務省のポータルサイト。学習、ビジネス、研究などに統計を活用する機能が提供されている。

GUI

Graphical User Interface の略称。コンピュータ画面でグラフィック表示を用いてユーザーに分かりやすい操作性を提供するユーザインターフェイス。

OpenID

様々な Web サイトに一つの ID でログインできるようにする認証の仕組み。あるいは ID。

OSS

Open Source Software（オープンソースソフトウェア）の略称。コンピュータプログラムのソースコードが無償で公開され、その改変を自由に行えるソフトウェア、あるいはそのようなソフトウェアの開発・利用の一つの形態。

PDF

Portable Document Format の略称。Adobe Systems Incorporated（アドビ システムズ社）によって開発された文書用電子フォーマット。印刷イメージの文書が表現できるため、Web 上での文書配布に広く用いられている。ISO（国際標準化機構）によって ISO 32000 という国際標準規格として承認されている。

RawData

統計において、データを作成・収集した組織から提供される生の（Raw）データ。表現形式としては CSV、XML などの様々な形式を取る。データ分析などを行うには、表示ソフトなど汎用的なツールや、データ提供組織によって配布される分析ツールなどを使用する。

RSS

RDF Site Summary の略称。Web 上のリソースに関するメタ情報／説明情報を表現する

ための RDF (Resource Description Framework) という W3C (World Wide Web Consotium) の規格を用い、Web サイトの見出し情報や要約情報を配信するためのデータ形式。RSS にはいくつかのバージョンが存在し、正式名称も異なる場合があるが、本システムでは、広く普及している RSS1.0 を使用している。

SaaS

Software as a Service の略称。クラウドコンピューティング環境における一つのサービス形態であり、ソフトウェアをユーザが必要とする時に必要な機能だけ提供するというもの。ユーザにはその分だけの使用料が課金される。

SDMX

Statistical Data and Metadata Exchangeの略称。統計データ交換のためのXMLタグセット標準を開発したイニシアティブ (<http://sdmx.org/>) の名称。具体的な統計用XMLタグセットを定めた仕様群のことを指しても使われ、本報告書では後者の統計用XMLタグセット仕様を表して用語として使っている。SDMX Standards としてVersion1.0 (2004 年 9 月) とVersion2.0 (2005 年 11 月) が制定済みであり、本報告書執筆時点でVerion2.1 が開発中。

STATS

経済産業省が構築した「調査統計システム」の通称であり、統計調査等業務における基幹業務 (統計調査の「企画」、「実施準備」、「実施」、「審査・集計」、「分析」、「公表・提供」) 及びこれを支援する名簿管理業務、並びにこれらの業務を処理するシステム。

WebAPI

Web 上のサイトが提供するサービス (プログラム) 処理の入出力インターフェイス。これを公開することにより、Web システムの開発者が独自のサイトを構築することができるようになる。

XML

Extensible Markup Language の略称。1998 年に W3C (World Wide Web Consotium) が勧告 (Recommendation : W3C での標準規格制定の最終決定版) として公開したマークアップ (<h1>、<title>等のテキスト情報で他のテキストを囲むことにより、囲まれたテキストデータの意味を表すもの) の書き方やタグセットの定義方法を定めたメタ言語。XML 規格は具体的なタグセットは規定していないので、実際に XML を活用するには、業界や適用領域ごとにタグセットを定義する必要がある。

アジャイル

英語では Agile であり、この意味である「敏しょう」「機敏」をソフトウェア開発にもたらし手法。開発者が密にコミュニケーションを取り合うことで、短期間での迅速なソフトウェア開発を行う。小さい単位でプログラムを開発し、その機能拡張を順次行うという方法が取られる。また、まずは動くプログラムを作り上げることを主眼としているため、作成される文書（仕様書）の量が少ないという特徴を持つ。

オープンガバメント

インターネットの双方向性等を活用することで政府機関による行政情報の公開や行政への市民参加を促進する政府の取り組み。

クラウド／クラウドコンピューティング

クラウドコンピューティングとは、コンピュータネットワークを運営する組織から、ハードウェア、ソフトウェア、サービス等の提供をインターネットを介して受け、利用した分だけの費用を支払うというコンピュータ処理の概念である。期限付きサービスとしてハードウェア、ソフトウェアを利用できることによって、システム導入期間の短縮、初期・運用コストの削減、サービスの迅速な開始が実現できる。更には、システムリソースの拡張にも柔軟に対応できる仕組みとして近年注目を集め、様々なサービスが提供され官民合わせその動向が注目されている。

マッシュアップ

Web 上で公開される複数の異なるサービスを組合せ・加工を行い、新たなサービスを構築する手法。Web2.0 と呼ばれるインターネット利用環境の潮流の中で出てきた Web コンテンツ提供の考え方。地図情報提供サービスと地図情報にマッピング可能な情報を組み合わせたサービスなどが多く出現している。

参照資料

以下に、本報告書で言及した資料やサイトなどの情報入手先としての URL を列挙する。

CiNii : 国立情報学研究所が運営する論文情報ナビゲータ

(<http://ci.nii.ac.jp/>)

Civic Commons : 米国の様々な政府機関でテクノロジーを共有することを目指すサイト

(<http://civiccommons.com/>)

DATA.gov : 米国における行政情報提供サイト

(<http://www.data.gov/>)

e-Stat : 統計情報提供のための総務省のポータルサイト

(<http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do>)

EUROSTAT : EU における統計情報提供サイト

(<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>)

NYC BigApps : ニューヨークのマッシュアップコンテスト

(<http://nycbigapps.com/>)

NYC Health and Demographics : NYC BigApps の昨年の受賞アプリケーション

(<http://maps.webfoot.com/demos/nyhealth/NYHealth.html>)

SDMX : 統計データの標準形式 (SDMX 形式) を推進するイニシアティブ

(<http://sdmx.org/>)

ウルフラム・アルファ (Wolfram Alpha) : 統計情報と地図情報のマッシュアップ例

(<http://www.wolframalpha.com/>)

新たな情報通信技術戦略 : 2010 年 5 月 11 日に IT 戦略本部がまとめた政策資料

(<http://www.kantei.go.jp/jp/singi/it2/100511honbun.pdf>)

新統計法：60 年振りに改定された統計法（平成十九年五月二十三日法律第五十三号）

(<http://www.stat.go.jp/index/seido/houbun2n.htm>)

統計情報の XML 利用環境の整備に関する調査研究に係る報告書：昨年度の報告書

(http://www.meti.go.jp/statistics/toppage/topics/kenkyuShiryo/hokokusho/pdf/xml_houkokusyo.pdf)

附録 A. 「データボックス オンラインアンケート」の結果一覧

「データボックス オンラインアンケート」の結果一覧

「データボックス」サイト内には設けられたオンラインの「アンケート」から得られた回答の集計結果を以下に記す。

【設問 1】 あなたの業種を選択してください。

[n=55]

(1) 農林漁業	0
(2) 鉱業	0
(3) 建設業	2
(4) 製造業	4
(5) 電気・ガス・熱供給・水道業	1
(6) 情報通信業	9
(7) 運輸業	0
(8) 卸売・小売業	2
(9) 金融・保険業	1
(10) 不動産業	1
(11) 飲食店, 宿泊業	0
(12) 医療, 福祉	2
(13) 教育, 学習支援業	0
(14) その他のサービス業	10
(15) 公務員	5
(16) 学生	2
(17) 家事従事者	1
(18) 無職	5
(19) その他	6
回答なし	4

【設問 2】 あなたの職務を選択してください。

[n=50]

(1) 管理職	10
(2) 技術職	15
(3) 事務職	14
(4) 営業販売職	5
(5) 研究職	3
(6) 学生	2
回答なし	1

【設問 3】 経済産業省を含む公的組織から公開されている統計データを利用していますか。

[n=54]

(1) 定期的に利用している	1
(2) 時々利用している	31
(3) 利用したことがない	20
(4) その他	2

【設問 4】 ご回答頂いた、利用したことのある統計データについてお聞きます。：利用頻度を左の選択肢の中から選んでください

[n=45]

(1) 毎日	0
(2) 週 1 回	1
(3) 月 1 回	5
(4) 四半期 1 回	15
(5) 半期 1 回以下	12
(6) その他	11
回答なし	1

【設問 5】 ご回答頂いた、利用したことのある統計データについてお聞きます。：最も多く利用する取得方法を左の選択肢の中から選んでください

[n=44]

(1) 紙媒体(報告書・リーフレット)	2
(2) 市販の CD-ROM	0
(3) ウェブサイト	35
(4) その他	6
回答なし	1

【設問 6】 ご回答頂いた、利用したことのある統計データについてお聞きます。：最も多く利用するデータ形式を左の選択肢の中から選んでください

[n=43]

(1) Excel	21
(2) CSV	7
(3) PDF	9
(4) その他	5
回答なし	1

【設問 7】 ご回答頂いた、利用したことのある統計データについてお聞きます。：利用目的を左の選択肢の中から選んでください

[n=42]

(1) 内部向け資料(事業計画、経営分析、業界景気判断等の基礎資料)作成のため	14
(2) 外部向け資料(公表資料、研究論文、教材、講演資料等)作成のため	11
(3) 二次利用(統計データの加工・再販等)のため	6
(4) その他	11

【設問 8】 本システムを何で知りましたか

[n=51]

(1) 経済産業省のウェブサイト	4
(2) 経済産業省 新着情報配信サービス	5
(3) 調査統計部メールマガジン	5
(4) 経済産業省以外のメールマガジン・メールニュース	4
(5) オープンガバメントラボウェブサイト	13
(6) テレビ・新聞・雑誌・ラジオ	0
(7) 専門雑誌・業界新聞	0
(8) IT 系ニュースサイト	2
(9) 行政系ニュースサイト	2
(10) Twitter	5
(11) ブログ	1
(12) 上司・知人からの紹介	7
(13) その他	3

【設問 9】 本システムの全般的な使い勝手や操作性はいかがでしたか

[n=50]

(1) 良かった	12
(2) やや良かった	13
(3) やや悪かった	20
(4) 悪かった	5

ーユーザによる自由記述コメント 5 件ー

1. サイトが重く、表示されるまでに時間がかかりすぎ、ストレスがある。まだできたばかりなので仕方がないと思うが、グラフがないデータが多く、何がメインコンテンツなのかよく分からない。統計の検索はしやすくなったが、データの閲覧製や利便性向上にはもう一歩努力が必要と思う。でもとても良い取り組みだと思います。
2. 資料を様々な属性でソートできる機能は必須だと思います。また、デフォルトで並べる順序は、
 - ・人気度の高い順
 - ・発行日が若い順
 などになっていると助かります。
3. アクセスしてもたびたび表示されない、あるいは更新されなかったりする。現状では業務に用いることのできる品質だとはとても言えない。ただ、各地各所ばらばらにあったデータ類を集約するという方向性は良いと思うので、今後の改善に努めてほしい。
4. 選択欄や記入欄が画面枠から外れていて見るのを飛ばしてしまう事が多々ありまし

た。本記入欄もその1つです。

5. サイト全体が見難い。また、いろいろなブラウザに対応してほしい。

【設問 10】 本システムの全般的な画面デザインはいかがでしたか

[n=50]

(1) 良かった	15
(2) やや良かった	22
(3) やや悪かった	8
(4) 悪かった	5

ーユーザによる自由記述コメント 3 件ー

1. 初めて観るサイトでも、どこに情報が配置されているか、分かりやすい。
2. 表示画面が大きくないと辛いですね
3. 17 インチ以上の画面であれば問題ないと思いますがノート PC 等の小さい画面になってくると表示される部分が少なくなり分かり辛くなると思います。

【設問 11】 本システムで関心を持った(便利と思った)機能は何ですか。(最大 3 つまで)

[n=48 (複数回答可のため集計総数は 92)]

(1) 検索機能	14
(2) グラフ機能	20
(3) ファイルダウンロード機能	14
(4) 詳細分析機能	5
(5) SDMX 形式のデータ提供	2
(6) コメント付与	7
(7) レイティング(評価)	6
(8) 画面カスタマイズ	0
(9) Web API 公開	2
(10) RSS 機能	3
(11) OpenID の活用	19
(12) その他	0

ーユーザによる自由記述コメント 2 件ー

1. グラフ機能はあると良いと思うが、重過ぎる。もっとサクサク表示されるならとても良いと思う。レイティングは何のためにあるのか良く分からなかった。
2. 他の方の意見も見ながら資料がダウンロードできるのが参考になるかと思いました。

【設問 12】 検索機能の充実は、統計データの利用を促進する要因になると思いますか。

[n=51]

(1) 大変そう思う	32
(2) ややそう思う	15
(3) あまり思わない	4
(4) まったく思わない	0

ーユーザによる自由記述コメント 2 件ー

1. 利用する人（主に仕事？）は以前から利用していただろうし、検索機能が充実しても、これまで利用しなかった人（統計データとかかわりの無かった人）が統計データを利用するようにはならないと思うので、あまり関係ないと思う。
2. 利用者が増えてくれば良く利用されるダウンロードデータ回数である程度決まってくると思いますが初めは目的のものを見つけるのに苦労するのかもしれない。

【設問 13】 グラフ化機能は、統計データの利用を促進する要因になると思いますか。

[n=53]

(1) 大変そう思う	25
(2) そう思う	19
(3) あまり思わない	7
(4) まったく思わない	1
回答なし	1

ーユーザによる自由記述コメント 2 件ー

1. パッと見た場合に利用可能なデータの全体像を見るのには良いかもしれませんが。詳細内容や各自が持つデータと合わせて使う場合にはレコード（行）ユニークにして使うのではないのでしょうか？
2. グラフ化機能ができたところで、もともと統計データを利用していなかった人が新たに統計データを利用するようにはならないと思う。もともと利用していた人が、自分でグラフを作る前に全体的な傾向をぱっと知ることができるという点で、利便性は向上しているといえる。

【設問 14】 詳細分析機能は、統計データの利用を促進する要因になると思いますか。

[n=49]

(1) 大変そう思う	21
(2) ややそう思う	22
(3) あまり思わない	5
(4) まったく思わない	1

ーユーザによる自由記述コメント 2 件ー

1. グラフ化機能と同様に、もともと統計データを利用していなかった人が利用するようにはならないと思う。また、統計データを利用する人は自分の目的に合わせて詳細分析を行うと思うので、詳細分析機能自体、あまり必要ないと思う。もちろん詳細分析はないよりはあったほうがいいかもしれないが、分析の切り口をすべてのユーザーの要望に合わせて作るとは難しいため、詳細分析に工数をかけても効果は薄いと思う。
2. ダウンロードして使う前の調査用には良いかもしれませんが詳細分析で全てをまかなうとは行かない様に思います。

【設問 15】 XML 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。

[n=50]

(1) 大変そう思う	23
(2) ややそう思う	18
(3) あまり思わない	7
(4) まったく思わない	2

ーユーザによる自由記述コメント 2 件ー

1. 統計データの利用者にとって、様々な形式でデータをダウンロードできることは非常に重要。
2. 汎用的で一般的なデータは利用促進には必要だと思います。

【設問 16】 Google や Yahoo の ID でログインできる機能(OpenID)が利用を促進する要因になると思いますか。

[n=51]

(1) 大変そう思う	28
(2) ややそう思う	13
(3) あまり思わない	8
(4) まったく思わない	2

ーユーザによる自由記述コメント 3 件ー

1. IdeaBox とは異なり、Databox では個人を区別しカスタマイズするような要件は、特段見当たらない。なので、現状ではログイン/ID を利用するだけの利点を感じられない。カスタマイズした分析解析結果や情報の保存…等？将来的に、ID が有用と感じられるような機能が付与されるのだろうか？あるいは API を利用してアクセスするアプリケーションで用いられるのだろうか？
2. 所属している組織にもよるのかもしれませんが単純に利用データが欲しいだけの場合と Web への書き込みが許可されている組織で扱いが変わってくるのではないかと思います。
3. そもそもなぜ統計データをダウンロードするのに ID 登録が必要なのか理解できない。統計局のウェブサイトにいけば ID 登録なしにすべてのデータが入手できるのに、なぜ？

【設問 17】 SDMX 形式の統計データの提供は統計データの利用を促進する要因になると思いますか。

[n=46]

(1) 大変そう思う	6
(2) ややそう思う	19
(3) あまり思わない	19
(4) まったく思わない	2

ーユーザによる自由記述コメント 1 件ー

1. 対応しているアプリケーション種類の数に左右されてくると思います。また XML や Excel、CSV 等の形式より利便性が高い必要もあるかと思います。

【設問 18】 今後も本システムを利用したいと思いますか

[n=50]

(1) 是非使いたい	14
(2) 使いたい	28
(3) あまり使いたくない	5
(4) 使いたくない	1
(5) その他	2

ーユーザによる自由記述コメント 2 件ー

1. データ更新の頻度と内容によります。
2. 新データが更新されていけば面白いと思います。パフォーマンスは上がってほしいですが・・・。

【設問 19】 使いたい理由について、あてはまるものをお選びください。(最大 3 つまで)

[n=42 (複数回答可のため集計総数は 78)]

(1) 画面のデザインが優れている	4
(2) 応答速度が速い	5
(3) 検索機能が使いやすい	9
(4) 操作がわかりやすい	9
(5) 自分のやりたことができる	10
(6) 欲しい統計データがある	22
(7) 統計データを必要としている	19

ーユーザによる自由記述コメント 1 件ー

1. 選択肢の (7) の存在意義が不明。(7) に該当しない人はそもそも利用しないでしょう。

【設問 20】 使いたくない理由について、あてはまるものをお選びください。(最大 3 つまで)

[n=34 (複数回答可のため集計総数は 66)]

(1) 画面のデザインが悪い	9
(2) 応答速度が遅い	18
(3) 検索機能が使いにくい	6
(4) 操作がわかりにくい	15
(5) 自分のやりたことができない	5
(6) 欲しい統計データがない	8
(7) 統計データを必要としていない	4
回答なし	1

—ユーザによる自由記述コメント 1 件—

1. 応答速度が遅い点はできる限り早く改善して欲しい。ストレスがたまる。

【設問 21】 本システムに期待することは何ですか。(最大 3 つまで)

[n=50 (複数回答可のため集計総数は 112)]

(1) 検索機能の強化	11
(2) 画面カスタマイズ	4
(3) グラフ機能の強化	4
(4) 詳細分析機能の強化	11
(5) XML/SDMX 形式データの充実	4
(6) 過去の統計データの公開	24
(7) 統計データの時系列整備	23
(8) 公開データ粒度の細分化	12
(9) 品目・業種・調査項目等の拡充	10
(10) その他	8
回答なし	1

—ユーザによる自由記述コメント 3 件—

1. 統計は時系列比較が基本だと思います。過去データの整備、時系列整備と、時系列でのグラフ化機能の強化が進めばとても使いやすいサイトになると思います。
2. どのような統計情報があるのか一覧が欲しい。
3. 地域や業種等に関する大括りでの細分化は必要になってくると思います。

【設問 22】 統計データの利用を促進する施策について、ご意見がありましたらお願い致します

(ユーザからの回答)

[n=19]

- 『各省庁や省庁にぶら下がる外郭団体のデータも同列に扱って公開して欲しいです。』
- 『Google Fusion Tables はまだ実験段階っぽいですが、こうしたオープンなサービス上にデータを転用すると、他のデータと組み合わせた分析ができるため、用途が広がります。』
- 『・検索ワードに関連する統計情報を探し出すことができるよう、検索機能を強化して欲しい。例えば「自動車」と入れると工業統計と産業連関表がでるような感じで。
・地方自治体が所管する統計にも範囲が広がればなお良いと思います。例えば水道など。今だと地方自治体が所管する分野の統計は各自治体から個別に集めなければならず、大変に手間です。』
- 『最小単位で公開し要望枠を設けて一定期間等で検討改善していくのが良さそうに思えました。』
- 『統計データ利用のためにログインする必要があるのでしょうか？』
- 『他省庁や自治体の統計にも是非広げて欲しい。いわゆる統計法上の統計だけでなく、政府が所有する調査結果データ、科学技術関係のデータなど「政府保有情報」に範囲を広げてほしい。レコメンデーション機能を付加して欲しい。』
- 『どのような統計データを取得しているかを、世間に広めることが利用者を増やす・促進することに繋がると思います。また、幅広い統計データを取得し、様々なニーズにいつでも応えられるようにすると、利用者は喜びます。』
- 『「最新データの反映」「表示速度の高速化」これらが＜継続的に実施＞されることを望みます。』
- 『◎メルマガを通じたニュースリリースやHPの新着情報などで情報を得る機会が多いので、検索については、必要な時検索エンジンまたは行政のHP内検索等を利用し行っていたが、今後は、L a b sを通じて行っていきたい。コンテンツを充実されんことを希望します。』
- 『最新データと過去のデータを比較しやすくしてほしい』
- 『国の統計データを各省の HP など閲覧していますが、こういったデータをどの省が取りまとめているのか分からない時があります。各省の統計データがすべて閲覧可能で、かつキーワード検索などで必要とする情報にすぐ辿りつけるサイトを作るのはどうでしょうか？』

- 『1つの場所に集められていることだけでも、非常に価値があります。ただし、企業内の全員が多頻度で使うものではないかもしれません。継続的にデータが更新されて、いつも使えることで、しだいに利用者は増えてくると思います。(間違いなく)』
- 『情報のトレーサビリティとして、どういう方法で統計データを得たか。具体的に(紙ベースアンケート、web、電話、メール、質問票形式、質問文、対象の決め方、機械式無差別コール(時間帯は?)、任意抽出(母集団は?)、街頭(どこか)等、業界紙上(何誌)等)わかるようにしてほしい。提供形式としては、国産ではないが、実質標準として Excel 形式は必須では?』
- 『特になし』
- 『サイトの使いやすさ、検索機能を充実させてほしい。』
- 『国への要望を言える、それを議員及び職員へ効率よく伝達できるシステムを今後とも開発していただきたく存じ上げます。』
- 『あらゆる詳細情報もデータ化し、欲しいと思ったデータを確実に取り出せるようにすれば、利用者の裾野が広がるし、複合的に利用しようとする人も増えると思います。』
- 『レイティング(☆)だけでなく、実際に引用された(ダウンロードされた)回数も表示したり、利用者のコメントも書き込めるようになっていて、とても助かります。また、資料間の関係(引用や参照関係)や、この資料を見た人はこんな資料を見てますよというようなナビゲーション[※]もあると便利です。アマゾンなどを参考に、是非検討してみてください。期待しています!』
(※「レコメンデーション」の意)
- 『ある程度系統立てた整理が必要。現状、雑多に並んでいるような印象を受けたが、見たい情報にある程度近い内容に絞り込まれていれば「ついでにチェックしようかな」とも思えるものの、あれもあります。これもあります状態になってしまうと探すだけで辟易してしまう。結果、検索してピンポイント到達の方向に行ってしまうと、人目に触れず埋もれてしまうもので溢れかえってしまうことになる。もっとも、これは取捨選択代行を示しているのでは無い。運営側が重み付けまでをしてしまうと、意見誘導にまで行ってしまうことになるため、そこは慎重に対応する。』

附録 B. オープンガバメント系サイトのツールやマッシュアップ事例

オープンガバメント系サイトのツールやマッシュアップ事例

概要：

米国 data.gov、英国 data.gov.uk、ニューヨーク市アプリコンテストの事例を調査した。
公開されている事例は、施設・設備（介護施設、育児施設、駐車場、学校、医療、薬局、歯医者、不動産、郵便ポスト、ゴミ処理場、等）の所在地を、地図上にマップしたものが多い。
治安・犯罪、求人・求職、投資強度等、大気汚染の情報を地図上にマップした事例もある。
現在地の位置情報と連動させた iPhone アプリは多い。

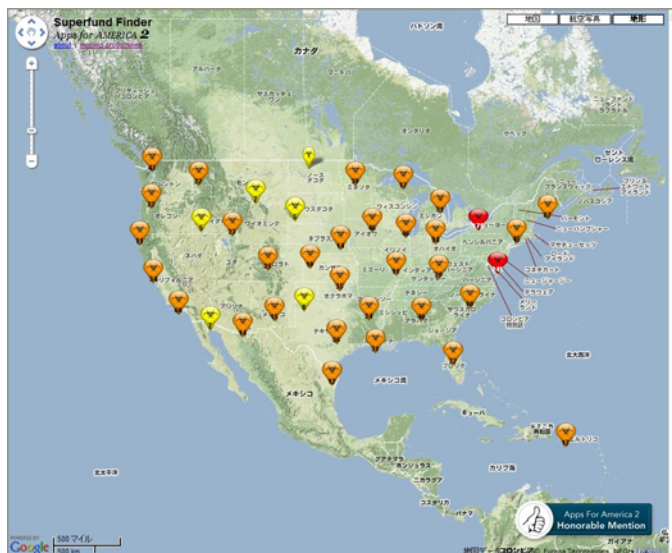
1. 米国 DATA.gov <http://www.data.gov/>

2. Superfund Finder

アプリケーションURL <http://checkitnsee.com/>

データソースURL <http://checkitnsee.com/SuperfundFinderSource.tgz>

EPA National Priorities List(NPL)に関する有害廃棄物サイトは一般に Superfund サイトと呼ばれる。Superfund は 1980 年の「Comprehensive Environmental Response, Compensation, and Liability Act」(CERCLA)の一般名になっている。Superfund Finder は Superfund サイトの位置情報を対話的に見える化することにより、国全体の Superfund サイトの配置について気づきを与えることを目的としている。



3. Interactive USDA Plant Hardiness Zone Map for the Continental United States

アプリケーションURL <http://www.hardiness-zone-map.com/>

米国では樹木と植物の耐寒性を表すのに、Z1 から Z11 までの zone number で表される。この地図は 1990 USDA Hardiness Zone Map に基づいている。ゾーンのボタンをクリックすると、地図上にそのゾーンが表示される。



4. [National Obesity Comparison Tool](#)

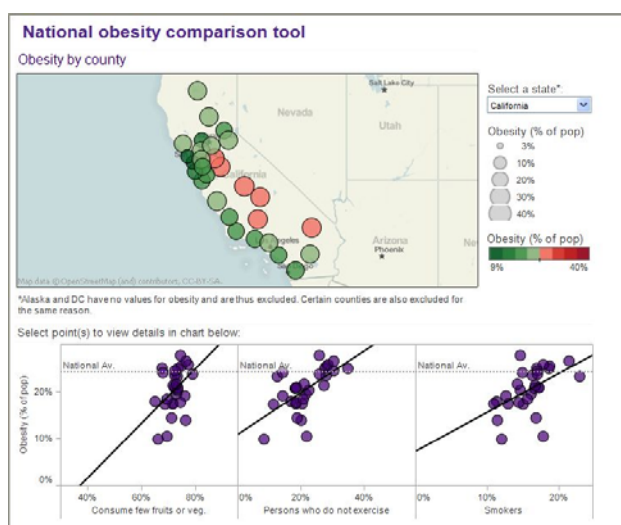
アプリケーション URL

<http://public.tableausoftware.com/views/contributorstoobesity/Eatyourvegetables?:embed=yes&:toolbar=yes>

肥満率のデータを地図上にマッピングしている。地図上でズームイン・ズームアウトが可能。また、肥満率と他の要因(野菜やフルーツを摂るか、運動をするか、タバコを吸うか)などの相関グラフがある。

グラフエンジンは、「Tableau (タブロー)」を使用している。

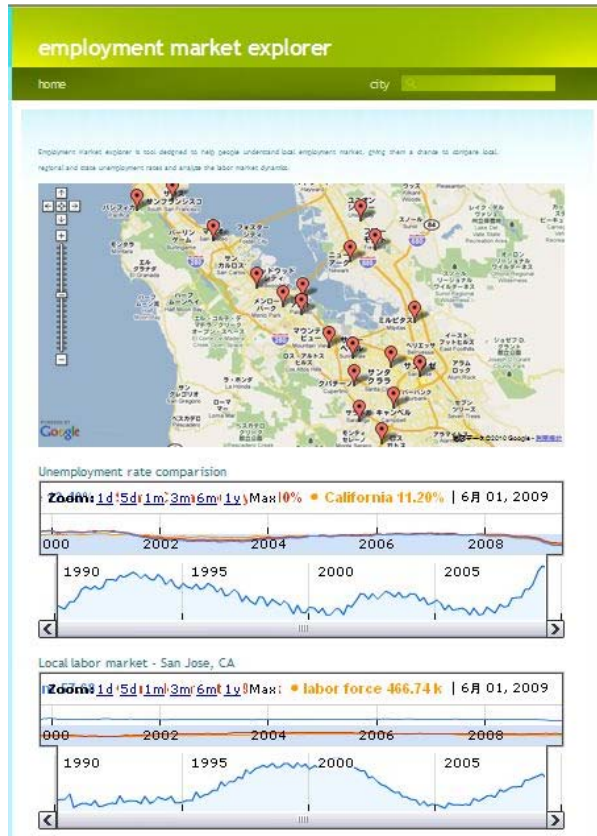
<http://www.tableausoftware.com/public/>



5. [Employment Market Explorer](#)

アプリケーションURL <http://pujaplicaciones.javeriana.edu.co/Employment/>

ローカル、地域、州の失業率を比べて、労働市場力学を分析できるようにすることで、ローカルの雇用市場の理解できるように設計された。



6. [DataMasher](#)

URL <http://www.datamasher.org/>

複数のデータの関係をグラフ化できるツール。データセットを選択、演算子を選択、別のデータセットを選択の3ステップでマッシュアップが完成する。

州別のランキングを地図に表示する、州別のランキングを一覧表にする、マッシュアップについて議論する、マッシュアップをレーティングする、等が可能。

マッシュアップ例

Percent of Adults Who Smoke [plus] Adult overweight obesity rate

<http://www.datamasher.org/mash-ups/percent-adults-who-smoke-plus-adult-overweight-obesity-rate#map>



7. Clean Air Status and Trends Network (CASTNET)

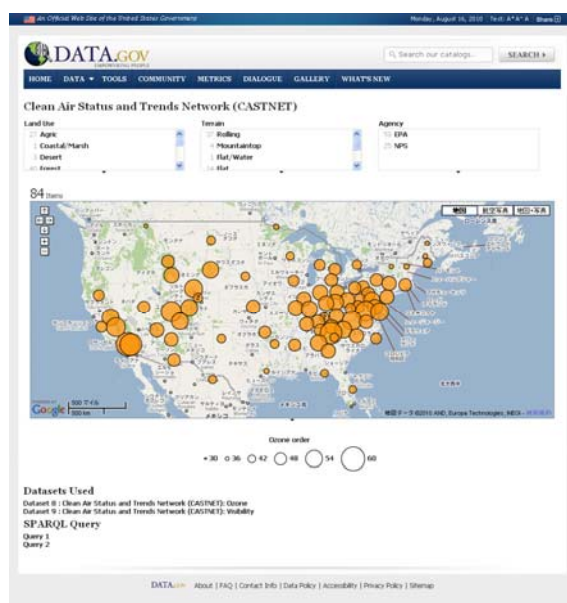
アプリケーションURL <http://www.data.gov/semantic/Castnet/html/exhibit>

データソース

Dataset 8 : Clean Air Status and Trends Network (CASTNET): Ozone

Dataset 9 : Clean Air Status and Trends Network (CASTNET): Visibility

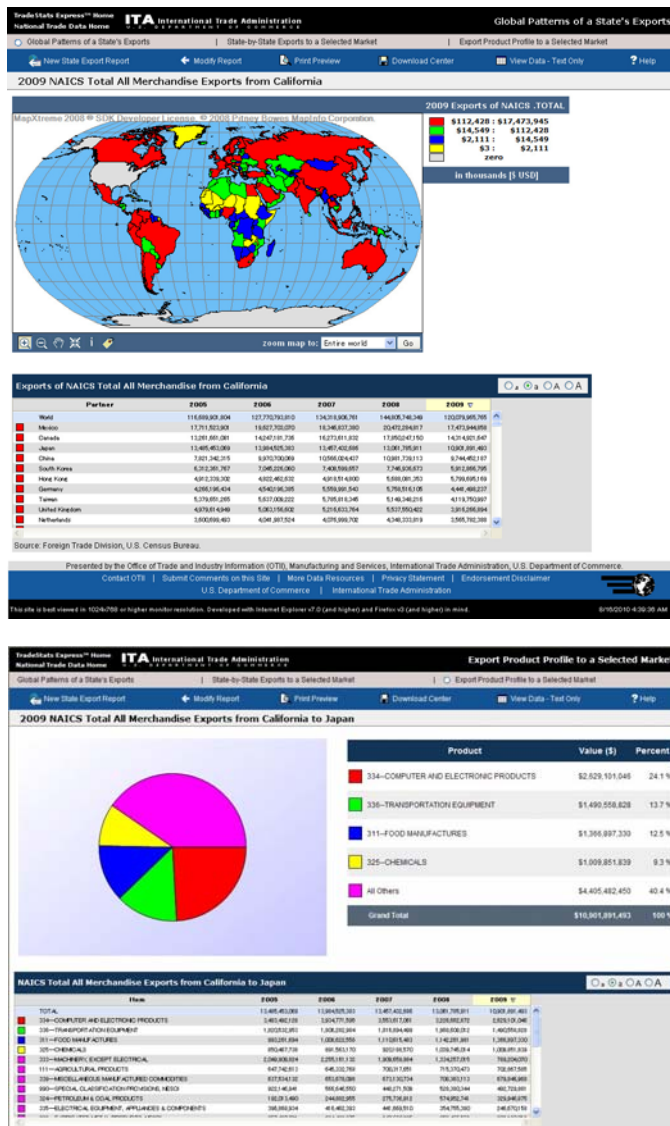
オゾンの濃度(ppb: part per billion)を地図にマッピングした。土地の用途と地形で絞り込みができる。



8. TradeStats Express™ - State Export Data

アプリケーションURL <http://tse.export.gov/TSE/TSEReports.aspx?DATA=SED>

ある州から各国への輸出額、各州からある国への輸出額、ある州からある国へ輸出された製品の3種類の検索ができる。



9. Quick Stats (National Agriculture Statistics Service)

アプリケーションURL <http://quickstats.nass.usda.gov/>

データソース United States Department of Agriculture

Keyword、Commodity(Sector、Group、Commodity)、Location、Time を条件に絞り込む

だ結果を、表示する。

Source	Year	Frequency	Location	State	State Fips	Data Item	Domain	Domain Category	Value
SURVEY	1918	DEC	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	6210000
SURVEY	2002	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	1004902000
SURVEY	1997	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	913770000
SURVEY	1992	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	866110000
SURVEY	1987	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	828360000
SURVEY	1982	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	852072000
SURVEY	1978	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	816360000
SURVEY	1974	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	781971000
SURVEY	2008	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	930708000
SURVEY	1998	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	935880000
SURVEY	1999	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	972210000
SURVEY	2000	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	979648000
SURVEY	2001	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	970121000
SURVEY	2003	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	992876000
SURVEY	2004	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	919919000
SURVEY	2005	YEAR	NATIONAL	US TOTAL	99	ICE CREAM - PRODUCTION, MEASURED IN GALLONS	TOTAL	NOT SPECIFIED	959941000

例：アイスクリーム生産量(単位：ガロン)

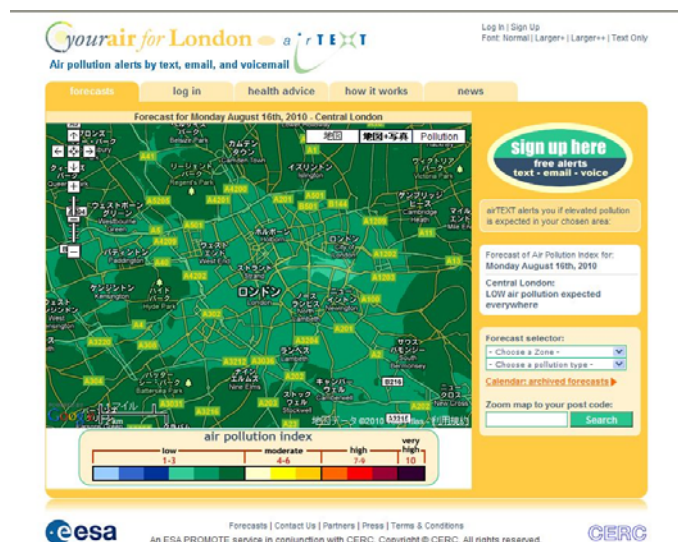
10. 英国 DATA.gov.uk <http://data.gov.uk/>

11. airTEXT

アプリケーションURL：<http://www.airtext.info>

airTEXT はロンドンで住んでいるか、または働いて、喘息、肺気腫、気管支炎、心臓病または狭心症に苦しむ人々のための大気汚染情報サービス。あるゾーンの 10 分の 1 以上で大気汚染レベルが MODERATE か、より高いレベルに達すると予測されるとき、汚染レベルが上がる可能性を警告する無料の SMS メッセージ、ボイスメールまたはメールを該当者に

送信する。 また、サイト上の詳細な大気汚染予測地図を見ることができる。



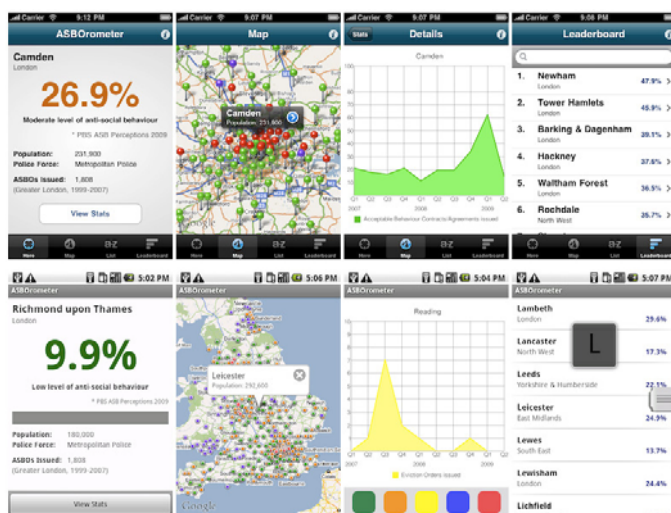
12. ASBOrometer

アプリケーションURL <http://asborometer.com/>

データセット ASBO CDRP survey full dataset Oct 03 to March 09

データソースURL <http://data.gov.uk/dataset/jsjd-asbo-data>

ASBOrometer は、イングランドおよびウェールズ内で現在位置における反社会的行為のレベルを測定して、主要なローカルの ASB 統計へのアクセスを可能にするモバイルアプリケーション。iPhone と Android で利用可能。iTunes App Store や Android Market で無料公開中。



13. Where Does My Waste Go?

アプリケーション URL <http://matt.chadburn.co.uk/projects/waste/authorities/>

各カウンティーが再生利用不可能な家庭廃棄物を埋め立てごみ処理した場所に関する 2004 年から 2009 年までのデータベース。埋立て地デポの地図、推定処理費用、人口 1 人あたりの平均埋立て容量、カウンティー毎にブレイクダウンされているすべての埋立てのインデックスを含む。

Where Does My Waste Go?

A database of where your [county council](#) landfilled your non-recyclable household waste from 2004 to 2009. [What is this?](#)

[Authorities](#) » Devon

Tonnage by year

2005	2006	2007	2008	2009
- 189,538 tonnes - tonnage per head 0.29 £ - cost 46,033,216 £	- 229,659 tonnes - tonnage per head 0.39 £ - cost 47,342,686 £	- 220,261 tonnes - tonnage per head 0.29 £ - cost 47,248,262 £	- 206,476 tonnes - tonnage per head 0.27 £ - cost 46,607,232 £	- 149,451 tonnes - tonnage per head 0.20 £ - cost 44,782,432 £

Landfill sites

In which counties did Devon council dump their waste?

- Cornwall = 15,262
- Devon = 234,411
- Unknown or exempt = 744,512

Locations of the landfill sites.



Returns (47 records)

Non-hazardous returns by Devon council.

- Landfilled 66,468 tonnes in to [Cornwall](#) in 2005
- Landfilled 63,475 tonnes in to [Cornwall](#) in 2005
- Landfilled 58,595 tonnes in to [Cornwall](#) in 2005
- Landfilled 58,523 tonnes in to [Cornwall](#) in 2006
- Landfilled 60,197 tonnes in to [Cornwall](#) in 2006

14. expendituremap

アプリケーション URL <http://ishortman.com/projects/expendituremap/>

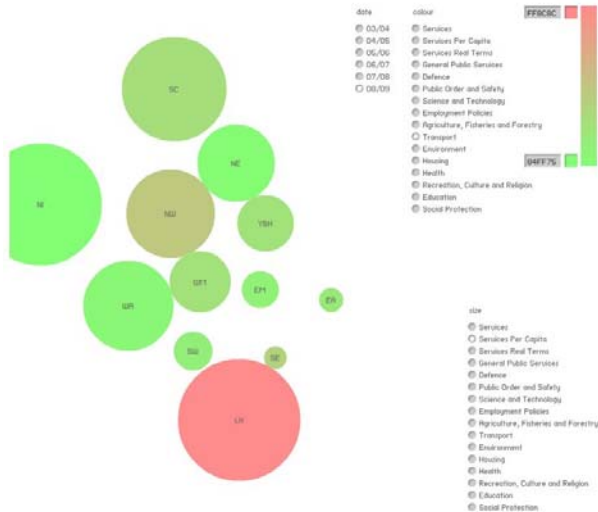
expendituremap はボーダフォン後援の懸賞当選者の 2 件のうちの 1 つ。英国の地域と公費データをビジュアライズする対話的な地図。 サービス、ディフェンス、公安、科学技術、農業、輸送などから 2 つのカテゴリを選択して、各地方を表す円の大きさと色で比較できる。 また、2 つの異なったカテゴリにおける公共の支出の数年間の変化を比較できる。

expendituremap

Rewired State – National Hack the Government Day 2010

View our project page at Rewired State and data.gov.uk.

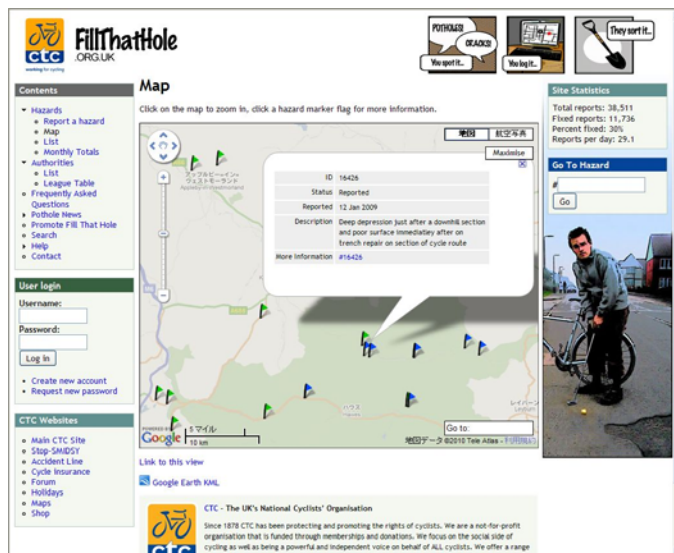
RS



15. Fill That Hole

アプリケーション URL <http://www.fillthathole.org.uk/>

CTC(英国のサイクリストの組織)が主催する、道路の穴を報告するサイト。安全に自転車に乗れるようにするために、危険な道路の穴を報告し、その穴が補修されたかを示す。



16. Find BRB (Residuary) Ltd. property

アプリケーション URL <http://www.adambrookes.co.uk/brbr/property>

1993 年の鉄道法により、旧 British Rail は分割売却された。British Railways Board は、現在は政府所有の会社である BRB (Residuary)株式会社となつて存続している。BRBR は英国、スコットランド、およびウェールズの廃止された鉄道線路に関連している建造物に責任がある。このツールで、郵便番号を入力すると BRBR によって所有されている資産を探すことができる。

adam brookes.co.uk

home tools contact

Find BRB (Residuary) Ltd. property

Under the 1993 Railways Act, the old British Rail was split up and sold off. The remaining functions of the British Railways Board are now discharged by the Government owned company, [BRB \(Residuary\) Ltd.](#)

BRB (Residuary) Ltd. is responsible for structures associated with closed railway lines in England, Scotland and Wales. It also manages the disposal of land and buildings which are deemed by the Department for Transport to be not required to be retained for current or future railway purposes.

Any problems or suggestions regarding this tool? Please [contact me](#).

Search Postcode

L20 3LT

Find

or Ref

Results

- [R6B/6](#)
0.9 mile
- [HUS/42A W](#)
0.9 mile
- [HUS/43E](#)
0.9 mile
- [HUS/43W](#)
0.9 mile
- [HUS/41A W](#)
1 mile
- [HUS/42A E](#)
1 mile
- [HUS/42L](#)
1 mile
- [HUS/42](#)
1 mile
- [R6B/2](#)
1 mile
- [R6B/8](#)
1 mile

HUS/43W

0.9 mile

name: Selwyn Street

type: Tunnel

status: Public Road & Non-BR

Property

router: Huskisson Branch

grid ref: 53551943

View in Google Maps

17. Unofficial National Jobcentre Plus

アプリケーション URL <http://home.zois.co.uk/jcpnational.html>

郵便番号または地域を入力すると、その地域の求人の一覧を表示するサイト。

ZOIS

JCP Office

Hammermith
Glen House, 22 Glasthorne Rd, Hammermith, London, W6 0PP.
020 8210 8500

This is a list of vacancies that we've noted for a single [Jobcentre Plus](#) office. It's the last 48 hours worth, the field added is what we gleaned from the web-site.

We've obtained these vacancies from the [Jobcentre Direct](#) web-site generally at the end of the working day, so at maximum some of these entries will be about a day out of date. You can bookmark this page and subsequent visits will allow you see what's been posted in future 48 hours periods. Alternatively, you can [subscribe](#) to an RSS feed for this office and [request an e-mail feed](#). There's more [explanation](#) elsewhere.

Postings matching code "JLN" made in the last 48 hours.

Vacancy	Location	Added
NIGHT DOMESTIC CARE OFFICER	HAMMERSMITH LONDON, W6 W6	2010-08-17
CARE COORDINATOR	HAMMERSMITH LONDON, W6 W6	2010-08-17
DOMESTIC CARE OFFICER	HAMMERSMITH LONDON, W6 W6	2010-08-17
WETA CARE SERVICE MANAGER	HAMMERSMITH LONDON, W6 W6	2010-08-17
ACCREDITED ADMINISTRATOR	LONDON, W6 W6	2010-08-17
NON STAFF NURSE 100-85-N-00	HAMMERSMITH LONDON, W6 W6	2010-08-17
REHABILITATION ASSISTANT 100-CLCH-012	LONDON, W6 W6	2010-08-17
RAMPSHIRE BAR ASSISTANT RAMPSHIRE MARKET	HAMMERSMITH LONDON W6 W6	2010-08-17
SECURITY GUARD	LONDON W14 W14	2010-08-17
NON STAFF Nurse 100-85-N-01	LONDON, W6 W6	2010-08-16
BAR ATTENDANT	WEST KENSINGTON LONDON W14 W14	2010-08-16
MUSCOP ELECTRIC METER OPERATORS	LONDON W6	2010-08-16
NON STAFF NURSE 100-85-TA1	HAMMERSMITH LONDON, W6 W6	2010-08-16
CARE WORKER	LONDON, W6 W6	2010-07-15

Generated 2010-08-18

Site: jcpoffice.php, v 1.7 2010/07/29 19:19 null van Exp 3

18. Historic Fenlands Mashup

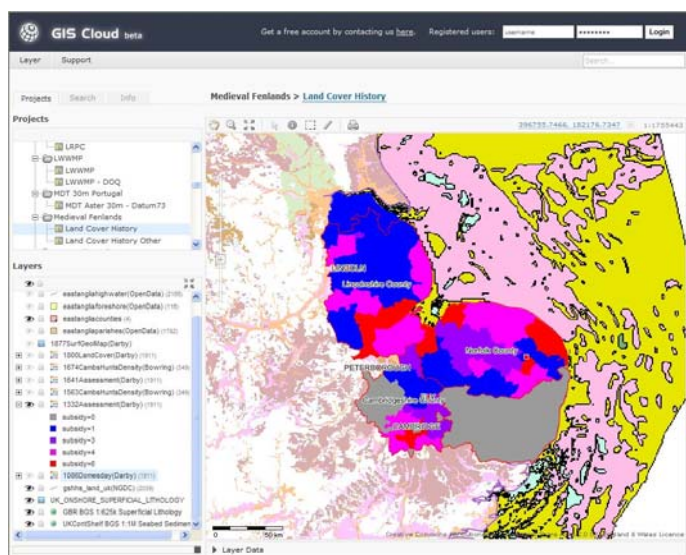
アプリケーション URL

<http://www.giscloud.com/map/3186/medieval-fenlands/land-cover-history>

Domesday(ウィリアム 1 世が行った検地)以来の土地被覆の地理的な歴史と、東アングリアの表層地質学に関するマッシュアップ。

以下のデータセットに基づいている：

- 選択された土地被覆は、中世以降の H.C.ダービーの刊行物より。
- 管理的および地理的な境界は、英国の英国陸地測量部公開データより。
- 地表およびオフショアの地質は、学英地質調査局ウェブマッピングサービスより。
- グローバルな海岸線は、US NGDC GSHHS より。



19. Jobseekers' Allowance claimant data

アプリケーション URL

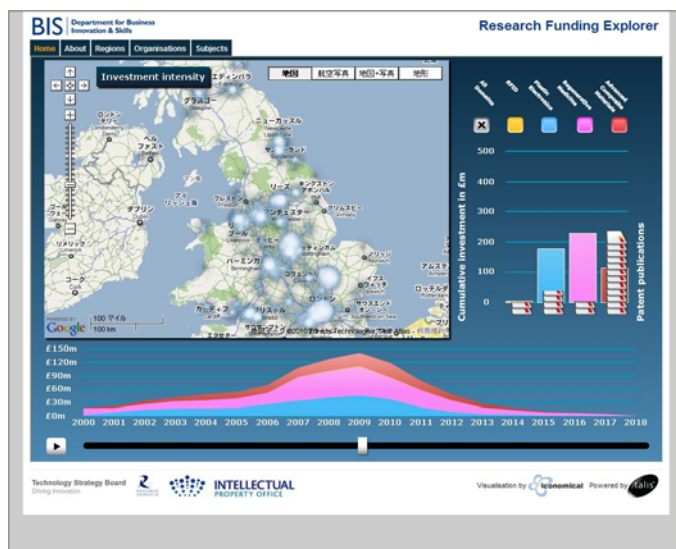
[URL:http://www.wmro.org/noncms/instantatlas/JSA-claimant-count-LAD-2_NW/atlas.html](http://www.wmro.org/noncms/instantatlas/JSA-claimant-count-LAD-2_NW/atlas.html)

データセット

Proportion of the working age population in West Midlands claiming Jobseekers Allowance

ウェストミッドランズにおいて求職者手当(jobseekers allowance)を申請している生産年齢人口の割合を示している対話的な地図。ある地域の月別の申請者の割合(%)と前年からの変化を見ることができる。

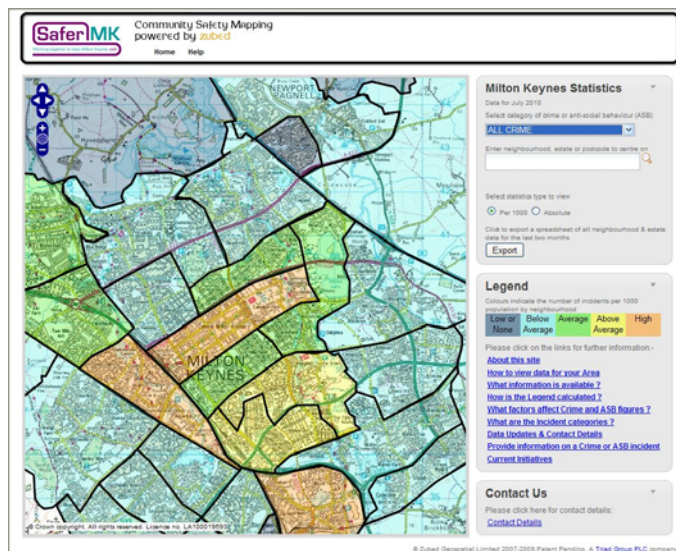
英国が特に投資している 4 つの優秀な技術領域での活動を示す特許と研究基金からリンクされたデータのビジュアル化。Making Public Data Public イニシアチブの一部。



22. SaferMK

アプリケーション URL <http://safermk.zubed.com/>

犯罪および ASB データを見ることができる犯罪減少アプリケーション。犯罪のカテゴリ別に地区のデータが色分けされている。

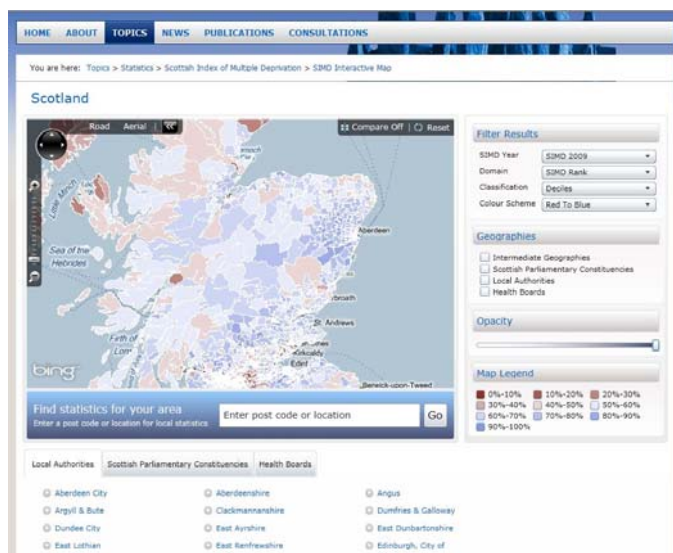


23. Scottish Index of Multiple Deprivation

アプリケーション URL <http://simd.scotland.gov.uk/map>

スコットランド政府関連の統計を集めたポータルサイト。

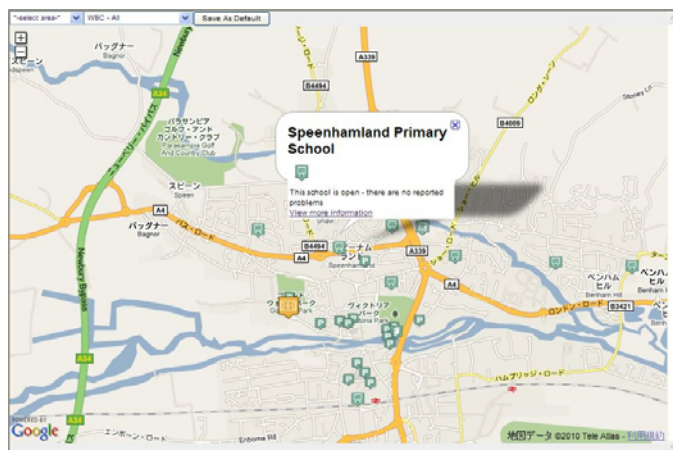
例：SIMD(スコットランドの複合減退指標：各地域の相対的豊かさを、データに基づき数値化した指数)マップ



24. West Berkshire Service Status Map

アプリケーションURL http://www.westberks.gov.uk/xml/service_status.html

西バークシャーのケアホーム、図書館、レジャーセンター、学校、廃棄物サービス場所がわかる対話的な Google Map。

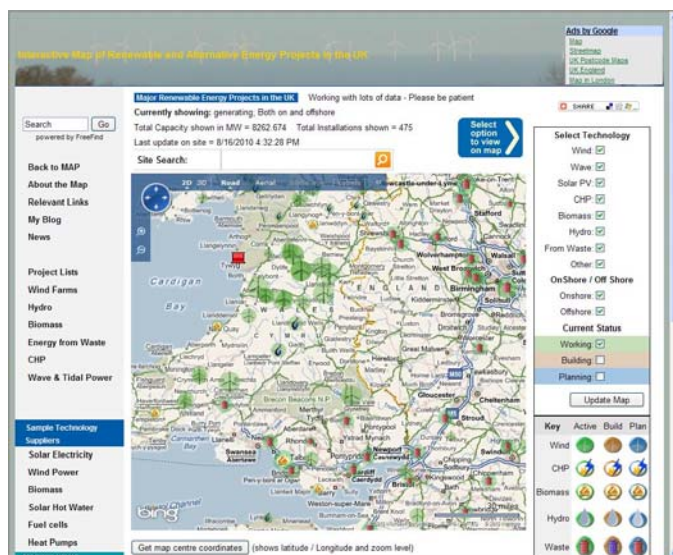


25. Renewable Energy Map

アプリケーション URL <http://www.renewables-map.co.uk>

データセット Renewables map

英国の主な再生可能エネルギーと代替エネルギープロジェクトの一覧。風力、地熱、波力、バイオマス、熱電併給(CHP)、PV(太陽光発電)を含む 1MW 以上のすべてのプロジェクトと、1MW 未満の多くの革新的なプロジェクトが対象になっている。



26 . ニューヨーク・アプリ・コンテスト

<http://www.nycbigapps.com/>

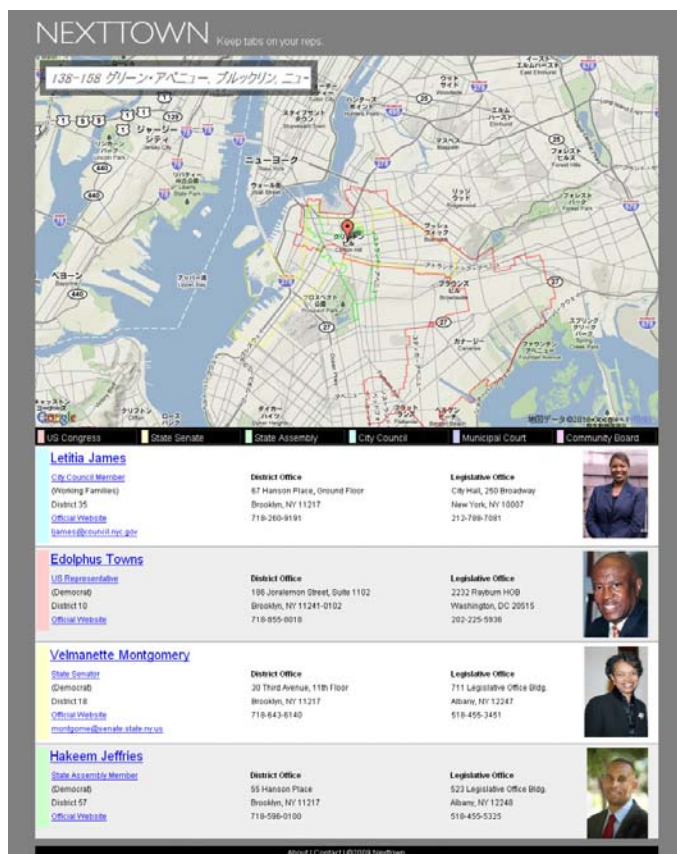
27. Nexttown

アプリケーション URL <http://www.nexttown.org>

データセット

GIS: State Assembly Districts, GIS: City Council Districts, GIS: Congressional Districts, GIS: State Senate Districts, GIS: Election Districts, GIS: Municipal Court Districts

Nexttown は、代議士の名前とコミュニティに対して何をしているかを一瞬で理解できるサイト。また、町の政治的な傾向を見やすく表示する。

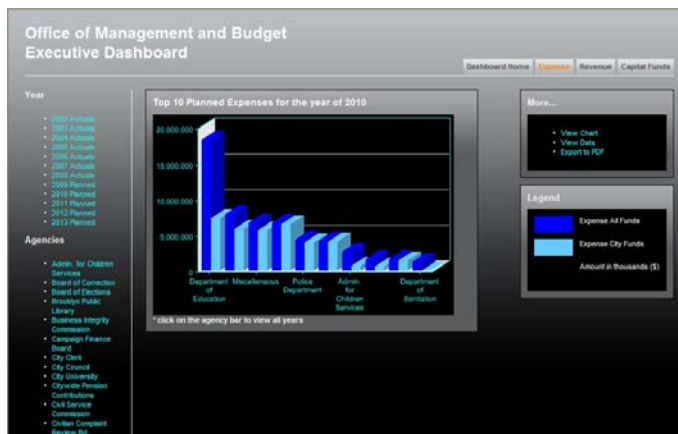


28. OMB Executive Dashboard

アプリケーション URL <http://www.cfcrow.com/omb/>

データセット Financial plan statements

ニューヨーク市の 80 以上の市の機関に対する収入と支出をグラフィカルに表示する OMB(Office of Management and Budget ; OMB、行政管理予算局)のサイト。OMB は、市の財政状況と市のサービスとプログラムの効率に影響する問題において、市長の Preliminary Budget および Executive Budget を準備すること、および市の財政的な健全性と市のサービスやプログラムの効率に影響する問題について、市長にアドバイスする責任がある。

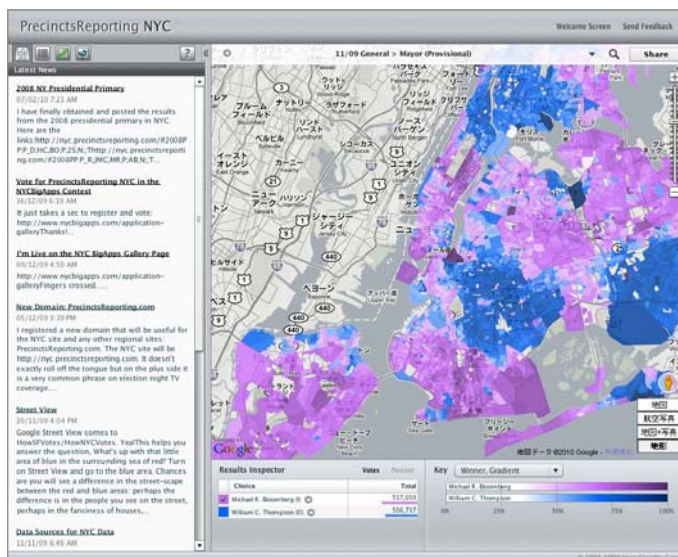


29. PrecinctsReporting NYC

アプリケーション URL <http://nyc.precinctsreporting.com>

データセット Borough, Election District and Assembly District GIS boundary data

PrecinctsReporting NYC は、ニューヨーク市のブロックごとに歴史的な選挙結果をマップする対話的なツール。2009 年の市長選を含む 2005 年以來のほとんどの市と州の選挙が対象となっている。



30. Eat Clean.ly

アプリケーション URL <http://clean.ly>

データセット Restaurant inspection results

Eat Clean.ly は、位置情報、料理の種類、NY 市の検査結果に基づいてレストランを表示する。検査結果が高得点のレストランは緑色、平均的な得点のレストランは黄色、65 点以下

は赤で表示される。

Clean.ly

Eat Cleanly gives you easy access to restaurant inspections around you. Try it now by entering a restaurant name, category, or your neighborhood or address in the box to the right.

Currently, Eat Cleanly is only available in the New York City area with plans to expand to other US areas in the near future, so check back soon if we're not yet in your neighborhood.

Search

Address

GO

Recently inspected in New York

Rating	Restaurant Name	Address
66	PIPERS KILT	4946 Broadway
81	POLISH NATIONAL HOME	261 Driggs Ave
61	SANSOOGAPSAN	171-10 Northern Blvd
64	Wardella's Italian Cuisine	2016 Victory Blvd
48	MOHITO'S	227 E 116th St
81	MUSSO'S CAPTAIN TABLE	860 2nd Ave
66	KELLY'S PUB	149-11 41st Ave
91	Teriyaki Boy Incorporated	218 E 10th St
53	ESTRELLA BAR & RESTAURANT	1470 Jerome Ave
44	Ariana's Catering	106 New Deep Plaza
100	GAF-D-CORP	180 Varick St
100	LAUT	15 E 17th St

Travel in Japan with Visa
Exclusive campaign for cardholders Check privileges & information now
[www.visa-nova.jp](#)

Popular places to eat in New York

Manhattan

Neighborhood	Neighborhood
Chelsea	Chinatown
Columbus Circle	Downtown Manhattan
Flatiron	Gramercy
Inwood	Little Italy
Lower East Side	Midtown
SoHo	Times Square
TriBeCa	Union Square
Union Square	Upper East Side
Upper Manhattan	Upper West Side
Washington Heights	

Brooklyn

Neighborhood	Neighborhood
Brooklyn Heights	Carroll Gardens
DUMBO	Greenpoint
Park Slope	Williamsburg

Queens

Neighborhood	Neighborhood
Flushing	Forest Hills
Jackson Heights	

The City of New York can not vouch for the accuracy or completeness of data provided by this web site or application or for the usefulness or integrity of the web site

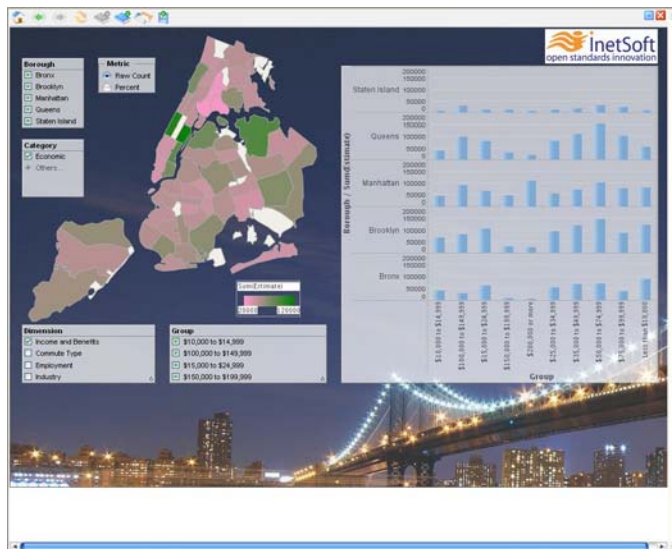
31. Analysis of District Data on an Interactive Map

アプリケーション URL <http://visualizefree.com/public.jsp?id=sYixthhp>

データセット

DCP - Demographic, Social, Economic, and Housing Profiles by Community District/PUMA

ニューヨーク市地区に関するデータを地図上で見る簡単で対話的なサイト。



32. NYC Health and Demographics

アプリケーションURL <http://maps.webfoot.com/demos/nyhealth/NYHealth.html>

データセット Community Health Survey

2000 年の米国国勢調査からの人口学的情報と、Community Health Survey からの結果を地図上で比べることができる。Google Map に色分けして表示される。健康レイヤー(例：「現在喫煙する人口のパーセント」)と、人口学的レイヤー(例：「平均収入」)の間で切り替え可能。

NYC Health and Demographics

by Webfoot Maps

On this map, you can compare or overlay information from the NYC Community Health Survey with demographic information from the US Census Bureau

Use the controls below to switch Health, Demographic, or Dots layers, and then to modify colours and opacity.

► How the area colour controls work

Find address

Health Layers:

% of population with high blood pressure ever

▼ Hide controls

Min Max

mm Colour mm000

17 Value 36

Mapping linear

<- transparent | opaque ->

▼ Hide data information

Percent of population who have ever had high blood pressure; data from New York City Department of Health and Mental Hygiene (2007). See [here](#).

Demographic Layers:



その他のウェブサイト

33. Open Government Data Initiative (OGDI)

URL <http://ogdisdk.cloudapp.net/>

オープンガバメントデータイニシアティブ(OGDI)は、マイクロソフトの Public Sector Developer Evangelism チームが主導している。OGDI は、政府機関から公開されるさまざまな公共データをより簡単に公表し利用するために Windows Azure プラットフォームを使用している。OGDI は使いやすくオープンな API でインターネット上にデータをウェブフレンドリーな形式で公表するために使用できるコードを含む、無料でオープンソースな'スターターキット'でもある。Silverlight、Flash、JavaScript、PHP、Python、Ruby、マッピング Web サイトなどのさまざまなクライアント技術から OGDI ベースの Web API にアクセスできる。

OGDI のサイトでは、検索したデータを一覧表形式、地図、棒グラフ、円グラフで表示できる。

以下のカテゴリに分類されている。

Census、Demographics、Economic Development、Education、Environment、FIPS、Government Operations、Health Care、Historical Outlook、Human Services、Infrastructure、Other、Per Diem Rates、Public Safety、Unemployment

以下のデータソースを対象としている。

Bureau of Labor Statistics、District of Columbia、General Services Administration、New America Foundation、U.S. Geological Survey

OGDI のアプリケーションの例

Unemployment Rates by County per Month from Jan 2009 to Feb 2010

<http://ogdisdk.cloudapp.net/DataBrowser/bls/UnEmpCountyMonth#param=NOFILTER--DataView--Results>



OPEN GOVERNMENT

data initiative

EXAMPLE

Home

Data

Developers

Unemployment Rates by County per Month from Jan 2009 to Feb 2010

Unemployment Rate by County per Month from Jan 2009 to Feb 2009

GET QUERY

DATE

RATING

VIEWS

Unemployment

4/25/2010

1

1

6155

BASE QUERY

<http://ogdi.cloudapp.net/v1/bits/UnEmpCountyMonth/>

FILTER EXPRESSION

Query Builder [Test Query](#)

+

+

enttyid (Gnd)

is equal to

AND

Run

Clear

Data View

Bar Chart

Pie Chart

Full query URL: <http://ogdi.cloudapp.net/v1/bits/UnEmpCountyMonth/>
(Click to view results as XML/Atom)

DOWNLOAD as: CSV

enttyid	id	lps_code	lps_state	lps_county	county_name	state	wherew	year	month	cliffes	labov	employed	total unemployed	total unemp
1	01001	01	001	Autauga County	AL			2009	1	24160	22279	1881	7.8	7.8
12873	01001	01	001	Autauga County	AL			2009	5	24056	22038	2018	8.4	8.4
16091	01001	01	001	Autauga County	AL			2009	6	24249	21946	2303	9.5	9.5
19309	01001	01	001	Autauga County	AL			2009	7	24507	22252	2255	9.4	9.4
22527	01001	01	001	Autauga County	AL			2009	8	24204	21987	2217	9.6	9.6
25745	01001	01	001	Autauga County	AL			2009	9	23695	21393	2302	9.7	9.7
28963	01001	01	001	Autauga County	AL			2009	10	23633	21448	2185	9.2	9.2
32181	01001	01	001	Autauga County	AL			2009	11	23519	21368	2151	9.1	9.1
3219	01001	01	001	Autauga County	AL			2009	2	24234	22148	2086	8.6	8.6
35399	01001	01	001	Autauga County	AL			2009	12	23360	21142	2218	9.5	9.5
38617	01001	01	001	Autauga County	AL			2010	1	23252	20890	2362	10.2	10.2
41836	01001	01	001	Autauga County	AL			2010	2	23115	20934	2181	9.4	9.4
6437	01001	01	001	Autauga County	AL			2009	3	24125	22073	2052	8.5	8.5
9655	01001	01	001	Autauga County	AL			2009	4	24129	22233	1896	7.9	7.9
12874	01003	01	003	Baldwin County	AL			2009	5	82304	75799	6605	7.9	7.9
16092	01003	01	003	Baldwin County	AL			2009	6	83763	76413	7350	8.8	8.8
19310	01003	01	003	Baldwin County	AL			2009	7	83709	76202	7507	9	9
2	01003	01	003	Baldwin County	AL			2009	1	79425	73435	5990	7.5	7.5

34. West Midlands Regional Observatory

URL <http://www.wmro.org/standardTemplate.aspx/Home/Resources>

ウェストミッドランズ地方のデータ、情報、インテリジェンスのポータルサイト。Resource Catalogue、Interactive maps、Local Economic Assessment resource guide、Guidelines for handling geographic data、Web service API 等のページが公開されている。

West Midlands Regional Observatory

This site provides a complete single point of access to data, information and intelligence about the West Midlands region.

Home | Contact Us | Research & Data Enquiries | Blog

Log in | Register with us

Enter search term: Go

Help with search

Resources

Summary Subject

Access to reliable information is essential for good decision making.

Helping decision makers in the West Midlands access reliable information is at the heart of our role. We work closely with organisations and individuals to identify and catalogue data and evidence related to the West Midlands, building a comprehensive, publicly-accessible online collection of West Midlands data and evidence.

Resource Catalogue

The Resource Catalogue is a directory on this website categorising reports, datasets, websites and other information resources published by a wide range of organisations.

The catalogue is searchable and you can filter search results by subject, date, type and geographical coverage.

We encourage you to [submit your resources](#) to the Catalogue. You'll need to be registered on this website ([it's quick and free to register](#)) to do so.

Interactive maps

Explore a variety of datasets using interactive maps. The geographical display of data helps to understand how separate issues affect the West Midlands region and local areas.

Recent resources

- Dataset: worklessness in the West Midlands: impact of demographics and multiple risk factors 17 August 2010
- Centro Annual Statistical Report 2009-2009 17 August 2010
- Weathering the storm - saving and making money in a changing climate 12 August 2010

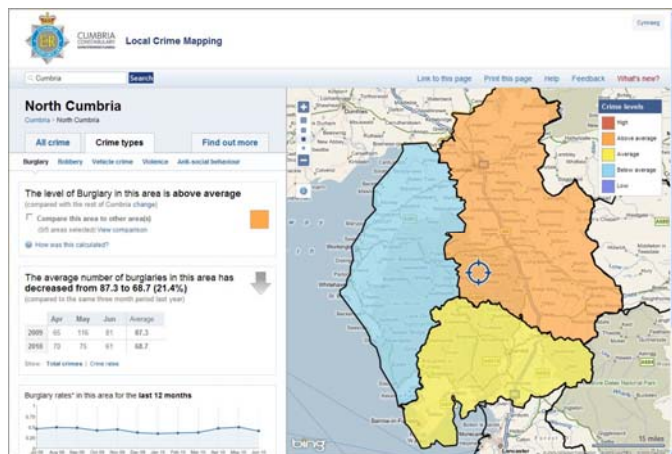
Search all resources

Keeps up to date

35. national crime map

アプリケーションURL <http://maps.police.uk/>

地域別の犯罪マップ。



36. UK Air Quality Archive

アプリケーションURL <http://www.airquality.co.uk/>

英国各地の大気汚染の情報ポータルサイト。各地の詳細情報へのリンクがある。



附録 C. SDMX 関連情報の概要

SDMX 関連情報の概要

1. 本文書について

1.1. 本文書の概要

本文書は「オープンガバメントを実現するシステムの可能性等についての調査研究 統計情報提供基盤（通称：データボックス）の機能評価・検証について」の補足資料として、SDMX 関連情報の概要を提供することを目的としている。本資料は、SDMX 仕様に関連するものについては、SDMX User Guide の抜粋抄訳のほか、SDMX.org サイトより入手可能な各種マニュアル、レポート、プレゼンテーション資料等を加味して作成している。

2. SDMX の背景

2.1. SDMX の概要

SDMX とは、BIS, IMF, OECD, 世銀, UN, Eurostat, 欧州中央銀行により発足した “Statistical Data and Metadata Exchange”（統計データおよびメタデータ交換）という取組みにより制定された、統計データおよびメタデータ交換の標準仕様を策定したものである。これは国際機関および所属各国において統計データならびにメタデータを効果的に交換・共有するためのプロセスを開発・利用することが主眼に置かれているものであって、このため、SDMX にはデータおよびメタデータの標準フォーマット、コンテンツのガイドライン、並びに I T アーキテクチャが包含されている。

2.2. SDMX が必要とされる背景

SDMX のような統一的な統計情報フォーマットが策定されるにあたっては、以下のような背景が認識されている。

- ・ インターネットの普及により、データの交換・共有は容易になっているにもかかわらず、フォーマット等は様々な種類が存在すること。統計データおよびメタデータの交換は複雑かつコストがかかるものであると考えられていること。
- ・ 各国際機関等によりアドホックなアプローチによりデータの交換・共有が行われている現状があるが、それらは一般的な問題解決となっているわけではない。
- ・ XML やウェブサービスのような新しいテクノロジーの潮流にたいして、問題解決のテクノロジーが必要とされてきていること。

また、平成21年度報告の「統計情報のXML利用環境の整備に関する調査研究に係る報告書」においても、XML データの整備に当たっては、下記の点に留意し重点的に取り組むとして

- ・ 統計データに付随するメタデータまでを含めた統計情報の活用を検討する。
また、その際にはISO など国際規格とも整合性の取れたデータ仕様、タグ仕様等規約類の検討・整備を行う。
- ・ 統計表の表章の整備を行い、統計データの利活用の促進を図る。
なお、表章の定義が容易に分るように整備することで、統計表を並べて分析する等、統計データの利活用の一層の促進が図られる。

という提言がなされているほか、統計 XML フォーマット策定に向けての検討項目においては、「デジタルデータの円滑な情報交換には、共通のデータフォーマットの利用が不可欠である。そのため、ターゲットとする分野に広く流通するフォーマットを採用することが肝要である。統計分野では先に上げた例のように OECD が策定する SDMX(<http://sdmx.org>) が欧州を中心に利用が広がっている。SDMX を統計情報基盤のデータとして採用できないか検討することが必要である」とされていることを受け、今プロジェクトでは SDMX に関する検討を行うものである。

2.3. 現在の SDMX 採用状況

SDMX による統計情報の公開は世界的には以下に示されるような状況にある。

- ・ 国連統計委員会では SDMX を統計情報の交換書式として推奨しており、SDMX は「国際的統計関連コミュニティにおけるデータおよびメタデータの交換および共有のための推奨標準」と認識されている。
- ・ 40 以上の機関が採用、もしくは採用を検討しており、より多くの国および国際機関の参加が望まれている。

2.4. SDMX の適用範囲

SDMX は以下の状況において適用が可能であるとされている。

- ・ SDMX 標準は 2 つ以上の統計情報を扱う組織でデータの交換・共有を行うことを目的としており、例えば、国家統計局、中央銀行、行政府などが挙げられる。これらの組織団体内外がデータおよびメタデータの交換・共有を行うときに、この標準規格を適用することは、統計情報がもつとも効率的に報告および共有が可能となりうる。
- ・ SDMX 標準は国内のシステムだけでなく、関係組織が複数にわたる場合や、国際的な統計データ共有・交換のために用いることも可能である。このような各組織が密接にかかわる場合に効果を上げることが期待されている。
- ・ SDMX 標準をデータ公開の標準仕様として採用することが可能である。これはデータおよびメタデータが構造的であることと、それが再利用可能な形

式で定義されているからであって、国際組織並びに国内組織で採用を検討すべきものである。

- **SDMX** を組織内での標準フォーマットとすることも可能である。これにより複数のツールを導入したり、複数のレイアウトを策定したりすることによるコストを避けることができるし、**SDMX** の情報モデルは統計情報を処理する組織での業務に適用することのメリットを享受しうる。

3. SDMX の概要

3.1. SDMX の構成要素

SDMX は単にデータの交換形式を定義するものではなく、データの利活用のための枠組みをも提供している。このため、構成要素としては以下に示されるように、情報モデルだけではなく、コンテンツ指向ガイドラインやデータ交換のための IT アーキテクチャも定義されている。

- **SDMX** 情報モデル
- コンテンツ指向ガイドライン
- データおよびメタデータ交換のための標準フォーマット
- データ交換のための IT アーキテクチャ
- **SDMX** ツール

3.2. SDMX 情報データモデル

SDMX は、統計データ、メタデータ、およびデータ交換プロセスをモデリングするための方法論を提供する。データおよび関連メタデータは「データ構造定義 (**Data Structure Definition**) (以下 **DSD** またはキーファミリーという)」により構造化される。**DSD** は「構造メッセージ」において、「国名」「年度」といった一連の次元 (ディメンション) と、「計測単位」「倍率」のような一連のデータ属性 (個別データに関するメタデータ) と、関連するコードリストにより、統計データの構造の定義を行う。

また、「参照メタデータ」において、自然言語で記述される当該データに関する内容説明、方法論、品質などに関する追加説明を付与するメタデータも定義に含めることができる。

3.3. コンテンツ指向ガイドライン

コンテンツ指向ガイドラインとは、特定の統計ドメインもしくは機関に依存せず、一般的となることを意図している、データの分類および概説のための推奨ガイドラインである。以下の 4 つのガイドによりこれは構成されている。

- ドメイン横断概念 (**Cross-Domain Concepts**)。異なる統計ドメインでデー

タを利用する場合に適用される、当該 DSD およびメタデータ構造（以下、MSD）で使用される共通概念の識別子を提供。

- ・ ドメイン横断コードリスト（Cross-Domain Code Lists）。異なる統計ドメインでデータを提供する場合に適用される、コードリストの集合を提供。
- ・ 統計主題ドメイン。UNECE による国際統計活動分類をベースとした統計ドメインの一覧。
- ・ メタデータ共通語彙（Metadata Common Vocabulary, MCV）。各統計機関による統計の語彙定義およびプロセスの定義。

3.4. データおよびメタデータ交換のための標準フォーマット

SDMX 標準では、XML をベースとしたデータ交換フォーマット（SDMX-ML）および EDIFACT(SDMX-EDI)を定めている。

3.5. データ交換のための IT アーキテクチャ

SDMX によるデータ交換は大きく分類して、PUSH 型、PULL 型、HUB 型をサポートしている。PUSH 型とはデータ提供者の組織から利用者の組織に向かってデータが転送されるモードであり、PULL 型とは利用者側の組織がデータ提供者の組織からデータを取得するモードである。HUB 型とは、中央の HUB が他のデータソースにクエリを行ったうえで必要なデータを収集するアーキテクチャとなっている。

また、SDMX による IT アーキテクチャは SDMX レジストリと呼ばれる、メタデータのレポジトリに関する仕様を定義している。これにより、データの必要が生じたときに、レジストリに対して要求を満たすデータやデータ提供者を発見することが可能となっており、データおよびメタデータの交換・共有を効果的に行うことを可能としている。

3.6. SDMX ツール

SDMX をサポートする前記各機関並びに他の組織により、SDMX 形式のデータを処理するためのツールが開発されている。

4. SDMX の利点

SDMX を採用するに当たっては、技術的要因、外部的要因を含め、以下のようなメリットがあると考えられる。

4.1. SDMX の技術的メリット

- ・ レポートに伴うデータ処理等にかかる負荷の軽減を図ること。
- ・ データおよびメタデータの整合性促進を図ること。

- ・ 統計処理の効率向上促進を図ること。
- ・ データ配布フォーマットの標準化を図ること。
- ・ データおよびメタデータの比較可能化の促進を図ること。

4.2. SDMX 採用に伴う外部的メリット

- ・ オープンソース標準・オープンソースツールの利用が可能であること。
- ・ トレーニング等の標準化が可能であること。
- ・ コンサルティング等のプロセスの標準化が可能であること。

5. SDMX のデータフォーマット

5.1. データフォーマットの考え方

SDMX のデータフォーマットは以下のような考えに基づいている。

例えば、CSVファイルであれば”100”, ”12345”, ”2469118”のように単純な値の列挙だが、単にこのような数値では何を表すのかが自明ではないので、以下をまず定義する必要があると考える。

- ・ 測定対象は何か → 例えば「総人口」
- ・ 測定の単位は何か → 例えば「1000人」
- ・ 対象となる国・地域は何か → 例えば「ABC国」
- ・ 測定がなされたのはいつか → “2001年1月1日時点”

	Exports and imports of United Kingdom per quarter (not seasonally adjusted, in millions of Euros)			
	Exports to / imports from Germany		Exports to / imports from United States	
	Exports	Imports	Exports	Imports
	Q:GB:N:2:100:DE	Q:GB:N:3:100:DE	Q:GB:N:2:100:US	Q:GB:N:3:100:US
2008q1	12,450	13,660	14,328	15,156
2008q2	12,480	13,670	14,330	15,150
2008q3	12,520	13,710	14,324	15,163

出典：SDMX User Guide

上記の表を例にすると、各セルの数値には下記のような属性があることが分かる。

- ・ 頻度 (M = monthly, Q=quarterly, A = annual)
- ・ 対象国 (GB = 英, DE = 独, US = 米)
- ・ 補正 (N = 季節補正なし, S = 季節補正あり)
- ・ 取引方向・タイプ (2=物品輸出, 3=物品輸入, 4=全体)
- ・ 対象/変数 (100= 物品, 200=サービス, 300=収入)

- ・ 相手国(GB = 英, DE = 独, US = 米)
- ・ 単位 (EUR = ユーロ)
- ・ 位取り(6 = 1 0 0 万)

このようなコード表を含めたデータの属性をメタデータで定義する形式をSDMXでは採用している。以下はメタデータで定義されうるコード表の例示であるが、具体的なサンプルに関しては、SDMXウェブサイト (<http://SDMX.org>) より取得することが可能となっている。

Position in key	Dimension/attribute name	Identifier	Presentation ³	Attachment level	Code list
1	Frequency	FREQ	A1		CL_FREQ
2	Reporting/reference area	REF_AREA	A2		CL_AREA
3	Adjustment	ADJUSTMENT	A1		CL_ADJUSTMENT
4	Data type for balance of payments statistics	DATA_TYPE_BOP	A1		CL_DATA_TYPE_BOP
5	Balance of payments topic	BOP_ITEM	A3		CL_BOP_ITEM
6	Vis-à-vis area	COUNT_AREA	A2		CL_AREA
	Unit of measure	UNIT_MEASURE	AN3	Data set	CL_UNIT_MEASURE
	Unit multiplier	UNIT_MULT	N1	Data set	CL_UNIT_MULT
	Compiling agency	CL_COMPILING_ORG	AN3	Sibling Group	CL_ORGANISATION

出典：SDMX User Guide

なお、サンプルは EUROSTAT で使用されている SDMX のサンプルであって、以下のファイルを含んでいる。

- 構造定義メッセージ
 - データ構造の定義
 - <EUROSTAT_STS.xml>.
- 一般データメッセージ
 - DSD に相当するデータ構造を内包した、データメッセージ
 - <EUROSTAT_STS_STS_IND_PROD_M_BE_generic.xml>
- コンパクトデータメッセージ
 - DSD に依存した構造のデータメッセージ
 - <EUROSTAT_STS_STS_IND_PROD_M_BE_compact.xml>
- クロスセクションデータメッセージ
 - <CrossSectionalSample.xml>
- クエリーメッセージ
 - クエリ情報

➤ < QuerySample.xml>

5.2. SDMX データモデル

SDMX のデータモデルとは簡単に以下のようにまとめられる。

5.2.1. 識別子概念

統計データに対して定義される概念（例えば、1.2953 という数値が、2006 年 11 月 23 日の USD・EURO 交換レートであるということ）。

5.2.2. データパッケージ

統計データをグループ化して扱うためのレベルであって、次の 4 つが定義されている：観測レベル、シリーズレベル、グループレベル、データセット／データフローレベル。観測レベルはある特定の現象に対する観測値を示し、シリーズレベルは一般的には定期的な間隔で観測される、ある特定の現象に関する観察を示し、グループレベルは例えば計測間隔が異なる以外は同一（例えば、四半期単位と年単位）のような、一連のシリーズデータの複合体を意味する。データセット／データフローレベルとは、例えばある統計ドメインをカバーするために、複数のグループより構成されるものである。

5.2.3. 次元と属性

ある特定のデータを同定および記述するものが次元（ディメンション）と呼ばれ、記述のみを行う識別子を属性（アトリビュート）と呼ぶ。

5.2.4. キー

キーとは上記の「次元」をシリーズ内で特定したり、グループ化したりする。キーの値はシリーズレベルに関連付けられる。一般的には計測間隔／頻度が第一の識別子概念となるものである。

5.2.5. コードリスト

ある次元が取り得る値をリスト化したものをコードリストと呼ぶ。このリストは言語に依存しない「コード」と、言語ごとの説明子により構成される。データの属性はコードによる場合もあるし、テキストによる説明の場合もある。

5.2.6. DSD

DSD は「キーファミリー」とも呼ばれ、あるデータセットに関する記述と識別を行う概念の集合である。DSD ではどの概念が次元であるか、もしくは属性であるかが定義されている。また、それぞれの概念が上記のデータパッケージ構造の内、どのレベルに属するか、あるいは必須か否かの定義を行う。コードリストは DSD の一部となっている。

5.3. コードリストの検討について

実際の業務への適用にあたっては、特にコードリストの定義に注意を払う必要があり、以下の項目も併せて検討することが推奨されるものである。

- **SDMX 技術標準（2.0）** では階層コードリストがサポートされている。
- データ構造定義およびメタデータ構造定義にはデータフォーマット内で使用されるコードリストが含まれている。既存の統計アプリケーションで使用しているコードリストが存在するときは、このコードリストをデータ構造定義およびメタデータ構造定義にマイグレートすることも可能である。
- コードリストの検討にあたっては既存のコードリストの構造を十分に吟味することが必要であり、単純にコードリストを抜き出すことですむ場合もあれば、詳細な構造分析やコードのクリーンアップが必要な場合もありうる。
- 同一の組織内であっても、複数のコードリストが、複数データセットに対して各々異なるコードリストを用いてほぼ同一のオペレーションを適用する状況が発生するような場合、**SDMX 技術標準**では、二つのコードリストの関係を記述するためのフォーマットを用いて、データを相互に交換するだけでなく、それら関連するコードリストを交換することも可能である。ただし、それぞれのコードリストのコード同一性の検査は必要である。

5.4. SDMX の多言語対応について

SDMX は、多言語対応となっている点に留意が必要である。

- **SDMX-ML**（XML による **SDMX** データ交換フォーマット）は、複数の言語をサポートしている。
- **DSD**（データ構造定義）においては、多言語表現はコードの定義に現れる場合が多く、コードに対するラベルは各言語で付与することができる。
- 文字コードは **UTF-8** が想定されている。

6. SDMX のメタデータ

6.1. メタデータの分類

SDMX におけるメタデータは、大きく二つ分類される。

- データ構造の一部である構造メッセージ定義
- データ構造にかかわらない、参照メタデータ定義
 - 概念メタデータ（概念の記述を行う）
 - 方法論メタデータ（当該データの生成に適用された方法論の記述）
 - 品質メタデータ（各統計データの品質に関する記述）

6.2. 参照メタデータ

この参照メタデータ定義は「発行元」「更新日付」など、各メタデータに関する構造を伴う「メタデータレポート」として構成することが可能となっている。メタデータレポートは統計データおよびメタデータに関連するほぼあらゆる事柄についての情報を提供することが可能である（例えば、データ公開の頻度や、データの粒度など）。

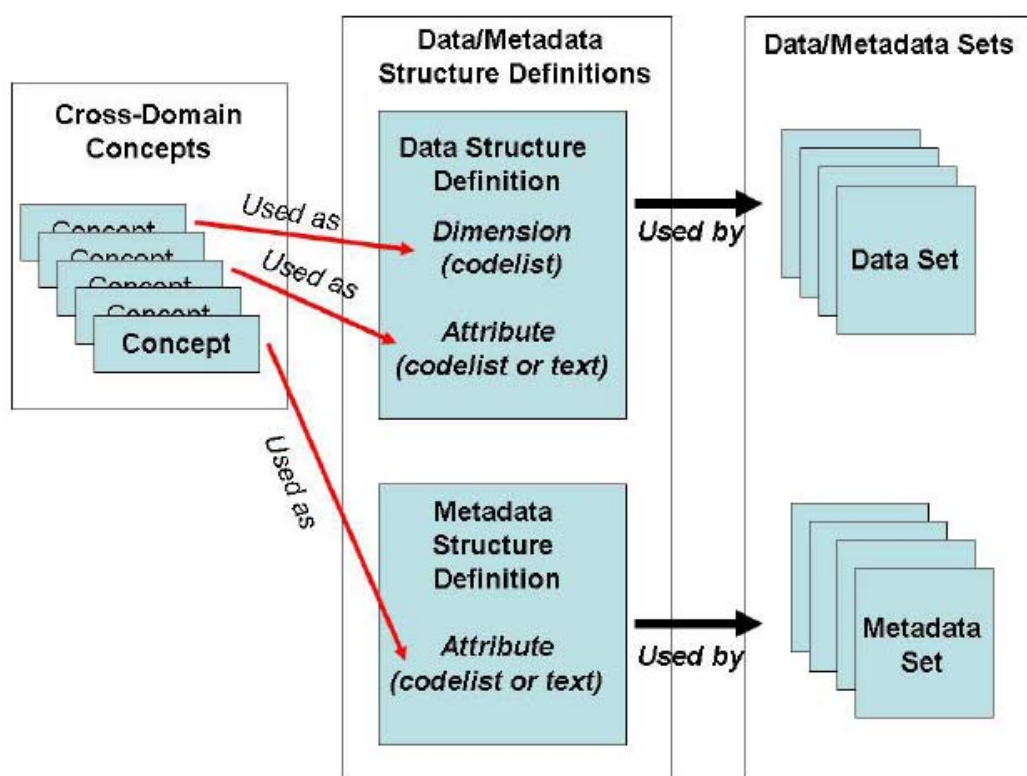
SDMX Reference Metadata Report (partial)	
Country: Bulgaria	
Domain name: Industrial Production Index	
1. Contact	
1.1 Contact organisation	National Statistical Institute
1.2 Contact organisation unit	Industrial Statistics Division, Department of Business Statistics
1.3 Contact name	Mrs. Guergana Maeva
1.4 Contact person function	State expert
2. Metadata update	
2.1 Metadata last certified	2008-04-04
2.2 Metadata last posted	2008-04-04
2.3 Metadata last update	2008-04-04
3. Statistical presentation	
3.1 Data description	The index measures the monthly change in the value of production, covering mining, manufacturing, electricity, water and gas supply.
3.2 Classification system	Classification of Economic Activities (NACE.BG), which complies with NACE Rev.1.1, is used for the classification of statistical units to the 4-digit level.
3.3 Sector coverage	The Industrial production index covers the mining and manufacturing industries, the production and distribution of electricity and steam, and natural gas and water supply. All resident market enterprises are in scope for the Production index.
3.4 Concepts and definitions (main variables)	The index aims to follow the monthly change in the value of industrial production. The current concepts and framework are those set forth by the EU Regulation 1165/98, concerning short-term statistics amended by Regulation (EC)1158/2005 that established data requirements in relation to coverage, periodicity, and timeliness, as well as the new 2005 Eurostat manual <i>Methodology of Short-term Business Statistics (interpretations and guidelines)</i> . The index measures the monthly change in the value of production, covering mining, manufacturing, electricity, water and gas supply.
3.5 Statistical unit	Statistical units are classified according to their principal activity at the class level (4-digit) in accordance to the National Classification of Economic Activities (NACE.BG), which complies with NACE rev.1.1.
3.6 Statistical population	

出典：SDMX User Guide

7. ドメイン横断概念

7.1. ドメイン横断概念の概略

ドメイン横断概念とは、コンテンツ指向ガイドラインにて示されている、SDMX 準拠のデータおよびメタデータを作成するために必要となるガイダンスのことである。現在 135 の統計に関する概念を包含しており、例えば、頻度、分野、通貨など、DSD に包含される構造メタデータに適用されるものが主となっているが、正確性、互換性、即時性など計測や品質にかかわる一般的な概念も含まれており、MSD に包含される概念も定義されている。



出典：SDMX User Guide

この図はドメイン横断概念、データおよびメタデータ構造定義、データおよびメタデータセットの関連を示したものである。この図に示されているのは、ドメイン横断概念がデータ構造定義に対しては次元の定義を提供しているということと、参照メタデータに対してはデータに関する属性情報を提供することができるということである。また、コード化されて表現される概念についてはコードリストによって関連付けられる必要がある。

8. SDMX レジストリ

8.1. SDMX レジストリの概略

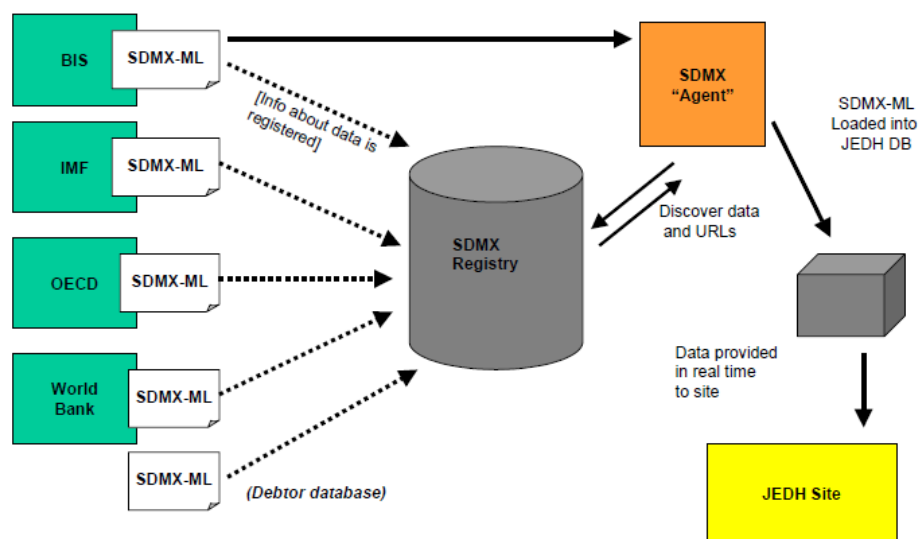
SDMX レジストリとは、構造メタデータ（DSD、コードリストなど）に関する情報、データセットの場所に関する情報、参照メタデータに関する情報を提供するためのアプリケーションであり、以下の機能を提供する。

- ・ メタデータレポジトリ：どのデータセットおよびメタデータセットが利用可能であるか、そのデータをどこから取得可能であるかの情報を提供する機能。
- ・ 提供メタデータレポジトリ：提供されるデータセットおよびメタデータセットに関する情報を提供する機能。例えば、更新頻度、コンテンツについて、アクセス方法などが挙げられる。
- ・ 構造メタデータレポジトリ：データセットおよびメタデータセットの構造に関する情報を提供する機能。例えば、使用されているコードリストや、適用されている概念などが挙げられる。
- ・ 通知設定を行い、ある特定のデータおよびメタデータが利用可能になった時に自動的に通知を行う機能。

また、SDMX レジストリは、ウェブサービスを利用した標準的インターフェースが定義された仕組みのことであり、ある特定の組織が集権的に一括管理しているものではない。上記の機能概要の通り、レジストリではデータおよびメタデータの分類・一覧化を行うもので、データおよびメタデータはどのレポジトリ・ウェブサーバ上にも展開可能である。

8.2. SDMX レジストリ実例 1

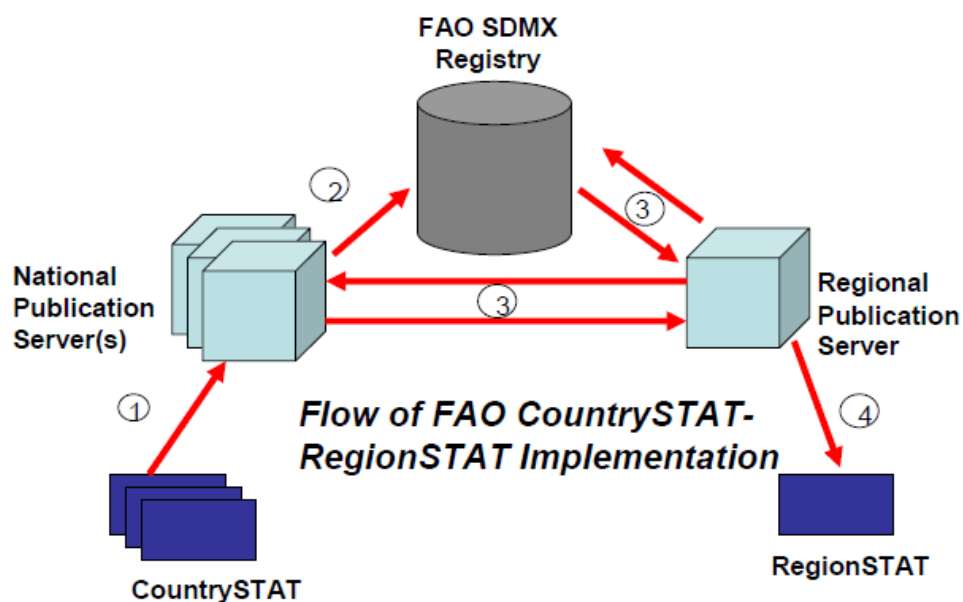
下記は SDMX レジストリの一例であるが、BIS や IMF が SDMX レジストリに対してデータおよびメタデータを登録し、JEDH（共同開発対外債務）のサイトから情報が利用可能になっていることがわかる。



出典：SDMX User Guide

8.3. SDMX レジストリ実例 2

また、FAO(国際連合食料農業機関)で構築されている SDMX レジストリの構造図は下図のようになっている。



出典：SDMX User Guide

9. SDMX ツール

9.1. ツールの入手

SDMXの利用および普及促進のため、各種のSDMXツールがSDMX設立団体および各種団体により公開されている。ツールに関する情報及びツール本体は <http://SDMXtools.org> および <http://www.metadatatechnology.com> などから利用可能となっている。本プロジェクトではSDMXのツール評価を実施した。

なお、ソフトウェアの実行環境は

ノートブック PC (ThinkPad X61) / メモリ 4G / Windows 7 (x86) / MS Office 2003 である。

これらのツールは、下記に挙げられている通り、多くがメタデータの編集、データの変換等の機能を提供しているが、SDMX データを直接処理可能な統計ツールの登場が期待される。なお、既存の統計ツール等で SDMX に対応したものは現在のところ確認されていない。

9.2. ツールの一覧

No.	名称	作者・団体	機能等
1	SDMX レジストリ	Metadata Technology	レジストリスキーマに指定されたレジストリ API の実装を提供。構造メタデータの登録およびクエリが可能。FLASH を多用した高機能な SDMX レジストリ構築ツール。
2	SDMX ツールスイート	Metadata Technology	SDMX1.0 および 2.0 XML フォーマットに対応した各種ツールセット。多機能かつ包括的な SDMX データツール。
3	メタデータエディタ	Metadata Technology	メタデータ構造定義およびメタデータレポートの生成を行う。SDMX-ML ファイルを生成可能。
4	DSD データベースツール	Metadata Technology	Microsoft Access ベースで構築された、SDMX データ構造定義をメンテナンスするツール。
5	Excel SDMX オーサリングツール	Metadata Technology	SDMX-ML2.0 構造メッセージを含む DSD から Excel のテンプレートを生成するためのツール。
6	ECB チェッカ	European Central Bank	SDMX-ML および SDMX-EDI のデータファイルを読み込み及びチェックを行うツール。

7	ビジュアル化フレームワーク	European Central Bank	アクションスクリプト3にて記述された、統計データおよびメタデータを視覚化するためのライブラリ。
8	データストラクチャウイザード	Eurostat	DSD メンテナンス、メタデータメンテナンス、SDMX データ入出力、メタデータ構造 GUI 等の機能を提供するツール。
9	SDMX コンバータ	Eurostat	CSV,FLF,SDMX-EDI,SDMX-ML ファイルフォーマット間で、DSD をベースとしてデータの変換を行うツール。
10	SDMX レジストリ	Eurostat	500 ページ以上のマニュアル完備かつソースコード公開の SDMX レジストリ構築パッケージ。

9.3. ツール1 (SDMX レジストリ)

作者・団体：Metadata Technology

ライセンス：明記なし

機能概要・評価：

レジストリスキーマに指定されたレジストリ API の実装を提供。このパッケージは構造メタデータの登録およびクエリが可能。提供される機能は以下の通り。

データおよびメタデータ構造定義、関連コード定義及び概念のサポート

エージェンシー、データプロバイダ、データ利用者に対応

データおよびメタデータフローに対応

カテゴリスキーム対応

階層コードリスト対応

データおよびメタデータの登録及びクエリ実行、通知登録、通知機能対応

KeyMaster GUI ツールによりレジストリの登録内容をメンテナンス可能 など

本体のプログラムについてはソースが提供されていないことと、実行のためのライセンス関係が不明であったため、実行環境の構築は行っていない。HTML 等のテンプレートや素材類を見る限り、FLASH を多用した高機能な SDMX レジストリ構築ツールであ

るように見受けられる。

9.4. ツール 2 (SDMX ツールスイート)

作者・団体：Metadata Technology

ライセンス：明記なし

機能概要・評価：

SDMX1.0 および 2.0 XML フォーマットに対応した各種ツールセット。

XML データバリデーションおよび変換機能を提供。

キーファミリー → コンパクト XML スキーマへの変換

キーファミリー → ユーティリティ XML スキーマへの変換

キーファミリー → クロスセクション XML スキーマへの変換

キーファミリー → HTML 表示

一般 XML データ形式 → コンパクト XML データ形式への変換

コンパクト XML データ形式 → 一般 XML データ形式への変換

コンパクト XML 形式およびデータの統合ツール

一般 XML データ形式 → ユーティリティ XML データ形式への変換

一般 XML データ → CSV への変換

SDMX1.0 と 2.0 形式間の相互変換

多機能かつ包括的な SDMX データツール。今回のプロジェクトでは生成 XML のバリデーションのために使用している。

9.5. ツール 3 (メタデータエディタ)

作者・団体：Metadata Technology

ライセンス：明記なし

機能概要・評価：

本ツールはメタデータ構造定義およびメタデータレポートの生成を行う。メタデータ構造定義に沿ってメタデータレポートがバリデートされた XML スキーマに合致した SDMX-ML ファイルを生成可能。

2008 年リリースのためか、現行の SDMX2.0 に対応していないようにも見受けられる。

EUROSTAT 等からダウンロードした SDMX を認識することができず、詳細の評価は不可能であった。

9.6. ツール 4 (DSD データベースツール)

作者・団体：Metadata Technology

ライセンス：明記なし

機能概要・評価：

本ツールは **SDMX** データ構造定義をメンテナンスするものであり、**SDMX1.0** および **2.0** をサポートする。**Microsoft Access** ベースで構築されており、同ソフトウェアが実行には必要となっている。

2007 年リリースでありながら、現行の **SDMX2.0** に対応していると明記されていて、内部的に **Java** も併用して作成されているように見受けられる。しかしながら、実際に **XML** ファイルを読み込んだときに **Unknown Protocol Error** が発生したため、実際のツール評価を行ってはいない。

9.7. ツール 5（Excel SDMX オーサリングツール）

作者・団体：Metadata Technology

ライセンス：明記なし

機能概要・評価：

SDMX-ML2.0 構造メッセージを含む **DSD** から **Excel** のテンプレートを生成するためのツール。生成したテンプレートにデータを入力し、**SDMX2.0** 一般データメッセージを生成可能となっている。

ダウンロード不能のため、評価は行っていないが、**Excel** は一般的なツールであり、広く使われているため、実行可能であれば有力なツールであると思われる。

9.8. ツール 6（ECB チェッカ）

作者・団体：European Central Bank

ライセンス：BSD

機能概要・評価：

SDMX-ML および **SDMX-EDI** のデータファイルを読み込み及びチェックを行うツール。**FAME** データベースに対して基本的なデータ操作を行うことが可能であるが、このツールは例外発生のため評価を実施できなかった。

9.9. ツール 7（ビジュアル化フレームワーク）

作者・団体：European Central Bank

ライセンス：BSD

機能概要・評価：

アクションスクリプト 3 にて記述された、統計データおよびメタデータを視覚化するためのライブラリである。現在 **ECB**, **FRB**, カナダ中央銀行, **OECD** のデータ視覚化要素に

使われている。ダウンロード可能ファイルは提供されていないことと、ライブラリであるため実行しての評価は行っていないが、SDMX レジストリのようなウェブサービスを構築する上でビジュアル化ライブラリは重要な要素であるため、実際にシステムを構築する際に検討すべきであると考えられる。

9.10. ツール 8 (データストラクチャウイザード)

作者・団体：Eurostat

ライセンス：EUPL (European Union Public License)

機能概要・評価：

DSD メンテナンス、メタデータメンテナンス、SDMX データ入出力、メタデータ構造 GUI 等の機能を提供するツール。また、階層コードリスト、データフロー、メタデータフロー、カテゴリスキーム等をサポート。

オンラインモードとオフラインモードを備え、SDMX レポジトリと通信が可能となっている。オフラインモードではコードリスト、DSD、MSD、階層コードリストほかの SDMX-ML の要素について編集および保存が可能となっている。また、DSD に定義された形式で XML データのテンプレートファイルを作成することが可能である。オンラインモードでは SDMX レジストリと接続し、クエリの実行やローカルレポジトリのアップデート、リモートレポジトリのアップデートが可能である。

9.11. ツール 9 (SDMX コンバータ)

作者・団体：Eurostat

ライセンス：EUPL (European Union Public License)

機能概要・評価：

CSV,FLF,SDMX-EDI,SDMX-ML ファイルフォーマット間で、DSD をベースとしてデータの変換を行うツール。

CSV ファイルの取り扱いができるため、既存システム等から CSV でデータが取得できる場合にデータを SDMX 形式に変換してデータの共有・交換を実施できるため、有望なツールとなりうる。

9.12. ツール 10 (SDMX レジストリ)

作者・団体：Eurostat

ライセンス：EUPL (European Union Public License)

機能概要・評価：

一通りの機能を備えた SDMX レジストリ構築パッケージ。500 ページ以上のマニュアル

完備かつソースコード公開。SDMX レジストリを構築する際に最も少ない工数で対応が可能になると思われる点で極めて有望なツールである。Metadata technology 作の SDMX レジストリとは別物である。

9.13. その他のツール

ほかにも、ISTAT 作成のツールおよびライブラリや、UNESCO 作成のライブラリが存在している。

10. 経済産業省統計データの SDMX 化試行

本プロジェクトでは、経済産業省発行の統計データの一部について、SDMX 化を試行した。

10.2. 対象データの検討

諸外国にて同等のデータが公開されていること、データ加工の容易性等の観点から、鉱工業指数（IIP）が選定された。したがって、本プロジェクトでの SDMX 化は以下を対象としている。

生産・出荷・在庫・在庫率指数（2003 年 1 月～2010 年 5 月）

10.3. データ構造の検討

SDMX 化を実施するに当たり、既存のデータの分析により、以下の構造を有していることが判明し、これに沿って DSD の設計を行っている。

集計軸

時間軸

月次、四半期、年、年度

分類

業種別

財別

品目別

出力データ

生産（付加価値額）

出荷

在庫（期末）

在庫（平均）

在庫率

その他

季節調整要素あり

品目ごとに付加生産ウェイトあり

10.4. メタデータの要素と考察

メタデータの設計に当っては、諸外国で先行例として実施されているデータとの互換性を高める目的で、Eurostat および FRB における IIP もしくは類似データのメタデータを参考に検討された。

データのステータス

CL_OBS_STATUS (観察ステータス)

通常/該当なし/データなし/計算不能 が定義されているため、ほぼそのまま流用可能。

データの頻度 (Frequency)

CL_FREQ

EUROSTAT では Annual, Daily, Half-Yearly, Monthly, Quarterly, Weekly を定義しているが、「年度」に相当する概念を定義していない。FRB では Annual(Jan) ... Annual (Dec)と年度の初月を定義しているのでこれを参考にしている。

通貨 (Currency)

本データでは通貨の概念を使用しないため、定義せず。

単位(Unit)

CL_UNIT

数値は基準年に対する指数であるので、これを定義しておく必要がある。例えば FRB では Index:_2005_100 (2005 年=100 とする) のように単位をコード化して定義しているため、これにならって定義を行っている。

調整 (Adjustment) :

CL_SA

調整要因として定義済みのものは下記に挙げられる通りであるが、このうち、“Neither seasonally or working day adjusted” および “Seasonally Adjusted”が季節調整の有無を意味するものとして採用可能であった。

Neither seasonally or working day adjusted

Seasonally adjusted, not working day adjusted

Trend

Working day adjusted, not seasonally adjusted

Seasonal factors

Working day and seasonally adjusted

Combined seasonal and calendar factors

品目番号、品目名称：

CL_IP_MAJOR_INDUSTRY_GROUPS_CODES

これらに関して、諸外国においても同等の概念は当然ながら存在し、また、国際的な品目リストについても定義が存在するが、変換が容易ではないことなどを考慮し、本プロジェクトのためのコードリスト(IIP 産業コード)として、改めて定義をしている。

10.5. データ変換の手法

SDMX 化を実施するに当たり、データ変換の手法に関しては以下の選択肢が存在した。

元データについて：

- ・ 経済産業省ウェブサイト掲出の XLS ファイル
- ・ 経済産業省ウェブサイト掲出の CSV ファイル

変換プログラム作成について：

- ・ データをいったん全件メモリ中にロードしたうえで DOM ツリーを構築し、XML 形式に変換して出力する方法。
- ・ データを読み込みながら逐次 XML 形式に変換しながら出力する方法。

本プロジェクトでは、IIP のデータに関しては元のフォーマットが比較的単純であることと、CSV 形式を取り扱うことの簡便さを考慮した上、本格的な変換プログラム（プログラム中で DOM 構築）の手法は取らず、比較的平易な逐次読み込み型のプログラムが作成された。

また、変換データのプログラム作成にあたっては、サンプルとなる SDMX ファイルを作成する必要があったが、これについては手動で作成した。

11 まとめ

今プロジェクトでは、SDMX に関連するものとして、以下を実施した。

- ・ SDMX 関連ドキュメントの精査
- ・ 諸外国における SDMX 化の例の精査
- ・ SDMX 既存ツールの検討
- ・ メタデータの検討

- ・ 既存データの **SDMX** 形式化

また、以下の課題が引き続き検討される必要があると思われる。

- ・ 国内統計データの **SDMX** 化に伴うメタデータの検討
- ・ **SDMX** 形式でデータを公開することによるメリットの検討

SDMX 関連ドキュメント

英語ではすでに **User Guide** が発行されていたり、カンファレンスが開催されているなど、活発なやりとりが行われているなかで、ガイダンスがすでに示されていたり、事例が提示されているなど、具体的に実施するための下地がすでに整備されつつあるように思われる。しかしながら、本邦においては日本語によるドキュメントや事例がほぼ存在しないことから、原文を参照することで仕様の把握等を行った。

このように、日本国内では日本語での文書が存在しないことや、日本語化されたツールが存在しないことなどから、現状では導入の敷居が非常に高いものとなっていると思われる。このため、ドキュメント及びツールの整備が **SDMX** の国内普及においては必要になるのではないと思われる。

諸外国の例

すでに **Eurostat** や **FRB** 等、海外の諸機関においてはすでにいくつか先行事例が存在する。ただ、各々のメタデータを比較検討すると、コードリスト等など形式にはばらつきがみられ、標準規格の運用の難しさを垣間見ることとなった。また、**Eurostat** 等においても、すべてのデータが **SDMX** 形式で公開されているわけではなく、部分的に **SDMX** 形式による公開となっているため、本形式によるデータ公開の対象とすべきデータの選定には引き続き検討すべき余地が多くあるように見受けられる。

SDMX ツール

すでに数多くのツールが存在していることは事実であるが、ツールそのものに関する情報や、それらの利活用のノウハウが決定的に不足しているのが現実である。また、利用可能なツールはいずれもがデータの編集・変換・メタデータの管理等の機能を提供するものであって、統計解析ツール等やデータ加工ツールで **SDMX** 形式に直接対応しているものは見受けられなかった。このため、実際のデータ利活用の現場では結局 **CSV** 等の既存の形式に変換することとなり、利活用の促進を阻害する要因となりうると思われる。フォーマットの普及とともにツールのさらなる充実が望まれる。

メタデータの検討

今回のプロジェクトにより、一部既存公開データについてメタデータの検討を行った。当

該データに必要なメタデータの設計は可能であることが判明した。しかしながら、以下についてさらに検討を続ける必要があると思われる。

- ・ メタデータ制定のプロセスおよび包含性
- ・ メタデータの維持

メタデータ制定のプロセス

本プロジェクトにおけるメタデータは実証実験であり、特定のデータにおいてのみ適用されるものであった。また、メタデータの策定は本プロジェクトの担当者内のみで検討が行われた。実証実験のレベルにおいては **SDMX** 形式利用の検討が主体であったが、**SDMX** 形式を今後展開するに当たっては、メタデータの形式について、データの発行を行う各組織がかかわる委員会等による検討が必要ではないかと思われる。

メタデータの維持

本プロジェクトにおけるメタデータは、現在公開されている時点の品目分類等に限定して策定が行われたが、今後 **SDMX** 形式での展開にあたっては各コード表等において改定が行われるため、過去のデータについても包含しうる、メタデータの維持プロセスが必要になってくるのではないかと思われる。

SDMX 形式によるデータ公開のメリット

SDMX に対応したツールがないので直接メリットを享受できないという点は先も述べたとおりであるが、標準規格であるため、データの利活用に対するメリットは極めて大きなものであると期待されるため、フォーマットの普及とツールのさらなる整備が望まれる。このため、日本国内でも先行的に **SDMX** 形式の普及促進を図る必要性があり、これにより実質的なメリットがさらに拡大するものと思われる。

また、**SDMX** 形式の展開においては、**SDMX** レジストリの設置も必要であろうと思われる。

12 参考資料

- [1] Overview of SDMX (SDMX Conference 2009 by Pedro Diaz Munoz)
- [2] Teaching SDMX – issues and best practices (Giuseppe Sindoni, ISTAT Italy)
- [3] SDMX and the DDI: Using the Right Tool for the Job (Arofan Gregory, Open Data Foundation)
- [4] SDMX User Guide Version 2009-01 (www.SDMX.org)
- [5] SDMX Tools: Introduction and Demonstration (Arofan Gregory, Chris Nelson)
- [6] SDMX Content-Oriented Guidelines (www.SDMX.org)