

# JSON-statによる統計データ 提供フォーマットの共通化 (プレゼンテーション資料)

STAT DASH グランプリ 2016  
行政サービス開拓部門応募作品  
佐藤 英人

<http://www.statld.org/hp/category/json-stat>

# JSON-stat とは

- **JSON-stat**

- 統計データ提供用のデータフォーマット
- 2011年にスペインの統計実務家 Xavier Badosa 氏が考案

- **JSON-statを利用した公的データの提供状況**

- ヨーロッパの6カ国や2つの国際機関が統計データ提供に利用中
  - ノルウェー政府 (Statistics Norway)、英国政府 (UK's Office for National Statistics)、スウェーデン政府 (Statistics Sweden)、デンマーク政府 (Statistics Denmark)、スペイン・ガリシア州政府 (Instituto Galego de Estatística)、アイルランド政府 (Central Statistics Office of Ireland)、国連欧州経済委員会 (United Nations Economic Commission for Europe)、ユーロスタット (Eurostat 欧州連合統計局) 2015年9月から

- **JSON-statを利用するアプリやツール**

- JSON-stat Format Viewer: 多次元表の閲覧
- 統計分析アプリやGoogle Visualization API 向けなど

# JSON-stat のイメージ

日本の人口(2010年)

男女別	人口
総数	128,057,352
男	62,327,737
女	65,729,615

分類軸「sex」は  
("T", "M", "F") の3つの値をもち、  
そのラベルは "総数", "男", "女" である。

多次元マトリックスの特性を活かして、統計値をコンパクトに記述する。

```
1) {
2)   "dataset": {
3)     "class": "dataset",
4)     "label": "日本の人口(2010年)",
5)     "dimension": {
6)       "id": [ "year", "sex", "place", "concept" ],
7)       "size": [ 1, 3, 1, 1 ],
8)       "role": { "time": [ "year" ], "geo": [ "place" ], "metric": [ "concept" ] },
9)       "year": {
10)        "label": "年",
11)        "category": { "label": { "o": "2010" } }
12)      },
13)      "sex": {
14)        "label": "男女別",
15)        "category": {
16)          "index": { "T": 0, "M": 1, "F": 2 },
17)          "label": { "T": "総数", "M": "男", "F": "女" }
18)        }
19)      },
20)      "place": {
21)        "label": "場所",
22)        "category": { "label": { "o": "日本" } }
23)      },
24)      "concept": {
25)        "label": "表章項目",
26)        "category": { "label": { "o": "人口" } }
27)      }
28)    },
29)    "value": [ 128057352, 62327737, 65729615 ]
30)  }
31) }
```

# JSON-stat で何ができるか？

## 実証実験

1. e-Statデータを JSON-stat形式に変換するアプリ (jsg-estat) の試作
2. Web上で公開されている統計データを JSON-stat形式に変換するアプリ(jsg-web)の試作
3. jsg-estat と jsg-web を利用して、異なるソースのデータ統合分析を実施

jsg: JSON-stat Getter の略

# e-StatをJSON-statに変換(jsg2-estat)

リクエストURLの問い合わせ文は、e-Statの統計データ取得の問い合わせ文と同じ(ただしappIdは不要)。

以下は、国勢調査の人口統計表(0003041389)から年齢階級別男女別人口データを取得する例です。

e-Statの統計データ取得

```
http://api.e-stat.go.jp/rest/2.0/app/json/getStatsData/  
?appId=XXXXXXXXX&statsDataId=0003041389&lvArea=1&cdCato1=00710&  
cdCato3=000&cdCato4From=200&cdCato4To=600
```



jsg2-estatを利用して、JSON-stat形式の統計データ取得

```
http://www.satolab.org/js/jsg2-estat/  
?statsDataId=0003041389&lvArea=1&cdCato1=00710&  
cdCato3=000&cdCato4From=200&cdCato4To=600
```

# e-StatデータをJSON-stat Viewerで利用

<http://www.satolab.org/js/jsg2-estat/?statsDataId=0003041389&lvArea=1&cdCato1=00710&cdCato3=000&cdCato4From=200&cdCato4To=600>

## JSON-stat Format Viewer

## 年齢階級別男女別人口

JSON-stat > Format > Viewer

**Dataset** **Dimension** **Category** **Data**

Samples: index.json galicia.json oecd-canada-col.json order.json collection.json galicia.csv

↑ 年齢(各歳), 国籍(総数及び日本人), 年齢別割合, 平均年齢及び年齢中位数, 男女別人口 全国, 市部 郡部, 都道府県, 市部

- ↓ 表章項目 (metric)
- ↓ 全域・人口集中地区2010 (classification)
- ↓ 男女別2010 (classification)
- ↓ 国籍2010 (classification)
- ↓ 年齢4 (classification)
- ↓ 地域(2010) (geo)
- ↓ 時間軸(年次) (time)
- ↓ Data

© XAVIER BADOSA (CC BY 3.0) | ICON SET BY ICOMOON (CC BY 3.0) | BACKGROUND IMAGE BY DRUI (CC BY-NC 2.0) | LAST UPDATE: 2015-12-26

↑ Data

Rows & Columns  
年齢4 ↔ 男女別2010

Filters  
全国 地域(2010)

Constants  
総数(国籍) 国籍2010  
2010年 時間軸(年次)  
人口 表章項目  
全域 全域・人口集中地区2010

	総数(男女別)	男	女
(再掲)0~4歳	5,296,748	2,710,581	2,586,167
(再掲)5~9歳	5,585,661	2,859,805	2,725,856
(再掲)10~14歳	5,921,035	3,031,943	2,889,092
(再掲)15~19歳	6,063,357	3,109,229	2,954,128
(再掲)20~24歳	6,426,433	3,266,240	3,160,193
(再掲)25~29歳	7,293,701	3,691,723	3,601,978
(再掲)30~34歳	8,341,497	4,221,011	4,120,486
(再掲)35~39歳	9,786,349	4,950,122	4,836,227
(再掲)40~44歳	8,741,865	4,400,375	4,341,490
(再掲)45~49歳	8,033,116	4,027,969	4,005,147
(再掲)50~54歳	7,644,400	3,800,576	3,843,824

# Web統計データをJSON-statに変換 (jsg2-web)

Web上で公開されている以下の3種が変換可能

1. HTMLテーブル + 変換用メタファイル  
 <table>タグを使って書かれたホームページ上の統計表
2. CSV ファイル + 変換用メタファイル
3. Excel ファイル + 変換用メタファイル

現在の変換用メタファイルは佐藤個人の試案で、国際的に認知されたものではなく、改良の余地が多々残されている。

# Web上の統計表と変換用メタデータ

[http://www.satolab.org/js/jsg2-samples/kawagoe\\_population.meta.json](http://www.satolab.org/js/jsg2-samples/kawagoe_population.meta.json)  
(変換用メタファイル)

<http://www.city.kawagoe.saitama.jp> の人口ページ



The screenshot shows the Kawagoe City website with a table of population data. The table has four columns: 年齢 (Age), 人口 (Population), 男 (Male), and 女 (Female). The data is for February 1, 2016.

年齢	人口	男	女
0～4歳	14,618	7,619	6,999
5～9歳	15,487	7,881	7,606
10～14歳	15,231	7,727	7,504
15～19歳	16,215	8,360	7,855
20～24歳	18,442	9,530	8,912
25～29歳	19,133	10,132	9,001
30～34歳	21,436	11,134	10,302
35～39歳	24,707	12,985	11,722
40～44歳	29,698	15,594	14,104
45～49歳	25,726	13,385	12,341
50～54歳	21,579	11,174	10,405
55～59歳	18,734	9,519	9,215
60～64歳	21,227	10,436	10,791
65～69歳	27,933	13,317	14,616

```
1) {
2)   "dataset": {
3)     "class": "dataset",
4)     "label": "川越市年齢別・男女別人口",
7)     "link": {
8)       "describes": [ {
9)         "type": "text/html",
10)        "href": "http://www.city.kawagoe.saitama.jp/shisei/
11)          toukeidata/jinkotokei/nenrei_danjo.html",
12)        "extension": {
13)          "table-location": { "selector": "table", "position": 0 },
14)          "property-location": {
15)            "updated": { "selector": "div.update p" }
16)          }
17)        }
18)      } ],
19)     "dimension": {
20)       "id": ["age", "sex", "time", "area", "concept"],
21)       "age": {
22)         "label": "年齢",
23)         "extension": { "category-location": { "colNum": 0 } }
24)       },
25)       "sex": {
26)         "label": "性",
27)         "extension": { "category-location": { "rowNum": 0 } }
28)       },
29)     },
30)   },
31) }
32) }
```

統計表ページのアドレス

更新日付の取得方法

年齢軸データは1列目

性別軸データは1行目



# WebデータをJSON-stat Viewerで利用

9

Web上のデータも、あたかもJSON-statを返す専用APIがあるかのように、使うことができる。

[http://www.satolab.org/js/jsg2-web/?url=http://www.satolab.org/js/jsg-samples/kawagoe\\_population.meta.json](http://www.satolab.org/js/jsg2-web/?url=http://www.satolab.org/js/jsg-samples/kawagoe_population.meta.json)

## JSON-stat Format Viewer

JSON-stat Format Viewer interface showing the URL input field and the 'connect' button. The interface includes tabs for Dataset, Dimension, Category, and Data. The URL input field contains the URL from the text above. Below the input field, there are sample files listed: index.json, galicia.json, oecd-canada-col.json, order.json, collection.json, galicia.csv. A list of dimensions is shown: 川越市年齢別・男女別人口, 年齢 (classification), 性 (classification), 時点 (time), 場所 (geo), 表章項目 (metric), and Data.

## 年齢階級別男女別人口

JSON-stat Viewer interface displaying the data table. The table shows population data for Kawagoe City, categorized by age group, sex, and population count. The table is titled 'Data' and has columns for 'Rows & Columns', 'Constants', and 'Data'. The 'Data' section shows a table with columns for '人口' (Population), '男' (Male), and '女' (Female). The rows represent age groups from 0~4歳 to 40~44歳.

	人口	男	女
0~4歳	14,618	7,619	6,999
5~9歳	15,487	7,881	7,606
10~14歳	15,231	7,727	7,504
15~19歳	16,215	8,360	7,855
20~24歳	18,442	9,530	8,912
25~29歳	19,133	10,132	9,001
30~34歳	21,436	11,134	10,302
35~39歳	24,707	12,985	11,722
40~44歳	29,698	15,594	14,104

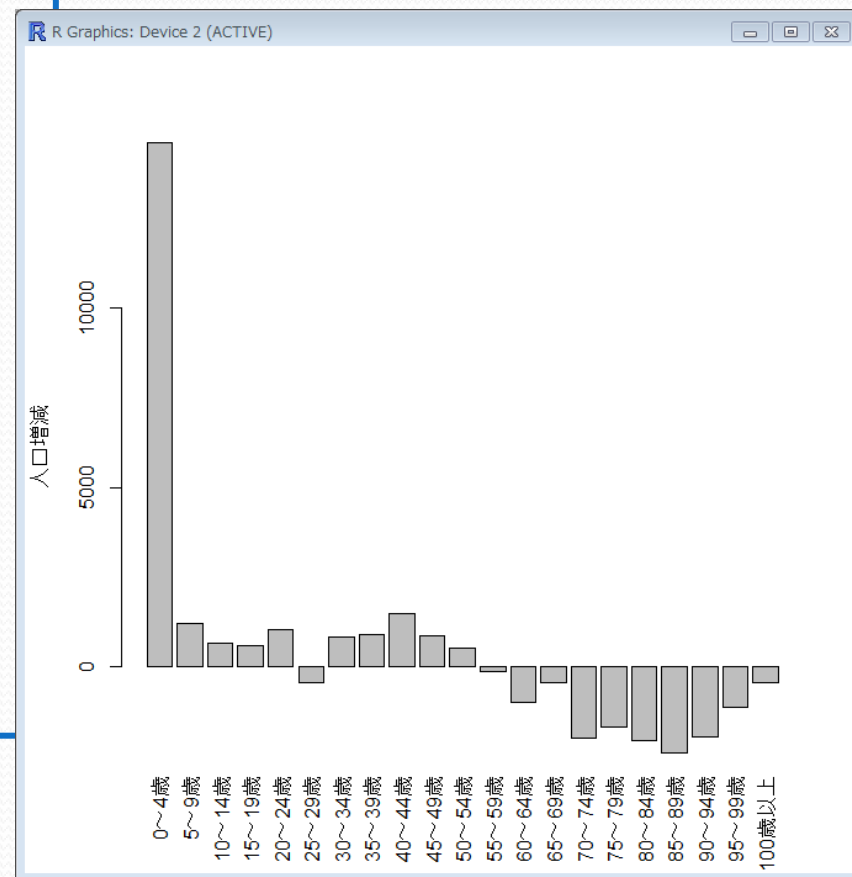
# jsg-estatとjsg-webを利用した データ統合

## 川越市の国勢調査以降の人口増減

```

1) > library(rjstat)
2) > # 川越市の最新人口 (川越市ホームページ)
3) > pop.url <- "http://www.satolab.org/js/jsg2-web/
    ?url=http://www.satolab.org/js/jsg2-samples/
    kawagoe_population.meta.json&dim_sex=0"
    # 性別は総数に固定(dim_sex=0)
4) > dss1 <- fromJSONstat(readLines(pop.url))
5) > pop <- dss1[[1]]
6) > popv <- pop[1:21, 6]
7) > # 川越市の国勢調査人口
8) > pop2010.url <- "http://www.satolab.org/js/jsg2-estat/
    ?statsDataId=0003041389&cdArea=11201
    &cdCato1=00710&cdCato2=000&cdCato3=000
    &cdCato4From=200&cdCato4To=600"
    # 地域は川越市(cdArea=11201)、性別は総数(cdCato2=000)
9) > dss2 <- fromJSONstat(readLines(pop2010.url))
10) > pop2010 <- dss2[[1]]
11) > # 国勢調査人口を5歳分ずらす
12) > pop2010v <- append(c(0), pop2010[1:20, 8])
13) > pop2010v[21] <- pop2010v[21] + pop2010[25, 8]
14) > # 両者の差をグラフで表す
15) > dpopv = popv - pop2010v
16) > barplot(dpopv, names.arg=pop[1:21, 1], ylab="人口増減", las = 3)
  
```

川越市HP → 最新時点の人口  
e-Stat → 国勢調査時点の人口



rjstat ライブラリを利用したR言語のプログラム

# 実証実験の結果

## 1. e-StatデータのJSON-stat化

- 200行足らずのプログラムで、ほぼ完璧に移し換えられた。

## 2. Web統計データのJSON-stat化

- HTML、CSV、Excelの統計表の多くを問題なく変換できた。

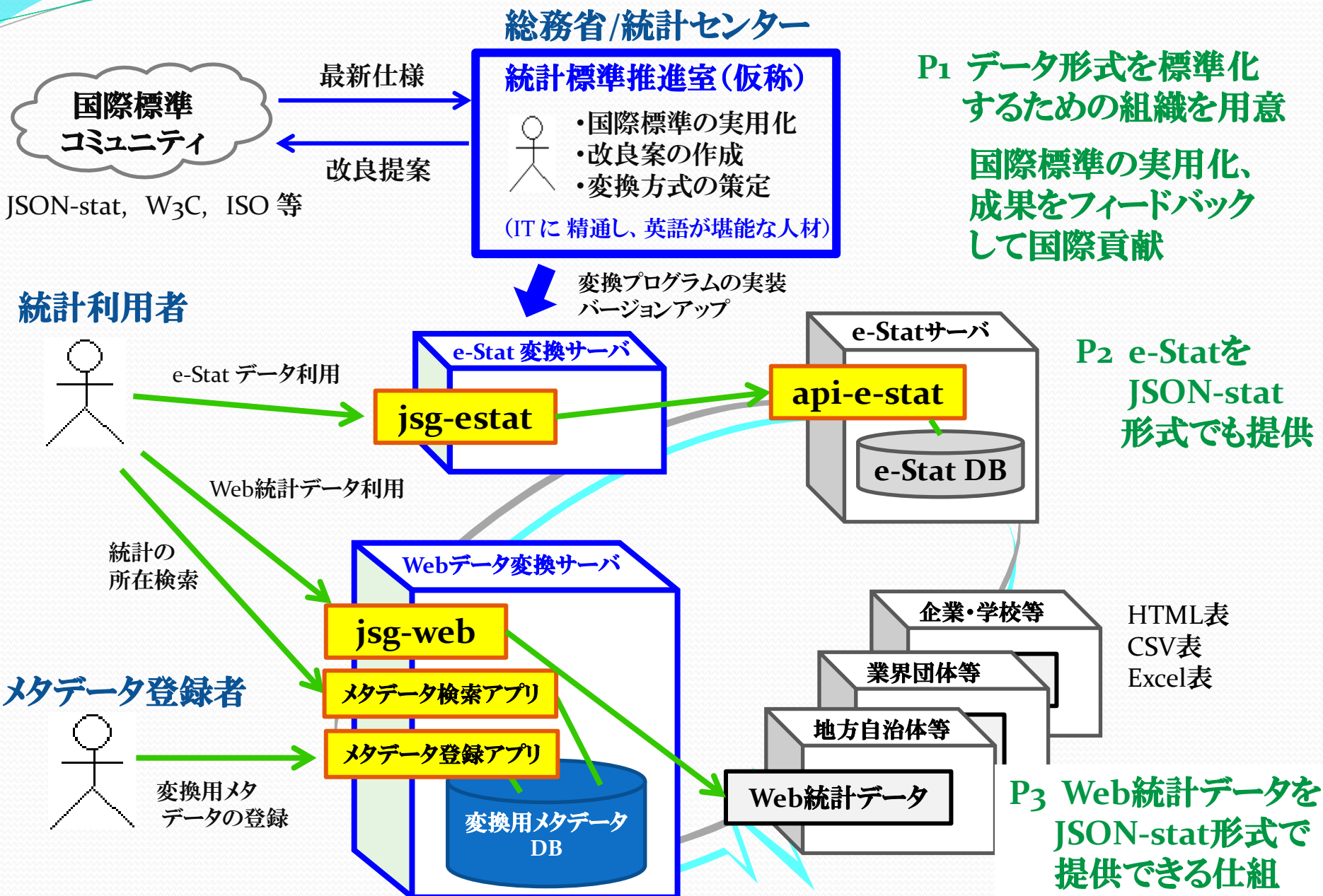
## 3. 複数ソースの統計データの統合利用

- 元データのデータ構造の違いを意識する必要はない。
- Rで利用する時はRのデータ構造のみ意識すれば良い。



**「共通フォーマットですべての統計データを利用する」  
JSON-statで現実的なものに**

# 提案の概要



# 提案の新しさと期待

JSON-statフォーマットでの統計データ提供

⇒ 日本では初

Web上の任意の統計データを同一フォーマットで利用可能に ⇒ おそらく世界でも初の試み

優れたデータ概念やアプリ



その改良や発展をフィードバック

➡ 統計データ利用技術の高度化に貢献

ご静聴ありがとうございました