

非識別加工情報の作成におけるサンプリングの導入
(佐藤構成員提出資料)

地方公共団体の個人情報保護条例に、非識別加工情報の導入が議論されている。その前提として、行政機関・独立行政法人等の個人情報保護法改正で導入される非識別加工情報と加工基準を同様にするのが前提になっている。基本としてはそれでよいが、一方で地方公共団体が保有する個人情報の特性を考えると同様にすると、対象個人が明確で、さらに悉皆性の高いデータに関してはサンプリングによる加工も行う必要がある。

さて「地方公共団体が保有するパーソナルデータに関する検討会報告書(案)」(24 ページ、2017年3月29日、事務局提出)において

これらの点について、検討会では、民間部門、国及び地方公共団体で非識別加工情報の加工の基準を同等の内容としつつも、1 地方公共団体ではその地域について網羅的に収集している個人情報が多いため、そのような個人情報についてより安全に加工するためには、サンプリングを行うことを基本とすべきではないか、2 市町村では個人情報の住所は同一市町村である場合が多く、事業者の活用方法、個人情報ファイルの性質等を勘案して、住所の加工方法を(県単位や市町村単位へ置き換えることに限らず)判断するなど、地方公共団体の保有する個人情報の特質を考慮する必要があるのでないかとの指摘があった。

のように指摘されている通り、地方公共団体が保有する個人情報の特徴を考えると、単純に行政機関・独立行政法人等個人情報保護法改正(以下、改正行個法)で導入される非識別加工情報や、改正個人情報保護法(以下、改正個情法)で導入される匿名加工情報の加工基準と単純に同じにすると、個人識別性などの問題が生じる可能性がある。地方公共団体が保有する個人情報の特徴は次の二つがあげられるだろう。

(1) 個人情報ファイルなどの個人情報を含むデータに対して、そこに含まれる個人情報の範囲が明確化されている、つまり対象地域や対象年齢などが明記されていることが多い。

(2) 所定条件に合致する住民全員、ある地域の住民大多数に関する個人情報などの網羅性が高い、つまり悉皆性の高いことが多い。

ここで改正個情法の対象となる民間事業者の場合、保有する個人情報は必ずしも公開されておらず、さらに網羅的に多数の個人情報を保有する可能性も低く、その意味では(1)と(2)ともに必ずしも当てはまらない。行政機関・独立行政法人の個人情報ファイルに関しては(1)と(2)の両方が当てはまるが、特に(2)に関しては、現状、個々の行政機関・独立行政法人が保有する個人情報ファイルなどは地方公共団体と比べると悉皆性は

幾分低く、改正行個法においても悉皆性の高いデータに関する対策は必要だが、その必要性は地方公共団体が保有する個人情報ファイルなどはさらに重要となる。(なお、その意味では、そもそも改正行個法の非識別加工情報の加工基準においても、悉皆性の高いデータに関してはサンプリングを導入する必要があるともいえる)

ここで悉皆性の高いデータのリスクを検討してみる。例として地方公共団体が保有、所定地域の住民大多数の検診データを想定してみる。そして地方公共団体が、その検診データを非識別加工情報に加工・提供したとする。さて、このときその非識別加工情報を何らかの方法で手に入れた第三者が、その地域の住民一人に関する断片的な情報を持っていたとする。そしてその非識別加工情報と断片情報の突合せにより、その非識別加工情報の中に、その断片的な情報と類似性が高いのは一人であることがわかった場合、その非識別加工情報に含まれる個人と、先述のその断片的な個人情報に関する個人は同一人物であることがわかる。そして、その第三者はその個人に関する非識別加工情報を利用する、その個人に関する、より詳細な個人情報を知り得ることになり、結果として個人識別リスクは高くなってしまう。

さてこのような悉皆性の高いデータの個人識別リスクは、例えば統計情報等の作成においてよく知られており、その統計情報では、同様のリスクを低減するためにサンプリング、一部の個人に関する情報を削除(レコード削除)することが広く利用されている。

同様に非識別加工情報の加工でもサンプリングを含めることを基礎とすべきである。そのサンプリングを導入すれば、仮にある個人に関する断片的な情報と非識別加工情報の突合せにより、非識別加工情報の中に類似した情報の個人がいたとしても、同様に類似した情報の個人は、サンプリングにより削除された情報に含まれていた可能性がある。その個人と断片的な情報の個人が同一とは断定できなくなり、結果として個人識別リスクを低減できる。

特に非識別加工情報は、提案に基づいて非識別加工情報の対象などを決まるが、仮に提案対象となる個人情報ファイルに悉皆性があっても、その提案においてその全件を非識別加工情報として提供を求められるケースはないといえない。そのとき地方公共団体が、個人識別リスクを下げる目的においては、全件データの加工・提供を避けるためにサンプリングを加工基準に入れることは不可欠だろう。

なお、サンプリングを行うことは利活用に大きな影響を与えず、むしろ非識別加工情報の作成を容易化するなど利活用の面でも有益である。その理由であるが、個人識別リスクを低減する目的からは、サンプリングにおける抽出率は高くてもよい(抽出率は半分以上、つまり半分以上を残したとして、一定の効果は期待できる)、つまり削除する情報は少なくできるので、非識別加工情報の有用性を下げることは少ない。さらにサン

プリングを導入することは個人情報非識別加工情報への加工を難しくすることもなく、むしろ単純化する効果が期待できる。その理由であるが、ここでは改正行個法のための非識別加工情報に関するガイドラインは、改正個情法の匿名加工情報に関するガイドライン（個人情報の保護に関する法律についてのガイドライン(匿名加工情報編)(平成28年個人情報保護委員会告示第9号))をベースにしているため、後者で説明すると規則19条5号の基準である、「前各号に掲げる措置のほか、個人情報に含まれる記述等と当該個人情報を含む個人情報データベース等を構成する他の個人情報に含まれる記述等との差異その他の当該個人情報データベース等の性質を勘案し、その結果を踏まえて適切な措置を講ずること。」においてその差異の基準を設定することも、その差異をデータから見つけることも技術的な難易度が高く、地方公共団体にとっての負担は非常に大きいのが現実である。しかし、規則19条の加工基準にサンプリングを加えることにより、仮に規則19条5号で対処が求められる差異が非識別加工情報に残ってしまった場合でも、サンプリングによりその差異と合致する特質をもつ個人すべてが、非識別加工情報に含まれるとは限らないことから(サンプリングにより削除された対象に含まれる可能性がある)、残ってしまった差異から個人を識別する可能性は低くなる。

悉皆性の高いデータに対してサンプリングを行うことは、個人識別のリスクを下げるとともに、非識別加工情報の利活用に与える影響は軽微であるといえる。さらに加工基準で求められる、差異の発見やそれに応じた加工を単純化することもできる。特に非識別加工情報作成に関する経験や技術が十分とは言えない、一部の地方公共団体においては非識別加工情報の作成におけるサンプリングの導入は、不可欠となるはずである。