

オープンデータ化のためのデータ作成に関する技術ガイド

オープンデータ流通推進コンソーシアム

2013.02.26.

目 次

1	概要	3
2	用語定義	4
3	規定範囲	7
3.1	表形式データ	7
3.2	文書形式データ	8
3.3	リアルタイムデータ	9
3.4	地理空間データ	10
4	技術レベル	12
4.1	レベル 1 (必須レベル)	12
4.2	レベル 2 (推奨レベル)	12
4.3	レベル 3 (理想レベル)	12
5	参照規定	13
6	表形式データのオープンデータ化に関する技術ガイド	14
6.1	定義	14
6.2	技術ガイド	16
6.2.1	表全体に関する要件 (レベル 1)	16
6.2.2	セルに関する要件 (レベル 1)	24
6.2.3	タイトルに関する要件 (レベル 1~2)	28
6.2.4	データの公開形式に関する要件 (レベル 2~レベル 3)	35
7	文書形式データのオープンデータ化に関する技術ガイド	38
7.1	定義	38
7.2	技術ガイド	39
7.2.1	レベル 1	39
7.2.2	レベル 2	41
7.2.3	レベル 3	43
8	地理空間データのオープンデータ化に関する技術ガイド	45
8.1	定義	45

8.2	技術ガイド	46
8.2.1	レベル 1	46
8.2.2	レベル 2	47
8.2.3	レベル 3	48
9	リアルタイムデータのオープンデータ化に関する技術ガイド	49
9.1	定義	49
9.2	技術ガイド	49
9.2.1	レベル 1	49
9.2.2	レベル 2	50
9.2.3	レベル 3	51
付録 A	表形式データの整形手順	52
A.1	表形式データをレベル 1 に整形する手順	52
A.1.1	複数のテーブルを分割	53
A.1.2	脚注・脚注番号・キャプションを削除	53
A.1.3	不必要なスペース・改行・カンマの除去	54
A.1.4	年の値を西暦で記載	55
A.1.5	セルの結合を解除	55
A.1.6	省略されたセルをコピー	55
A.1.7	タイトルを 1 行にまとめる	55
A.1.8	データの単位を明記	55
A.1.9	単位や記数単位を別の行に移行	56
A.2	表形式データをレベル 2 に整形する手順	57
付録 B	文書データの整形手順	60
B.1	文書データをレベル 1 に整形する手順	60
B.2	文書データをレベル 2 に整形する手順	61

1 概要

本書は、幅広いアプリケーションやサービスが有効に利活用するために、政府自治体、企業等、さまざまな組織が保持するデータをオープンデータ化するための技術的な要求事項、およびそれを実現するための手順を示す。

技術的な要求事項については、3段階のレベルを用意し、それを満たすための事項を明確化する。技術的な内容や対象とする組織を限定するものではないが、扱う対象となる規定範囲は、主に政府自治体が公表するデータを念頭においている。また本書は、データによる情報伝達機能や、情報伝達の曖昧性排除を最大化させることを目的としているのではなく、即座にオープンデータ化できることを最重要視し、データ作成の容易性に重点をおいている。従って、オープンデータ化に関与する主体のデータリテラシ向上に応じて、本書の内容も見直されるべきである。

2 用語定義

本書が使用する用語の定義を、表 1 に示す。

表 1: 用語定義

用語	意味
二次利用	公開されているデータを引用・転載・コピーして利用すること。
機械可読	コンピュータプログラムに代表される機械が、データを自動的に解読し、技術的に二次利用できること。
オープンデータ化	データを、機械可読な形式に加工すること。
表形式データ	行と列の、縦横 2 次元状に配列されたデータ。
文書形式データ	1 次元状に配列された文字を主な構成要素とし、一部図や費用などを含み、人間がそれを読むことによって人間に何らかの作用を与えることを目的としたデータ。
リアルタイムデータ	値が時刻に応じて刻々と変化するデータ。
地理空間データ	2 次元平面状の地図の地点や領域と関連づけられたデータ。
メタデータ	あるデータ自身について記述した、抽象度の高い付加データ。
オープンな形式	仕様が公開されており、それが国際標準化団体によって標準化されている形式。
表計算ソフト	数値データの集計・分析に用いられるアプリケーションソフトウェア。画面上に格子状のマス目を表示し、そのマス目にデータを入れることにより表を作成する機能を有する。
キャプション (Caption・表題)	表形式データ全体を表す短い説明。
カラム (Column)	表形式データの、縦方向の列。
ロウ (Row)	表形式データの、横方向の行。
セル (Cell)	表形式データの各項目。表計算ソフトでは、個々のマス目として表現される。
データセル	表形式データにおいて、データ本体の値が格納されるべきセル。
タイトル (Title・題目)	表形式データの、各カラムの冒頭。カラムに含まれるデータセルの内容や単位を説明する。
タイトル行	タイトルが配置された行。
データセット (Dataset)	機械がセルを取得する対象となる、表形式データの基本単位。表計算ソフトでは、1 シートにあたる。CSV 形式ファイルでは、1 ファイルにあたる。

テーブル (Table・表)	1 行以上からなるタイトル行, 1 行以上のデータセル, 0 行以上の脚注からなる, セルの集合.
脚注	表形式データに付与する, タイトルやデータセルに対する補助説明.
脚注番号	タイトルやデータセルに付与する, 脚注と結びつけるための番号.
単位	数値の基準となる, 約束された一定量. 「m」「g」に代表される物理単位, または「円」「ドル」に代表される貨幣単位である.
記数単位	データセルの値の桁を示す数. たとえば, 単位が「百万円」である場合, 記数単位は「1,000,000」である. 実際の値は, データセルの値に記数単位を乗じたものである.
データ型	機械が扱うデータの形式. 文字列型・整数型・実数型・日付型などを指す.
CSV (Comma Separated Values) [1]	表形式数値データの行を改行で区切り, セルを半角のカンマ「,」で区切る, テキストデータの表現形式. RFC 4180[1]により標準化されている.
XML (Extensible Markup Language) [3]	個別の目的に応じたマークアップ言語作成のため, 汎用的に使うことができる仕様, および仕様により策定される言語の名称.
RDF (Resource Description Framework) [4]	主語・述語・目的語の 3 つ組で物事を表現するモデル. Web 技術の標準化団体 World Wide Web Consortium (W3C) が標準化している.
見出し	文章において内容の要点を非常に短い言葉にまとめ, 本文より大きな字で章や節の最初に置かれる言葉. 大きい方から編 (部)・章・節・項・目といった名称が付けられる慣習がある.
プレインテキスト	コンピュータ上で文章を扱うための一般的なファイルフォーマット, または文字列の形式の 1 つ. 文字以外の情報, たとえば文字の色や大きさ, 形状, 文章に含まれる図表などの情報を含まない.
タグ	文章に対する構造 (章・節・図表など) や見栄え (色・大きさ・形状など) に関する指定, またはその指定方法.
マークアップ言語	文章の構造や見栄えに関する指定を, 文章とともにテキストファイルに記述するための言語.
ワープロソフト	コンピュータ上で動作する, 文章の入力・編集・印刷機能 (ワードプロセッサ機能) を実現したソフトウェア.
測地系	位置を緯度・経度・標高からなる座標で表すときの前提条件. 代表的なものとしては, WGS84 (World Geodetic System 1984) や, JGD2000 (Japan Geodetic Datum 2000) などがある.

GIS(Geographic Information System)	コンピュータ上に地図情報やさまざまな付加情報を持たせ，作成・保存・利用・管理し，地理情報を参照できるように表示・検索機能をもった，地理情報システムのこと．
GML (Geography Markup Language) [7]	地理的特徴を表現する XML ベースのマークアップ言語． Open Geospatial Consortium (OGC)によって開発され，現在は ISO 19136[7]として標準化されている．
KML[8]	3 次元地理空間情報の表示を管理するための，XML ベースのマークアップ言語．
shape 形式	米国の ESRI 社が規定する GIS 標準データフォーマット形式．事実上の業界標準であり，多くの企業が発売している GIS システムは，この形式をサポートしている．

3 規定範囲

本書は、以下の 4 種類のデータをオープンデータ化する手法に関する技術要件を規定する。

- 表形式データ
- 文書形式データ
- リアルタイムデータ
- 地理空間データ

それぞれのデータの詳細について、以下に記す。

3.1 表形式データ

表形式データとは、行と列の縦横 2 次元状に配列されたデータである。統計の結果、あるいは計測の結果得られた数値データである。

表形式データの例を図 1 に挙げる。

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 増減率 (平成12 ~17年) (%)	22年	人口性比 (女性100に 対する男性)	人口密度 (人/km ²)	人口 増減率 (平成12 ~22年) (%)
	(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
青森県	1,476	1,437	653	-2.6	1,373	88.9	142.4	-4.4
岩手県	1,416	1,385	407	-2.2	1,330	91.3	87.1	-4.0
宮城県	2,365	2,360	1,371	-0.2	2,348	94.3	322.3	-0.5
秋田県	1,189	1,146	385	-3.7	1,085	88.5	93.3	-5.2
山形県	1,244	1,216	504	-2.2	1,169	92.2	125.4	-3.9
福島県	2,127	2,091	806	-1.7	2,029	94.3	147.2	-3.0
茨城県	2,886	2,975	1,068	-0.4	2,970	99.3	487.2	-0.2
栃木県	2,005	2,017	860	0.6	2,008	98.6	313.3	-0.4
群馬県	2,025	2,024	801	-0.0	2,008	96.9	315.6	-0.8
埼玉県	6,938	7,054	5,566	1.7	7,195	100.6	1,894.2	2.0
千葉県	5,926	6,056	4,342	2.2	6,216	99.4	1,205.5	2.6
東京都	12,064	12,577	12,329	4.2	13,159	98.0	6,015.7	4.6
神奈川県	8,490	8,792	8,250	3.6	9,048	100.9	3,745.4	2.9
新潟県	2,476	2,431	1,139	-1.8	2,374	93.6	188.7	-2.3
富山県	1,121	1,112	398	-0.8	1,093	92.9	257.4	-1.7
石川県	1,181	1,174	573	-0.6	1,170	93.4	278.5	-0.4
福井県	829	822	333	-0.9	806	93.5	192.4	-1.9
山梨県	888	885	305	-0.4	863	95.9	193.3	-2.4
長野県	2,215	2,196	764	-0.8	2,152	94.6	158.7	-2.0
岐阜県	2,108	2,107	822	-0.1	2,081	93.6	185.9	-1.3
静岡県	3,767	3,792	2,216	0.7	3,765	97.0	483.9	-0.7
愛知県	7,043	7,255	5,480	3.0	7,411	99.9	1,434.8	2.2
三重県	1,857	1,867	762	0.5	1,855	95.0	321.0	-0.7
滋賀県	1,343	1,380	614	2.8	1,411	97.6	351.2	2.2
京都府	2,644	2,648	2,162	0.1	2,636	92.3	571.4	-0.4
大阪府	8,805	8,817	8,440	0.1	8,855	93.6	4,669.7	0.5
兵庫県	5,551	5,581	4,198	0.7	5,588	91.7	865.6	-0.0
奈良県	1,443	1,421	808	-1.5	1,401	90.0	378.5	-1.4
和歌山県	1,070	1,036	408	-3.2	1,002	88.8	212.0	-3.3
鳥取県	613	607	207	-1.0	589	91.1	167.8	-3.0
島根県	762	742	179	-2.5	717	91.6	107.0	-3.3
岡山県	1,951	1,957	856	0.3	1,945	92.2	273.5	-0.6
広島県	2,879	2,877	1,532	-0.1	2,861	93.3	337.4	-0.6
山口県	1,528	1,493	708	-2.3	1,451	89.2	237.4	-2.8
徳島県	824	810	260	-1.7	785	90.3	189.4	-3.0
香川県	1,023	1,012	330	-1.0	996	93.0	530.7	-1.6
愛媛県	1,493	1,468	748	-1.7	1,431	88.8	252.1	-2.5
高知県	814	796	341	-2.2	764	88.6	107.6	-4.0
福岡県	5,016	5,050	3,525	0.7	5,072	89.4	1,019.0	0.4
佐賀県	877	866	246	-1.2	850	89.0	348.3	-1.9
長門県	1,517	1,479	688	-2.5	1,427	87.5	347.5	-3.5
熊本県	1,859	1,842	805	-0.9	1,817	88.5	245.4	-1.3
大分県	1,221	1,210	537	-0.9	1,197	89.4	188.7	-1.1
宮崎県	1,170	1,153	512	-1.4	1,135	86.5	146.7	-1.5
鹿児島県	1,785	1,753	889	-1.8	1,705	87.6	185.7	-2.7
沖縄県	1,318	1,362	893	3.3	1,393	96.3	611.9	2.3

10月1日現在。① 人口密度の高い基本単位的区（人口密度が1km²当たり4,000人以上）が市町村の境域内で互いに隣接して、国勢調査時に人口5,000人以上を有する地域。② 算出に用いた国勢調査は、国土交通省国土情報「平成22年全国都道府県市区町村別国勢調査」による。また、海外未定地域については、総務省統計局に於いて国境を指定。a) 国勢調査。資料 総務省統計局統計課国勢調査課「国勢調査報告」「国勢調査」(ホームページ)

図 1: 表形式データの例¹

¹総務省統計局刊行、総務省統計研修所編集「日本の統計 2012」内「2-2 都道府県別人口と人口増加率」による。 <http://www.stat.go.jp/data/nihon/02.htm>

3.2 文書形式データ

文書形式データとは、1次元状に配列された文字を主な構成要素とし、一部図や費用などを含み、人間がそれを読むことによって人間に何らかの作用を与えることを目的としたデータである。例えば、白書や報告書である。

文書形式データの例を図 2 に挙げる。

第2節 グローバルに展開する ICT 市場

この10年間で、グローバル市場は開発途上国・新興国を中心として成長を遂げつつある。特に、その中のインターネットや携帯電話などのICT市場も拡大を続けており、潜在的な成長力に対し、世界的な注目が集まっている。こうした背景の下、先進国、途上国問わず、ICTに対する戦略的な取組が進められている。

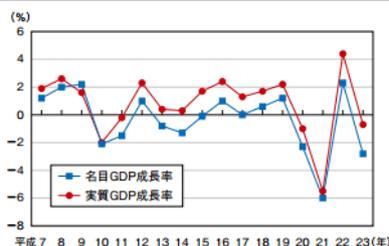
ここでは、グローバルに見たICT市場の状況を概観した上で、開発途上国において貧困からの脱出にICTが寄与している状況を分析する。そして、ICTの有する成長けん引力を踏まえ、各国・各地域がICTに対する戦略的な取組を強化している状況を分析する。

1 我が国社会経済の現状

(1) 我が国のポジションの低下

我が国経済の状況を実質及び名目GDP成長率の推移から見てみると、緩やかなデフレ状況が続く中、名目GDP成長率を実質GDP成長率が上回る状況が続いている(図表1-2-1-1)。近年では、平成20年及び平成21年には、リーマンショックの影響により、実質及び名目成長率いずれもマイナス成長になるなど大きな落ち込みがみられた。平成22年にはプラス成長(名目:2.3%、実質:4.4%)に回復したものの、平成23年は再びマイナス成長(名目:-2.8%、実質:-0.7%)となっている。

図表 1-2-1-1 我が国の実質GDP成長率及び名目GDP成長率の推移



内閣府 国民経済計算により作成
<http://www.esri.cao.go.jp/jp/sna/menu.html>

図 2: 文書データの例²

²総務省「平成24年版 情報通信白書」による。

<http://www.soumu.go.jp/johotsusintokei/whitepaper/index.html>

3.3 リアルタイムデータ

リアルタイムデータとは、時刻に応じて刻々と変化するデータである。例えば、センサーやメータによって観測されたデータである。また、気象庁が観測・公開している気象データや、スマートグリッドにおける電力メータのデータなども、リアルタイムデータである。

リアルタイムデータの例を図 3 に挙げる。



図 3: リアルタイムデータの例³

³気象庁「地震情報」による。

<http://www.jma.go.jp/jp/quake/20130209134635391-091343.html>

3.4 地理空間データ

地理空間データとは、2次元平面状の地図の地点や領域と関連づけられたデータである。たとえば、行政界や路線図である。

地理空間データの例を図4に挙げる。

```

<?xml version="1.0" encoding="UTF-8"?
<ksj:Dataset gml:id="N02Dataset"
  xmlns:ksj="http://nlftp.mlit.go.jp/ksj/schemas/ksj-app"
  xmlns:gml="http://www.opengis.net/gml/3.2"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=
    "http://nlftp.mlit.go.jp/ksj/schemas/ksj-appKsjAppSchema-N02-V2_0.xsd">
<gml:description>国土数値情報(鉄道)データ 2011 年度</gml:description>
<gml:boundedBy>
<gml:EnvelopeWithTimePeriod srsName="JGD2000 / (B, L)" frame="GC / JST">
<gml:lowerCorner>20 123</gml:lowerCorner>
<gml:upperCorner>46 154</gml:upperCorner>
<gml:beginPosition calendarEraName="西暦">1900</gml:beginPosition>
<gml:endPosition indeterminatePosition="unknown" />
</gml:EnvelopeWithTimePeriod></gml:boundedBy>
<gml:Curve gml:id="cv_rss1">
<gml:segments>
<gml:LineStringSegment>
<gml:posList>
26.21454 127.67948
26.21474 127.6797
26.2148 127.67975
26.21728 127.68217
26.21862 127.68357
26.21891 127.6839
26.21905 127.68419
</gml:posList>
</gml:LineStringSegment></gml:segments></gml:Curve>

```

図 4: 地理空間データの例⁴

⁴国土交通省「国土数値情報（鉄道データ）」による。
<http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02.html>

4 技術レベル

本書は、データが満たすべき技術要件を、3つのレベルに分けて規定する。その詳細を以下に記す。

4.1 レベル1（必須レベル）

レベル1（必須レベル）を満たすデータを見た人間は、機械にそのデータを解読させるための方針を立てられる。

4.2 レベル2（推奨レベル）

レベル2（推奨レベル）を満たすデータを取得した機械は、そのデータの項目（タイトル）・値・単位を正しく解読できる。

4.3 レベル3（理想レベル）

レベル3（理想レベル）を満たすデータを取得した機械は、そのデータの意味を理解でき、他のデータと自動的にマッシュアップできる。

5 参照規定

- [1] Y. Shafranovich. Common Format and MIME Type for Comma-Separated Values (CSV) Files. RFC 4180, 2005.
<http://www.ietf.org/rfc/rfc4180.txt>
- [2] Dave Raggett, Arnaud Le Hors, and Ian Jacobs. HTML 4.01 Specification. W3C Recommendation, 1999. <http://www.w3.org/TR/html401/>
- [3] Tim Bray, JeJean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau, and John Cowan. Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation, 2006. <http://www.w3.org/TR/xml11/>
- [4] Graham Klyne and Jeremy J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, 2004.
<http://www.w3.org/TR/rdf-concepts/>
- [5] Information technology -- Open Document Format for Office Applications (OpenDocument) v1.0. ISO 26300:2006.
- [6] Information technology -- Document description and processing languages -- Office Open XML File Formats. ISO 29500:2012.
- [7] Geographic information -- Geography Markup Language (GML). ISO 19136:2007.
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32554
- [8] Tim Wilson. OGC KML. Open Geospatial Consortium 07-147r2, 2008.
<http://www.opengeospatial.org/standards/kml>
- [9] 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 形式データ規格. 2013.

6 表形式データのオープンデータ化に関する技術ガイド

本章では、表形式データをオープンデータ化する際に、留意すべき点をレベルごとにまとめる。

6.1 定義

表形式データとは、行と列の縦横 2 次元状に配列されたデータである (図 5)。表形式データを構成する各要素の名称を、以下の通り定義する。

- キャプション (表題) :
 - 表形式データ全体を表す短い説明。
- カラム (Column) :
 - 表形式データの、縦方向の列。
- セル (Cell) :
 - 表形式データの各項目。表計算ソフトでは、個々のマス目として表現される。
- データセル (Data Cell) :
 - 表形式データにおいて、データ本体数値が格納されるべきセル。
- タイトル (Title・題目) :
 - 表形式データの、各カラムの冒頭。カラムに含まれるデータセルの内容や単位を説明する。
- タイトル行:
 - タイトルが配置された行。
- データセット (Dataset) :
 - 表形式データの、各カラムの冒頭。カラムに含まれるデータセルの内容や単位を説明する。
- テーブル (Table・表) :
 - 1 行以上からなるタイトル行, 1 行以上のデータセル, 0 行以上の脚注からなる, セルの集合。
- 脚注:
 - 表形式データに付与する, タイトルやデータセルに対する補助説明。
- 脚注番号:
 - タイトルやデータセルに付与する, 脚注と結びつけるための番号。
- 単位:
 - 数値の基準となる, 約束された一定量。「m」「g」に代表される物理単位, または「円」「ドル」に代表される貨幣単位である。
- 記数単位:

- ▶ データセルの値の桁を示す数. たとえば, 単位として「百万円」と書かれているカラムの記数単位は「1,000,000」である. 実際の値は, データセルの値に記数単位を乗じたものである.

キャプション(表題) → 2-2 都道府県別人口と人口増加率

タイトル行 → 都道府県

記数単位 → (1,000人)

ロウ(行) → 人口集地区

カラム(列) → 平成12年, 17年, 人口集地区, 人口増加率(平成12~17年), 22年, 人口性比(女性100に對する男), 人口密度(人/km²), 人口増加率(平成17~22年)

脚注番号 → a)

セル → 125.4

データセル → 6,015.7

脚注 → 10月1日現在。1) 人口密度の高い基本単位数(人口密度が1km²当たり約4,000人以上)が市区町村の境域内と互いに隣接して、国勢調査時に人口5,000人以上を有する地域。2) 算出に用いた面積は、国土交通省国土地理院「平成22年全国都道府県市区町村別面積調」による。また、境界未定地域については、総務省統計局において面積を推定。a) 青森県、色丹島、国後島、択捉島及び竹島を除き算出。資料 総務省統計局統計調査部国勢調査課「国勢調査報告」「国勢調査」(ホームページ)

都道府県	平成12年 (1,000人)	17年 (1,000人)	人口集地区 (1)	人口増加率 (平成12~17年) (%)	22年 (1,000人)	人口性比 (女性100に對する男)	人口密度 (人/km ²)	人口増加率 (平成17~22年) (%)
全 国	126,926	127,768	84,331	0.7	128,057	94.8	a)343.4	0.2
北海道	3,600	3,628	4,100	-1.0	3,306	89.1	170.2	-2.2
青 森 県	1,476	1,437	653	-2.6	1,373	88.9	142.4	-4.4
岩 手 県	1,416	1,385	407	-2.2	1,330	91.3	87.1	-4.0
宮 城 県	2,365	2,360	1,371	-0.2	2,348	94.3	322.3	-0.5
秋 田 県	1,189	1,146	386	-3.7	1,086	88.5	93.3	-5.2
山 形 県	1,244	1,216	504	-2.2	1,169	92.2	125.4	-3.9
福 島 県	2,127	2,091	806	-1.7	2,029	94.3	147.2	-3.0
茨 城 県	2,986	2,975	1,068	-0.4	2,970	93.3	487.2	-0.2
栃 木 県	2,005	2,017	860	0.6	2,008	98.6	313.3	-0.4
群 馬 県	2,025	2,024	801	-0.0	2,008	96.9	315.6	-0.8
埼 玉 県	6,938	7,054	5,566	1.7	7,195	100.6	1,894.2	2.0
千 葉 県	5,926	6,056	4,342	2.2	6,216	93.4	1,205.5	2.6
東 京 都	12,064	12,577	12,329	4.2	13,159	98.0	6,015.7	4.6
神 奈 川 県	8,490	8,792	8,250	3.6	9,048	100.9	3,745.4	2.9
愛 知 県	3,436	3,421	1,139	-0.4	3,423	93.6	189.7	-0.2
福 岡 県	5,016	5,050	3,525	0.7	5,072	89.4	1,019.0	0.4
佐 賀 県	877	866	246	-1.2	850	89.0	348.3	-1.9
長 崎 県	1,517	1,479	688	-2.5	1,427	87.5	347.5	-3.5
熊 本 県	1,859	1,842	805	-0.9	1,817	88.5	245.4	-1.3
大 分 県	1,221	1,210	537	-0.9	1,197	89.4	188.7	-1.1
宮 崎 県	1,170	1,153	512	-1.4	1,135	88.5	146.7	-1.5
鹿 児 島 県	1,786	1,753	689	-1.8	1,706	87.6	185.7	-2.7
沖 縄 県	1,318	1,362	893	3.3	1,393	96.3	611.9	2.3

図 5: 表形式データの各要素の名称定義

6.2 技術ガイド

本節では、表形式データを機械可読にする際に留意すべき点を挙げる。

6.2.1 表全体に関する要件（レベル1）

ガイド1(レベル1: 必須レベル)

1つのデータセットには、1つの表のみを含む。(複数個の表を含まない)

【解説】

図6のデータセットは、複数の表を含んでいる。このようなデータセットをコンピュータが解読するためには、表の切れ目を扱う必要があり、解読手順が複雑になる。このため、1つのデータセットは、1つの表のみを持つべきである。複数の表が必要である場合は、その数だけのデータセットに分割すべきである(図7)。

1. 砕石及び再生骨材の生産・出荷・月末在庫						2-1. 原石、コンクリート塊等 再生骨材原料	
用途	品 種 別	生産数量	販売数量	在庫数量	原石、コンクリート塊等	採取・受入	
砕 石	単 粒 度 砕 石	1.012	1.014	1.041	原 石	砂 岩	1.014
	道 路 用 粒 度 調 整 砕 石	1.035	1.019	1.081		安 山 岩	1.036
	ク ラ ッ シ ャ ー ラ ン	1.040	1.028	1.059		輝 緑 岩	1.000
	ス ク リ ー ニ ン グ ス	1.011	1.009	1.007		粘 板 岩	1.008
	そ の 他	1.039	1.027	1.030		玄 武 岩	1.000
再 生 骨 材	単 粒 度 砕 石	1.010	1.011	1.043		粗 面 岩	1.000
	コ ン ク リ ー ト 用 砕 石	1.044	1.047	1.018		花 こ う 岩	1.164
	割 栗 石	1.029	1.021	1.034		石 灰 石	1.001
	そ の 他	1.010	1.011	1.011		け い 石	1.001
	出 荷 金 額 (千 円)		1.029			そ の 他	1.048
道 路 用 ・ そ の 他 用	1.099	1.085	1.164	コ ン ク リ ー ト 塊 等	再 生 骨 材 原 料	1.123	
コ ン ク リ ー ト 用	1.000	1.000	1.000				
出 荷 金 額 (千 円)		1.081					

2-2. 電力・燃料			4. 設備公称能力、自家発電能力		
区 分	消 費 量		区 分	能 力	
電 力	1.032		一 次 ク ラ ッ シ ャ ー (t / 時 間)	1.039	
灯 油	1.062		砕 石 製 造 乾 式	1.035	
軽 油	1.024		プ ラ ント (t / 時 間)	湿 式	1.075
重 油	1.055		自 家 発 電 (k w)	1.017	

図6: 1つのデータセルに複数の表がある(ガイド1を満たさない)例⁵

⁵経済産業省「平成24年7-9月砕石動態統計調査」による。
<http://www.meti.go.jp/statistics/sei/saiseiki/result-2.html>

1. 砕石及び再生骨材の生産・出荷・月末在庫				
用途	品 種 別	生産数量	販売数量	在庫数量
砕石	単粒度砕石	1.012	1.014	1.041
	粒度調整砕石	1.035	1.019	1.081
	道路用クラッシャーラン	1.040	1.028	1.059
	スクリーニングス	1.011	1.009	1.007
	その他	1.039	1.027	1.030
	コンクリート用単粒度砕石	1.010	1.011	1.043
	砕砂	1.044	1.047	1.018
	割薬石	1.029	1.021	1.034
	その他	1.010	1.011	1.011
	出荷金額(千円)			1.029
再生骨材	道路用・その他用	1.099	1.085	1.184
	コンクリート用	1.000	1.000	1.000
	出荷金額(千円)			1.081

採取・受入	1.014
	1.038
	1.000
	1.008
	1.000
	1.000
	1.184
	1.001
	1.001
	1.048
コンクリート塊等再生骨材原料	1.123

燃料消費量	1.032	設備公称能力、自家発電能力	
	1.062		
軽油	1.024	分能力	
重油	1.055	クラッシャー(t/時間)	1.039
		乾式	1.035
		湿式	1.075
		自家発電(kw)	1.017

図 7: 図 6 の表を分割(ガイド 1 を満たす)

ガイド2(レベル1: 必須レベル)

セルに、整形のためのスペース・改行、および位取りのためのカンマを含めない。

【解説】

図8の都道府県カラムにある「青森」「岩手」などのデータセルは、整形のための空白を含んでいる。データセルに含まれる空白や改行に意味があるのか否かは、機械は判別できない。また、数値データには位取りのためのカンマが含まれている。カンマを除かなければ、機械はそのデータは数値として認識できない。このため、機械の解読に必要な空白や改行、カンマは、除くべきである(図9)。

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 中 区 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)	人口密度 (人/km ²) 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	2,431	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	-1.7
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	-1.3
静岡	3,767	3,792	2,216	0.7	3,765	97.0	483.9	-0.7
愛知	7,043	7,255	5,480	3.0	7,411	99.9	1,434.8	2.2
三重	1,857	1,867	762	0.5	1,855	95.0	321.0	-0.7
滋賀	1,343	1,380	614	2.8	1,411	97.6	351.2	2.2
京都	2,644	2,648	2,162	0.1	2,636	92.3	571.4	-0.4
大阪	8,805	8,817	8,440	0.1	8,865	93.6	4,669.7	0.5
兵庫	5,551	5,591	4,198	0.7	5,588	91.7	665.6	-0.0
奈良	1,443	1,421	908	-1.5	1,401	90.0	379.5	-1.4
和歌山	1,070	1,036	408	-3.2	1,002	88.8	212.0	-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	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
徳島	824	810	260	-1.7	785	90.3	189.4	-3.0
香川	1,023	1,012	330	-1.0	996	93.0	530.7	-1.6
愛媛	1,493	1,468	748	-1.7	1,431	88.8	252.1	-2.5
高知	814	796	341	-2.2	764	88.6	107.6	-4.0
福岡	5,016	5,050	3,525	0.7	5,072	89.4	1,019.0	0.4
佐賀	877	866	246	-1.2	850	89.0	348.3	-1.9
長崎	1,517	1,479	688	-2.5	1,427	87.5	347.5	-3.5
熊本	1,859	1,842	805	-0.9	1,817	88.5	245.4	-1.3
大分	1,221	1,210	537	-0.9	1,197	89.4	188.7	-1.1
宮崎	1,170	1,153	512	-1.4	1,135	88.5	146.7	-1.5
鹿児島	1,786	1,753	689	-1.8	1,706	87.6	185.7	-2.7
沖縄	1,318	1,362	893	3.3	1,393	96.3	611.9	2.3

10月1日現在。 1) 人口密度の高い基本単位区(人口密度が1km²当たり約4,000人以上)が市区町村の境界内で互いに隣接して、国勢調査時に人口5,000人以上を有する地域。 2) 算出に用いた面積は、国土交通省国土地理院「平成22年全国都道府県市区町村別面積調」による。また、境界未定地域については、総務省統計局において面積を推定。
a) 青森県、岩手県、宮城県、秋田県、山形県、福島県、茨城県、栃木県、群馬県、埼玉県、千葉県、東京都、神奈川県、新潟県、富山県、石川県、福井県、山梨県、長野県、岐阜県、静岡県、愛知県、三重県、滋賀県、京都府、大阪府、兵庫県、奈良県、和歌山県、鳥取県、島根県、岡山県、広島県、山口県、徳島県、香川県、愛媛県、高知県、福岡県、佐賀県、長崎県、熊本県、大分県、宮崎県、鹿児島県、沖縄県を除き算出。
資料: 総務省統計局統計調査部国勢調査課「国勢調査報告」(「国勢調査」ホームページ)

図8: セルに整形のための空白・改行・カンマを含む(ガイド2を満たさない)例

2-2都道府県別人口と人口増加率

都道府県	平成12年 (1000 人)	17年(1000 人)	人口集中 地区 (1)	人口増減率 (平成12~ 17年) (%)	22年 (1,000 人)	人口性比 (女性100 に対する 男性)	人口密度 (人/ km ²)	人口増減率 (平成17~ 22年) (%)
全国	126926	127768	84331	0.7	128057	94.8	a)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	2386	2375	1068	-0.4	2370	93.3	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
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	1857	1867	762	0.5	1855	95.0	321.0	-0.7
滋賀	1343	1380	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	908	-1.5	1401	90.0	379.5	-1.4
和歌山	1070	1036	408	-3.2	1002	88.8	212.0	-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	1493	708	-2.3	1451	89.2	237.4	-2.8
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	814	796	341	-2.2	764	88.6	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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

10月1日現在。1)人口密度の高い集中地区(人口密度が1km²当たり約4,000人以上)が市区町村の域内を互いに隣接して、国勢調査時に人口5,000人以上を有する地域。2)算出に用いた面積は、国土交通省国土地理院「平成22年全国都道府県市区町村別面積調」による。また、境界未定地域については、総務省統計局において面積を推定。a)宮崎県、色丹島、国後島、択捉島及び竹島を除き算出。
資料総務省統計局統計調査部国勢調査課「国勢調査報告」「国勢調査」(ホームページ)

図 9: 整形のためのスペース・改行・カンマを除去(ガイド 2 を満たす)

ガイド3(レベル1: 必須レベル)

年の値には、西暦表記を備える。

【解説】

図10の年次カラムは和暦で書かれている。機械は、年の値を数値の大小により比較する。従って、年が経過するごとに、その年の値が単調増加する、つまりに西暦を利用することが望ましい(図11)。和暦を併記してもよい。

2-17 出生・死亡数と婚姻・離婚件数

年次	出生数 (L.千人)	死亡数 (L.千人)	婚姻件数 (L.千件)	離婚件数 (L.千件)	人口1,000につき 1)					乳児死亡率 (出生1,000につき)
					出生率	死亡率	自然増減率	婚姻率	離婚率	
昭和34年	1,626	690	847	72	17.5	7.4	10.1	9.1	0.78	33.7
35	1,606	707	866	69	17.2	7.6	9.6	9.3	0.74	30.7
36	1,589	696	890	69	16.9	7.4	9.5	9.4	0.74	28.6
37	1,619	710	928	71	17.0	7.5	9.5	9.8	0.75	26.4
38	1,660	671	938	70	17.3	7.0	10.3	9.7	0.73	23.2
39	1,717	673	963	72	17.7	6.9	10.7	9.9	0.74	20.4
40	1,824	700	955	77	18.6	7.1	11.4	9.7	0.79	18.5
41	1,961	670	940	79	19.7	6.8	12.0	9.5	0.80	19.3
42	1,936	675	953	83	19.4	6.8	12.7	9.6	0.84	14.9
43	1,872	687	956	87	18.6	6.8	11.8	9.5	0.87	15.3
44	1,890	694	984	91	18.5	6.8	11.7	9.6	0.89	14.2
45	1,934	713	1,029	96	18.8	6.9	11.8	10.0	0.93	13.1
46	2,001	685	1,091	104	19.2	6.6	12.6	10.5	0.99	12.4
47	2,039	684	1,100	108	19.3	6.5	12.8	10.4	1.02	11.7
48	2,092	709	1,072	112	19.4	6.6	12.8	9.9	1.04	11.3
49	2,030	711	1,000	114	18.6	6.5	12.1	9.1	1.04	10.8
50	1,901	702	942	119	17.1	6.3	10.8	8.5	1.07	10.0
51	1,833	703	872	125	16.3	6.3	10.0	7.8	1.11	9.3
52	1,755	690	821	129	15.5	6.1	9.4	7.2	1.14	8.9
53	1,709	696	793	132	14.9	6.1	8.8	6.9	1.15	8.4
54	1,643	690	789	135	14.2	6.0	8.3	6.8	1.17	7.9
55	1,577	723	775	142	13.6	6.2	7.3	6.7	1.22	7.5
56	1,529	720	777	154	13.0	6.1	6.9	6.6	1.32	7.1
57	1,515	712	781	164	12.8	6.0	6.8	6.6	1.39	6.6
58	1,509	740	763	179	12.7	6.2	6.5	6.4	1.51	6.2
59	1,490	740	740	179	12.5	6.2	6.3	6.2	1.50	6.0
60	1,432	752	736	167	11.9	6.3	5.6	6.1	1.39	5.5
61	1,383	751	711	166	11.4	6.2	5.2	5.9	1.37	5.2
62	1,347	751	696	158	11.1	6.2	4.9	5.7	1.30	5.0
63	1,314	793	708	154	10.8	6.5	4.3	5.8	1.26	4.8
平成元年	1,247	789	708	158	10.2	6.4	3.7	5.8	1.29	4.6
2	1,222	820	722	158	10.0	6.7	3.3	5.9	1.28	4.6
3	1,223	830	742	169	9.9	6.7	3.2	6.0	1.37	4.4
4	1,209	857	754	179	9.8	6.9	2.9	6.1	1.45	4.5
5	1,188	879	793	188	9.6	7.1	2.5	6.4	1.52	4.3
6	1,238	876	783	195	10.0	7.1	2.9	6.3	1.57	4.2
7	1,187	922	792	199	9.6	7.4	2.1	6.4	1.60	4.3
8	1,207	896	795	207	9.7	7.2	2.5	6.4	1.66	3.8
9	1,192	913	776	223	9.5	7.3	2.2	6.2	1.78	3.7
10	1,203	936	785	243	9.6	7.5	2.1	6.3	1.94	3.6
11	1,178	982	762	251	9.4	7.8	1.6	6.1	2.00	3.4
12	1,191	962	798	264	9.5	7.7	1.8	6.4	2.10	3.2
13	1,171	970	800	286	9.3	7.7	1.6	6.4	2.27	3.1
14	1,154	982	757	290	9.2	7.8	1.4	6.0	2.30	3.0
15	1,124	1,015	740	284	8.9	8.0	0.9	5.9	2.25	3.0
16	1,111	1,029	720	271	8.8	8.2	0.7	5.7	2.15	2.8
17	1,063	1,084	714	262	8.4	8.6	-0.2	5.7	2.08	2.8
18	1,093	1,084	731	257	8.7	8.6	0.1	5.8	2.04	2.6
19	1,090	1,108	720	255	8.6	8.8	-0.1	5.7	2.02	2.6
20	1,091	1,142	726	251	8.7	9.1	-0.4	5.8	1.99	2.6
21	1,070	1,142	708	253	8.5	9.1	-0.6	5.6	2.01	2.4
22	1,071	1,197	700	251	8.5	9.5	-1.0	5.5	1.99	2.3

「人口動態調査」による。日本において出生した日本人に属するもの。昭和47年以前は出生数を除く。1) 昭和47年以前は総人口【日本に在住する外国人を含む】を分母に用いている。
資料：厚生労働省大臣官房統計情報課人口動態・家族統計課「人口動態調査」【ホームページ】

図10: 年が和暦で書かれている(ガイド3を満たさない)例⁶

⁶ 総務省統計局刊行、総務省統計研修所編集「日本の統計 2012」内「2-17 出生・死亡数と婚姻・離婚件数」による。 <http://www.stat.go.jp/data/nihon/02.htm>

2-17 出生・死亡数と婚姻・離婚件数

年次	年次 (西暦)	出生数 (L.000人)	死亡数 (L.000人)	婚姻件数 (L.000件)	離婚件数 (L.000件)	人口1,000につき 1)					乳児 死亡率 (出生 1,000 につき)	
						出生率	死亡率	自然 増減率	婚姻率	離婚率		
昭和	34	1959	1,626	690	847	72	17.5	7.4	10.1	9.1	0.78	33.7
	35	1960	1,606	707	866	69	17.2	7.6	9.6	9.3	0.74	30.7
	36	1961	1,589	696	890	69	16.9	7.4	9.5	9.4	0.74	28.6
	37	1962	1,619	710	928	71	17.0	7.5	9.5	9.8	0.75	26.4
	38	1963	1,660	671	938	70	17.3	7.0	10.3	9.7	0.73	23.2
	39	1964	1,717	673	963	72	17.7	6.9	10.7	9.9	0.74	20.4
	40	1965	1,824	700	955	77	18.6	7.1	11.4	9.7	0.79	18.5
	41	1966	1,361	670	940	79	13.7	6.8	7.0	9.5	0.80	19.3
	42	1967	1,936	675	953	83	19.4	6.8	12.7	9.6	0.84	14.9
	43	1968	1,872	687	956	87	18.6	6.8	11.8	9.5	0.87	15.3
	44	1969	1,890	694	984	91	18.5	6.8	11.7	9.6	0.89	14.2
	45	1970	1,934	713	1,029	96	18.8	6.9	11.8	10.0	0.93	13.1
	46	1971	2,001	685	1,091	104	19.2	6.6	12.6	10.5	0.99	12.4
	47	1972	2,039	684	1,100	108	19.3	6.5	12.8	10.4	1.02	11.7
	48	1973	2,092	709	1,072	112	19.4	6.6	12.8	9.9	1.04	11.3
	49	1974	2,030	711	1,000	114	18.6	6.5	12.1	9.1	1.04	10.8
	50	1975	1,901	702	942	119	17.1	6.3	10.8	8.5	1.07	10.0
	51	1976	1,833	703	872	125	16.3	6.3	10.0	7.8	1.11	9.3
	52	1977	1,755	690	821	129	15.5	6.1	9.4	7.2	1.14	8.9
	53	1978	1,709	696	793	132	14.9	6.1	8.8	6.9	1.15	8.4
	54	1979	1,643	690	789	135	14.2	6.0	8.3	6.8	1.17	7.9
	55	1980	1,577	723	775	142	13.6	6.2	7.3	6.7	1.22	7.5
	56	1981	1,529	720	777	154	13.0	6.1	6.9	6.6	1.32	7.1
	57	1982	1,515	712	781	164	12.8	6.0	6.8	6.6	1.39	6.6
	58	1983	1,509	740	763	179	12.7	6.2	6.5	6.4	1.51	6.2
	59	1984	1,490	740	740	179	12.5	6.2	6.3	6.2	1.50	6.0
	60	1985	1,432	752	736	167	11.9	6.3	5.6	6.1	1.39	5.5
	61	1986	1,383	751	711	166	11.4	6.2	5.2	5.9	1.37	5.2
	62	1987	1,347	751	696	158	11.1	6.2	4.9	5.7	1.30	5.0
	63	1988	1,314	793	708	154	10.8	6.5	4.3	5.8	1.26	4.8
平成	元	1989	1,247	789	708	158	10.2	6.4	3.7	5.8	1.29	4.6
	2	1990	1,222	820	722	158	10.0	6.7	3.3	5.9	1.28	4.6
	3	1991	1,223	830	742	169	9.9	6.7	3.2	6.0	1.37	4.4
	4	1992	1,209	857	754	179	9.8	6.9	2.9	6.1	1.45	4.5
	5	1993	1,188	879	793	188	9.6	7.1	2.5	6.4	1.52	4.3
	6	1994	1,238	876	783	195	10.0	7.1	2.9	6.3	1.57	4.2
	7	1995	1,187	922	792	199	9.6	7.4	2.1	6.4	1.60	4.3
	8	1996	1,207	896	795	207	9.7	7.2	2.5	6.4	1.66	3.8
	9	1997	1,192	913	776	223	9.5	7.3	2.2	6.2	1.78	3.7
	10	1998	1,203	936	785	243	9.6	7.5	2.1	6.3	1.94	3.6
	11	1999	1,178	982	762	251	9.4	7.8	1.6	6.1	2.00	3.4
	12	2000	1,191	962	798	264	9.5	7.7	1.8	6.4	2.10	3.2
	13	2001	1,171	970	800	286	9.3	7.7	1.6	6.4	2.27	3.1
	14	2002	1,154	982	757	290	9.2	7.8	1.4	6.0	2.30	3.0
	15	2003	1,124	1,015	740	284	8.9	8.0	0.9	5.9	2.25	3.0
	16	2004	1,111	1,029	720	271	8.8	8.2	0.7	5.7	2.15	2.8
	17	2005	1,063	1,084	714	262	8.4	8.6	-0.2	5.7	2.08	2.8
	18	2006	1,093	1,084	731	257	8.7	8.6	0.1	5.8	2.04	2.6
	19	2007	1,090	1,108	720	255	8.6	8.8	-0.1	5.7	2.02	2.6
	20	2008	1,091	1,142	726	251	8.7	9.1	-0.4	5.8	1.99	2.6
	21	2009	1,070	1,142	708	253	8.5	9.1	-0.6	5.6	2.01	2.4
	22	2010	1,071	1,197	700	251	8.5	9.5	-1.0	5.5	1.99	2.3

「人口動態調査」による。日本において発生した日本人に関するもの。昭和47年以前は出生数を除く。① 昭和41年以前は総人口【日本に在住する外国人を含む】を分母に用いている。
 資料 厚生労働省大臣官房統計情報部人口動態・婚姻統計課「人口動態調査」【ホームページ】

図 11: 西暦のカラムを追加(ガイド 3 を満たす)

ガイド4(レベル1: 必須レベル)

数値やタイトル・単位以外の情報が、セルに含まれない。

【解説】

図12の人口密度カラムの全国の値は「a) 343.4」である。このセルには、値である「343.4」と注釈番号である「a)」の両方が含まれている。機械がこのセルを解釈するには、事前に注釈番号「a)」を除かなければならない。このため、機械に解釈させるべき数値やタイトル以外の情報を、セルに持たせるべきではない(図13)。

2-2 都道府県別人口と人口増加率

都道府県	2000年 (1000人)	2005年 (1000人)	人口集中 地区 1)	人口増減率 (2000~ 2005年) (%)	2010年 (1000人)	人口性比 (女性100に 対する男性 性)	人口密度 (人/km ²) 2)	人口増減率 (2005~ 2010年) (%)
全国	126926	127768	84331	0.7	128057	94.8	a)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	2385	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
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	-1.3
長崎	1517	1479	688	-2.5	1427	87.5	347.5	-3.5
熊本	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

10月1日現在。 1) 人口密度の高い基本単位区(人口密度が1km²当たり約4,000人以上)が市区町村の圏域内で互いに隣接して、国勢調査時に人口5,000人以上を有する地域。 2) 算出に用いた面積は、国土交通省国土地理院「平成22年全国都道府県市区町村別面積調」による。また、境界未定地域については、総務省統計局において面積を推定。
a) 宮崎県、色丹島、国後島、根室島及び竹島を除き算出。
資料: 総務省統計局統計調査部国勢調査課「国勢調査報告」「国勢調査」(ホームページ)

図12: セルにキャプション・注釈・注釈番号を含む(ガイド4を満たさない)例

都道府県	2000年 (1000人)	2005年 (1000人)	人口集中 地区	人口増減率 (2000～ 2005年) (%)	2005年 (1000人)	人口性比 (女性100に 対する男 性)	人口密度 (人/kin ²)	人口増減率 (2000～ 2005年) (%)
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	2385	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	877	868	246	-1.2	850	89.0	348.3	-1.3
長崎	1517	1479	688	-2.5	1427	87.5	347.5	-3.5
熊本	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 13: キャプション・脚注・脚注番号を除去(ガイド 4 を満たす)

6.2.2 セルに関する要件（レベル 1）

ガイド 5(レベル 1: 必須レベル)

すべてのセルが、他のセルと結合されていない。

【解説】

図 14 のタイトルのセルが結合されている。結合されたセルは、原則すべて同じ値を持つ。これを機械が解読するためには、結合されたセルの値と、結合されている範囲を認識する必要がある。また、この状態のデータセットを CSV に変換すると、セルの結合は解除され、1 つのセルを除いて値は空になる。このため、セルは結合せず、同じ値を記載する方が機械は解読しやすい（図 15）。

第1-1-1-1 図 世界実質GDP成長率の推移

		世界	先進国	新興国及び途上国
2008	Q1	3.4	0.8	7.4
	Q2	1.9	-0.7	5.9
	Q3	-0.6	-3.2	3.3
	Q4	-6.9	-8.7	-4.1
2009	Q1	-5.7	-8.2	-2.3
	Q2	4.5	0.9	9.6
	Q3	5.1	2.2	9.4
	Q4	5.5	3.8	8.1
2010	Q1	5.8	4.0	8.4
	Q2	5.5	4.2	7.2
	Q3	3.7	2.1	5.9
	Q4	4.7	1.6	8.9
2011	Q1	3.7	1.3	6.9
	Q2	3.2	0.8	6.3
	Q3	3.3	1.9	5.2
	Q4	3.1	1.2	5.5
2012	Q1	2.6	0.6	5.1
	Q2	3.4	1.1	6.3
	Q3	3.7	1.6	6.4
	Q4	3.9	1.8	6.6
2013	Q1	3.9	2.0	6.1
	Q2	4.1	2.1	6.5
	Q3	3.8	2.1	5.9
	Q4	4.4	2.3	7.0

備考: 季節調整済み。
資料: IMF「WEO, January 2012 update」から作成。

図 14: セルが結合されている(ガイド 5 を満たさない)例⁷

⁷ 経済産業省「通商白書 2012」による。
<http://www.meti.go.jp/report/tshuhaku2012/>

第1-1-1-1 図 世界実質GDP成長率の推移

		世界	先進国	新興国及び途上国
2008	Q1	3.4	0.8	7.4
2008	Q2	1.9	-0.7	5.9
2008	Q3	-0.6	-3.2	3.3
2008	Q4	-6.9	-8.7	-4.1
2009	Q1	-5.7	-8.2	-2.3
2009	Q2	4.5	0.9	9.6
2009	Q3	5.1	2.2	9.4
2009	Q4	5.5	3.8	8.1
2010	Q1	5.8	4.0	8.4
2010	Q2	5.5	4.2	7.2
2010	Q3	3.7	2.1	5.9
2010	Q4	4.7	1.6	8.9
2011	Q1	3.7	1.3	6.9
2011	Q2	3.2	0.8	6.3
2011	Q3	3.3	1.9	5.2
2011	Q4	3.1	1.2	5.5
2012	Q1	2.6	0.6	5.1
2012	Q2	3.4	1.1	6.3
2012	Q3	3.7	1.6	6.4
2012	Q4	3.9	1.8	6.6
2013	Q1	3.9	2.0	6.1
2013	Q2	4.1	2.1	6.5
2013	Q3	3.8	2.1	5.9
2013	Q4	4.4	2.3	7.0

備考: 季節調整済み。
資料: IMF「WEO, January 2012 update」から作成。

図 15: 年カラムのセル結合を解除(ガイド5を満たす)

ガイド6(レベル1: 必須レベル)

値が存在しない場合を除き、データセルが空白でない。(省略されていない)

【解説】

図 16 の最後の 11 行の第 1 列は、空白である。人間はこの部分のデータセルに「平成 24 年」が省略されていることがわかるが、機械には分からない。

従って、このデータの機械可読を容易にするためには、値を省略しない、つまり値が存在しない場合を除き、データセルを空白にすべきでない(図 17)。

		コンビニエンスストア販売額等								
		Sales value of convenience stores and number of establishments							(単位:百万円)	
									(million yen)	
年	月	合 計					サービス 売上高	年(度、期、 月)末の 店舗数	Year and Month	
		Total	商品販売額							
		of	FF・日配食品	加工食品	非食品	of	of			
		Goods	Fast foods & Dailyfoods	Processed Foods	Non-Foods	Services	Establishments			
平成23年	1月	663,767	600,700	208,607	175,528	216,565	26,087	42,583		
	2	626,787	600,700	208,607	175,528	216,565	26,087	42,583	Feb.	
	3	711,626	683,539	224,032	202,102	257,405	28,087	42,412	Mar.	
	4	663,047	636,746	225,987	193,347	217,412	26,301	42,401	Apr.	
	5	719,221	690,761	234,184	201,083	255,494	28,460	42,473	May	
	6	741,911	711,601	239,107	204,499	267,995	30,310	42,518	Jun.	
	7	820,406	785,258	271,966	231,760	281,532	35,148	42,660	Jul.	
	8	809,938	775,848	271,767	228,690	275,391	34,090	42,956	Aug.	
	9	747,231	717,795	250,821	207,724	259,250	29,436	42,939	Sep.	
	10	758,669	725,295	255,003	208,034	262,258	33,374	43,027	Oct.	
	11	731,449	697,699	238,149	204,014	255,536	33,750	43,215	Nov.	
	12	780,638	740,699	262,386	204,131	274,182	39,939	43,373	Dec.	
平成24年	1月	723,545	691,927	240,339	197,698	253,890	31,618	45,466	Jan. 2012	
	2	708,233	669,117	237,692	187,656	243,769	39,116	45,731	Feb.	
	3	771,513	735,582	262,021	206,208	267,353	35,931	45,753	Mar.	
	4	759,825	724,155	258,107	204,797	261,251	35,670	45,956	Apr.	
	5	791,989	756,730	264,580	217,931	274,219	35,259	46,245	May	
	6	781,819	744,370	265,490	210,673	268,207	37,449	46,359	Jun.	
	7	860,401	822,572	296,845	240,648	285,079	37,829	46,603	Jul.	
	8	867,181	832,399	300,324	245,090	286,985	34,782	47,052	Aug.	
	9	797,471	766,436	276,545	220,618	269,273	31,035	47,115	Sep.	
	10	805,700	771,231	276,105	228,441	266,685	34,469	47,198	Oct.	
	11	774,089	739,019	262,358	210,806	265,855	35,070	47,550	Nov.	
	12	835,422	790,845	288,595	218,277	283,973	44,577	47,801	Dec.	

図 16: 年のデータセル値が省略されている(ガイド6を満たさない)例⁸

⁸ 経済産業省「コンビニエンスストア商品別販売額等及び前年(度, 同期, 同月)比」による。 <http://www.meti.go.jp/statistics/tyo/syoudou/result-2/>

コンビニエンスストア販売額等								(単位:百万円)	
Sales value of convenience stores and number of establishments								(million yen)	
年	月	合計	商品販売額			サービス売上高	年(度、期、月)末の店舗数	Year and Month	
			FF・日配食品	加工食品	非食品				
		Total	Sales of Goods			Sales of Services	Number of Establishments		
			Fast foods & Dailyfoods	Processed Foods	Non-Foods				
平成23年	2月	626,787	600,700	208,607	175,528	216,565	26,087	42,583	Feb. 2011
平成23年	3月	711,626	683,539	224,032	202,102	257,405	28,087	42,412	Mar. 2011
平成23年	4月	663,047	636,746	225,987	193,347	217,412	26,301	42,401	Apr. 2011
平成23年	5月	719,221	690,761	234,184	201,083	255,484	28,460	42,473	May 2011
平成23年	6月	741,911	711,601	239,107	204,499	267,995	30,310	42,518	Jun. 2011
平成23年	7月	820,406	785,258	271,966	231,760	281,532	35,148	42,660	Jul. 2011
平成23年	8月	809,938	775,848	271,767	228,690	275,391	34,090	42,856	Aug. 2011
平成23年	9月	747,231	717,795	250,821	207,724	259,250	29,436	42,939	Sep. 2011
平成23年	10月	758,669	725,295	255,003	208,034	262,258	33,374	43,027	Oct. 2011
平成23年	11月	731,449	697,699	238,149	204,014	255,536	33,750	43,215	Nov. 2011
平成23年	12月	780,638	740,699	262,386	204,131	274,182	39,939	43,373	Dec. 2011
平成24年	1月	723,545	691,827	240,339	197,698	253,890	31,618	45,466	Jan. 2012
平成24年	2月	708,233	669,117	237,692	187,656	243,769	39,116	45,731	Feb. 2012
平成24年	3月	771,513	735,582	262,021	206,208	267,353	35,931	45,753	Mar. 2012
平成24年	4月	759,825	724,155	258,107	204,797	261,251	35,670	45,956	Apr. 2012
平成24年	5月	791,989	756,730	264,580	217,931	274,219	35,259	46,245	May 2012
平成24年	6月	781,819	744,370	265,490	210,673	268,207	37,448	46,359	Jun. 2012
平成24年	7月	860,401	822,572	296,845	240,648	285,079	37,829	46,603	Jul. 2012
平成24年	8月	867,181	832,399	300,324	245,090	286,985	34,782	47,052	Aug. 2012
平成24年	9月	797,471	766,436	276,545	220,618	269,273	31,035	47,115	Sep. 2012
平成24年	10月	805,700	771,231	276,105	228,441	266,685	34,469	47,198	Oct. 2012
平成24年	11月	774,089	739,019	262,358	210,806	265,855	35,070	47,550	Nov. 2012
平成24年	12月	835,422	790,845	288,595	218,277	283,973	44,577	47,801	Dec. 2012

図 17: 省略されている語句を補う(ガイド 6 を満たす)

6.2.3 タイトルに関する要件（レベル 1~2）

6.2.3.1 レベル 1

ガイド 7(レベル 1: 必須レベル)

データセルの内容を示すタイトルが 1 行で構成されている。

【解説】

図 18 のタイトルは構造化されており、2 行からなっている。左から 4 番目のカラムは、「2005 年の人口集中地区の人口」という意味であるが、これを機械は解読できない。タイトルの文言を工夫して、カラムのタイトルを 1 行で表現するべきである（図 19）。

都道府県	2000年 (1000人)	2005年 (1000人)	人口集中 地区	人口増減率 (2000~ 2005年) (%)	2005年 (1000人)	人口性比 (女性100に 対する男 性)	人口密度 (人/km ²)	人口増減率 (2000~ 2005年) (%)
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	1225	1221	1129	-0.3	1217	97.2	188.7	-0.3
富山	1025	1025	1025	0.0	1025	98.5	188.7	0.0
石川	1025	1025	1025	0.0	1025	98.5	188.7	0.0
福井	1025	1025	1025	0.0	1025	98.5	188.7	0.0
山梨	1025	1025	1025	0.0	1025	98.5	188.7	0.0
長野	1517	1479	688	-2.5	1427	87.5	347.5	-3.5
岐阜	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 18: タイトルが複数行からなる(ガイド 7 を満たさない)例

都道府県	2000年 (1000人)	2005年 (1000人)	2005年の 人口集中 地区	2000~2005 年の人口増 減率 (%)	2010年 (1000人)	2010年の 人口性比 (女性100に 対する男)	2010年の人 口密度 (人/km ²)	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	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

徳島	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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 19: タイトルを 1 行にまとめる(ガイド 7 を満たす)

ガイド 8(レベル 1: 必須レベル)

データの単位が明記されている。

【解説】

データの単位（物理単位・貨幣単位）は、データ処理に必須である。このため、カラムにはデータの単位を明記すべきである。ただし、「人」「個」等の助数詞は日本語独特であり、機械が解読する時にも必要ない（図 20）。これらを「〇〇の人数」「〇〇の個数」というように、タイトルに明記すれば、理解できる（図 21）。これは、該当する表形式データを英語等の他言語に翻訳するときにも有用である。

都道府県	2000年 (1000人)	2005年 (1000人)	2005年の 人口集中 地区	2000~2005 年の人口増 減率 (%)	2010年 (1000人)	2010年の 人口性比 (女性100に 対する男)	2010年の人 口密度 (人/km ²)	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	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

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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1796	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 20: タイトルに単位がない, または助数詞が含まれる(ガイド 8 を満たさない)例

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に對する男性)	2010年の人口密度(人/㎢)	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	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	93.3	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

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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 21: タイトルの文言を修正し助数詞を除去(ガイド 8 を満たす)

ガイド9(レベル1: 必須レベル)

データセルの内容, 単位, 記数単位を示すタイトルが, それぞれ別の行に記載されている。

【解説】

図 22 の左から 2 番目のカラムは, 平成 12 年の人口値であり, 単位は[1,000 人]である。つまり, データセルに記載されている値の 1,000 倍が実際の値である。単位や記数単位 (この例では「1,000」) は, 機械がデータを解読する際に必要であるため, 確実に取得できることが望ましい。タイトルに単位や記数単位が含まれている場合, 機械がそれらを解読するには, タイトルから単位や記数単位を抽出する必要がある。これは煩雑であるから, タイトルに単位や記数単位を含まず, 別の行に記述すべきである (図 23)。

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に對する男性)	2010年の人口密度 (人/㎢)	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	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

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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 22: タイトルの内容・単位・記数単位が 1 行に併記されている(ガイド 9 を満たさない)例

都道府県	2000年の人口	2005年の人口	2005年の人口集中地区	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
	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	93.3	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

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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 23: 単位・記数単位を別の行に移転(ガイド 9 を満たす)

本ガイドまでの条件は、機械がデータを解読するために必要不可欠な事項である。つまり、本節までに記した条件を満たしたデータを見た人間は、機械にデータを解読させるための方針を立てられる。よって、これらをすべて満たす表形式データをレベル 1 とする。

6.2.3.2 レベル 2

ガイド 10 (レベル 2: 推奨レベル)

タイトルやデータ型が、機械可読な形で記述されている。

【解説】

単位やデータ型を、定型フォーマットで記述することにより、複数のデータセットを機械が同様に解読できる。具体的な記法については、オープンデータ化のための CSV 形式データ規格[9]が規定する。

6.2.4 データの公開形式に関する要件（レベル2~レベル3）

6.2.4.1 レベル2

ガイド11(レベル2: 推奨レベル)

データセットが、オープンなフォーマットで提供されている。

【解説】

仕様が公開され、それが標準化されたフォーマット、すなわちオープンなフォーマットは、解読するツールが広く普及しており機械可読しやすい。このため、データセットはオープンなフォーマット、たとえばCSV[1]やXML[3]、OpenDocument[5]形式で提供されるべきである。

本ガイドまでのすべてを満たしたデータを取得した機械は、単位やカラムのタイトル、データの値を正しく解読できる。よって、本ガイドまでのすべてを満たす表形式データを、レベル2とする。

6.2.4.2 レベル 3

ガイド 12(レベル 3: 理想レベル)

データに対する、利用者が理解できる説明が、XML や RDF[4]によりメタデータ記述されている。かつ、その説明記述がデータをリンクしている。

【解説】

格納されているデータがどのようなものであるか、タイトルや作成者などの情報をメタデータとして記述し、データにリンクさせる。このようなメタデータから、機械は必要なデータを検索する、あるいはデータの意味を知ることができる。

ガイド 13(レベル 3: 理想レベル)

データ本体が、XML や RDF によりメタデータ記述されている。

【解説】

データ自体をメタデータで記述すれば、機械はデータの意味を理解できる。これを利用して、機械が複数のデータの意味を理解して、それらをマッシュアップできる。このような応用を可能にするためには、データ自体をメタデータ記述することが望ましい。

データを取得した機械がデータの意味を理解するためには、データ自体をメタデータ化するか、またはデータの定義をメタデータ化すればよい。よって、ガイド 11 までのすべてに加えてガイド 12 またはガイド 13 を満たす表形式データを、レベル 3 とする。

7 文書形式データのオープンデータ化に関する技術ガイド

本章では、文書形式データをオープンデータ化する際に留意すべき点を、レベルごとにまとめる。

7.1 定義

文書形式データとは、1次元状に配列された文字を主な構成要素とし、一部図や費用などを含み、人間がそれを読むことによって人間に何らかの作用を与えることを目的としたデータである。文書形式データに関連する用語を、以下の通り定義する。

- 見出し:
 - 文章において内容の要点を非常に短い言葉にまとめ、本文より大きな字で章や節の最初に置かれる言葉。大きい方から編（部）・章・節・項・目といった名称が付けられる慣習がある。
- プレインテキスト:
 - コンピュータ上で文章を扱うための一般的なファイルフォーマット、または文字列の形式の 1 つ。文字情報以外の情報、たとえば文字の色や大きさ、形状、文章に含まれる図表などの情報を含まない。
- タグ:
 - 文章に対する構造（章・節・図表など）や見栄え（色・大きさ・形状など）に関する指定。
- マークアップ言語:
 - 文章の構造や見栄えに関する指定を、文章とともにテキストファイルに記述するための言語。

7.2 技術ガイド

本節では、文書形式データを機械可読にする際に留意すべき点を挙げる。

7.2.1 レベル 1

ガイド 1(レベル 1: 必須レベル)

文章に存在する部・章・節・図表などの構造が、明快に分かる形で記述されている。

【解説】

文章は、部・章・節・段落・図表などの構造を持っている。たとえば機械が文章の第 1 章を抽出したいとするならば、第 1 章が文章のどの部分にあるのか分からなければならない。このためには、少なくとも見出し等の機能を利用して、部・章・節・段落・図表などのタイトルが区別されている必要がある (図 24)。

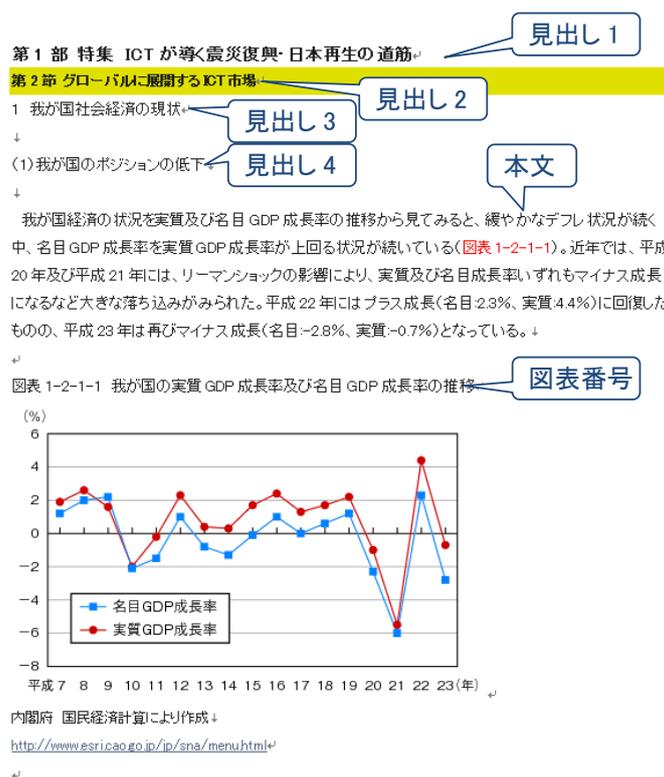


図 24: 見出しを利用して文章を執筆⁹

⁹ 図中の文章は、総務省「平成 24 年版 情報通信白書」より引用。
<http://www.soumu.go.jp/johotsusintokei/whitepaper/>

ガイド 2(レベル 1: 必須レベル)

文章内に、整形のための空白・改行を含めない。

【解説】

文章に含まれる空白・改行が有意であるか否かを、機械は判断できない。文書の解析や読み上げを行う際に、これらの空白・改行が支障となる。このため、機械の解読に必要なのない空白や改行、カンマは、事前に除くべきである。

7.2.2 レベル 2

ガイド 3(レベル 2: 推奨レベル)

オープンに利用できるデータフォーマットで公開する。

【解説】

仕様が公開され、それが標準化されたフォーマット、すなわちオープンなフォーマットは、解読するツールが広く普及しており機械可読しやすい。このため、文書形式データはオープンなフォーマットで公開されるべきである。

具体的には、プレインテキストにタグを挿入した XML 形式や HTML 形式のようなマークアップ形式が望ましいが、仕様が国際標準化されている Open Document Format (.odt) [5]や Office Open XML 形式 (.docx) [6]でもよい。

ガイド 4(レベル 2: 推奨レベル)

文書形式データが図表を含む場合、それらを構成するレベル 1 以上の表形式データが添付されている。

【解説】

図表やグラフを多く含む文書の、それら図表やグラフを形成した元になる表形式データが、機械可読なフォーマットで取得できるならば、それらのデータを利用したマッシュアップが容易になる。

7.2.3 レベル 3

ガイド 5(レベル 3: 理想レベル)

図表やグラフとそのキャプションが結びつくように、文書形式データが構成されている。

【解説】

全文検索を行う場合に、検索した文字列と関係のない画像が見つかることがある。これは、現在の HTML 等の文書フォーマットでは、図のタイトルと本体を結びつけて管理できないためことに起因する。つまり、図の近くにある文字列が、図を説明しているとは限らないためである。

ガイド 6(レベル 3: 理想レベル)

文章の構造を利用者が理解できるような説明がメタデータとして記述され、文書にリンクされている。

【解説】

格納されているデータがどのようなものであるか、タイトルや作成者などの情報をメタデータとして記述し、データにリンクさせる。このようなメタデータを収集したデータカタログから、機械は、必要なデータを検索したり、データの意味を知ったりできる。

8 地理空間データのオープンデータ化に関する技術ガイド

地理空間データをオープンデータ化する際に、留意すべき点をレベルごとにまとめる。

8.1 定義

地理空間データとは、2次元平面状の地図の地点や領域と関連づけられたデータである。地理空間データに関連する用語を、以下の通り定義する。

- 測地系：
 - 位置を緯度・経度・標高からなる座標で表すときの前提条件であり、測地座標系（直交座標系としての定義）・準拋楕円体（地球をモデル化する楕円体）・ジオイド面（標高の基準）の3つからなる。代表的なものとしては、WGS84 (World Geodetic System 1984) や、JGD2000 (Japan Geodetic Datum 2000) などがある。
- GIS (Geographic Information System) :
 - コンピュータ上に地図情報やさまざまな付加情報を持たせ、作成・保存・利用・管理し、地理情報を参照できるように表示・検索機能をもった、地理情報システムのこと。
- GML (Geography Markup Language) :
 - 地理的特徴を表現する XML ベースのマークアップ言語。Open Geospatial Consortium (OGC)によって開発され、現在は ISO 19136[7]として標準化されている。
- KML
 - 3次元地理空間情報の表示を管理するための、XML ベースのマークアップ言語 [8]。
- shape
 - 米国の ESRI 社が規定する GIS 標準データフォーマット形式。事実上の業界標準であり、多くの企業が発売している GIS システムは、この形式をサポートしている。

8.2 技術ガイド

本節では，地理空間データを機械可読にする際に留意すべき点を挙げる．

8.2.1 レベル 1

ガイド 1(レベル 1: 必須レベル)

測地系が明記されている．

【解説】

地理空間データにおいて，ものの位置を機械が解読できるように記述することは重要である．特に，位置に関する情報は，一般に緯度・経度・標高を用いて表されるが，緯度・経度・標高を表す基準は，複数存在している．この基準を測地系という．つまり，位置に関する情報を記述する際は，準拠している測地系を明記しなければならない．

8.2.2 レベル2

ガイド2(レベル2: 推奨レベル)

地理空間情報を記述するために広く使われているフォーマット測地系が明記されている。

【解説】

地理空間データは、GML[7]、KML[8]、shape 形式などが広く使われている。これらのフォーマットに基づく解析・処理ツールが広く普及しているため、これらのフォーマットに従って記述された地理空間データは、機械可読されやすい。

8.2.3 レベル3

ガイド3(レベル3: 理想レベル)

データに対する、利用者が理解できる説明が、XML や RDF によりメタデータ記述されており、かつその説明がデータをリンクしている。

【解説】

格納されているデータがどのようなものであるか、タイトルや作成者などの情報をメタデータとして記述し、データにリンクさせる。このようなメタデータを収集したデータカタログから、機械は、必要なデータを検索したり、データの意味を知ったりできる。

9 リアルタイムデータのオープンデータ化に関する技術ガイド

本章では、リアルタイムデータをオープンデータ化する際に、留意すべき点をレベルごとにまとめる。

9.1 定義

リアルタイムデータとは、値が時刻に応じて刻々と変化するデータである。

9.2 技術ガイド

本節では、リアルタイムデータを機械可読にする際に留意すべき点を挙げる。リアルタイムデータに関連する用語を、以下の通り定義する。

9.2.1 レベル 1

ガイド 1(レベル 1: 必須レベル)

表形式データ・地理空間データのレベル 2 以上に準拠した形式のデータを、ファイルとして取得できる。

【解説】

リアルタイムデータの取得周期は、表形式データや地理空間データに比べて短い。このため、人間が、機械にデータを解読させるための方針を立てるためには、取得するデータ自体が機械可読になっている、つまりレベル 2 以上になっていなければならない。

9.2.2 レベル2

ガイド2(レベル2: 推奨レベル)

リアルタイムデータの最新値・差分を取得する手法が提供されている。

【解説】

リアルタイムデータの最新値や差分値は、機械がリアルタイムデータを解析するうえで必要不可欠な情報である。これらを取得する手法が確立されていることが望ましい。

9.2.3 レベル 3

ガイド 3(レベル 3: 理想レベル)

リアルタイムデータの最新値や差分を取得するための、メタデータ記述に対応したデータ取得規約が提供されている。またはメタデータ記述されたリアルタイムデータを取得する手法が提供されている。

【解説】

リアルタイムデータがメタデータ記述に対応しているならば、機械は取得したデータの意味を理解できる。このためには、メタデータ記述に対応したデータ取得規約（たとえば情報流通基盤システム外部仕様書）や、リアルタイムな RDF データの提供手法が提供されていることが望ましい。

付録 A 表形式データの整形手順

本章では、表形式データをレベル 1 およびレベル 2 に整形する手順を示す。

A.1 表形式データをレベル 1 に整形する手順

図 25 を例に、表形式データをレベル 1 に整形する手順を示す。

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 区 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)	人口密度 (人/km ²) 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
青森県	1,476	1,437	653	-2.6	1,373	88.9	142.4	-4.4
岩手県	1,416	1,385	407	-2.2	1,330	91.3	87.1	-4.0
宮城県	2,365	2,360	1,371	-0.2	2,348	94.3	322.3	-0.5
秋田県	1,189	1,146	386	-3.7	1,086	88.5	93.3	-5.2
山形県	1,244	1,216	504	-2.2	1,169	92.2	125.4	-3.9
福島県	2,127	2,091	806	-1.7	2,029	94.3	147.2	-3.0
茨城県	2,986	2,975	1,068	-0.4	2,970	99.3	487.2	-0.2
栃木県	2,005	2,017	860	0.6	2,008	98.6	313.3	-0.4
群馬県	2,025	2,024	801	-0.0	2,008	96.9	315.6	-0.8
埼玉県	6,938	7,054	5,566	1.7	7,195	100.6	1,894.2	2.0
千葉県	5,926	6,056	4,342	2.2	6,216	99.4	1,205.5	2.6
東京都	12,064	12,577	12,329	4.2	13,159	98.0	6,015.7	4.6
神奈川県	8,490	8,792	8,250	3.6	9,048	100.9	3,745.4	2.9
新潟県	2,476	2,431	1,139	-1.8	2,374	93.6	188.7	-2.3
富山県	1,121	1,112	398	-0.8	1,093	92.9	257.4	-1.7
石川県	1,181	1,174	573	-0.6	1,170	93.4	279.5	-0.4
福井県	829	822	333	-0.9	806	93.5	192.4	-1.9
山梨県	888	885	305	-0.4	863	95.9	193.3	-2.4
長野県	2,215	2,196	764	-0.8	2,152	94.6	158.7	-2.0
岐阜県	2,108	2,107	822	-0.1	2,081	93.6	195.9	-1.3
静岡県	3,767	3,792	2,216	0.7	3,765	97.0	483.9	-0.7
愛知県	7,043	7,255	5,480	3.0	7,411	99.9	1,434.8	2.2
三重県	1,857	1,867	762	0.5	1,855	95.0	321.0	-0.7
滋賀県	1,343	1,380	614	2.8	1,411	97.6	351.2	2.2
京都府	2,644	2,648	2,162	0.1	2,636	92.3	571.4	-0.4
大阪府	8,805	8,817	8,440	0.1	8,865	93.6	4,669.7	0.5
兵庫県	5,551	5,591	4,198	0.7	5,588	91.7	665.6	-0.0
奈良県	1,443	1,421	908	-1.5	1,401	90.0	379.5	-1.4
和歌山県	1,070	1,036	408	-3.2	1,002	88.8	212.0	-3.3
鳥取県	613	607	207	-1.0	589	91.1	167.8	-3.0
島根県	762	742	179	-2.5	717	91.6	107.0	-3.3
岡山県	1,951	1,957	856	0.3	1,945	92.2	273.5	-0.6
広島県	2,879	2,877	1,832	-0.1	2,861	93.3	337.4	-0.6
山口県	1,528	1,493	708	-2.3	1,451	89.2	237.4	-2.8
徳島県	824	810	260	-1.7	785	90.3	189.4	-3.0
香川県	1,023	1,012	330	-1.0	996	93.0	530.7	-1.6
愛媛県	1,493	1,468	748	-1.7	1,431	88.8	252.1	-2.5
高知県	814	796	341	-2.2	764	88.6	107.6	-4.0
福岡県	5,016	5,050	3,525	0.7	5,072	89.4	1,019.0	0.4
佐賀県	877	866	246	-1.2	850	89.0	348.3	-1.9
長門県	1,517	1,479	688	-2.5	1,427	87.5	347.5	-3.5
熊本県	1,859	1,842	805	-0.9	1,817	88.5	245.4	-1.3
大分県	1,221	1,210	537	-0.9	1,197	89.4	188.7	-1.1
宮崎県	1,170	1,153	512	-1.4	1,135	88.5	146.7	-1.5
鹿児島県	1,786	1,753	689	-1.8	1,706	87.6	185.7	-2.7
沖縄県	1,318	1,362	893	3.3	1,393	96.3	611.9	2.3

10月1日現在。 1) 人口密度の高い基本単位区(人口密度が1km²当たり約4,000人以上)が市区町村の圏域内で互いに隣接して、国勢調査時に人口5,000人以上を有する地域。 2) 算出に用いた数値は、国土交通省国土地理院「平成22年全国都道府県市区町村別面積調」による。また、境界未定地域については、総務省統計局において面積を推定。 a) 青森県、色丹島、国後島、択捉島及び竹島を除き算出。
資料: 総務省統計局統計調査部国勢統計課「国勢調査報告」「国勢調査」(ホームページ)

図 25: 整形前のオリジナルデータ¹⁰

まず、表形式データのレベル 1 が満たすべき条件のうち、図 25 が満たしていない箇所を

¹⁰ 総務省統計局刊行、総務省統計研修所編集「日本の統計 2012」内「都道府県別人口と人口増加率」による。 <http://www.stat.go.jp/data/nihon/02.htm>

列記する。その結果は表 2 の通りである。

表 2: 図 25 の条件確認結果

項目	条件	評価
(1)	1 つのデータセットに、1 種類の表形式データ (1 つのテーブル) が掲載されている。	○
(2)	整形のためのスペース・改行、位取りのカンマを含まない。	×
(3)	年の値を西暦で表記している。	×
(4)	数値やタイトル以外の情報 (ラベル・注釈など) が、テーブルに含まれない。	×
(5)	すべてのデータセルが、他のデータセルと結合されていない。	○
(6)	値がない場合を除き、データセルの値が空白でない。	○
(7)	カラムのタイトルが 1 行で構成されている。	×
(8)	データの単位が明記されている。助数詞は含めない。	×
(9)	カラムのタイトルに、単位や記数単位が含まれない。	×

それぞれの項目について、条件を満たしていない箇所を Microsoft Excel を利用して整形する手法を記す。

A.1.1 複数のテーブルを分割

新しいシートをテーブルの個数分作成し、それぞれのシートにテーブルを移動させる。これにより、1 つのデータセットに 1 つのテーブルを掲載することができる。

A.1.2 脚注・脚注番号・キャプションを削除

セルの値として脚注・脚注番号・キャプションが記載されている場合は、それを取り除く。

脚注番号がセルの書式設定として付与されている場合は、セルの書式設定メニューを利用して除去する。Microsoft Excel 2007 以降であれば、「ホーム」メニューの「セル」タブにある「書式」メニュー (図 26) を利用する。Microsoft Excel 2003 以前であれば、「書式」→「セル」メニューを利用する。「セルの書式設定」ウィンドウの「分類」項目が「ユーザー定義」になっているので、これを「数値」に変更すれば、脚注番号を除去できる。

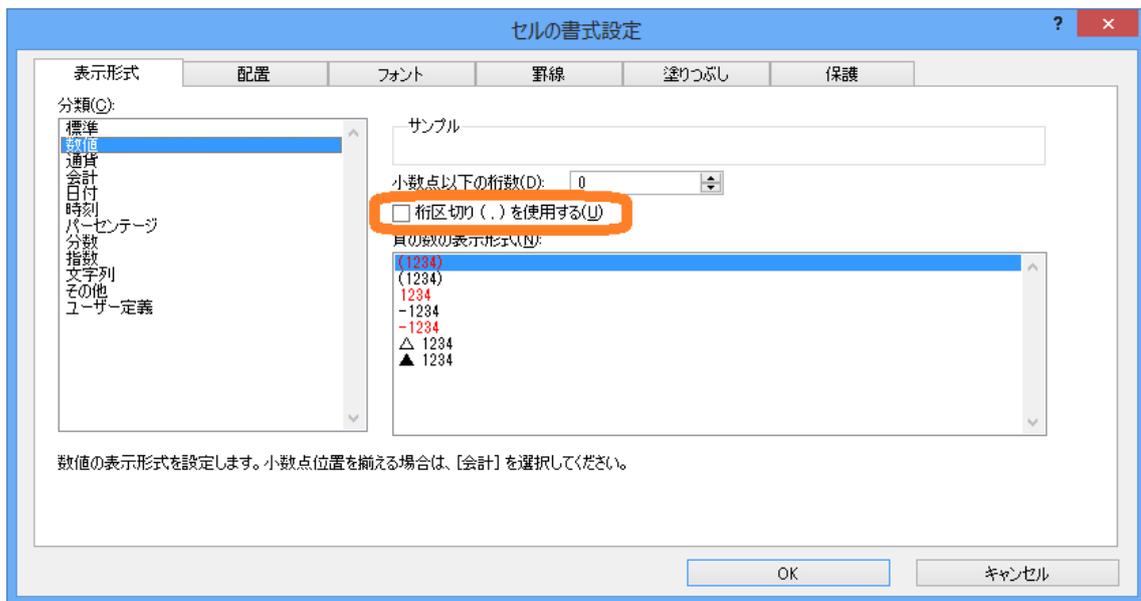


図 26: セルの書式設定ウィンドウ (Microsoft Excel 2007)

A.1.3 不要なスペース・改行・カンマの除去

不要なスペース・改行・カンマがカラム全体、行全体、または複数のセルにまたがっている場合は、セルの置換機能を利用して、スペースと改行を除去できる。

Microsoft Excel 2007 以降であれば、除去対象のカラム全体、行全体、または複数のセルを選択し、ホームメニューの「編集」タブにある「検索と選択」というメニュー (図 27) を選択する。Microsoft Excel 2003 以前であれば、「編集」→「置換」メニューを選択する。検索する文字列欄に空白を入力し、置換する文字列欄空にして「置換」ボタンを押すと、スペースを除去できる。

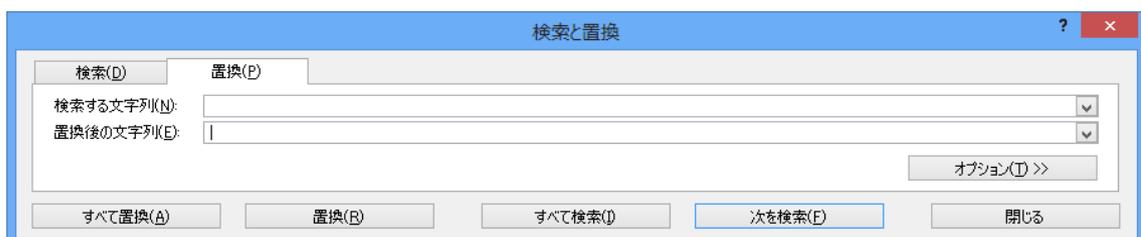


図 27: 検索と置換ウィンドウ (Microsoft Excel 2007)

位取りのためのカンマは、A.1.2 節と同じ方法、つまりセルの書式設定メニューを利用して除去する。「セルの書式設定」ウィンドウの「分類」項目から「数値」を選択し、右側にある「桁区切りを使用する」チェックボックスを外せば、位取りのためのカンマを除去できる。

この作業が完了した時点で、データセットは図 28 のようになる。

都道府県	2000年 (1000人)	2005年 (1000人)	人口集中 地区	人口増減率 (2000～ 2005年) (%)	2005年 (1000人)	人口性比 (女性100に 対する男性)	人口密度 (人/kin ²)	人口増減率 (2000～ 2005年) (%)
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	2385	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	8938	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	8732	8250	3.6	9048	100.9	3745.4	2.9
新潟	877	868	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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 28: 不必要なスペース・改行・カンマ・キャプション・脚注を除いたデータセット

A.1.4 年の値を西暦で記載

西暦の年を記載するためには、2つの方法がある。

- 和暦を記載しているセルを書き換える。
- 和暦を記載しているカラムの隣に、西暦を記載するカラムを追加する。

今回の例では、前者の方法をとる。

A.1.5 セルの結合を解除

セルの結合を解除する。解除した結果生じた空白セルには、解除前に記載されていた値をコピーする。

A.1.6 省略されたセルをコピー

前行と同じ値であるため記載が省略されているセルには、前行の値をコピーする。

A.1.7 タイトルを1行にまとめる

図 25 のタイトルは構造を持っており、セルの結合を利用してその階層を表現している。これを1行で表現するために、タイトルの文言を変更する。たとえば、左から4番目のセルのタイトルを「2005年の人口集中の人口」とする。

A.1.8 データの単位を明記

図 25 の左から2番目、3番目のカラムの単位は「1,000人」である。「人」は助数詞であるので、それを示す文言をタイトルに追加して「人」を除去する。たとえば、2番目、3番

目のカラムのタイトルを、それぞれ「2000年の人口」「2005年の人口」とすればよい。

A.1.9 単位や記数単位を別の行に移行

図 25 の左から 2 番目, 3 番目のカラムの単位「1,000 人」は, タイトルに含まれている。このため, タイトルの下に 2 行追加し, そこに単位と記数単位を記述する。

これまで整形を行った結果は図 29 の通りである。これは, 表形式データのレベル 1 を満たしている。

都道府県	2000年の人口	2005年の人口	2005年の人口集中地区	2000~2005年の人口増減率	2010年の人口	2010年の人口住比 (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
	1000	1000	1000	%	1000		/km2	%
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
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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図 29: 整形完了後のデータシート

A.2 表形式データをレベル2に整形する手順

表形式データをレベル2に整形する手順は、[9]の付録Aを参照のこと(図30, 図31).

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	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
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	1859	1842	805	-0.9	1817	88.5	245.4	-1.3
大分	1221	1210	537	-0.9	1197	89.4	188.7	-1.1
宮崎	1170	1153	512	-1.4	1135	88.5	146.7	-1.5
鹿児島	1786	1753	689	-1.8	1706	87.6	185.7	-2.7
沖縄	1318	1362	893	3.3	1393	96.3	611.9	2.3

図30: レベル1表形式データにヘッダを追加

```

@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,,,,,
,,,%,,,/km2,%
@@Datatype,,,,,
xsd:string,xsd:integer,xsd:integer,xsd:integer,xsd:double,xsd:integer,xsd:double,xsd:d
oule,xsd:double
,,,,,
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
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,1857,1867,762,0.5,1855,95.0,321.0,-0.7
滋賀,1343,1380,614,2.8,1411,97.6,351.2,2.2
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,1443,1421,908,-1.5,1401,90.0,379.5,-1.4
和歌山,1070,1036,408,-3.2,1002,88.8,212.0,-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,1528,1493,708,-2.3,1451,89.2,237.4,-2.8
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,814,796,341,-2.2,764,88.6,107.6,-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,1859,1842,805,-0.9,1817,88.5,245.4,-1.3
大分,1221,1210,537,-0.9,1197,89.4,188.7,-1.1
宮崎,1170,1153,512,-1.4,1135,88.5,146.7,-1.5
鹿児島,1786,1753,689,-1.8,1706,87.6,185.7,-2.7
沖縄,1318,1362,893,3.3,1393,96.3,611.9,2.3

```

図 31: 図 30 を CSV 形式で出力

付録 B 文書データの整形手順

本章では、文書データをレベル 1 およびレベル 2 に整形する手順を示す。

B.1 文書データをレベル 1 に整形する手順

Microsoft Word を利用して文書データをレベル 1 に整形するためには、下記作業を行えばよい。

1. 部・章・節などの構造と、見出しレベルと対応づける。
たとえば、部は「見出し 1」、節は「見出し 3」、小節は「見出し 3」、小々節は「見出し 4」、図表タイトルは「図表番号」に対応づける (図 32)。
2. 対応づけた規則に従って、文章を執筆する。その際、整形のために空白や改行を挿入しないように留意する。

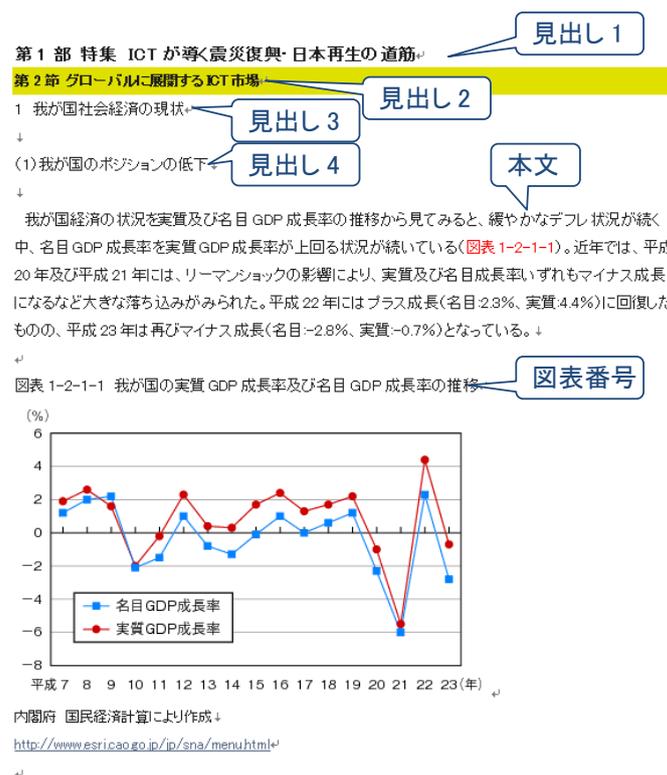


図 32: 文章の構造と見出しを対応付ける例¹¹

¹¹ 総務省「平成 24 年版 情報通信白書」による。
<http://www.soumu.go.jp/johotsusintokei/whitepaper/index.html>

B.2 文書データをレベル2に整形する手順

Microsoft Word を利用して文書データをレベル2に整形するためには、以下の作業を行えばよい。

1. 編集した文書を、HTML形式やXML形式で書き出す。
Microsoft Word であれば「ファイル」→「名前をつけて保存」の順に選択し、「ファイルの種類」を「Web ページ (.html)」に指定すれば、HTML形式で書き出す。
2. 生成したファイルに、図表を構成する表形式データのリンクを追加する。