

狛江市オープンデータ作成マニュアル

オープンデータの意義とCSVファイルの作成

第1.0版

狛江市

○オープンデータとは

国の「オープンデータ基本方針」において、オープンデータとは、国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工・編集・再配布）できるよう、以下のいずれにも該当する形で公開されたデータを指します。

1. 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
2. 機械判読に適したもの
3. 無償で利用できるもの

つまり、紙の帳票や台帳ではなく、テキストやCSVといったコンピュータで扱いやすいデータであって、複製や改変、再配布などの二次利用（商用利用を含む）の許可を明示したものが「オープンデータ」にあたります。

「データ」という言葉からは、数字や文字を羅列したものとといったイメージが浮かびますが、写真や画像のファイルもデータに含まれるため、二次利用が可能な状態で公開すれば広義の「オープンデータ」と呼ぶことができます。

スマートフォン、タブレット端末、SNSの普及等を背景に、多種多様な情報を相互に連携させて新たな価値を生み出すことが期待されており、特に国や自治体が保有する公共データが、国民や企業に利用されやすい形で公開されることが国内外でも求められています。

こうした背景を受けて平成28年12月14日に公布・施行された「官民データ活用推進基本法（官民データ法）」第11条では、国、地方公共団体が保有する官民データについて国民がインターネット等を通じて容易に利用ができるよう措置を講じることが義務付けられました。また、この官民データ法に基づき、令和2年7月17日に閣議決定された「世界最先端IT国家創造宣言・官民データ活用推進基本計画」においても、地域におけるオープンデータの利活用促進が掲げられています。

ポイント！

次の2点を満たしたデータを「オープンデータ」と呼ぶ。

- ・機械判読に適したデータ形式（コンピュータで扱いやすいデータ）であること
- ・商用利用も含め、複製や改変などの二次利用を許可した状態で公開していること

○オープンデータの推進は必要なのか

国の「オープンデータ基本方針」では、公共データの活用を促進する意義・目的を、次の通り整理しています。

1. 国民参加・官民共同の推進を通じた諸課題の解決、経済の活性化
2. 行政の高度化・効率化
3. 透明性・信頼性の向上

地方自治体においてオープンデータに取り組む際には、上記の意義に加えて、公共データの公開と利活用により「地域の課題を解決する」という視点も大切になります。

前提として、行政機関が業務で作成したデータは元をただせば税金で作られたものであるため、公開可能なものであれば市民、国民の公共財として活用されるべきです。

このような考え方を「Open by Default（オープンバイデフォルト：個人情報や安全保障に関わる情報以外は公開を原則とする）」といい、オープンガバメント（開かれた政府）の流れと相まって欧米をはじめ世界中に広まっています。

地方自治体は、限りある予算や人員で業務を遂行しています。行政が地域の課題に対し、今後も継続的に取り組んでいくためには、業務を効率化し、より優先度の高い業務に注力していくことが求められおり、オープンデータの取り組みに合わせて、各担当部署が紙ベースで管理していたデータを電子化したり・同じようなデータを集約して一元管理したり、ホームページ等への掲載手続を見直すことで業務の効率化につなげることができます。

また、現在は日本全体で急速な高齢化や人口減少が進み、社会的・地域的な課題が増加しつつあるなか、市民ニーズは拡大・細分化し続けており、もはや行政組織だけで地域の課題や要望全てに対応していくことは困難です。

従来のように地方自治体が主体となって地方課題の解決を担うのではなく、市民や地域社会が一体となって地域の課題を解決し、新たな活力や経済活動を生み出す地盤を整えるため、オープンデータという形で情報を提供し、市民が積極的に行政に参画し地域をよりよくしていくための流れを生み出していくことが必要なのです。

ポイント！

次の理由から、行政はオープンデータに取り組むべきである。

- ・行政として作成したデータはオープンバイデフォルトの精神で公開する必要がある
- ・データを通して地域の状況や課題を共有することで、地域全体で課題を解決していくきっかけとなる

○オープンデータ化を進めることで何が起きるのか

これまで行政機関だけが持っていたデータを広く市民や地域に共有することで、データの可視化・分析などを通して市民の行政参画や市民協働の動きを促進させることができます。また、ICTスキルを持った市民や地域コミュニティによって、データを活用した地域課題解決のためのアプリ・サービスの開発が期待されるほか、民間企業等によるデータ活用によって新たなビジネスが生まれ、地域経済が活性化することが期待されます。

ポイント！

オープンデータを推進することで、次のような動きが期待される。

- 地域とデータ（情報）を共有することで、市民協働の動きが促進される
- ICT技術を活用した、地域課題解決のための新しいサービスが生まれる可能性が広がる
- 民間企業等によるビジネスの創出で地域経済の活性化が期待される

○一見意味のないデータでも公開する必要があるのか

一見して意味がないと思われるデータでも、加工したり、他のデータと組み合わせることによって様々な価値を生み出すことができます。

データの価値や意味は、データを出す側（行政側）が判断するものではなく、利用者が判断するものですので、行政の価値観や常識に当てはめて考えるのではなく、「可能なものから速やかに」オープンデータ化を進めてください。

ポイント！

- 一見意味のないデータでも、加工や組み合わせによって価値を生み出せる
- 行政はデータの意味や活用方法は考えず、ただ公開していくだけで良い

○オープンデータは「情報公開制度」とはどう違うのか

情報公開制度はいわゆる「行政機関の保有する情報の公開に関する法律（情報公開法）」に基づいて行うもので、請求権者から情報の開示請求を受けた場合、請求された情報に限って情報を開示する受け身の仕組みです。

これに対してオープンデータは税金で生み出された情報は、国民の資産であり、行政はその情報を預かっている管理人であることから、原則として公開可能な情報は最初から全て公開し、常に利用可能な状態（Open by Default）になっているべきという考え方で取り組みが進められています。

そのため、オープンデータは単なる情報提供ではなく、データの活用を促すため機械可読な形式かつ二次利用が可能なルールでの公開が前提となっています。これにより、ただの数値の羅列であるデータをグラフなどで可視化したり、アプリやサービスの基礎データとして組み込むことが民間企業や市民の手で自由に行うことができるのです。

ポイント！

オープンデータと情報公開制度との違いは次のとおり。

- ・情報公開制度は開示請求を受けて情報を公開するが、オープンデータは請求等の行為なく最初から公開することを前提としている
- ・オープンデータとして公開される情報は「データの活用（二次利用）」を前提としているため、民間企業や市民の手で自由に加工・活用することができる

○データの公開は誰が主体になって行うのか

データの公開判断や公開のためオープンデータとする作業は、そのデータを保有・管理している所管課で行います。

全体的な方針の決定や、データの整形に関する技術的なサポート、東京都カタログサイトへの公開については情報政策課で行いますので困ったときはお問い合わせください。

○保有しているデータをオープンデータにするにはどうしたら良いのか

所管課で保有しているデータをオープンデータにするためには、P12からの資料を参考にして機械判読が可能なデータであることを意識してデータ整備をする必要があります。困ったときには情報政策課にお問い合わせください。

○公開したデータに誤りがあったらどうするのか

誤りを指摘してくれた方に対し、データ不備についての謝罪と、指摘への感謝の意を伝えたいので、データを正しいものに修正します。

公的な行政の文書やデータを公開する上では間違いがないことが当然ではありますが、人間が作るものである以上、絶対に間違いがないことを保証することはできません。

そのため、情報を行政内部でクローズドにしておくのではなく積極的にオープンデータとして公開し、外部の目からも正誤をチェックしていただける状態にしましょう。

内部だけでは気付けない不備を外部から指摘してもらえる可能性が生まれることで、行政側としてもデータの誤りを早期に正す機会が得られることとなります。

また、東京都カタログサイトの「東京都オープンデータ利用規約」の中では、「コンテンツ提供者は、コンテンツについて様々な注意を払っていますが、その内容の完全性・正確性・有用性・安全性等については、いかなる保証も行いません。」として、無保証、免責事項を定めています。

ポイント！

- ・公的なデータであっても人間が作っている以上、絶対に間違いがないものを作るのは不可能である
- ・誤りを指摘されたら「嫌だなあ」ではなく「早く誤りを修正できて良かった！」と考える

○オープンデータにすべきでない情報の例はどんなものか

行政機関が扱う情報のなかには、広く公開すべきでない情報も当然含まれています。例えば「個人情報」はオープンデータにすべきではなく、「データを公開することで明らかに第三者が不利益を被る情報」も同様です。

上記の例も含め、情報公開制度における「不開示情報」に定められているものはオープンデータ化の対象にはなりません。

ただし、不開示情報がデータ全体のうちごく一部であれば、その部分を削除したデータをオープンデータ化することが望ましいです。

ポイント！

- 個人情報や第三者が不利益を被る情報、不開示情報は公開できない
- 不開示にあたる情報がデータ全体のうちごく一部であれば、その部分を削除してオープンデータ化を行う

〇インターネット上の情報は、そもそも自由に使って良いのでは

文書やデータには作成された時点で著作権が発生するため、公共的な情報であっても許可なく二次利用を行うことはできません。

これはウェブサイトにおいても同様で、一般的な自治体ウェブサイトではサイトポリシーが掲載されており、本市ウェブサイトでも、著作権法上認められた場合を除いて無断で転載・複製等を行うことは禁止する旨の記載がなされています。

もちろん悪意のある利用者ならば、サイトポリシーや利用規約など一切関係なくウェブサイト上のデータを無断でダウンロードして使うこともできますが、善意で利用しようとしている方はきちんと許可を取る手間と時間をかけることになり（許可されるかどうかは分からない）、それが面倒でデータ活用を止めてしまう場合もあります。これは非常にもったいなく、不公平な状態です。

そのため、善意ある方に、手間と時間をかけずに自由なデータ利用を促すためには、クリエイティブ・コモンズ・ライセンスなどの表示によって、二次利用が可能なオープンデータであると明示することが望ましい状態です。

ポイント！

- ・行政組織が作成した公共的な情報であっても、ウェブサイト上に掲載されているものを無断で転載・複製等の二次利用することは禁止されている
- ・ウェブサイト上のデータは技術的には誰でも取得し使うことができるので、悪意を持ったユーザーはいくらでも無断使用できる。ウェブ上のデータに使用制限を課すのは、善意のユーザーに負担をかけるだけであるため、クリエイティブ・コモンズ・ライセンス等を表示し自由に使うてもらうことが望ましい

○公開データを勝手に販売されてしまったらどうするのか

まず、公開しているデータがそのまま販売されるというケースはほぼ無いと言えるでしょう。なかなか見つからない貴重な情報はそれだけで価値がありますが、オープンデータとして無料で手に入るような情報はそのまま販売しても価値がありません。もし販売するならば、例えば複数の情報を組み合わせて冊子として分かりやすくまとめるなど、何かしらの付加価値をつけることが前提となります。

また、万一仮に情報がそのまま売れたとしても、「(購入者に)情報が伝わる」という意味ではメリットが発生しており、購入しなければ、購入者はその情報を知りえなかったことから、「情報を広めた」という部分に対して付加価値が認められるべきです。

そもそもモノや情報の価値は提供する側ではなく受け取る側によって変わるため、どんなものであろうと受け手側が価値を認めれば対価は支払われるべきです。その価値判断および契約は民法にもとづいて直接の販売者と購入者の間で行われるものであって、本市がそこに関わることはありませんし、むしろ関わるべきではありません。

例えばオークションなどで希覯本が定価の数倍で取引されている場合がありますが、これは購入者がそれだけの価値を認めて販売者に正当な対価として支払うもので、この取引について著者が何らかの関わりや責任を負うことはありません。行政機関が公開するオープンデータについても同様の考え方となります。

ポイント！

- 公開している情報をただ販売される、というケースはほぼ考えられない
- データの販売は販売者と購入者との間でのみ成立する商行為であり、市としては一切関係しない

○公開したデータが悪用・改ざんされる危険性はないのか

データの悪用・改ざんは常に起こりうるもので、オープンデータに限った問題ではありません。

例えばウェブサイトなどを通してインターネット上に公開されたデータは、ダウンロードや悪用、不正コピー、改ざんなどが容易に可能であり、それらを技術的に防ぐことは非常に困難です。

また紙媒体で発信されたデータであったとしても、記載内容をマスキングして上書きしたり切り貼りすれば、やはり改ざん等は可能であり、本来悪用や改ざんのリスクはオープンデータでなくても常に存在しており、むしろ「紙媒体なら絶対安心！」という考え方は逆に危険です。

つまりオープンデータであるかどうかに関係なく、悪意のある利用者であれば勝手にデータを取得して悪用してしまうのです。

上記のとおり、技術的に改ざん等の悪用を防ぐことはできませんが、「職員が定期的にネットや紙媒体を検索・閲覧し改ざんされた情報が出回っていないかチェックする」、「善意ある人からの発見連絡を受ける」という方法で事後対応は可能です。しかし職員負荷を考えると、前者は現実的ではありません。

正しい情報がオープンデータとして地域に広く周知されていれば、誤った情報が流通された場合でも原本として市で直接公開しているデータとの比較がしやすくなり、善意ある人からの改ざん情報等の発見連絡の確率が高まります。

つまりオープンデータ化によっていろいろな方の監視の目が集まることで、悪意ある人間の行動が発覚しやすくなり結果的に改ざん等のリスク対策となるのです。

ポイント！

- データの不適切な利用や改ざんは、オープンデータだけに発生するものではない
- オープンデータに限らず情報の悪用は可能であり、それを技術的に防ぐ手段はない
- 正しいデータを広く公開することで外部の目が集まり、改ざん等のリスク対策になる

○「二次利用を許可する」ということは、データの不適切な利用や改ざんも認めるのか

データの二次利用を許可する＝データの不適切な利用も含めて全て認める、というわけではありません。不適切な利用・望まない利用に対しては改善指示や訴えを行うことができます。

オープンデータとして二次利用を許可する場合、一般的に「クリエイティブ・コモンズ・ライセンス (<http://creativecommons.jp/licenses/>)」を適用しますが、東京都のカタログサイトもこのライセンスを適用しています。

このライセンスにおいては「著作権人格権」は著作者（市）が保持するため、意に反する利用・改変を受けた場合は著作権人格権に基づき改善指示や訴えを行います。

またデータの不適切な利用によって市が直接的・間接的な被害を被った場合は、オープンデータであるかどうかに関わらず法律において対応すべきです。

ポイント！

- ・「データの二次利用の許可＝データの不適切な利用等を許可する」という意味ではない
- ・データの不適切な利用や市・第三者への被害については法律において対処すべきである

○公開したデータはどのように使われるのか

例えば、これまで数値の羅列だけで表現されていたデータをグラフ化したり地図上に表示、経年での数値の変化などを視覚的に表現するようなアプリケーションが作られています。

ただしこれはほんの一例で、今後の技術革新や時代の変化とともに、これまでに想像もできなかった新しいサービスができる可能性もあります。

いずれにせよ、行政の役割は「市民サービスとしてオープンデータで情報を公開すること」までで、先々の活用方法や可能性はデータを使ってくれる方が知恵を絞って考える部分です。データを公開する時点では、活用方法や意味についてあまり深く考えこまないようにしましょう。

ポイント！

- ・データをグラフ化したり、地図上に表示したりといった視覚化サービスに使われている（一例）
- ・先々どう使われるか、何ができ上がるかは予測できないので、深く考えなくても良い

○オープンデータを進めると情報格差（デジタルデバイド）が広がり、高齢者の方などは情報を得にくくなってしまおうのでは？

「オープンデータ化」は必ずしも「情報発信の電子化」とイコールではありません。

公開・蓄積される情報自体は機械処理できるようなデータ形式が推奨されていますが、情報の発信形式や伝達手段は個人の利便性や伝わりやすさに合わせた多様なものであるべきです。

むしろデータ化が進むことによって、いろいろな伝達手段を広く選択することができます。例えば紙に印刷された情報は、「コピーして配る」「口頭で伝える」といった伝達手段しか取れず、情報伝達に手間と時間がかかりますが、データとして蓄積されていれば、印刷する・メールで送る・音声として読み上げる・自動で多言語翻訳するなど様々な加工が瞬時にできるため、様々な立場の方に適した形で素早く情報を提供することができます。

ポイント！

- ・「オープンデータ化＝情報発信の電子化」ではない。伝達方法は様々な手段があって然るべき
- ・情報がデータとして蓄積されていれば、ICT 技術によって様々な形に加工して伝達できる

会津若松市のオープンデータQ&Aのコラムより

- ・オープンデータは不公平？

大量のデータを視覚化・分析したり、新たなサービスを生み出すためには一定以上のICTに関する知識が必要になります。そのため「オープンデータはICT を活用できる個人や企業のためだけの施策」だと捉えられることがありますが、それは誤りです。

もしあなたが今現在ICT 技術をうまく活用できなくても、個人の努力により技術の習得は可能であり、学習の機会や方法論はいくらでも手に入れることができます。

もしくは、身近にICT を活用できる方がいれば、その方と一緒にデータ活用を進めていくのも方法の一つでしょう。

オープンデータは「結果均等」ではなく「機会均等」の施策です。機会は誰にでも平等に与えられていますが、それを活用してどのような結果を掴むかは個人の考え方や努力次第なのです。

機械判読可能なデータの作成方法

第1章 機械判読可能なデータの作成 Excel 形式による統計表の作成

第1節 データ形式における留意点

チェック項目1-1：ファイル形式はExcelファイルかCSVとなっているか

統計データを機械判読可能なデータとして作成する場合、そのファイル形式は、データベース形式として公表する場合を除いて、仕様が公開・標準化されているもの、または国際的に標準化されているフォーマットであることが必要で、具体的には、Excelの保存形式であるxlsx (xls) やCSV、XML、Open Document Format (.ods) などが挙げられます。

また、統計表の解説や注釈、図が入った報道発表資料や報告書ベースのファイル (PDF や印刷することを目的として作成した表形式データ) で掲載する場合、それとは別に基データを機械判読可能な状態で登録するようにすることが必要です。

第2節 Excel ファイルによる統計表のレイアウト

本節では、ファイル形式による統計表における留意点を以下で説明していきます。統計データを作成する際には、本節で示すチェック項目を一つずつ踏まえ改善していくことで、従来のPDF形式による統計表や、印刷して報告書として提供することを主眼にした表形式データを、機械判読可能なデータとして利用できる統計表に改善することが可能となります。

第1項 データ・項目の取扱い

チェック項目1-2：ヘッダは1行の構成となっているか

ファイルのヘッダが、例の様に結合されて2行となっている場合、このようなファイルを機械が判読するためには、ヘッダとデータの切れ目を確認する必要が生じ、判読手順が複雑になってしまいます。このため、ヘッダについては1行で構成する必要があります。

例

修正前				修正後			
ヘッダが結合して2行となっている				ヘッダを1行にまとめた状態			
月	気温			月	A地区の気温	B地区の気温	C地区の気温
	A地区	B地区	C地区				
1月	18	17	18	1月	18	17	18
2月	18	15	17	2月	18	15	17
3月	20	22	20	3月	20	22	20
4月	24	22	23	4月	24	22	23

□チェック項目1-3：1セル1データとなっているか

1つのセルに複数のデータが入力されていると、計算や昇順・降順の並べ替え、コピーペーストやグラフ化等加工編集する場合に多くの手作業やプログラムの作成が必要となり、データとしてすぐに利用できないことから、1セルには1つのデータの入力とすること。

例1

修正前		修正後		
1セルに複数のデータが入力されている		1セル1データとして入力した状態		
	全国	年度	仕入額（全国）	出荷額（全国）
仕入額	373（平成27年度）、434（平成28年度）、549（平成29年度）、638（平成30年度）、741（平成31年度）	平成27年度	373	973
出荷額	973（平成27年度）、1234（平成28年度）、1449（平成29年度）、1738（平成30年度）、1841（平成31年度）	平成28年度	434	1234
		平成29年度	549	1449
		平成30年度	638	1738
		平成31年度	741	1841

※分かりやすい例として、この表にしたが、修正の際には西暦表示を追記することが好ましい

仕入額と出荷額のセルに複数のデータが入力されている。

このような場合は、年度ごとに列・行を分けて、各セルにデータを分離して入力すること。

例2

修正前				修正後			
1セルに複数のデータが入力されている				1セル1データとして入力した状態			
	企業等数	売上金額（費用総額）		種別	企業数等（件）	売上金額（円）	費用総額（円）
総計	1188389	391445	(355943)	総数	1188389	391445	355943
A	50384	69565	(63883)	A	50384	69565	63883
B	154138	50468	(45332)	B	154138	50468	45332
C	86522	44098	(33210)	C	86522	44098	33210
D	85983	22258	(20237)	D	85983	22258	20237
E	580003	37652	(35016)	E	580003	37652	35016
F	27456	15288	(14633)	F	27456	15288	14633
G	119085	115499	(110735)	G	119085	115499	110735
H	84818	36617	(32897)	H	84818	36617	32897

売上金額と費用総額が1セルに入力されているため、例えば、売上金額に占める費用総額の割合を計算したい場合、修正前の場合、「()」内の数字を分離してから、割り算を行う必要が生じます。一方、修正後の場合では、費用総額列の値を売上金額列の値で割るだけでよく、単純な処理で計算が可能となります。

そのため、1セル1データの入力とすることが重要です。

□チェック項目1-4：数値データは数値属性とし、文字列を含まないこと

数値データに、「円」、「¥」、「kg」、「ト」などや「マイナス記号（▲）」を文字列として入力すると、Excel では数値ではなく文字列として扱われてしまうため、関数等で計算ができなくなるほか、昇順・降順等の並べ替えも正確にできない場合があります。

また、人間が見やすくするための工夫として、千円単位を示す「,（カンマ）」を文字列として直接入力している場合、関数によっては正確に計算できない場合があるほか、この様な Excel ファイルを csv 化すると、csv はカンマでデータを区切っているため、想定していない様な動きをしたり、おかしいデータが作成されたりする場合があります。

そのため、数値データは数値属性として、文字列を含まないようにすることが重要です。

（補足）

Excel の場合「書式設定」により体裁を整えることは可能であるため、データは数値として扱いつつ、見た目の体裁を整えることができます。

ただし、書式設定を施した Excel ファイルを csv 化すると、おかしいデータが作成される場合があるため、この機能は印刷を意識した報告書ベースの Excel にのみ使用する等、使い分ける必要があります。

例1 数値データ内に文字列が含まれる場合

修正前				修正後			
「円」、「▲（マイナス表記）」、「,（カンマ）」が文字列として入力されている				数値データを数値属性として入力した状態			
	単価	前回差分	生産台数		単価	前回差分	生産台数
サンプル1	10,030円	130	12,000	サンプル1	10030	130	12000
サンプル2	9,100円	▲200	29,000	サンプル2	9100	-200	29000
サンプル3	8,020円	▲350	37,000	サンプル3	8020	-350	37000
サンプル4	7,500円	500	43,000	サンプル4	7500	500	43000
SUM関数	0	630	0	SUM関数	34650	80	121000
+(加算演算)	#VALUE!	#VALUE!	121000	+(加算演算)	34650	80	121000
↑黄色の行は関数で合計を表示した例				↑黄色の行は関数で合計を表示した例			

「円」、「▲（マイナス表記）」、「,（カンマ）」が文字列として入力されているため、関数計算を行うとエラーとなります。

数値データは修正後の様に数値属性として入力することで、関数計算が可能となります。

例2 数値データ内に空白がある場合

修正前			修正後		
桁区切りのために空白が含まれている			空白を除いた状態		
	仕入台数	在庫台数		仕入台数	在庫台数
サンプル1	5 000	1 030	サンプル1	5000	1300
サンプル2	4 000	1 030	サンプル2	4000	1800
サンプル3	3 000	1 000	サンプル3	3000	1000
サンプル4	2 000	1 600	サンプル4	2000	1600

数値データに、空白が含まれる場合、数値ではなく文字列として扱われることから、関数計算を行うとエラーとなり、昇順・降順等の並び替えも正確にできなくなります。

そのため、数値データは修正後の様に数値属性とし、文字列を含まないことが重要です。

例3 数値データ内に注釈・脚注が含まれる場合

修正前			修正後		
セルに注釈が含まれている			注釈を削除した状態		
	商品 A	商品 B		商品 A	商品 B
1月	151 a)	139	1月	151	139
2月	182	201 b)	2月	182	201
3月	131	123	3月	131	123

※ a)は冷凍、b)は調理済

修正前の様に注釈を含むファイルは、人間がデータを見て解釈するためには必要となりますが、数値データに注釈やそれを表す記号が含まれている場合、数値ではなく文字列として扱われることから、同じ列について数値だけのデータと、文字列を含めたデータが混在してしまうこととなります。

機械によってデータを処理する際には、1つの列には同じ種類のデータ形式（数値や文字列など）であることを前提としているため、注釈・脚注等の文字列が混在すると、その列は数値を含めて、全て文字列として扱われてしまい、その後の処理でおかしなデータが作成されることがあります。

そのため、注釈や脚注については、例えば、オープンデータが掲載されているホームページ中の「メタ情報」などに記載をします。

□チェック項目1-5：セルの結合をしていないか

表形式のデータのうち、特にレコード形式のデータを機械判読可能なデータとして利用するためには、1件のデータを、横1列（レコード）で入力又は縦1列（カラム）で入力する必要があります。

レコード及びカラムはそれぞれ独立しており、レコードの場合は上下の並べ替えをしてもデータの意味が変わらず、カラムの場合は左右で入れ替えても意味が変わらないようにセルの結合又は不必要な分離を行わないようにします。

例1

修正前		修正後																											
セルが結合（又は分離）されている		セルの結合を解除した状態																											
	<table border="1"> <thead> <tr> <th>管理職</th> <th>従業員数（上段は正社員、下段はパート）</th> </tr> </thead> <tbody> <tr> <td>第一営業所</td> <td>3</td> </tr> <tr> <td></td> <td>15</td> </tr> <tr> <td></td> <td>2</td> </tr> </tbody> </table>	管理職	従業員数（上段は正社員、下段はパート）	第一営業所	3		15		2		<table border="1"> <thead> <tr> <th>管理職</th> <th>従業員数（正社員）</th> <th>従業員数（パート）</th> </tr> </thead> <tbody> <tr> <td>第一営業所</td> <td>3</td> <td>15</td> </tr> <tr> <td></td> <td></td> <td>2</td> </tr> </tbody> </table>	管理職	従業員数（正社員）	従業員数（パート）	第一営業所	3	15			2									
管理職	従業員数（上段は正社員、下段はパート）																												
第一営業所	3																												
	15																												
	2																												
管理職	従業員数（正社員）	従業員数（パート）																											
第一営業所	3	15																											
		2																											
	<table border="1"> <thead> <tr> <th>市区町村</th> <th>生産本数</th> </tr> </thead> <tbody> <tr> <td>ちよだく</td> <td>58408</td> </tr> <tr> <td>千代田区</td> <td></td> </tr> <tr> <td>ちゅうおうく</td> <td>141183</td> </tr> <tr> <td>中央区</td> <td></td> </tr> <tr> <td>みなとく</td> <td>243283</td> </tr> <tr> <td>港区</td> <td></td> </tr> </tbody> </table>	市区町村	生産本数	ちよだく	58408	千代田区		ちゅうおうく	141183	中央区		みなとく	243283	港区			<table border="1"> <thead> <tr> <th>市区町村</th> <th>ふりがな</th> <th>生産本数</th> </tr> </thead> <tbody> <tr> <td>千代田区</td> <td>ちよだく</td> <td>58408</td> </tr> <tr> <td>中央区</td> <td>ちゅうおうく</td> <td>141183</td> </tr> <tr> <td>港区</td> <td>みなとく</td> <td>243283</td> </tr> </tbody> </table>	市区町村	ふりがな	生産本数	千代田区	ちよだく	58408	中央区	ちゅうおうく	141183	港区	みなとく	243283
市区町村	生産本数																												
ちよだく	58408																												
千代田区																													
ちゅうおうく	141183																												
中央区																													
みなとく	243283																												
港区																													
市区町村	ふりがな	生産本数																											
千代田区	ちよだく	58408																											
中央区	ちゅうおうく	141183																											
港区	みなとく	243283																											

セルが結合（又は分離）されてしまうと、機械判読に適さなくなるので、1件のデータは、横1列で表記し、セルの結合又は不必要な分離を行わないことが重要です。

例2

修正前					修正後					
セルが結合されている					セルの結合を解除した状態					
都道府県	エリア	市区町村	導入台数	増減数	都道府県	エリア	市区町村	導入台数	増減数	
東京都	特別区部	千代田区	58406	11291	東京都	特別区部	千代田区	58406	11291	
	特別区部	中央区	141183	18421	東京都	特別区部	中央区	141183	18421	
	特別区部	港区	243283	38152	東京都	特別区部	港区	243283	38152	
	Σ					Σ				
	特別区部	葛飾区	442913	327	東京都	特別区部	葛飾区	442913	327	
	特別区部	江戸川区	681298	2331	東京都	特別区部	江戸川区	681298	2331	
	市町村	八王子市	577513	-2540	東京都	市町村	八王子市	577513	-2540	
	市町村	立川市	176295	-3373	東京都	市町村	立川市	176295	-3373	
	市町村	武蔵野市	144730	5996	東京都	市町村	武蔵野市	144730	5996	
	市町村	三鷹市	186936	853	東京都	市町村	三鷹市	186936	853	
市町村	青梅市	137381	-1958	東京都	市町村	青梅市	137381	-1958		
市町村	府中市	260274	4768	東京都	市町村	府中市	260274	4768		
Σ					Σ					
市町村	青ヶ島村	-	-	東京都	市町村	青ヶ島村	-	-		
市町村	小笠原村	-	-	東京都	市町村	小笠原村	-	-		

セルを結合した場合、並べ替えができなかったり、処理がエラーとなったり、グラフ化ができなくなったりします。

また、次の例のとおり、範囲の選択がしづらかったり、コピーペーストができない状態となります。

例：市区町村と「導入台数」をコピーすべく、範囲選択しても…

市区町村	導入台数	増減数
西東京市	200012	3501
瑞穂町	33445	-52
日の出町	17446	796
檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-

セルが結合されている場合、青ヶ島村以降は不要な「増減数」まで選択されてしまう…

市区町村	導入台数	増減数
西東京市	200012	3501
瑞穂町	33445	-52
日の出町	17446	796
檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-


別の表からコピーして貼付けようとしても…

檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	2749	-134
小笠原村	1891	2

エラーになって貼付けできない…

檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-

Microsoft Excel

 この操作は結合したセルには行えません。

ヘルプの表示(E) >>

OK

□チェック項目1-6：スペースや改行等で体裁を整えていないか

スペースや改行等で体裁を整えた場合、データの検索性が低下するほか、複数の表を横断的に利用する場合においても支障が生じる可能性があります。

そのため、体裁を整えるためのスペースや改行等は削除する必要があります。

例1

修正前	修正後																														
スペースで体裁を整えている	スペースによる整形を解除した状態																														
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>地域名</th> <th>出荷本数</th> <th>在庫本数</th> </tr> </thead> <tbody> <tr> <td>津□□市</td> <td style="text-align: center;">429</td> <td style="text-align: center;">756</td> </tr> <tr> <td>四日市市</td> <td style="text-align: center;">321</td> <td style="text-align: center;">648</td> </tr> <tr> <td>伊勢市</td> <td style="text-align: center;">384</td> <td style="text-align: center;">438</td> </tr> <tr> <td>松阪市</td> <td style="text-align: center;">408</td> <td style="text-align: center;">775</td> </tr> </tbody> </table>	地域名	出荷本数	在庫本数	津□□市	429	756	四日市市	321	648	伊勢市	384	438	松阪市	408	775	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>地域名</th> <th>出荷本数</th> <th>在庫本数</th> </tr> </thead> <tbody> <tr> <td>津市</td> <td style="text-align: center;">429</td> <td style="text-align: center;">756</td> </tr> <tr> <td>四日市市</td> <td style="text-align: center;">321</td> <td style="text-align: center;">648</td> </tr> <tr> <td>伊勢市</td> <td style="text-align: center;">384</td> <td style="text-align: center;">438</td> </tr> <tr> <td>松阪市</td> <td style="text-align: center;">408</td> <td style="text-align: center;">775</td> </tr> </tbody> </table>	地域名	出荷本数	在庫本数	津市	429	756	四日市市	321	648	伊勢市	384	438	松阪市	408	775
地域名	出荷本数	在庫本数																													
津□□市	429	756																													
四日市市	321	648																													
伊勢市	384	438																													
松阪市	408	775																													
地域名	出荷本数	在庫本数																													
津市	429	756																													
四日市市	321	648																													
伊勢市	384	438																													
松阪市	408	775																													
	住所コードを併記した状態																														
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>地域コード</th> <th>地域名</th> <th>出荷本数</th> <th>在庫本数</th> </tr> </thead> <tbody> <tr> <td>24201</td> <td>津□□市</td> <td style="text-align: center;">429</td> <td style="text-align: center;">141183</td> </tr> <tr> <td>24202</td> <td>四日市市</td> <td style="text-align: center;">321</td> <td style="text-align: center;">243283</td> </tr> <tr> <td>24203</td> <td>伊勢市</td> <td style="text-align: center;">384</td> <td style="text-align: center;">333560</td> </tr> <tr> <td>24204</td> <td>松阪市</td> <td style="text-align: center;">408</td> <td style="text-align: center;">219724</td> </tr> </tbody> </table>	地域コード	地域名	出荷本数	在庫本数	24201	津□□市	429	141183	24202	四日市市	321	243283	24203	伊勢市	384	333560	24204	松阪市	408	219724										
地域コード	地域名	出荷本数	在庫本数																												
24201	津□□市	429	141183																												
24202	四日市市	321	243283																												
24203	伊勢市	384	333560																												
24204	松阪市	408	219724																												

住所などを文字列で入力した場合、表示順が実際の業務で取り扱っている状態と同じにならないことがあるので、例えば地域コード等の住所識別番号を併記することで機械での処理がスムーズになることがあります。

例2

修正前				修正後			
スペースで体裁を整えている				スペースを削除した状態			
分類	総数	事業所数	企業数	分類	総数	事業所数	企業数
合計	900	450	450	合計	900	450	450
A	200	100	100	A	200	100	100
B	300	150	150	B	300	150	150
C	400	200	200	C	400	200	200

分類の「A」以降の項目が総計の内数であることを表すために、「□A」とスペースを挿入することで体裁を整えています。本来の情報とは無関係な情報が入力されているために、関数の VLOOKUP 等を使用して他の統計データと結合しようとしても、同じ分類がマッチングせずに処理ができないなどの可能性があります。

そのため、体裁のためスペースによる整形をしないことが重要です。

※この表で機械判読を行う場合、A・B・Cの合計は機械処理で自動計算することができるので、合計欄は必要なく削除するという判断もできます。人間が表を確認する場合、合計欄があった方が利便性が増しますが、機械にとっては必要がないのです。

例3

修正前			修正後		
改行で体裁を整えている			改行を解除した状態		
薬剤名	出荷本数	単価	薬剤名	出荷本数	単価
鎮静剤 A-1	429	756	鎮静剤 A-1	429	756
鎮静剤 A-2	321	648	鎮静剤 A-2	321	648
鎮静剤 A-3	384	438	鎮静剤 A-3	384	438
鎮静剤 A-4	408	775	鎮静剤 A-4	408	775

項目の体裁を整えるため改行が使用されていますが、その改行に意味があるのかを機械は判別することが出来ません。

そのため、体裁のために改行を使用して整形をしないことが重要です。

□チェック項目1-7：項目名等を省略していないか

人間であれば文脈等を判断して省略されている部分の意味を判断できますが、機械ではその判断をすることができません。

そのため、項目名等を省略しないことが重要です。

例

修正前			修正後		
同じ名称を空白で省略している			省略せずに入力した状態		
薬剤名	出荷本数	在庫本数	薬剤名	出荷本数	在庫本数
鎮静剤 A-1	429	756	鎮静剤 A-1	429	756
2	321	648	鎮静剤 A-2	321	648
3	384	438	鎮静剤 A-3	384	438
4	408	775	鎮静剤 A-4	408	775

人間の目で見ると「鎮静剤A」が全ての項目につくので「1」以降について表記を省略していますが、機械にとっては「2」以下の項目名が何を意味するのか分からない構造になっているため、省略せずに入力することが必要です。

□チェック項目1-8：数式を使用している場合は、数値データに修正しているか

例

修正前					修正後								
合計が数式で入力されたままになっている					数値データに修正した状態								
都道府県	市区町村	合計	男	女	都道府県	市区町村	合計	男	女				
東京都	千代田区	20000	10000	10000	東京都	千代田区	20000	10000	10000				
東京都	中央区	6000	2000	4000	東京都	中央区	6000	2000	4000				
東京都	港区	8000	3000	5000	東京都	港区	8000	3000	5000				
東京都	新宿区	5000	4000	1000	東京都	新宿区	5000	4000	1000				
東京都	文京区	7000	5000	2000	東京都	文京区	7000	5000	2000				
<table border="1"> <tr> <td>f_x</td> <td>=SUM(Q100:R100)</td> </tr> </table>					f_x	=SUM(Q100:R100)	<table border="1"> <tr> <td>f_x</td> <td>20000</td> </tr> </table>					f_x	20000
f_x	=SUM(Q100:R100)												
f_x	20000												

数式を使用してセルの値を自動計算している場合、並べ替え等を行った場合、正確な値が表示されなくなる可能性があります。

そのため、「コピー」「値を貼り付け」の機能等でセルのデータは値のみとすることが重要です。

□チェック項目1-9：オブジェクトを使用してまとめていないか

例

修正前			修正後		
オブジェクトを使用している			オブジェクトを削除した状態		
名称	内容	数量	名称	内容	数量
商品 A	サンプル 1	100	商品 A	サンプル 1	100
	サンプル 2	200	商品 A	サンプル 2	200
	サンプル 3	300	商品 A	サンプル 3	300

セルを結合した上で、括弧等のオブジェクトを使用して体裁を整えていますが、機械による判読には適していません。

この場合、オブジェクトを削除した上で、それぞれのセルにデータを入力することが必要です。

□チェック項目1-10：データの単位を記載しているか

例

修正前			修正後		
単位が表記されていない			単位を入力した状態		
薬剤名	出荷本数	単価	薬剤名	出荷本数 (本)	単価 (円)
鎮静剤 A - 1	429	756	鎮静剤 A - 1	429	756
鎮静剤 A - 2	321	648	鎮静剤 A - 2	321	648
鎮静剤 A - 3	384	438	鎮静剤 A - 3	384	438
鎮静剤 A - 4	408	775	鎮静剤 A - 4	408	775

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とから、単位が含まれるような項目については、その項目の単位を入力することが重要です。

□チェック項目1-11：機種依存文字を使用していないか

例

修正前				修正後			
機種依存文字を使用している				機種依存文字を削除した状態			
	収納済額	収納未済額	収納未済額割合	項目	収入済額 (A)	収入未済額 (B)	収入未済額割合 (B/A)
	①	②	②/①	サンプルA	1000	100	0.1
サンプルA	1000	100	0.1	サンプルB	2000	200	0.1
サンプルB	2000	200	0.1	サンプルC	3000	300	0.1
サンプルC	3000	300	0.1	サンプルD	4000	400	0.1
サンプルD	4000	400	0.1				

機種依存文字は利用者の環境によっては正しく表示されない可能性があります。
そのため、機種依存文字は使用しないこと。

□チェック項目1-12：西暦表記又は和暦に西暦の併記がされているか

例

修正前			修正後		
和暦のみで表記している			西暦で表記した状態		
年次 (暦年)	出荷本数	在庫本数	西暦 (暦年)	出荷本数	在庫本数
平成29年	429	141183	2017年	429	141183
平成30年	321	243283	2018年	321	243283
平成31年	384	333560	2019年	384	333560
令和2年	408	219724	2020年	408	219724
			西暦に和暦を併記した状態		
西暦 (暦年)	和暦 (暦年)	出荷本数	在庫本数		
2017年	平成29年	429	141183		
2018年	平成30年	321	243283		
2019年	平成31年	384	333560		
2020年	令和2年	408	219724		

機械処理においては、「年」については値の大小により認識することが多いため、和暦の表示のみでは、元号が切り替わる際に手作業で西暦へ変換する必要が生じることから、西暦表記への変更、もしくは西暦と和暦を併記することが重要です。

第2項 表の構成の取扱い

□チェック項目2-1：データが分断されていないか

空白列などを追加したり、Excel の1シートに複数の表頭・表側を設定した場合、データが分断されてしまい、機械判読に支障をきたす可能性があります。


そのため、不必要な表の分離は行わないこと。

また、表頭、表側、欄外等に、表の印刷や成形を意識した不必要な空白行（列）を追加しないことも重要です。

例1

修正前								
データが分断されている								
都道府県	エリア	市区町村	導入台数	増減数		材料	仕入単価	仕入額
			台			台	円	円
東京都	特別区部	千代田区	58406	11291		429	28	12012
東京都	特別区部	中央区	141183	18421		321	35	11235
東京都	特別区部	港区	243283	38152		384	42	16128

修正後								
空白の列を削除した状態								
都道府県	エリア	市区町村	導入台数	増減数	材料	仕入単価	仕入額	
			台		台	円	円	
東京都	特別区部	千代田区	58406	11291	429	28	12012	
東京都	特別区部	中央区	141183	18421	321	35	11235	
東京都	特別区部	港区	243283	38152	384	42	16128	



データが分断されているため、空白列を削除し、データが分断されないようにすること。

例2

修正前				修正後			
データが分断されている				空白の行を削除した状態			
都道府県	市区町村	人口	増減数	都道府県	市区町村	人口	増減数
東京都	千代田区	58406	11291	東京都	千代田区	58406	11291
東京都	中央区	141183	18421	東京都	中央区	141183	18421
東京都	港区	243283	38152	東京都	港区	243283	38152
}				東京都	新宿区	333560	7251
東京都	中野区	328215	13465	東京都	中野区	328215	13465
東京都	杉並区	563997	14428	東京都	杉並区	563997	14428
東京都	豊島区	291167	6489	東京都	豊島区	291167	6489
}				東京都	北区	341076	5532
都道府県	市区町村	人口	増減数	東京都	荒川区	212264	8968
東京都	北区	341076	5532	東京都	板橋区	561916	26092
東京都	荒川区	212264	8968	東京都	練馬区	721722	5598
東京都	板橋区	561916	26092	東京都	足立区	670122	-13304
}				東京都	葛飾区	442913	327
東京都	府中市	260274	4768	東京都	調布市	229061	5488
東京都	昭島市	111539	-758	東京都	町田市	432348	5332
東京都	調布市	229061	5468	東京都	小金井市	121396	2544

空白行でデータが分断されているため、空白行を削除し、データが分断されないようにすること。

例3

修正前											
紙面に収めるために表を折り返している											
給与所得者数						給与所得者数					
	3月末	6月末	9月末	12月末	年間月平均		3月末	6月末	9月末	12月末	年間月平均
	千人	千人	千人	千人	千人		千人	千人	千人	千人	千人
平成7年分	55982	55935	55673	55354	55736	平成13年分	64379	64325	64024	63657	64036
平成8年分	55987	56809	56038	56189	56189	平成14年分	64385	65330	64444	64311	64617
平成9年分	56050			163	56689	平成15年分	65378			152	65192
平成10年分	58948			142	58958	平成16年分	67790			158	67112
平成11年分	59264		58300	58108	58761	平成17年分	68154	68280	67045	68824	67575
平成12年分	61190	60901	59276	59114	60151	平成18年分	70513	70036	68167	67981	69174
収録範囲 1						収録範囲 2					
給与所得者数						給与所得者数					
	3月末	6月末	9月末	12月末	年間月平均		3月末	6月末	9月末	12月末	年間月平均
	千人	千人	千人	千人	千人		千人	千人	千人	千人	千人
平成19年分	61580	61529	61240	60889	61310	平成25年分	67178	67122	66808	66425	66889
平成20年分	61586	62490	61642	61715	61808	平成26年分	67184	68171	67246	67108	67427
平成21年分	62535			09	62358	平成27年分	68220			56	68027
平成22年分	64843			86	64194	平成28年分	70738			30	70030
平成23年分	65190	67111	64130	63819	64637	平成29年分	71117	71249	69960	69730	70513
平成24年分	66991	66991	65204	65025	66166	平成30年分	73579	73081	71131	70937	72181
収録範囲 3						収録範囲 4					
修正後											
収録範囲を1つにした状態											
		給与所得者数		給与所得者数	給与所得者数	給与所得者数	給与所得者数				
		3月末		6月末	9月末	12月末	年間月平均				
		千人		千人	千人	千人	千人				
1995年	平成7年分	55982	55935	55673	55354	55736					
1996年	平成8年分	55987	56809	56038	55923	56189					
}											
2017年	平成29年分	71117	71249	69960	69730	70513					
2018年	平成30年分	73579	73081	71131	70937	72181					

資料として紙面1枚に収めるために表を加工して折り返しているため、実際に機械判読により処理を行おうとすると、正確な計算や並べ替えなどができない場合があります。

収録範囲を1つにして、データが分断されないようにすることが重要です。

□チェック項目2-2：1シートに複数の表が掲載されていないか

例

修正前	修正後																																																																																
1シートに複数の表が掲載されている	1つの表を1シートに分割した状態																																																																																
<table border="1"> <thead> <tr> <th>都道府県</th> <th>市町村</th> <th>参加者(人)</th> <th>男(人)</th> <th>女(人)</th> </tr> </thead> <tbody> <tr> <td>東京都</td> <td>千代田区</td> <td>30</td> <td>29</td> <td>1</td> </tr> <tr> <td>東京都</td> <td>中央区</td> <td>31</td> <td>17</td> <td>14</td> </tr> <tr> <td>東京都</td> <td>港区</td> <td>23</td> <td>13</td> <td>10</td> </tr> <tr> <td>東京都</td> <td>新宿区</td> <td>25</td> <td>10</td> <td>15</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>月</th> <th>施設A(人)</th> <th>施設B(人)</th> <th>施設C(人)</th> <th>西暦(暦年)</th> <th>参加者(人)</th> </tr> </thead> <tbody> <tr> <td>1月</td> <td>56</td> <td>55</td> <td>14</td> <td>2018年</td> <td>2340</td> </tr> <tr> <td>2月</td> <td>45</td> <td>23</td> <td>23</td> <td>2019年</td> <td>3129</td> </tr> <tr> <td>3月</td> <td>34</td> <td>11</td> <td>45</td> <td>2020年</td> <td>432</td> </tr> <tr> <td>4月</td> <td>67</td> <td>113</td> <td>36</td> <td>2021年</td> <td>1893</td> </tr> </tbody> </table> <div style="border: 1px solid gray; padding: 2px;"> 参加者まとめ + </div>	都道府県	市町村	参加者(人)	男(人)	女(人)	東京都	千代田区	30	29	1	東京都	中央区	31	17	14	東京都	港区	23	13	10	東京都	新宿区	25	10	15	月	施設A(人)	施設B(人)	施設C(人)	西暦(暦年)	参加者(人)	1月	56	55	14	2018年	2340	2月	45	23	23	2019年	3129	3月	34	11	45	2020年	432	4月	67	113	36	2021年	1893	<table border="1"> <thead> <tr> <th>都道府県</th> <th>市区町村</th> <th>参加者(人)</th> <th>男(人)</th> <th>女(人)</th> </tr> </thead> <tbody> <tr> <td>東京都</td> <td>千代田区</td> <td>30</td> <td>29</td> <td>1</td> </tr> <tr> <td>東京都</td> <td>中央区</td> <td>31</td> <td>17</td> <td>14</td> </tr> <tr> <td>東京都</td> <td>港区</td> <td>23</td> <td>13</td> <td>10</td> </tr> <tr> <td>東京都</td> <td>新宿区</td> <td>25</td> <td>10</td> <td>15</td> </tr> </tbody> </table> <div style="border: 1px solid gray; padding: 2px;"> 市区町村別 施設別 年別 </div>	都道府県	市区町村	参加者(人)	男(人)	女(人)	東京都	千代田区	30	29	1	東京都	中央区	31	17	14	東京都	港区	23	13	10	東京都	新宿区	25	10	15
都道府県	市町村	参加者(人)	男(人)	女(人)																																																																													
東京都	千代田区	30	29	1																																																																													
東京都	中央区	31	17	14																																																																													
東京都	港区	23	13	10																																																																													
東京都	新宿区	25	10	15																																																																													
月	施設A(人)	施設B(人)	施設C(人)	西暦(暦年)	参加者(人)																																																																												
1月	56	55	14	2018年	2340																																																																												
2月	45	23	23	2019年	3129																																																																												
3月	34	11	45	2020年	432																																																																												
4月	67	113	36	2021年	1893																																																																												
都道府県	市区町村	参加者(人)	男(人)	女(人)																																																																													
東京都	千代田区	30	29	1																																																																													
東京都	中央区	31	17	14																																																																													
東京都	港区	23	13	10																																																																													
東京都	新宿区	25	10	15																																																																													

修正前は1つのシートに複数の表がまとめて掲載されてしまっているため、実際に機械判読により処理を行おうとすると、正確な計算や並べ替えなどができない場合があります。

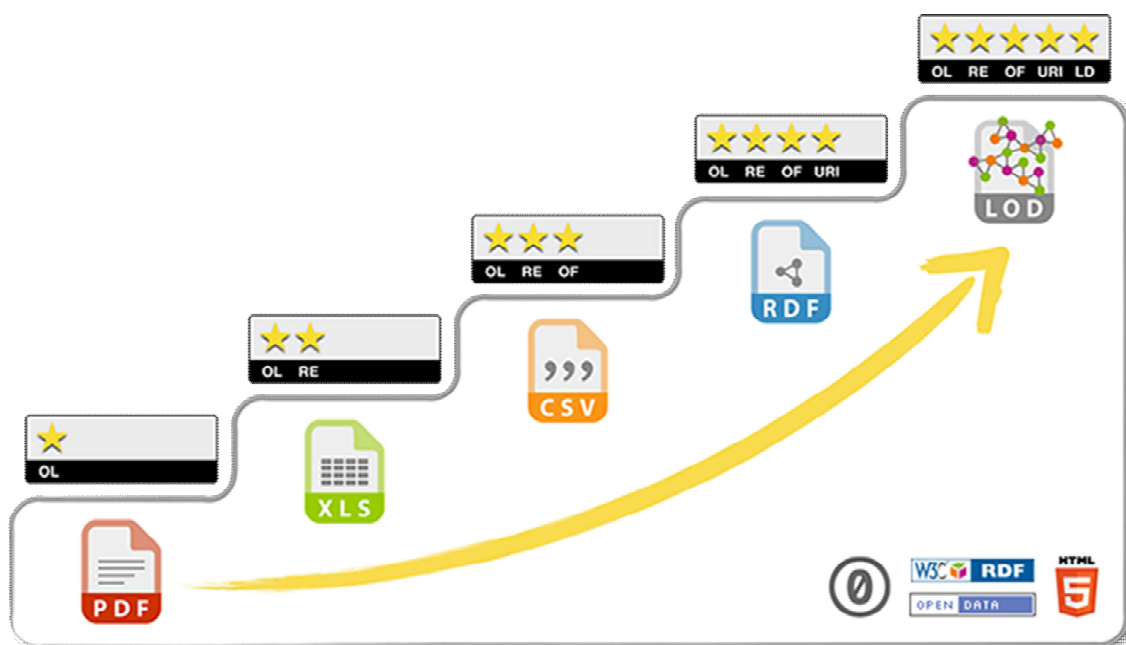
※CSV化する場合には各シートを更に分割し1ファイルにする必要があります。

第3節 機械判読性の高い CSV データの作成について

Web の父であるティム・バーナーズ＝リーは、オープンデータの形式を以下の 5 つのレベルに分けています。

- ・★は、オープンライセンスでデータを公開されていることを表しています。ファイル形式などは問われません。
- ・★★は、データが構造化されていることを表しています。Excel などが典型的な例になります。
- ・★★★は、特定のソフトに縛られることなく誰でも利用できる形式であることを表しています。例えば Excel のデータを CSV に変換して公開すると、星3つのレベルになります。
- ・★★★★は、データを一意に識別できるように識別子として URI (Unified Resource Indicator) を使用していることを表します。データを表現する方法としては、RDF (Resource Description Framework) が用いられます。
- ・★★★★★は、データとデータを結合したリンクト・オープンデータとなっていることを表しています。データが相互にリンクされることで、データのウェブができあがります。

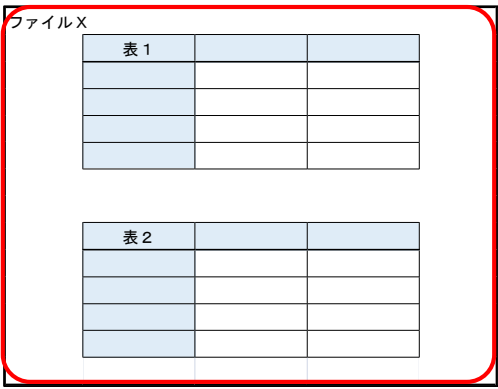

自治体がオープンデータを公開する際の一つの目安としては、★★★★の CSV のレベルを目標にするのが良いとされています。



(「5★オープンデータ」HP より)

□チェック項目3-1：1つのファイルは、1種類の表から構成されているか

例

修正前	修正後
<p data-bbox="304 465 724 497">1つのファイルに複数の表がある</p>  <p>The diagram shows a single file labeled 'ファイルX' containing two tables. The first table, '表1', has 4 rows and 3 columns. The second table, '表2', also has 4 rows and 3 columns. Both tables are highlighted with a red border.</p>	<p data-bbox="940 465 1225 497">ファイルを2つに分割</p>  <p>The diagram shows two separate files. 'ファイルX' contains '表1' (4 rows, 3 columns) and 'ファイルY' contains '表2' (4 rows, 3 columns). Both files and their respective tables are highlighted with a red border.</p>

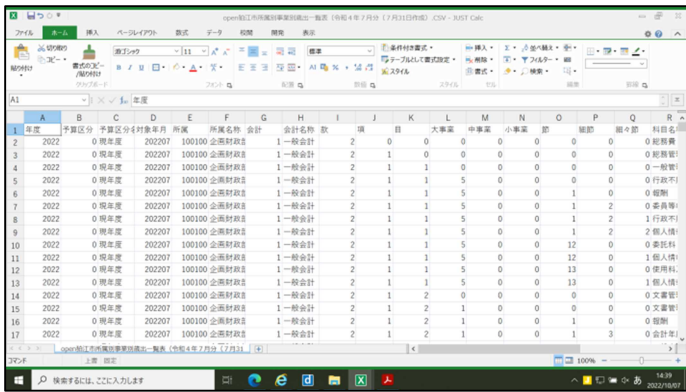
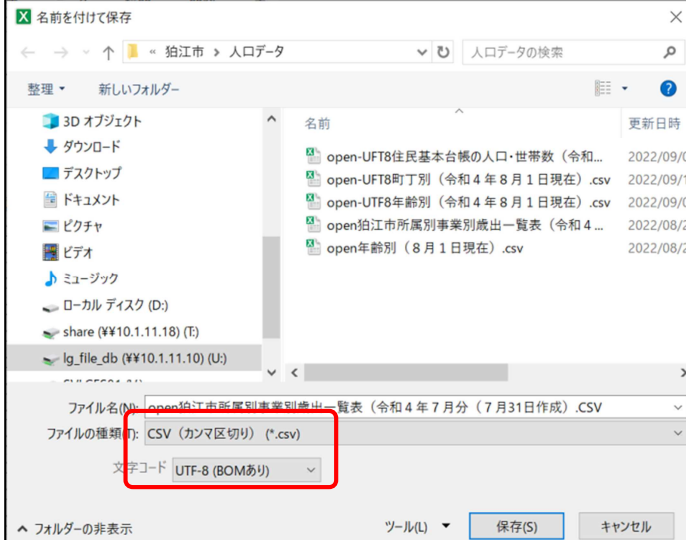
内容的に前述のチェック項目2-2と似ていますが、違う考え方になります。

オープンデータを Excel ファイル形式で公開する場合はファイル内に複数シートある形での公開も可能ですが、CSV 化して公開する場合はファイルにある先頭のシートの情報しか持たなくなるため、オープンデータ化したい表ごとにファイルを分割する必要があります。

□チェック項目3-2：国際的に広く利用されている文字コードを使用し、それを明記しているか

日本語を記述する文字コードには、JIS (ISO-2022-JP)、Shift-JIS、EUC、UTF-8 など複数の文字コードが存在します。このため、記述されている文字コードが明記されていなければ、機械判読が困難となります。

なお、データの国際的な展開や他の規格との整合性を考慮すると UTF-8 を利用することが望ましいとされています。

JUST Calc での CSV (UFT-8) 化	手順
	<p>①CSV 化したい Calc を開いて「ファイル」タブをクリック</p> <p>②ファイルの「名前を付けて保存」を選択します。</p>
	<p>③名前を付けて保存のダイアログが出るので、「ファイルの種類」を「CSV (カンマ区切り) (*.csv)」に変更、「文字コード」を「UTF-8 (BOMあり)」に変更</p> <p>④保存ボタンをクリック</p>


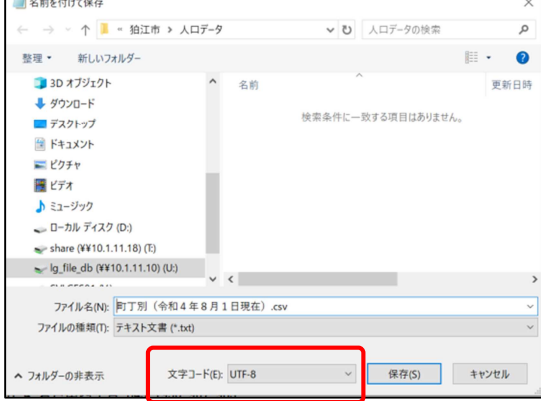
※Excel2016 の場合は、「名前を付けて保存」まで同じように操作しますが、保存時の拡張子の選択用プルダウンメニューで「CSV UTF-8 (コンマ区切り) (*.csv)」を選択して保存します。

※UTF-8が選択できない場合、メニューにない場合

一旦、保存形式を「CSV (コンマ区切り) (*csv)」に選択して保存します。

「ブックの一部の機能が失われる可能性があります」という警告が出た場合は、「はい」を選択してください。

その後、メモ帳の機能を使用してUTF-8へ文字コードを変換します。

メモ帳を使用した UTF-8 への変換	手順
	<p>①文字コードを変換したいファイルを開いて、「ファイル」タブをクリック</p> <p>②ファイルの「名前を付けて保存」を選択します。</p>
	<p>③名前を付けて保存のダイアログが出るので、「ファイルの種類」はそのまま、「文字コード」を「UTF-8」に変更</p> <p>④保存ボタンをクリック</p>

Excel 形式の元ファイルをオープンデータ化する

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
所属別事業別歳出一覧表																
1 2022 0 現年度 '202207 '100100 企画財政部 '1 一般会計 '02 '00 '00 '00 '00 '00 '00 '00 '00 '00																
3 (令和 4 年 7 月分)																
4 令和 4 年 7 月 31 日 現在																
5																
年度	予算区分	予算区分名	対象年月	所属	所属名称	会計	会計名称	款	項	目	大	中	小	節	細	科目
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'00	'00	00	00	00	00	00	00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'00	00	00	00	00	00	00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'00	'00	'00	'00	'00	'00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'00	'00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'01	'00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'01	'02
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'01	'02
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'01	'02
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'12	'00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'12	'00
2022	0	現年度	'202207	'100100	企画財政部	'1	一般会計	'02	'01	'01	'05	'00	'00	'00	'13	'00

処理後の CSV ファイルを Excel で表示

A	B	C	D	E	F	G	H	I
年度	予算区分	予算区分名	対象年月	所属 (組織番号)	所属名称	会計	会計名称	款
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02
2022	0	現年度	202207	100100	企画財政部	'1	一般会計	'02

同ファイルをメモ帳で表示

```
年度、予算区分、予算区分名称、対象年月、所属 (組織番号)、所属名称
(円)、計 (円)、本月分 (円)、累計 (円)、予算残額 (円)、執行率
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.0.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
2022.0.現年度.202207.100100.企画財政部・政策室.1.一般会計.2.1.
```


(参考資料)

- 「オープンデータをはじめよう～ 地方公共団体のための最初の手引書 ～」
内閣官房 情報通信技術 (IT) 総合戦略室 作成資料
- 新潟市オープンデータ CSV ファイル変換・作成マニュアル
新潟市 作成資料
- オープンデータ Q&A
会津若松市オープンデータ推進検討チーム 作成資料
- 統計表における機械判読可能なデータ作成に関する表記方法
総務省統計局 作成資料
- オープンデータに関する Q&A
一般社団法人デジタル地方創生機構ホームページ