



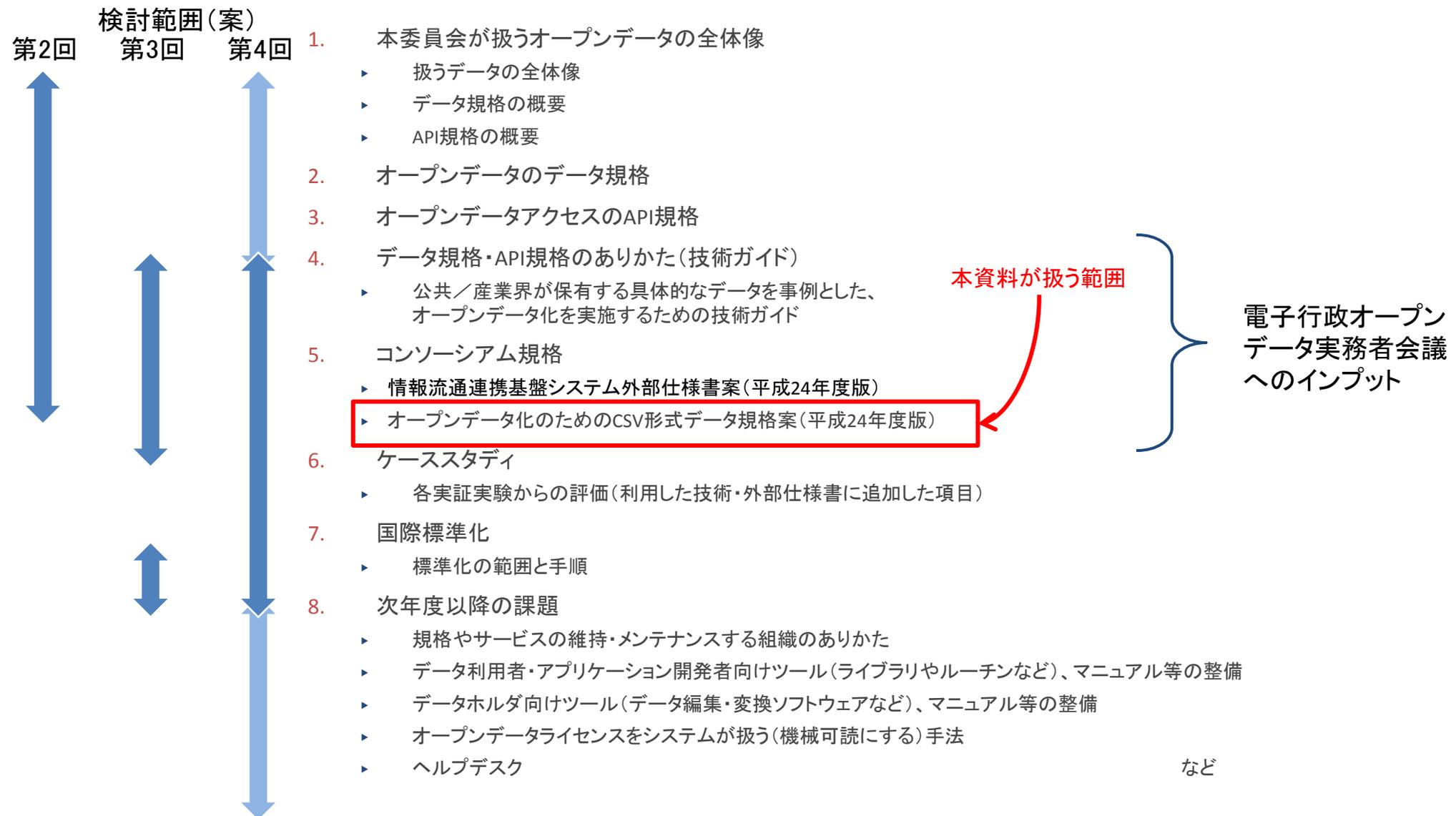
第三回 技術委員会資料

オープンデータ流通推進コンソーシアム オープンデータ化のための CSVデータ規格案

2013.2.26

オープンデータ流通推進コンソーシアム 事務局

技術委員会の論点(再掲)



CSVデータ規格の概要

■ 目的(Objective)

- ▶ 公共データの機械可読性を高めるために、データセットのキャプションやタイトル、単位を付与するためのCSV形式を規定する。
◇つまり、技術ガイドの表形式データレベル1(必須レベル)をレベル2(推奨レベル)にするための規格である。

■ 用語定義(Terms Definition)

- ▶ 「技術ガイド」と同じ。

■ 対象(Scope)

- ▶ 技術ガイドに記載されている、表形式データ・レベル1の条件を満たすデータを対象とする。
具体的には、下記を満たすCSVデータが対象である。
- ▶ テーブル全体に対する要件
 1. 1つのデータシートには、1種類の表のみを含むべきである。
 2. セルに、整形のためのスペース・改行、位取りのカンマを含めるべきではない。 ← (5),(9)
 3. 年の値には、西暦表記を備えるべきである。
 4. 数値やタイトル・単位以外の情報(ラベル・注釈など)を、セルに含めるべきではない。
- ▶ セルに関する条件
 5. すべてのセルが、他のセルと結合されているべきではない。
 6. 値がない場合を除き、データセルの値を空白にすべきではない。
- ▶ タイトルに関する条件
 7. データの内容を示すタイトルは、1行で構成すべきである。
 8. データの単位を明記すべきである。
 9. データセルの内容・単位・記数単位を示すタイトルを、それぞれ別の行に記載すべきである。

CSVデータ規格が定義するヘッダ

■ データセルの先頭に、下記のヘッダを付与することにより、キャプション、タイトル、およびセルに関する情報を機械可読にする。

- ▶ @で始まるヘッダは、1行で定義する。
- ▶ @@で始まるヘッダは、2行で定義する。

ヘッダ名	意味
@Caption	データセットのキャプション
@Creator	データセットの作成者
@Date	データセットの公開日
@Language	データセットの基本言語
@@Title	タイトル
@@Unit	カラムの単位(物理単位・貨幣単位)
@@Baseval	カラムの記数単位
@@Datatype	カラムのデータタイプ

メタデータとして広く使われている。

データセットを定義・解釈するために必須となる項目である。

- ▶ Notation3^(*1)記法に準じて、ヘッダの先頭にアットマーク”@”を付与している。

(*1) RDFの主語・述語・目的語を

<主語> <述語> <目的語>.

という形で表現する表記法。ネームスペース定義などを@prefix宣言として冒頭に記述できる。

ヘッダの規定 1/2

■ @Caption: データセットのキャプション

- ▶ 形式は「@Caption,キャプション名,言語コード」
- ▶ 言語コードはISO639-1に基づく値(2文字の言語コード)であり、省略可能。
◇省略した場合は、@Languageヘッダが指定する言語コードと同じ。

■ @Creator: データセットの作成者

- ▶ 形式は「@Caption,作成者名,言語コード」
- ▶ 言語コードはISO639-1に基づく値(2文字の言語コード)であり、省略可能。
◇省略した場合は、@Languageヘッダが指定する言語コードと同じ。

■ @Date: データセットの公開日

- ▶ 形式は「@Date,公開日」
- ▶ 日付はISO8610に基づく値(例: 2013-02-26)。

■ @Language: データセットの基本言語

- ▶ 形式は「@Language,言語コード」
- ▶ 言語コードはISO639-1に基づく値(2文字の言語コード)。

ヘッダの規定 2/2

■ @@Title: タイトル行

- ▶ 1行目の形式は「@@Title,言語コード」
- ▶ 2行目の形式は各タイトルの名称

■ @@Unit: カラムの単位

- ▶ 1行目の形式は「@@Unit,言語コード」
- ▶ 2行目の形式は各カラムの単位
 - ◇単位に記数単位を含めてはならない。
 - ◇物理単位のべき乗数は、そのままテキストで記述する(例: m/s²)。

■ @@Baseval: カラムの記数単位

- ▶ 1行目の形式は「@@Baseval」
- ▶ 2行目の形式は各カラムの記数単位。
 - ◇値を省略した場合は、「1」が指定されていると見なす。

■ @@Datatype: カラムのデータタイプ

- ▶ 1行目の形式は「@Datatype」
- ▶ 2行目の形式は、XML Schemaに基づくデータタイプ値(例: 整数であればxsd:integer)。

ヘッダの追加手順 1/3

1. 表形式データを、技術ガイド・表形式データのレベル1の条件を満たすように整形する。

2-2 都道府県別人口と人口増加率

都道府県	平成12年	17年	人口集中地 人口密度 (1)	人口増減率 (平成12 ~17年) (%)	22年	人口性比 (女性100に 対する男性) (2)	人口密度 (人/km ²)	人口増減率 (平成17 ~22年) (%)
	(1,000人)	(1,000人)			(1,000人)			
全国	126,926	127,768	84,331	0.7	128,057	94.8	a)343.4	0.2
北海道	5,683	5,628	4,108	-1.0	5,506	89.7	70.2	-2.2
青森	1,476	1,437	653	-2.6	1,373	88.9	142.4	-4.4
岩手	1,416	1,385	407	-2.2	1,330	91.3	87.1	-4.0
宮城	2,365	2,360	1,371	-0.2	2,348	94.3	322.3	-0.5
秋田	1,189	1,146	386	-3.7	1,086	88.5	93.3	-5.2
山形	1,244	1,216	504	-2.2	1,169	92.2	125.4	-3.9
福島	2,127	2,091	806	-1.7	2,029	94.3	147.2	-3.0
茨城	2,986	2,975	1,068	-0.4	2,970	99.3	487.2	-0.2
栃木	2,005	2,017	860	0.6	2,008	98.6	313.3	-0.4
群馬	2,025	2,024	801	-0.0	2,008	96.9	315.6	-0.8
埼玉	6,938	7,054	5,566	1.7	7,195	100.6	1,894.2	2.0
千葉	5,926	6,056	4,342	2.2	6,216	99.4	1,205.5	2.6
東京	12,064	12,577	12,329	4.2	13,159	98.0	6,015.7	4.6
神奈川	8,490	8,792	8,250	3.6	9,048	100.9	3,745.4	2.9
新潟	2,476	2,434	1,139	-1.8	2,374	93.6	188.7	-2.3
富山	1,121	1,112	398	-0.8	1,093	92.9	257.4	-0.7
石川	1,181	1,174	573	-0.6	1,170	93.4	279.5	-0.4
福井	829	822	333	-0.9	806	93.5	192.4	-1.9
山梨	888	885	305	-0.4	863	95.9	193.3	-2.4
長野	2,215	2,196	764	-0.8	2,152	94.6	158.7	-2.0
岐阜	2,108	2,107	822	-0.1	2,081	93.6	195.9	-0.7
愛知	3,767	3,792	2,216	0.7	3,765	97.0	483.9	-1.3
三重	1,704	1,755	5,480	3.0	7,411	99.9	1,434.8	2.2
滋賀	1,857	1,867	762	0.5	1,855	95.0	321.0	-0.7
京都	2,343	1,380	614	2.8	1,411	97.6	351.2	2.2
大阪	2,644	2,648	2,162	0.1	2,636	92.3	571.4	-0.4
兵庫	8,805	8,817	8,440	0.1	8,865	93.6	4,869.7	0.5
奈良	5,551	5,591	4,196	0.7	5,588	91.7	665.6	-0.0
和歌山	1,443	1,421	908	-1.5	1,401	90.0	379.5	-1.4
鳥取	1,070	1,036	408	-3.2	1,002	88.8	212.0	-3.3
徳島	613	607	207	-1.0	589	91.1	167.8	-3.0
高松	762	742	179	-2.5	717	91.6	107.0	-3.3
岡山	1,951	1,957	856	0.3	1,945	92.2	273.5	-0.6
広島	2,879	2,877	1,832	-0.1	2,861	93.3	337.4	-0.6
山口	1,528	1,493	708	-2.3	1,451	89.2	237.4	-2.8
香川	824	810	260	-1.7	785	90.3	189.4	-3.0
愛媛	1,023	1,012	330	-1.0	996	93.0	530.7	-1.6
高知	1,493	1,468	748	-1.7	1,431	88.8	252.1	-2.5
福岡	814	796	341	-2.2	764	88.6	107.6	-4.0
佐賀	5,016	5,050	3,525	0.7	5,072	89.4	1,019.0	0.4
熊本	877	866	246	-1.2	850	89.0	348.3	-1.9
大分	1,517	1,479	688	-2.5	1,427	87.5	347.5	-3.5
宮崎	1,859	1,842	805	-0.9	1,817	88.5	245.4	-1.3
鹿児島	1,221	1,210	537	-0.9	1,197	89.4	188.7	-1.1
沖縄	1,170	1,153	512	-1.4	1,135	88.5	146.7	-1.5
	1,786	1,753	689	-1.8	1,706	87.6	185.7	-2.7
	1,318	1,362	893	3.3	1,393	96.3	611.9	2.3



都道府県	2000年の人口	2005年の人口	2005年の人口集中地区の人口	2000~2005年の人口増減率	2010年の人口	2010年の人口性比 (女性100に 対する男)	2010年の人口密度	2005~2010年の人口増減率
	1000	1000	1000	%	1000		/km ²	%
全国	126926	127768	84331	0.7	128057	94.8	343.4	0.2
北海道	5683	5628	4108	-1.0	5506	89.7	70.2	-2.2
青森	1476	1437	653	-2.6	1373	88.9	142.4	-4.4
岩手	1416	1385	407	-2.2	1330	91.3	87.1	-4.0
宮城	2365	2360	1371	-0.2	2348	94.3	322.3	-0.5
秋田	1189	1146	386	-3.7	1086	88.5	93.3	-5.2
山形	1244	1216	504	-2.2	1169	92.2	125.4	-3.9
福島	2127	2091	806	-1.7	2029	94.3	147.2	-3.0
茨城	2986	2975	1068	-0.4	2970	99.3	487.2	-0.2
栃木	2005	2017	860	0.6	2008	98.6	313.3	-0.4
群馬	2025	2024	801	-0.0	2008	96.9	315.6	-0.8
埼玉	6938	7054	5566	1.7	7195	100.6	1894.2	2.0
千葉	5926	6056	4342	2.2	6216	99.4	1205.5	2.6
東京	12064	12577	12329	4.2	13159	98.0	6015.7	4.6
神奈川	8490	8792	8250	3.6	9048	100.9	3745.4	2.9
新潟	2476	2431	1139	-1.8	2374	93.6	188.7	-2.3
富山	1121	1112	398	-0.8	1093	92.9	257.4	-1.7
石川	1181	1174	573	-0.6	1170	93.4	279.5	-0.4
福井	829	822	333	-0.9	806	93.5	192.4	-1.9
山梨	888	885	305	-0.4	863	95.9	193.3	-2.4
長野	2215	2196	764	-0.8	2152	94.6	158.7	-2.0
岐阜	2108	2107	822	-0.1	2081	93.6	195.9	-1.3
愛知	3767	3792	2216	0.7	3765	97.0	483.9	-1.3
三重	1704	1755	5480	3.0	7411	99.9	1434.8	2.2
滋賀	1857	1867	762	0.5	1855	95.0	321.0	-0.7
京都	1343	1380	614	2.8	1411	97.6	351.2	2.2
大阪	2644	2648	2162	0.1	2636	92.3	571.4	-0.4
兵庫	8805	8817	8440	0.1	8865	93.6	4669.7	0.5
奈良	5551	5591	4196	0.7	5588	91.7	665.6	-0.0
和歌山	1443	1421	908	-1.5	1401	90.0	379.5	-1.4
鳥取	1070	1036	408	-3.2	1002	88.8	212.0	-3.3
徳島	613	607	207	-1.0	589	91.1	167.8	-3.0
高松	762	742	179	-2.5	717	91.6	107.0	-3.3
岡山	1951	1957	856	0.3	1945	92.2	273.5	-0.6
広島	2879	2877	1832	-0.1	2861	93.3	337.4	-0.6
山口	1528	1493	708	-2.3	1451	89.2	237.4	-2.8
香川	824	810	260	-1.7	785	90.3	189.4	-3.0
愛媛	1023	1012	330	-1.0	996	93.0	530.7	-1.6
高知	1493	1468	748	-1.7	1431	88.8	252.1	-2.5
福岡	814	796	341	-2.2	764	88.6	107.6	-4.0
佐賀	5016	5050	3525	0.7	5072	89.4	1019.0	0.4
熊本	877	866	246	-1.2	850	89.0	348.3	-1.9
大分	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
宮崎	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
鹿児島	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
沖縄	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
	1318	1362	893	3.3	1393	96.3	611.9	2.3

レベル1表形式データ形式に整形

オリジナルデータ

ヘッダの追加手順 2/3

2. カラムタイトルの上下に、必要なヘッダを追加する。

▶ 追加するヘッダとその箇所

追加するヘッダ名	追加する箇所	追加する値
@Caption	先頭行	都道府県別人口と人口増加率
@Creator	第2行	総務省統計局
@Date	第3行	2010-10-01
@Language	第4行	ja
@@Title	タイトル行の1行上	1行目: ja 2行目: (カラムのタイトルを示す行自体)
@@Unit	カラムの単位(物理単位・貨幣単位)が記載されている行の1行上	1行目: ja 2行目: (カラムの単位を示す行自体)
@@Baseval	カラムの記数単位が記載されている行の1行上	1行目: (なし) 2行目: (カラムの記数単位を示す行自体)
@@Datatype	カラムのデータタイプが記載されている行の1行上	1行目: (なし) 2行目: 第1列はxsd:string 整数値の列はxsd:integer それ以外の列はxsd:double

▶ Microsoft Excelで編集する際の留意事項

◇ヘッダ名の先頭に、アポストロフィ「'」をつける。

◆ 理由: @で始まる文字列をMicrosoft Excelが関数と判断するため。

◇日付値「2010-10-01」の先頭にアポストロフィ「'」をつける。

◆ Microsoft Excelが日付値を自動的に「2010/10/01」等の形式に変換してしまうため。

ヘッダの追加手順 3/3

3. CSV形式で出力する

都道府県	2000年の人口 (×1000)	2005年の人口 (×1000)	2005年の人口集中地区	2000~2005年の人口増減率 (%)	2010年の人口 (×1000)	2010年の人口性比 (女性100に対する男性)	2010年の人口密度	2005~2010年の人口増減率 (%)
全国	126926	127768	84331	0.7	128057	94.8	343.4	0.2
北海道	5693	5628	4108	-1.0	5506	99.7	70.2	-2.2
青森	1476	1437	653	-2.6	1373	88.3	142.4	-4.4
岩手	1416	1385	407	-2.2	1330	91.3	87.1	-4.0
宮城	2365	2360	1371	-0.2	2348	94.3	322.3	-0.5
秋田	1189	1146	386	-3.7	1086	89.5	93.3	-5.2
山形	1244	1216	504	-2.2	1169	92.2	125.4	-3.9
福島	2127	2091	806	-1.7	2029	94.3	147.2	-3.0
茨城	2986	2975	1068	-0.4	2970	99.3	487.2	-0.2
栃木	2005	2017	860	0.6	2008	96.6	313.3	-0.4
群馬	2025	2024	801	-0.0	2008	96.9	315.6	-0.6
埼玉	6938	7054	5566	1.7	7195	100.6	1894.2	2.0
千葉	5926	6056	4342	2.2	6216	99.4	1205.5	2.6
東京	12064	12577	12329	4.2	13159	98.0	6015.7	4.6
神奈川	8490	8792	8250	3.6	9048	100.9	3745.4	2.9
新潟	2476	2431	1139	-1.9	2374	93.6	189.7	-2.3
富山	1121	1112	398	-0.8	1093	92.9	257.4	-1.7
石川	1181	1174	573	-0.6	1170	93.4	279.5	-0.4
福井	829	822	333	-0.9	806	93.5	192.4	-1.9
山梨	888	885	305	-0.4	863	95.9	193.3	-2.4
長野	2215	2196	764	-0.9	2152	94.6	159.7	-2.0
岐阜	2108	2107	822	-0.1	2081	93.6	195.9	-1.3
静岡	3767	3792	2216	0.7	3765	97.0	483.9	-0.7
愛知	7043	7255	5480	3.0	7411	99.9	1434.8	2.2
三重	1657	1667	762	0.5	1655	95.0	321.0	-0.7
滋賀	1343	1360	614	2.8	1411	97.6	351.2	2.2
京都	2644	2648	2162	0.1	2636	92.3	571.4	-0.4
大阪	8805	8817	8440	0.1	8865	93.6	4669.7	0.5
兵庫	5551	5591	4198	0.7	5588	91.7	665.6	-0.0
奈良	1443	1421	308	-1.5	1401	90.0	379.5	-1.4
和歌山	1070	1036	408	-3.2	1002	90.8	212.9	-3.3
鳥取	613	607	207	-1.0	589	91.1	167.8	-3.0
島根	762	742	179	-2.5	717	91.6	107.0	-3.3
岡山	1951	1957	856	0.3	1945	92.2	273.5	-0.6
広島	2879	2877	1832	-0.1	2861	93.3	337.4	-0.6
山口	1528	1489	708	-2.9	1451	89.2	237.4	-2.9
徳島	824	810	260	-1.7	785	90.3	189.4	-3.0
香川	1023	1012	330	-1.0	996	93.0	530.7	-1.6
愛媛	1493	1468	748	-1.7	1431	88.8	252.1	-2.5
高知	914	796	341	-2.2	764	86.6	107.6	-4.0
福岡	5016	5050	3525	0.7	5072	89.4	1019.0	0.4
佐賀	877	866	246	-1.2	850	89.0	348.3	-1.9
長崎	1517	1479	688	-2.5	1427	87.5	347.5	-3.5
熊本	1959	1842	805	-0.9	1817	86.5	245.4	-1.9
大分	1221	1210	537	-0.9	1197	86.4	199.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1759	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

手順2の結果の表形式データ



```

@Caption,都道府県別人口と人口増加率,ja,.....
@Creator,総務省統計局,ja,.....
@Date,2010-10-01,.....
@Language,ja,.....
@Title,ja,.....
都道府県,2000年の人口,2005年の人口,2005年の人口集中地区の人口,2000~2005年の人口増減率,2010年の人口,2010年の人口性比
(女性100に対する男性),2010年の人口密度,2005~2010年の人口増減率
@@Baseval,.....
,1000,1000,1000,,1000,,
@@Unit,ja,.....
% ,/km2 %
@@Datatype,.....
xsd:string,xsd:integer,xsd:integer,xsd:integer,xsd:integer,xsd:integer,xsd:double,xsd:double,xsd:double
.....
全国,126926,127768,84331,0.7,128057,94.8,343.4,0.2
北海道,5693,5628,4108,-1.0,5506,99.7,70.2,-2.2
青森,1476,1437,653,-2.6,1373,88.3,142.4,-4.4
岩手,1416,1385,407,-2.2,1330,91.3,87.1,-4.0
宮城,2365,2360,1371,-0.2,2348,94.3,322.3,-0.5
秋田,1189,1146,386,-3.7,1086,89.5,93.3,-5.2
山形,1244,1216,504,-2.2,1169,92.2,125.4,-3.9
福島,2127,2091,806,-1.7,2029,94.3,147.2,-3.0
茨城,2986,2975,1068,-0.4,2970,99.3,487.2,-0.2
栃木,2005,2017,860,0.6,2008,96.6,313.3,-0.4
群馬,2025,2024,801,-0.0,2008,96.9,315.6,-0.6
埼玉,6938,7054,5566,1.7,7195,100.6,1894.2,2.0
千葉,5926,6056,4342,2.2,6216,99.4,1205.5,2.6
東京,12064,12577,12329,4.2,13159,98.0,6015.7,4.6
神奈川,8490,8792,8250,3.6,9048,100.9,3745.4,2.9
新潟,2476,2431,1139,-1.9,2374,93.6,189.7,-2.3
富山,1121,1112,398,-0.8,1093,92.9,257.4,-1.7
石川,1181,1174,573,-0.6,1170,93.4,279.5,-0.4
福井,829,822,333,-0.9,806,93.5,192.4,-1.9
山梨,888,885,305,-0.4,863,95.9,193.3,-2.4
長野,2215,2196,764,-0.9,2152,94.6,159.7,-2.0
岐阜,2108,2107,822,-0.1,2081,93.6,195.9,-1.3
静岡,3767,3792,2216,0.7,3765,97.0,483.9,-0.7
愛知,7043,7255,5480,3.0,7411,99.9,1434.8,2.2
三重,1657,1667,762,0.5,1655,95.0,321.0,-0.7
滋賀,1343,1360,614,2.8,1411,97.6,351.2,2.2
京都,2644,2648,2162,0.1,2636,92.3,571.4,-0.4
大阪,8805,8817,8440,0.1,8865,93.6,4669.7,0.5
兵庫,5551,5591,4198,0.7,5588,91.7,665.6,-0.0
奈良,1443,1421,308,-1.5,1401,90.0,379.5,-1.4
和歌山,1070,1036,408,-3.2,1002,90.8,212.9,-3.3
鳥取,613,607,207,-1.0,589,91.1,167.8,-3.0
島根,762,742,179,-2.5,717,91.6,107.0,-3.3
岡山,1951,1957,856,0.3,1945,92.2,273.5,-0.6
広島,2879,2877,1832,-0.1,2861,93.3,337.4,-0.6
山口,1528,1489,708,-2.9,1451,89.2,237.4,-2.9
徳島,824,810,260,-1.7,785,90.3,189.4,-3.0
香川,1023,1012,330,-1.0,996,93.0,530.7,-1.6
愛媛,1493,1468,748,-1.7,1431,88.8,252.1,-2.5
高知,914,796,341,-2.2,764,86.6,107.6,-4.0
福岡,5016,5050,3525,0.7,5072,89.4,1019.0,0.4
佐賀,877,866,246,-1.2,850,89.0,348.3,-1.9
長崎,1517,1479,688,-2.5,1427,87.5,347.5,-3.5
熊本,1959,1842,805,-0.9,1817,86.5,245.4,-1.9
大分,1221,1210,537,-0.9,1197,86.4,199.7,-1.1
宮崎,1170,1153,512,-1.4,1135,88.5,146.7,-1.5
鹿児島,1786,1759,689,-1.8,1706,87.6,185.7,-2.7
沖縄,1318,1362,893,3.3,1393,96.3,611.9,2.3
    
```

出力されたCSVデータ



OPEN DATA

オープンデータ流通推進コンソーシアム