

取りまとめに向けた論点について

令和3年11月17日

総務省統計改革実行推進室

取りまとめに向けた検討対象の範囲

公的統計におけるビッグデータ活用の現状と今後

- ビッグデータには多種多様なデータが含まれているところ、現在、公的統計においては、「Web掲載データ」、「人工衛星画像データ」、「POSデータ」（販売時点情報）、「携帯基地局情報」などのビッグデータの活用、又は活用の検討が進んでいる。
- コロナ禍においては、民間企業等から提供されたデータを基に国が公表したデータと実態に乖離があるとの指摘を受けたこと、データ提供を受けるに当たり官民連携協定の締結に時間を要したことなど、民間データの活用に係る様々な課題があることが判明した。
- 今後の公的統計におけるビッグデータ活用の裾野の拡大、更なる推進に当たっては、民間企業等が保有するデータの提供を官側が個別に受ける場合について、官民双方における課題を整理・検討することが有効と考えられる。

取りまとめに向けた検討対象の範囲

- 今回の取りまとめに向けては、上記の「民間企業等が保有するデータの提供を官側が個別に受ける場合」を念頭に、公的統計におけるビッグデータの活用の更なる推進に当たっての論点について、検討を深めることとしたい。
- ※ 将来的には、分野別など、データの特성에応じた課題の設定・検討等も考えられる。

1 ビッグデータが有する偏り、特徴・特性

現状・課題

- 一般的に、ビッグデータはデータの作成方法や作成プロセス等に起因する偏りや特徴・特性を有している。そのため、データの活用にあたっては、それぞれのビッグデータの偏り等を十分に把握した上で検討を進める必要がある。
- 一方で、ビッグデータの偏り等に関する情報の開示は必ずしも十分ではないとの指摘がある。



取りまとめに向けた論点

- ビッグデータを活用するにあたって官側が確認すべき事項（データホルダーに提供いただくことが望ましい事項）にはどのような項目があるか。
（例：データの項目・レイアウト、データ形式、データ定義、データソース、データ作成方法（プロセス）、提供データの範囲・精度や偏り、秘匿処理の有無・方法、利活用実績、データのカバレッジ等の変化や提供が中断する可能性及びその場合の取扱い、サンプルデータの有無、その他特記・留意事項 など）
- 上記の項目をどのように確認すべきか。
（確認した情報の公表の可否、確認の具体的な方法 など）

2 安定的・継続的な民間データの提供

現状・課題

- ビッグデータを公的統計に活用していくためには、安定的かつ同質なデータを継続的に得る必要がある。
- 一般的に、ビッグデータには、データの生成・提供が中断する、カバレッジ等の変化により同質なデータが継続的に生成されなくなる、などのリスクがある。



取りまとめに向けた論点

- データを安定的かつ継続的に得るためにはどのような取組が必要か。
- データ提供に係る協定書や契約書等において、担保すべき事項は何か。
(例：データの提供期間、提供を中断する場合のルール、カバレッジ等の変化が生じる場合の取扱いのルール（従来の方法と並行してデータを提供するなど） など)
- データを安定的かつ継続的に得るために、データ提供者へのインセンティブとして何が考えられるか。(例：データ提供の対価の支払い、研究レポートの発行 など)
- 最終的にデータの提供が中断した場合にどのように備えておくべきか。

3 人材育成と体制、システム・技術等の整備

現状・課題

- ビッグデータを公的統計に活用していくためには、データを扱うスキルや知識を有する人材を育成・確保するなど、ビッグデータを扱うための体制を整備し、組織的にノウハウの蓄積を行う必要がある。
- また、ビッグデータを継続して活用していくためには、取り扱うデータの種類や量、集計方法等に応じたシステムを整備する必要がある。



取りまとめに向けた論点

- ビッグデータを適切に取り扱うためには、どのような人材・スキルが必要か。
(例：データ分析スキル、システム関連スキル、民間企業等との交渉・調整スキル など)
- 必要な人材を育成・確保するためにはどのような取組が必要か。
(例：研修の実施・受講、官民の人材交流、研究機関への出向 など)