

自治体等の公表する統計データの 公開方式についての一考察

日本都市学会 研究発表

2013年10月27日 26日

石田 修二（四日市大学非常勤講師）

(C) Shuji ISHIDA 2013

1

石田修二とは

- 四日市大学，三重大学非常勤講師
- 担当している科目は，情報リテラシー，情報政策
- 授業の予習で，表計算ソフトや統計解析ソフトを利用する機会が多数
- 総務省統計局や自治体の統計情報のページは私のお気に入り

最後の2つは今回の発表の動機でもある

2

このお話は

- 自治体のホームページで公開されている数値情報がどのようなファイル形式で公開されているのか現状を確認する
- その現状に対する私の意見・提言
- 会場の皆様から意見を伺いたい
 - 利用者の立場から
 - 提供者の立場から

3

数値情報への関心の高まり

- KEYWORD : オープンデータ, オープンガバメント
- 統計学ブーム
 - 本屋では統計本がいっぱい
 - 『統計学が最強の学問である』（西内啓著, ダイヤモンド社）がロングセラーに
 - Amazonで調べても今月だけで統計, データ分析本が10作品登場
 - 学習指導要領で統計分野が重視される
 - 次のスライド参照

4

学習指導要領（中学・数学）

- 目的に応じて資料を収集し,コンピュータを用いたりするなどして表やグラフに整理し...
- コンピュータの活用
 - 大量の資料を整理する場合や大きな数, 端数のある数を扱う場合などには, **コンピュータなどを利用して作業の効率化を図り**, 処理した結果を基に資料の傾向を読み取ることに重点を置いて指導できるようにする。

(C) Shuji ISHIDA 2013

5

（コンピュータの側でも） 統計ソフトが花盛り

- Microsoft Excel, SPSS, SAS, JMP, S言語, Mathematica → 老舗
- LibreOffice (OpenOffice.org) のCalc, Gnumeric, R, Gretl → 最近の注目
- Web上で統計処理（グラフ化など）してくれるサービスも増えてきた
 - そういったサイトを自分で作るケースも増えてきた

6

私もささやかながらデータの可視化に取り組んでます

ブログ > いろいろ > (関連: データ統計処理 | サイト内検索)

石田修二のオープンデータ

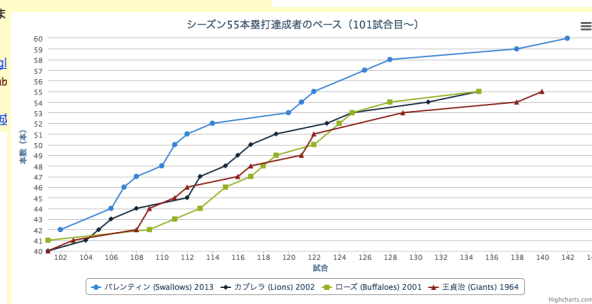
ExcelやR等でグラフを作ったり、統計学、データ分析の勉強をするために作成した元データを
ます。5 star Open Data と言えば3つ星くらいの取り組みになるでしょうか。

ていないので、お目当てのデータがあるかどうかは、ブラウザのページ検索機能で調べて下さい

メントではcsv形式でのダウンロードもできま
トメニューで確認してみてください。

求) シーズン55本塁打達成者のベース (Google
ン (Balentien), アレックス・カブレラ (Cob

ighchartsでグラフ化したもの: 55本塁打達成



(C) Shuji ISHIDA 2013

7

ストレスなく処理できるデータが欲しい

- 編集, 加工が容易な形でデータを公開してくれる都市のサイトのはデータに関心ある人には魅力的
 - 研究者だけじゃない, 市民, 企業にとっても
- 私の視点
 - データファイルの形式 (いわゆる「データ交換フォーマット」)
 - データの中身
 - 多くのソフトで利用できるかどうか

8

国際交流の視点でも...

- 数字もコミュニケーションツールの一つ
- 使いやすい統計データを公開することは、国際交流にも役立つ？
 - (と私は思っている)

(C) Shuji ISHIDA 2013

9

では、公開の現状はいかに

- HTML形式
- PDF形式
- **Excel (xls) ファイル形式**
 - 配布されている統計データの多くはこのタイプ
 - Excelは多くのコンピュータに使われており、開くのが楽
 - だからこそ、他のソフトで使う際には問題も多い

10

高松市もオールExcel

- はっきりと統計資料はExcel形式と宣言
<http://www.city.takamatsu.kagawa.jp/935.html>
 – 丁寧に Excel Viewer へのリンク付き

※ ご注意

統計資料はMicrosoft Excel 形式です。
 Microsoft Excel をお持ちでない場合は、Microsoft Excel Viewer を下記からダウンロードしてご利用ください。

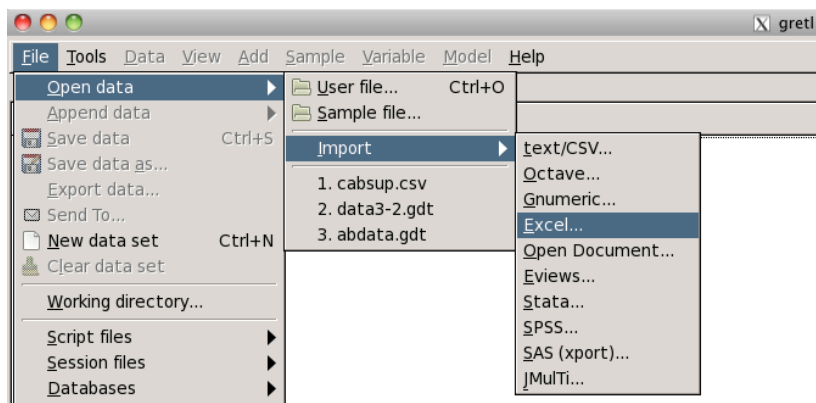
▶ [ダウンロードセンター Microsoft Excel Viewer](#)

(C) Shuji ISHIDA 2013

11

確かに、Excelファイル対応の ソフトは多い（1）

- Gretlの場合（xls形式のみ）



12

確かに，Excelファイル対応のソフトは多い（2）

- **SPSS**の場合
 （ [機能一覧] - [ファイル管理] より）
 - Excelのものと同様のテキストウィザードを使用して、ASCIIデータを容易に読み込み/定義
 - Excel 2007ファイルの読み込み/書き出し
 - Excel、Lotus 1-2-3、dBASER へ/からのファイルの変換
 - データファイルを最新バージョンのExcelにエクスポート

Excelとの親和性

13

セルの結合

- 不必要な部分での空白
- **セルの結合**

印刷するにはいいが、そのままでは他のソフトに読み込ませられない

第 1 表 産業(大分類)、経営組織(4区分)別民営事業所数、男女別従業者数及び1事業所当たり従業者数(名古屋市)

産業大分類	総数		個人				法人				会社				会社以外			
	事業所数	従業者数	事業所数	男性	女性	従業者数	男性	女性	事業所数	従業者数	男性	女性	従業者数	男性	女性	従業者数		
全産業(公務を除く、事業内容等不詳を含む)	236,283		
A~B 全産業(公務を除く)	122,153	1,495,881	868,596	596,247	46,729	131,348	54,513	76,835	86,835	1,272,055	752,998	512,050	74,917	1,138,808	695,616	435,894	5,918	133,987
A~B 農林漁業	53	489	341	339	-	-	-	-	52	626	298	328	43	301	189	112	9	135

15

不要な空白

G3							
自然動態							
A	B	C	D	E	F	G	H
平成25年9月1日 現在の三重県の人口							
3	区分	面積	人口		自	然	動
4		(Km2)	総数	男	女	出生	死亡
5	総数	5,777.31	1,829,237	890,426	938,811	1,270	1,490
6	市計	4,032.93	1,599,948	779,936	820,012	1,116	1,288
7	郡計	1,744.38	229,289	110,490	118,799	154	202
8	津市	710.81	281,663	136,450	145,213	189	243
9	四日市市	205.58	306,621	151,977	154,644	262	202
10	伊勢市	208.53	128,221	60,456	67,765	63	139
11	松阪市	623.77	166,821	80,366	86,455	104	135
12	桑名市	136.61	140,737	69,054	71,683	107	93
							差

(C) Shuji ISHIDA 2013

16

印刷イメージで公開

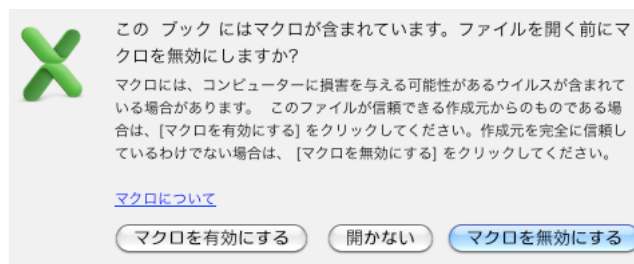
並び替えやグラフを作るのに、まとめなければならない。
→手間がかかる！

通称町名(中部)	総世帯	人口	通称町名(富津部)	総世帯	人口	通称町名(富田)	総世帯	人口						
世帯数(世帯数)	人数(人)	男(人)	女(人)	世帯数(世帯数)	人数(人)	男(人)	女(人)	世帯数(世帯数)	人数(人)	男(人)	女(人)			
北勢	165	582	170	212	440	232	211	鴨部第一	109	266	136	130		
津市	217	509	256	237	411	1096	538	462	鴨部第二	165	472	241	231	
磯部町	507	1137	544	593	1111	255	140	115	鴨部第三	95	214	109	105	
中津	387	828	402	424	253	629	324	305	鴨部第四	183	378	200	176	
西津一丁目	418	730	372	359	天方集賢二丁目	591	1639	807	832	大宇夜毎	370	915	466	449
西津二丁目	359	531	275	262	天方集賢三丁目	3	3	0	0	四日市町	91	170	71	99
西勢町	263	631	212	218	西勢町	172	385	202	183	富田一丁目	340	763	392	371
西勢	287	511	247	264	宇野	432	1074	529	545	富田二丁目	302	756	356	400
坂本一丁目	615	1134	531	603	富州磯部	254	621	287	334	富田三丁目	497	1109	551	558
坂本二丁目	389	914	451	463	富田一色町	805	1831	862	969	富田四丁目	258	709	346	363
北勢	219	615	266	269	富田二丁目	4	4	1	3	富田集賢	189	409	207	202
中勢	189	403	191	212	松濱町	476	1127	568	568	富田浜町	309	609	319	290
八幡町	168	363	165	198						富田浜元町	140	335	163	172
沖の島町	214	544	266	278						東富田町	655	1473	681	792
志保町	137	309	138	141						東浜町	172	363	189	204
北浜町	190	383	185	192						丸の内町一區	43	99	47	52
宇野	76	155	79	76						丸の内町二區	138	288	133	155
新々町	37	131	65	66						丸の内町三區	201	411	218	193
新町	173	361	177	184						南富田町	328	609	393	416

17

Excelマクロ

- 開いたらマクロの警告メッセージが
– ウィルス?
- セキュリティの面でも危険性が高い



18

「ネ申Excel」

- 「神Excel」（カミエクセル）と読む
- 「ネ申」は「神」の隠語であるが、紙への出力しか考えていないExcelの意味も含んでいる（奥村晴彦「ネ申Excel」）
- ネットで一部に浸透している
 - （例）公務員が公開するネ申Excelが日本の生産性を落としていく話 (<http://togetter.com/li/469017>)



(C) Shuji ISHIDA 2013

19

現状のExcel形式ファイルは

- 開くのは楽だ
- ダウンロード後
 - 印刷する → 適している。綺麗に印刷できる
 - 編集, 加工する → 手間がかかる。適しているとは言い難い
- データとしての再利用が困難な形で公開されている

(C) Shuji ISHIDA 2013

20

- 確かにExcelを持っている人は多数派ですが...
- Excelにだけ対応すればいい, というのは手間もかからないし, 効率的かもしれない → 公平性の観点では...
 - 私はExcelをあまり使いません。でも, 税金ちゃんと払ってます

(C) Shuji ISHIDA 2013

21

ここまですを踏まえて

- どの形式にも、一長一短あり
 - どれが飛び抜けていいというわけではない
- しかし、いつまでも、Excelを使った今のやり方**だけ**でいいの？
 - 他にもいい形式があれば、使ってみては
 - 特に地方自治体における統計データの出し方について、検討する必要があるのでは

22

Excelではないファイル形式での公開例

- 会津若松市：ODF形式に移行中

The screenshot shows the website of Aizuwakamatsu City. The URL is www.city.aizuwakamatsu.fukushima.jp/docs/2008021400340/. The page title is "Aizuwakamatsu City 会津若松市". There are navigation links for "市民の方へ" (For Citizens) and "事業者の方へ" (For Business Owners). A notice box contains the following text:

オープンドキュメント形式文書を利用するには

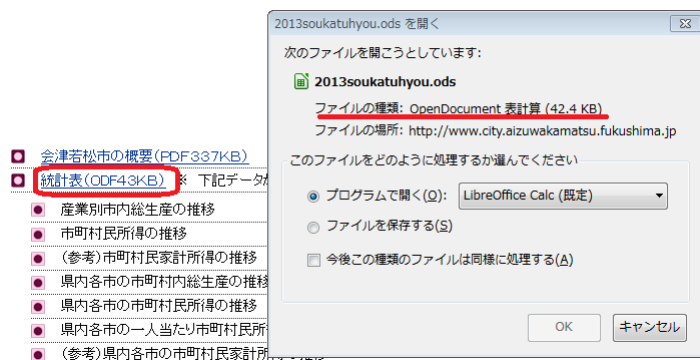
現在会津若松市では、公文書を国際標準であるオープンドキュメント形式に移行しています。

本市ウェブサイトに掲載されている、オープンドキュメント形式(ODF)の文書を利用するには以下の手順をご覧ください。

- LibreOfficeで編集する
- OpenOffice.orgで編集する
- Microsoft Officeで編集する

23

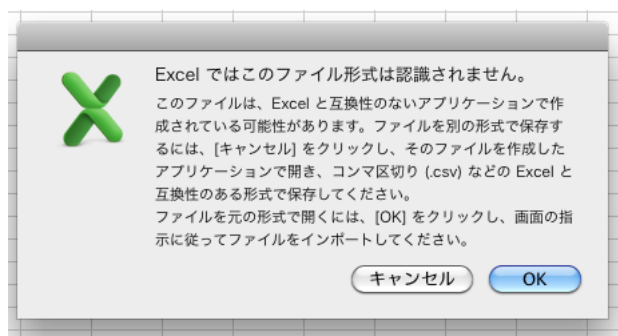
会津若松市の統計表はODF形式



24

ODF形式は本当にオープンか

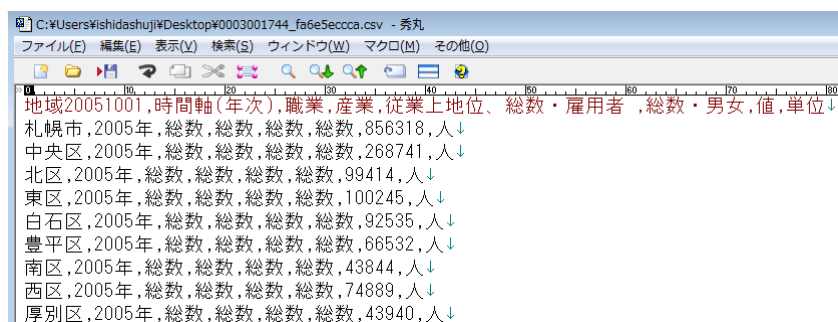
- MacのExcel 2011では開けませんでした



25

CSV形式でも提供してもらえませんか

主に表計算ソフトやデータベースソフトがデータを保存するときに使う形式だが、汎用性が高く、多くの電子手帳やワープロソフトなどでも利用できるため、異なる種類のアプリケーションソフト間のデータ交換に使われることも多い。(e-Words, <http://e-words.jp/w/CSV.html> より)



26

政府もCSVを推してます (1)

国民へ発信する重要情報のファイル形式について」
<https://www.lasdec.or.jp/cms/12,22060,84.html>

国民へ発信する重要情報のファイル形式について

標記の件につき、全国地方公共団体へ発信しましたこと、お知らせします。

国民へ発信する重要情報のファイル形式について

下記につき、当センター関係団体より情報提供いただきましたので、地方公共団体へ [CEPTOARメーリングリスト](#) を通じて展開したことをお知らせします。

なお、本記事は、被災して情報を取得することが困難な状況にある地方公共団体に向け、お知らせするものです。

通知日 平成23年3月18日 20時0分

件名

(参考) 国民へ発信する重要情報のファイル形式について

内容要旨

震災発生から時間経過とともに、全国民が注目・閲覧を要する重要コンテンツ。政府、地方公共団体から提供される情報の多くが、PDFやExcelファイル形式で情報配信されております。インターネットを通じて多数の国民に円滑に閲覧していただきたいところですが、アクセスが集中し、PDFやExcelファイル形式の場合、容量が大きく、サーバー・回線リソースを圧迫し、重要情報が閲覧できない事象が頻出しています。

27

具体的なファイルの公開方法

1. アップロードするファイルはPDFだけではなく、誰もが見ることができ、小容量な「HTML形式（テキスト）」のファイルも公開する。
2. 紙資料のスキャンファイルはPDFではなく、誰もが見ることができて、再利用が容易な「JPEG形式（静止画像データ）」とする。
3. 表形式データのファイルは、容量が大きくなるExcel形式ではなく、どのソフトでも開くことができ、再利用が容易な「CSV形式（表計算ソフトなどで開くことを前提に、データをカンマ(,)や、スペース、tab等で区切って並べたテキスト形式）」とする。

28

政府もCSVを推してます（2）

[高度情報通信ネットワーク社会推進戦略本部第14回 電子行政に関するタスクフォース](#)

- 携帯電話等からのアクセスやネットワークの負荷軽減、情報の2次的な利用など考慮し、情報は、PDFやWord、Excel等の形式のみではなく、html、csv等でも提供することを推奨

タイトル	概要
国民へ発信する重要情報のファイル形式について (3/18 LASDECより全国各地公共団体へ通知)	より多くの方に簡易に情報を受け取っていただけるようアップロードするファイルのデータ形式についてご一考いただき、国民への円滑な情報提供をご検討いただきたい。具体的なファイルの公開方法として下記を例示。 ・情報は、PDFのみではなくHTML形式のファイルも公開する。 ・紙資料のスキャンファイルはPDFではなく、JPEGとする。 ・形式データのファイルは、Excel形式ではなく、CSVとする。
震災関連情報の提供に係る協力依頼 (3/22 内閣広報官より各府省へ通知)	・各府省ホームページ等への掲載を通じて、迅速かつ国民に分かりやすい形で必要な情報の発信に努めていただきたい。 ・今後、被災地での情報収集手段として、携帯電話の重要性が格段に増すと考えられることから、速やかに携帯電話向けのホームページを用意していただきたい。また、同様の観点から、PDF等のページについては、HTML形式のページへ変更していただきたい。
東北地方太平洋沖地震等に係る情報提供のファイル形式について(3/29 総務省より各府省へ通知)	・ホームページにおいて情報提供を行う場合には、極力PDF等の容量の大きいファイル形式のみによるのではなくhtmlやcsv等の比較的回路負担の少ないファイル形式を用いるなど、国民等に対し円滑な情報提供を図る観点から、掲載するファイル形式について御検討いただきたい。
東北地方太平洋沖地震等に係る情報発信のデータ形式について(3/30 経済産業省より社団法人日本経済団体連合会へ通知)	・ホームページにおいて情報発信を行う場合には、極力PDF等自動処理がしにくいデータ形式のみによらずHTML/CSV等の自動処理に適したデータ形式を併用したり、別途オープンな情報発信APIを整備するなど、データを提供する方法を配慮いただきたい。

29

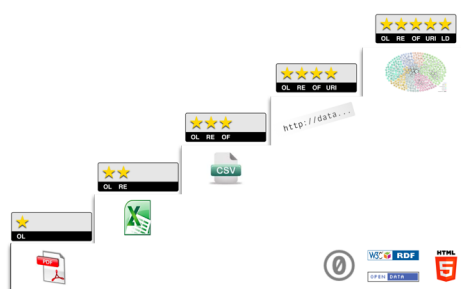
オープンデータのための5つ星スキーム

1. どんな形式でもいいのでまずはWeb上に公開
2. データを構造化データとして公開
3. ExcelよりCSV
4. 物事を示すのにURIを使え
5. データのコンテキストを提供するために他のデータにリンクしよう

5stardata.info

5 ★ Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



30

データ交換フォーマット，議論中

- 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）
 - 電子行政オープンデータ実務者会議
 - データワーキンググループ
- オープンデータ流通推進コンソーシアム
- 経済産業省 IT融合フォーラム

31

データの作成方法

- 「電子行政オープンデータ推進のためのロードマップ（案）」及び「二次利用の促進のための府省のデータ公開に関する基本的考え方（ガイドライン）（案）」に関するパブリックコメントの募集について（<http://www.kantei.go.jp/jp/singi/it2/info/h250524.html>）
- 「数値（表）、文章、地理空間情報のデータ作成にあたっての留意事項（ガイドライン）」が参考になる

意見募集対象

- [電子行政オープンデータ推進のためのロードマップ（案）本文](#)
- [電子行政オープンデータ推進のためのロードマップ（案）別紙](#)
- [二次利用の促進のための府省のデータ公開に関する基本的考え方（ガイドライン）（案）](#)
- [数値（表）、文章、地理空間情報のデータ作成にあたっての留意事項（ガイドライン別添）（案）](#)

※最終的な決定内容は、本案から変更される可能性があります。

(C) Shuji ISHIDA 2013

32

（ガイドラインの） 「留意事項」を一部紹介

- データセルに整形や位取りのための文字(スペース、改行、カンマ等)を含めない。【留意事項2】(p.4-p.5)
- すべてのセルは、他のセルと結合しない。【留意事項5】(p.8)
- 値が存在しない場合を除き、データセルを空白にしない。(データ値を省略しない)【留意事項6】(p.9)

これが全く正しいというわけではない。
しかし、処理し易いデータ形式を作る
指針にはなる

(C) Shuji ISHIDA 2013

33

おわりに

- データ交換フォーマットにCSV形式を追加できませんか？
 - どういったソフトでも読める
- 印刷としてではなく，再利用も考慮した公開はできませんか？
 - しかし，今までやり慣れてきたことを変えるにはコストがかかる
 - データのクレンジングは利用者側の仕事になるのか
 - 大学の出番？

34

統計データのCSV化サービスの例

- 統計くん (<http://toukei.in/>)
 - 政府の統計データをCSVファイルにしてくれる
 - 次世代統計利用システムのAPIを利用



35