토론문 : 익명화, 비식별화, 개인정보에 관하여

전응준(유미 법무법인 변호사)

1. 용어의 정의

1.1 익명화

○ Anonymisation
(6) “Rendering anonymous” shall mean the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense and effort(BDSG 3 (6))

○ Data are anonymised if all identifying elements have been eliminated from a set of personal data. No element may be left in the information which could, by exercising reasonable effort, serve to re-identify the person(s) concerned. Where data have been successfully anonymised, they are no longer personal data.(데이터보호법 핸드북 44면)

○ EU 95년 Directive는 본문 조항에서 익명화에 관한 정의를 두고 있지는 않으나 서문에서 익명화 정보가 데이터 규제의 대상이 되지 않음을 기술함

1.2 가명화

○ GDPR에서 가명화의 정의와 일정 영역의 법적 규율을 시도하고 있음

○ 'pseudonymisation' means the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person(GDPR 4 (3a))

○ “Aliasing(pseudonymization)” shall mean replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult. (BDSG 3 (6a))

○ Personal information contains identifiers, such as a name, date of birth, sex and address. When personal information is pseudonymised, the identifiers are replaced by one pseudonym. Pseudonymisation is achieved, for instance, by encryption of the identifiers in personal data(데이터보호 핸드북 45면)

Example: The sentence “Charles Spencer, born 3 April 1967, is the father of a family of four children, two boys and two girls” can, for instance, be pseudonymised as follows:
“C.S. 1967 is the father of a family of four children, two boys and two girls”; or
“324 is the father of a family of four children, two boys and two girls”; or
“YESz320l is the father of a family of four children, two boys and two girls”.

○ 방통위 가이드라인이 EU에서 Pseudonymous data를 익명화로 보고 마치 개인정보보호 관계 규정을 전부 면제하는 것으로 이해하는 것은 오류에 가까움.

1.3 비식별화

○ 미국 Department of Education
- Anonymization [of data] refers to the process of data de-identification which produces de-identified data, where individual records cannot be linked back to an original student record system or to other individual records from the same source, because they do not include a record code needed to link the records.
- De-identification [of data] refers to the process of removing or obscuring any personally identifiable information from student records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them.
- While it may not be possible to remove the disclosure risk completely, de-identification is considered successful when there is no reasonable basis to believe that the remaining information in the records can be used to identify an individual
출처 : http://ptac.ed.gov/sites/default/files/data_deidentification_terms.pdf

○ 주로 미국에서 비식별화(de-identification)이라는 용어를 사용하고 익명화(anonymization)와 일부 구분되는 것으로도 보이나, 의식적으로 명확한 용어정의를 하고 있다고 보이지는 아니함.

○ 개인정보보호법 제18조 제2항 제4호

② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.
4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

- ‘특정 개인을 알아볼 수 없는 형태’이면 개인정보성을 상실한다고 볼 수 있으므로 위 제4호를 주의적 규정으로 이해할 수도 있음. 그러나 제18조의 전체적인 지위를 볼 때 제4호는 제3자 제공의 예외를 둔 열거적 규정으로 보는 것이 자연스럽기 때문에 ‘특정 개인을 알아볼 수 없는 형태’는 가명처리, 총계처리, 데이터 마스킹, 범주화, 데이터 일부 값 삭제등의 비식별화 처리라고 이해하는 것이 타당(사견).

2. 비식별화 정보의 개인정보 해당성

2.1 EU 데이터 보호법 핸드북 2014

○ Data are personal data if they relate to an identified or at least identifiable person, the data subject.

○ A person is identifiable if additional information can be obtained without unreasonable effort, allowing the identification of the data subject.

○ Data are anonymised if they no longer contain any identifiers; they are pseudonymised if the identifiers are encrypted.

○ In contrast to anonymised data, pseudonymised data are personal data.

2.2 GDPR

○ EU 95년 Directive는 pseudonymization에 대한 규율이 없었음.
- 독일 BDSG등 회원국 개별법에서는 일부 규정이 있음

○ GDPR(안)은 pseudonymized data에 대한 일부 규정을 마련하고 있음
- 규정안이 계속 수정되면서 일부 문언 변경이 있음

○ 2014년 GDPR(안)
- Article 10
1. If the data processed by a controller do not permit the controller or processor to directly or indirectly identify a natural person, or consist only of pseudonymous data, the controller shall not process or acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation.
2. Where the data controller is unable to comply with a provision of this Regulation because of paragraph 1, the controller shall not be obliged to comply with that particular provision of this Regulation. Where as a consequence the data controller is unable to comply with a request of the data subject, it shall inform the data subject accordingly.

○ 2015. 12. GDPR안
- Article 10 Processing not requiring identification
1. If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.
2. Where, in such cases the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 18 do not apply except where the data subject, for the purpose of exercising his or her rights under these articles, provides additional information enabling his or her identification.

- 제23조 : pseudonymisation을 appropriate technical and organisational measure로 인식함(data protection by design and by default)
- 제30조 : 데이터 처리의 보호조치로서 pseudonymasation 예시
- 제32조 : 암호화 등 데이터를 인식불가능하게 하였다면 데이터 유출(data breach)시 정보주체에게 통지 면제
- 제38조 : code of conduct로서 pseudonymisation 강조

2.3 HIPAA(Health Insurance Portability and Accountability Act)

○ Since 1996 the US Federal Health Insurance Portability and Accountability Act of 1996 (HIPAA) differentiate between:
- Protected Health Information (Patient name + medical records = sensitive personal data)
- Limited Data Set (a form of pseudonymized data still covered by HIPAA)
- De-identified Data (a form of anonymized data not any longer covered by HIPAA)
- 다만 비식별화 정보, LDS의 기준에 관하여 명확한 통계적, 실증적 근거가 있다고 보이지는 아니함

○ 의료정보관기관(covered entity)가 취할 수 있는 비식별화 방법
- 아래의 2가지 경우에 해당하면 HIPPA 프라이버시 규율에서 벗어남
- 관련 전문가의 개별적인 판단(very small risk) : 통계학적, 과학적 원칙을 적용
- 식별자들 18개를 데이터에서 제거하는 방법(safe harbor)

○ Limited Data Set(LDS)
- HIPAA 프라이버시 규칙에서 부분적인 규율면제
- 제한적인 요건 하에서 정보주체의 동의, 허가 없이 이용하고 제공할 수 있음
- 18개 식별자 중 직접적 식별자에 해당하는 16가지 식별자를 제거하거나(나머지 간접적 식

별자는 생년월일, 치료나 처방일자, 일부 지리적 정보라고 함)
 - 데이터를 제공받는 자와 데이터이용계약(data use agreement)를 체결하면 LDS 인정

3. 우리 법체계상의 익명 정보, 비식별화 정보

○ 개인정보의 범위를 해석할 때 결합가능한 다른 정보의 보유가능성을 합리적인 범위내에서 제한 해석할 필요가 있다고 생각됨
 - EU 데이터 보호규정도 같은 입장에 서 있다고 보임
 - 다른 정보의 획득가능성을 합리적인 범위내에서 제한하지 않고 데이터의 속성에 따른 결합가능성의 용이성만을 따지는 것은 타당하지 아니함.

○ 원본 데이터로 회복불가능한 익명화 정보는 개인정보법의 대상이 아님. 문제는 이론상으로 보면 원본 데이터로 연결이 불가능한 익명화 정보는 없다는 점에 있음.
 - 독일 BDSG, EU 핸드북과 같이, 합리적 수단, 비용, 시간 범위내에서 원본 데이터로 복원할 수 없는 익명화 정보는 더 이상 개인정보로 보지 않는 해석론이 우리 법 해석에서도 가능하다고 생각됨(사견)

○ 일본 법과 같은 '익명가공정보', '특정성 저감 데이터' 등의 개념 정의 도입이 바람직하겠지만, 단기간내에 불가능하다면 법 해석론으로 개인정보보호법의 규율을 받지 않는 익명화 정보를 정의할 수 있지 않을까 함.
 - 일본 법은 익명가공정보에 대하여 제3자 제공시 정보주체의 동의를 받지 않아도 되도록 설계하였는데, 제3자 제공은 전면적인 데이터 유통을 가능하게 하므로 초기 단계에서는 수집 및 이용과 제3자 제공의 경우를 분리하여 규율하는 것이 정책적으로 바람직할 수 있다는 생각을 해 봄.

○ 미국 FTC 동의의결 사례를 참조
 - 형사처벌을 규정하고 있는 우리 법률 체계상 사법조치보다는 행정처분이 유연한 대응을 가져올 수 있음.

○ 통계법에 의하여 수집된 정보에 대하여는 개인정보보호법이 적용되지 않으므로(개인정보보호법 제58조) 이를 기초로 공공영역에서 빅데이터 분석은 일정 부분은 가능하지 않을까 함.

4. 위험성의 상존

○ 우리나라의 약학정보원 사례
 - 비식별화 논리가 무분별하게 이루어지는 제3자 제공의 근거가 될 수 있음
 - 위 사건은 개인정보처리자가 취한 비식별화(암호화) 정도가 실제 현실에서 쉽게 복호화될 수 있었던 사안으로 추측됨

○ 이론적 논의 외에 비식별성 방법론(알고리듬)과 재식별성 정도에 관한 통계적, 실증적 연구가 필요함