

統計データの整備に係る基本方針

令和 3 年 7 月 9 日
統計企画会議申合せ
統計調査等業務最適化推進協議会決定

1 目的

「公的統計の整備に関する基本的な計画」（令和 2 年 6 月 2 日閣議決定。以下「第Ⅲ期基本計画」という。）及び「官民データ活用推進基本計画」（令和元年 6 月 14 日閣議決定）では、証拠に基づく政策立案（Evidence-Based Policy Making。以下「EBPM」という。）に向けた取組の一環として、統計データを利活用可能な形で提供するとともに、利活用基盤の整備・強化を図ることが重要であるとしている。

本方針は、政府統計共同利用システムで提供する統計データを、利活用可能な形で作成及び登録するに当たって、各種政策決定等で掲げる諸要件を整理するとともに、その整備に必要な考え方や実施方法を統一的な観点で明確化及び具体化し、各府省が確実かつ円滑に取り組むことにより、EBPM の推進を始めとする取組の達成に寄与することを目的とする。

2 背景

現在、コンピュータ・アプリケーション（以下単に「コンピュータ」という。）の能力は、情報通信技術の進展に合わせて向上・高度化しており、インターネットを経由したデータの自動取得や、大量かつ多様なデータを高速に処理することが可能になった。

一方、これまでの報告書を基本とした統計表の作成及び提供は、画面上の表示又は印刷して人間が読むことを念頭に置いた形で行われていることから、利用者が統計データとして再利用する際に、人手でデータを再入力する必要があることや、行政機関ごとに異なるコード体系で提供されているため、情報の収集や整理などの前処理に多くの時間が必要になるという課題が生じている。

これに対して、コンピュータによる機械判読性と検索性を確保したデータでは、データの収集や各種コードによるデータの横断的な利用の自動化や、複数の行政機関や民間のデータを組み合わせて処理・利用することが容易となる。このため、EBPM、政策評価、学術研究、産業創造等に積極的に寄与することが可能となることから、人間の視認性だけでなく、機械判読や検索性も考慮した情報提供が求められるようになってきている。

このような状況から、現在、様々な政策決定等において統計データの利活用可能な形での提供及びこれに関連する利活用基盤の整備・強化が求められており、各府省は、所管する統計について取り組むこととなっている。

3 基本理念

本方針では、上記背景で挙げた課題に対応すべく、「公的統計の整備に関する基本的な計画」（第Ⅱ期基本計画。平成26年3月25日閣議決定）の基本的な視点及び「第Ⅲ期基本計画」の取組として示された内容、「統計改革推進会議最終取りまとめ」（平成29年5月19日統計改革推進会議決定）並びに「オープンデータ基本指針」（平成29年5月30日高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定）等で掲げられている以下について取り組む。

- ① 統計データのオープン化の推進
- ② ユーザーによる加工（前処理）コストの引下げ
- ③ データ検索の利便性向上
- ④ 統計の品質向上

4 用語の定義・解説

（1）統計データ

本方針において、「統計データ」とは、集計・統計処理された情報及びそれらに関連する付帯情報（以下「メタデータ」という。「（4）メタデータ」を参照）をいう。

（2）利活用可能な形

本方針において、「利活用可能な形」とは、機械判読が可能な（機械判読に適した）状態をいう。

（3）機械判読可能な統計データ

本方針において、「機械判読可能な統計データ」とは、汎用的なコンピュータが一定の条件の基に自動的に解読し、再利用（加工、編集、引用、転載、複製、改変等）できる統計データのことであり、以下に掲げる要件の全てを満たすものをいう。

機械判読可能な統計データの要件

- | |
|--|
| <ol style="list-style-type: none">① コンピュータが、値を機械的に判読できること② コンピュータが、メタデータを機械的に判読できること③ コンピュータが、値とメタデータの関連性（構造）を機械的に判読できること④ 特定のアプリケーションに依存しないデータ形式で入手できること |
|--|

各府省は、機械判読可能な統計データとして、所管する統計の種類に応じてxlsx、csv、統計データベースのいずれかの形式で作成する。

（4）メタデータ

本方針において、「メタデータ」とは、値そのものではなく、値のデータに付随し、値のデータに意味付けをする情報、値のデータ自身について解説する情報をいう。データの利用や解釈、データの効率的な管理や検索のための情報となることから、機械的に

判読可能とするほか、適切に整備・管理することが重要となる。

(5) 統計情報データベース

本方針において、「統計情報データベース」とは、統計データを構造化し、機械判読できるよう蓄積したものであり、時間軸、地域などのメタデータでの検索やデータの抽出のほか、グラフの表示、表レイアウトの変更、データのダウンロードが可能なものをいう。また、API (Application Programming Interface) 機能を用いたプログラムによるデータの自動取得を可能としているなど、機械判読可能な統計データに最も適した形式となる。

5 基本原則

各府省は、基本理念で掲げた取組に向けて、本方針及び「8 実施方法等」に掲げる各ガイドライン類に基づき、以下に掲げる事項を実施する。

- ① 機械判読可能な統計データを作成し、提供する。
なお、これを満たす場合、別途人間が読むことを念頭に置いた統計データを作成・提供することを妨げない。
- ② メタデータを整備し、管理する。
- ③ 継続的な取組であることを勘案し、各府省におけるリソースを考慮した上で実施体制を構築する。
- ④ 所管する統計の周期や特性、システム改修等を考慮し、計画的に取り組む。

6 対象統計・適用時期

(1) 機械判読可能な統計データの作成、提供の対象となる統計

機械判読可能な統計データの作成、提供は、政府統計共同利用システムを利用して提供する全ての統計データを対象とする（一つの統計表が、スプレッドシートとして複数のファイル・シートに分かれている場合、全てのファイル・シートが対象となる。）。

(2) データ形式と対象統計との関係

各府省は、機械判読可能な統計データの作成、提供及びメタデータの整備、管理に当たって、所管する統計の種類に応じて下表のとおり対応する。

データ形式と対象統計の関係

データ形式	基幹統計及び特定一般統計	特定一般統計を除く一般統計、加工統計及び業務統計
(※1、2) { xlsx によるスプレッド シート csv	DB として作成するため任意 (時間を要する等やむを得ず 整備していない場合は必須)	必須 (DB を整備済みであれば任意)
(※3) { 統計情報データベース (DB)	必須	可能な限り対応 (※4)
メタデータ		

※1 機械判読が可能となるよう整備したもの。

※2 昨今の高度情報通信ネットワーク社会推進戦略本部によるオープンデータに係る検討を踏まえ、csv 以上のフォーマットで掲載することを目指す（本方針においては、csv に変換しても適切にデータを表示・利用できるよう整備したスプレッドシートを含む。）。

※3 メタデータは、統計情報データベースを作成する際に必要となる。

※4 ICT の進展等を踏まえ継続して対応を検討するほか、既に DB 化済みの統計は継続して対応する。

(3) 適用時期

基本原則を始めとする本方針の取組は、本方針及び「8 実施方法等」で掲げるガイドライン類の施行日以降に作成・提供する統計データに適用する。

なお、過去に提供した統計データへの適用を妨げるものではない。

7 計画的な取組に当たっての留意点

機械判読可能な統計データの作成・提供及びメタデータの整備・管理を計画的に取り組むに当たって、以下について留意する。

- ① 関係する政策決定、各種ガイドライン、マニュアルで掲げる取組と整合性を図りつつ実施する。
- ② 大規模なシステム・プログラムの改修が必要になる場合は、システム更改のライフサイクルに応じる等費用対効果を考慮して実施する。
- ③ 対象となる統計が多量である等全ての統計に対応するために時間を要する場合は、総括表等特にニーズが高いものから優先的に着手する。
- ④ 機械判読可能な統計データと、印刷して人間が読むことを念頭に置いた形式とを併せて作成する場合は、統計作成プロセスを見直し効率化を図るなど第Ⅲ期基本計画で掲げる業務見直し（BPR）と併せて検討する。

8 実施方法等

機械判読可能な統計データの作成・提供及びメタデータの整備・管理を実施する際に必要な考え方や、具体的な作業内容を明確化及び具体化した方針・ガイドラインを以下のとおり定める。

各府省は、当該方針・ガイドラインを基に機械判読可能な統計データの作成・提供及びメタデータの整備・管理に取り組む。

(1) 「統計表における機械判読可能な統計データ作成に関する表記方法」(令和2年12月18日統計企画会議申合せ)

政府統計共同利用システムに統計データをスプレッドシート(及び csv)で掲載する場合の表記方法について掲載する。

(2) 「メタデータの整備に関するガイドライン(仮称)」

国際標準等を踏まえた統一的な観点を基に、メタデータを整備するに当たっての具体的な設計の観点、手法、留意点、メタデータを管理するためのシステムの利用について掲載する。

(3) 「統計情報データベースの整備に関するガイドライン(仮称)」

統計情報データベースを整備するに当たっての具体的な設計の観点、手法、留意点を掲載する。

(4) その他

上記以外に係る留意点等を掲載する。

9 その他

(1) フォローアップ

各府省は、本方針で掲げる内容に計画的に取り組むところ、着実な実施を図るべく、進捗状況のフォローアップを実施する。

(2) 方針の見直し

本方針は、本取組の対応状況、統計関連法制やオープンデータの見直しの動向、国際的な取組の動向等を踏まえ、随時見直しを行う。

附 則

本方針は、令和3年7月9日から施行する。