개인정보 비식별화 방향

2017.07

www.cisp.or.kr
qna.cisp@gmail.com

정보화사회실천연합

목차

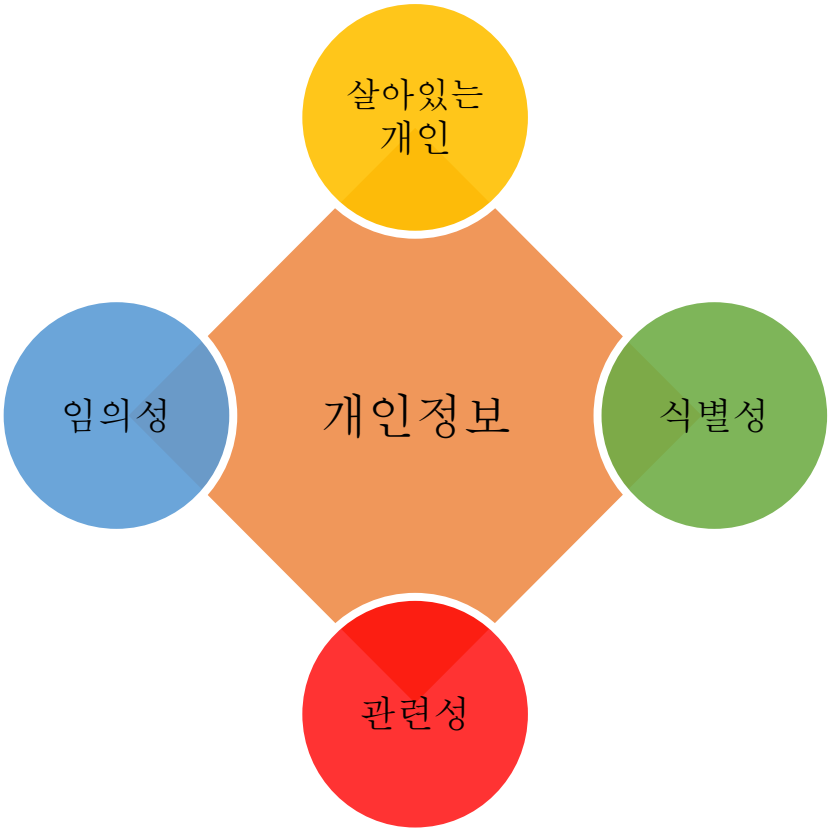
1. 개인정보

1. 개인정보

개인정보 보호법

「개인정보 보호법」 제2조 제1호

"개인정보"란 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다.



1. 개인정보

EU GDPR(일반 데이터 보호 규정), 2016

「General Data Protection Regulation」

Article 4. (1) 'personal data' means any information relating to an identified or identifiable natural person 'data subject'; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

개인정보는 식별된 또는 식별할 수 있는 개인(정보주체)에 관한 모든 정보를 의미한다. 식별할 수 있는 개인이란 특히 이름, 식별번호, 위치정보, 온라인 식별자와 같은 식별자(identifier) 또는 그 사람의 물리적・심리적・유전적・정신적・경제적・문화적・사회적인 정체성(identity)에 특정된 하나 이상의 요소를 참조하여 직접 또는 간접으로(directly or indirectly) 식별될 수 있는 자를 의미한다.

「EU 제29조 작업반 해설서」

III. ANALYSIS OF THE DEFINITION OF "PERSONAL DATA" ACCORDING TO THE DATA PROTECTION DIRECTIVE

3. THIRD ELEMENT: "IDENTIFIED OR IDENTIFIABLE"[NATURAL PERSON]

Means to identify – " to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person"

"개인이 식별가능한지 여부를 결정하기 위해서는, 개인을 식별하기 위해 개인정보처리자 또는 제3자에 의해 사용되는 합리적으로 가능하다고 생각되는 모든 수단을 고려하여야 한다."

1. 개인정보

일본 개인정보의 보호에 관한 법률, 2017

「개인정보의 보호에 관한 법률 제2조 제1항」

第２条 この法律において「個人情報」とは、生存する個人に関する情報であって、次の各号のいずれかに該当するものをいう。

一 当該情報に含まれる氏名、生年月日その他の記述等（文書、図画若しくは電磁的記録（電磁的方式（電子的方式、磁気的方式その他人の知覚によっては認識することができない方式をいう。次項第２号において同じ。）

생존하는 개인에 관한 정보로서, 당해 정보에 포함되어 있는 성명, 생년월일 그 밖의 기술(記述) 등에 의하여 특정의 개인을 식별할 수 있는 것(다른 정보와 용이하게 조합[照合][서로 맞추어 봄]할 수 있고, 그로써 특정의 개인을 식별할 수 있게 되는 것을 포함한다)

- 그 정보 자체로 식별되는 경우(식별성)
- 다른 정보와 용이하게 照合하여 식별이 가능한 경우(조합성)

1. 개인정보

개인정보와 비식별화 정보

※ 식별 가능한(Identifiable)의 의미

'식별가능한' 의 개념 요소

- 단일성 (Single-out)
- 연결성 (Likability)
- 추론성 (inference)

- 개인정보처리자 또는 정보주체 외 사람이
- 합리적으로 가능한 모든 수단을 이용하여
- 정보주체를 구별할 구체적인 가능성을 입수하는 경우

참고 : EU GDPR

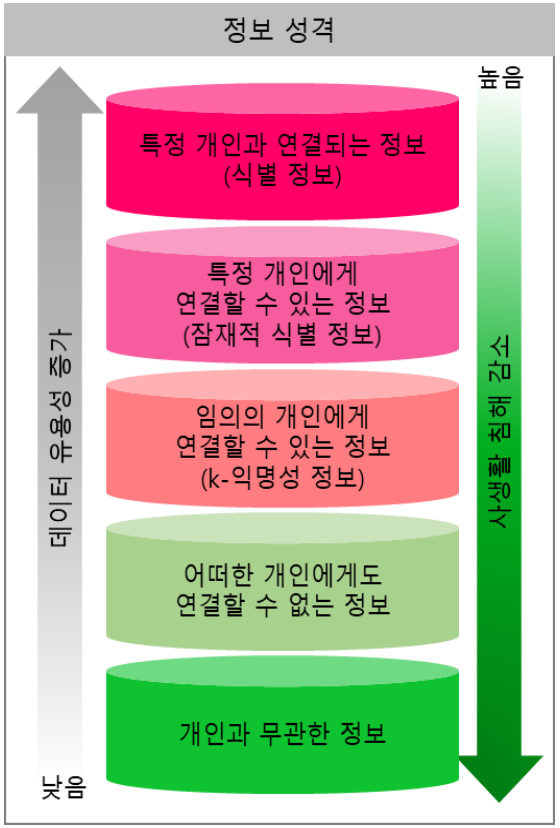
개인정보 Personal Data

단일성 : O
연결성 : O
추론성 : O

비식별화(De-identification)

Pseudonymization (가명화)Data	Anonymization (익명화)Data
단일성 : O 연결성 : X 추론성 : X	단일성 : X 연결성 : X 추론성 : X

※ 개인정보 식별가능성 스펙트럼



참고 : NISTIR 8053

2. 비식별화

빅데이터

왜 하는가?

Prediction

예측은 복잡한 현상을 다양한
각도로 풀어서 논리적으로 해명함

Measurement

측정은 일정한 기준을 가지고
물건의 양을 수치화하는 것

**분석은 측정하는 것이
아니라 예측하는 것이다**

빅데이터에 필요한 정보 수준

Prediction의 결과는 **Measurement** 수준의 입력값 보다

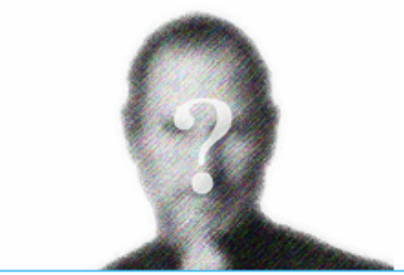
Analysis 기술의 차이가 **좌우** 한다.

Measurement 수준의 데이터를 원하는 것은 **Prediction** 이 목적이 아니라 다른데 있다고 본다.

2. 비식별화

비식별화 정보와 익명정보

- Anonymous Data : 정보의 수집단계에서 근원적으로 개인을 식별할 수 없는 형태로 수집한 정보
- de-identification Data : 개인을 식별할 수 있는 상태에서 수집한 정보를 비식별화 과정을 통하여 개인을 식별할 수 없게 처리한 정보

익명 정보		비식별화 정보
	≠	
Anonymous Data	vs	de-identification Data
특정개인을 식별할 수 없는 형태로 정보를 수집한 자료		개인을 식별할 수 있는 상태로 수집한 정보를 비식별화 과정을 통하여 개인을 식별할 수 없게 처리한 자료

Pseudonymous data ≠ Pseudonymisation data
Anonymous data ≠ Anonymisation data

결과는 유사할 수 있어도 개인에게 주는 영향은 다르다.

2. 비식별화

비식별화 정보

Pseudonymous data, Anonymous data의 정의

「EU 제29조 작업반 의견서」

Pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.

Anonymisation can be a result of processing personal data with the aim of irreversibly preventing identification of the data

→ **Pseudonymisation :**
Single out, NOT linkable
즉 연결성을 낮추는 처리

→ **Anonymisation :**
Irreversible
즉 되돌릴 수 없게 처리

2. 비식별화

비식별화 정보

일본 개인 정보 보호법 지침 (익명 가공 정보 편) 2016.11

「匿名加工情報」とは、個人情報を個人情報の区分に応じて定められた措置を講じて特定の個人を識別することができないように加工して得られる個人に関する情報であって、当該個人情報を復元して特定の個人を再識別することができないようにしたものをいう。

"익명 가공 정보"란 다음의 각호에 해당하는 개인정보 구분에 대응하고 해당 각호에 정하는 조치를 취하여 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻는 개인에 관한 정보로서, 해당 개인정보를 복원할 수 없도록 한 것을 말한다.

1.제1항 제1호에 해당하는 개인정보 : 해당 개인정보에 포함된 기술(記述) 등의 일부를 삭제하는 것(해당 일부 기술 등을 복원할 수 있는 규칙성을 가지지 않은 방법으로 다른 기술 등으로 대체하는 것을 포함)

2.제1항 제2호에 해당하는 개인정보 : 해당 개인정보에 포함되는 개인 식별 부호의 전부를 삭제하는 것(해당 개인 식별 부호를 복원할 수 있는 규칙성을 가지지 않은 방법으로 다른 기술(記述) 등으로 대체하는 것을 포함)

일본 개인정보보호법 제 36 조 (제 3 항)

개인 정보 취급 사업자는 익명 가공 정보를 생성 할 때 개인 정보 보호위원회 규칙으로 정한 바에 따라 당해 익명 가공 정보에 포함되는 개인 정보의 항목을 공표하여야 한다.

(익명 가공 정보 편) 제 21 조
개인 정보 취급 사업자는 익명 가공 정보를 생성 할 때 (※ 1) 익명 가공 정보 작성 후 지체없이 (※ 2) 인터넷 등을 이용하여 해당 익명 가공 정보에 포함 된 일반적인 개인에 관한 항목을 공표 (※ 3)해야한다.

[개인 정보의 항목 사례]
사례) "성명 · 성별 · 생년월일 · 구매 내역"중 이름을 삭제 한 후, 생년월일, 일반화 구매 내역에서 특이 값 등을 삭제하는 등 가공하여 "성별 · 출생 · 구매 내역" 대한 익명 가공 정보로 작성한 경우 공표 항목은 "성별", "출생", "구매 내역" 이다.

2. 비식별화

비식별 가이드라인 비교

	익명화 기준	익명화 기법
EU	【EU 29 Working Party Opinion】 ▪ Single out (개인식별) ▪ Linkability (연결가능성) ▪ Inference (추론) ☞ 합리적 수단을 사용하여 위 3요소를 달성할 수 없는 경우 익명화 충족	【EU 29 Working Party Opinion】 1. 무작위화 (Randomization) - 잡음추가, 대체, 차분 프라이버시 2. 일반화 (Generalization) - 총계처리, k-,익명성, l-다양성, t-근접성 ※ 익명처리 기법은 재식별 위험성을 항상 내재하고 있으므로, 결과의 안전성을 제고하기 위해 여러 기법을 적용할 때 면밀한 설계 권고
	【가명처리 기준】 개인 기록부(data set)내에 있는 어떤 속성 (일반적으로 고유속성)을 다른 속성으로 교체하는 것 - 데이터 셋의 익명화를 달성 할 수 없음	【가명처리 기법】 1. 비밀키 암호화(Secret Key) 2. 키를 사용하는 키 해쉬 3. 키 해쉬에서 키 삭제 4. 토큰화
일본	【개인정보보호법에 대한 지침 (익명가공 정보편)】 <익명가공정보 조치 기준> 개인정보를 개인정보의 구분에 맞추어 정해진 조치를 취해 특정 개인을 식별할 수 없도록 가공한 뒤 얻어지는 정보로서 통상의 방법을 통해 복원하여도 특정 개인을 재식별 할 수 없어야 함	【개인정보보호법에 대한 지침 (익명가공정보편)】 1. 삭제(개인연결번호 등 삭제) 2. 일반화(수치변경 등), 라운딩 3. 범주화 4. 대체 5. 잡음 추가 등
한국	【개인정보 비식별조치 가이드라인】 <비식별조치 기준> 비식별 조치된 정보에 대해, 재식별 시도 가능성 분석, 재식별 시 영향분석 등을 통해 산출된 기준값에 대한 적정성 평가 충족한 경우 ☞ 현재의 수단 등 고려시, 재식별이 불가능한 수준	【개인정보 비식별 조치 가이드라인】 1. 17개 비식별 조치 방법을 선택 사용 2. k-익명성, l-다양성, t-근접성 등 프라이버시 모델을 적용한 적정성 평가를 충족하는 비식별 조치

(출처 : KISA)

2. 비식별화

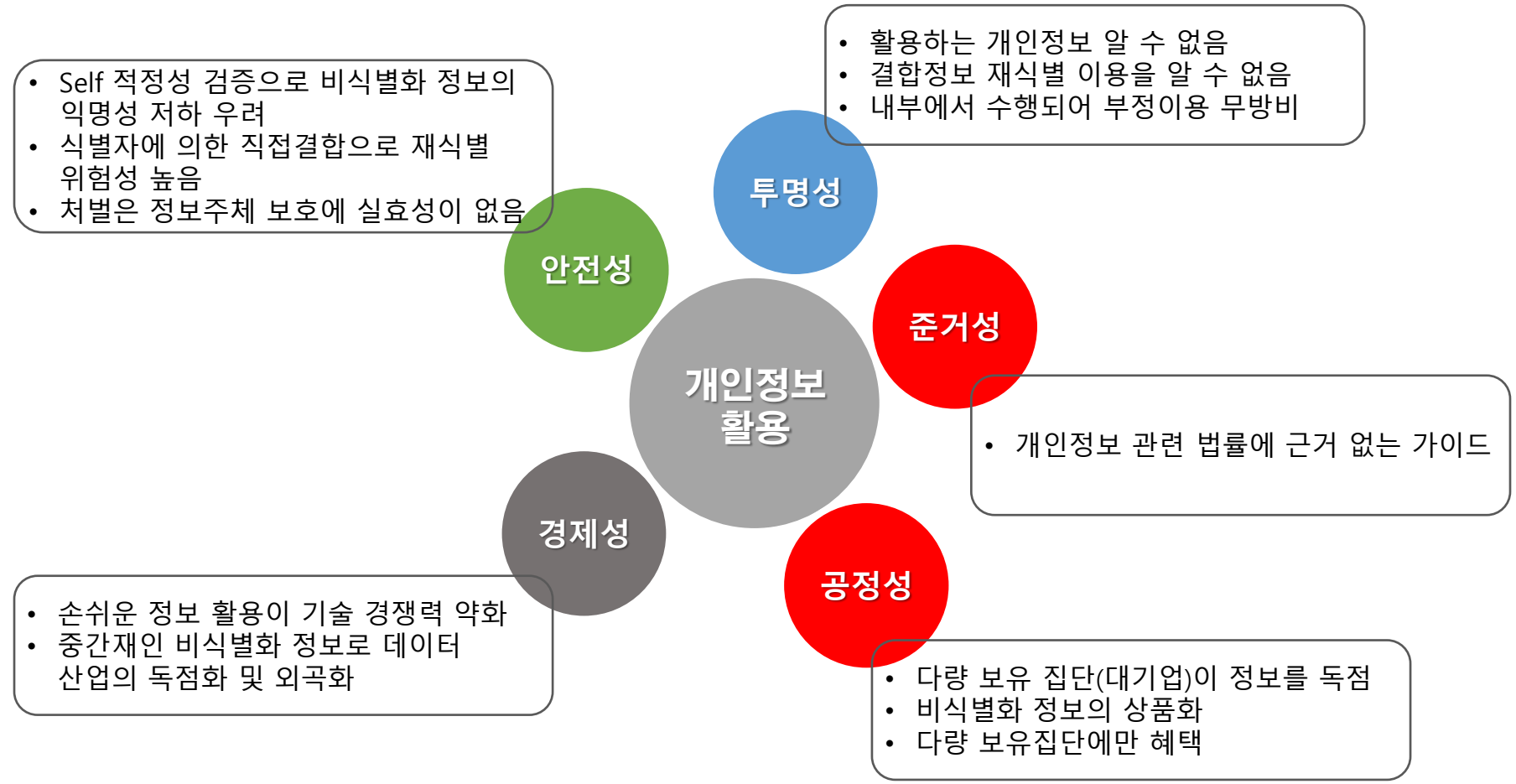
비식별 가이드라인 비교

항목/가이드	EU ENISA guidline (2015.12)	영국 UKAN guidline (2016.09)	미국 NIST 800-188 (2016.12)	호주 (2014.04)	일본 (2016.11)	한국 (2016.06)
용어	Anonymisation	Anonymisation	De-identification	De-identification	Anonymisation	De-identification
법적근거	O	O	O	O	O	X
비식별 정보 개념	O	O	O	O	O	O
비식별 처리 수준	X	X	O	X	X	X
비식별 기법	O	O	X	O	O	O
적정성 평가	권고 수준	X	X	X	권고 수준	결합시 평가
정보 공개	불특정 다수	불특정 다수	불특정 다수	불특정 다수	불특정 다수	조건부 제3자 제공(계약관계)

3. 가이드의 문제점

3. 가이드의 문제점

비식별화 조치 가이드의 문제점



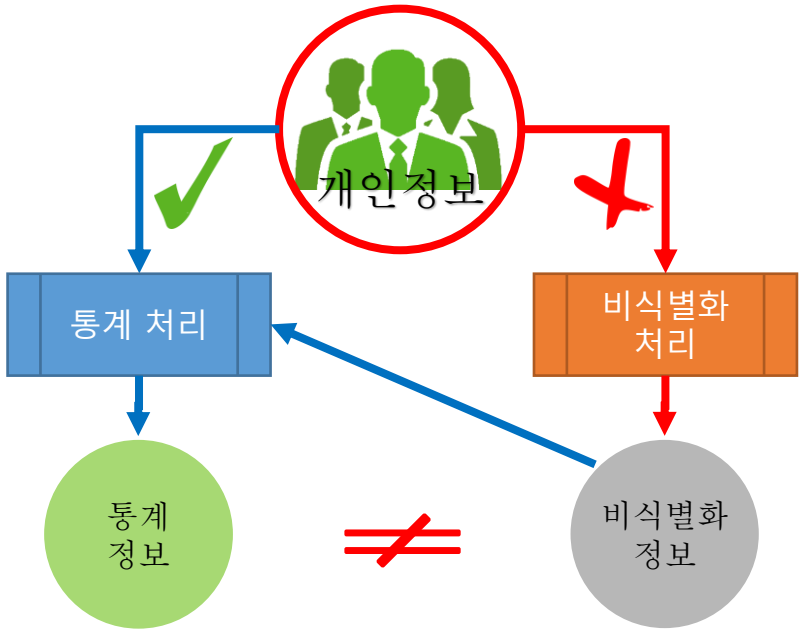
3. 가이드의 문제점

비식별화의 근거 미비

비식별화 조치는 개인정보의 이용 범위 인가?

제18조(개인정보의 목적 외 이용·제공 제한)

②-4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우



통계처리와 비식별화 처리는 다르다

3. 가이드의 문제점

비식별화의 적정성 판단

식별 가능성의 판단을 활용 주체(기업)가 판단

- 개인정보 해당 여부의 판단 기준

iv) (개인을 알아볼 수 있는 정보)이므로 특정 개인을 알아보기 어려운 정보는 개인정보가 아님 '알아볼 수 있는'의 주체는 해당 정보를 처리하는 자(정보의 제공 관계에 있어서는 제공받은 자를 포함)이며, 정보를 처리하는 자의 입장에서 개인을 알아볼 수 없다면 그 정보는 개인정보에 해당하지 않음

처리 및 활용을 활용 주체가 스스로

- 적정성 판단을 정보처리주체가 스스로 검증을 한다.
- 어떤 개인정보를 활용하는지 공개가 안된다.
- 비식별 정보의 사용 목적이 공개되지 않는다.

모든 것이 SELF

(출처 : 개인정보 비식별조치 가이드라인)

3. 가이드의 문제점

비식별화의 충족성 검증

"Pseudonymization만으로는 비식별화 조치에 해당하지 않음" 누가 검증?

▣ 비식별 조치 방법

◎ 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 마스킹 등 여러가지 기법을 단독 또는 복합적으로 활용

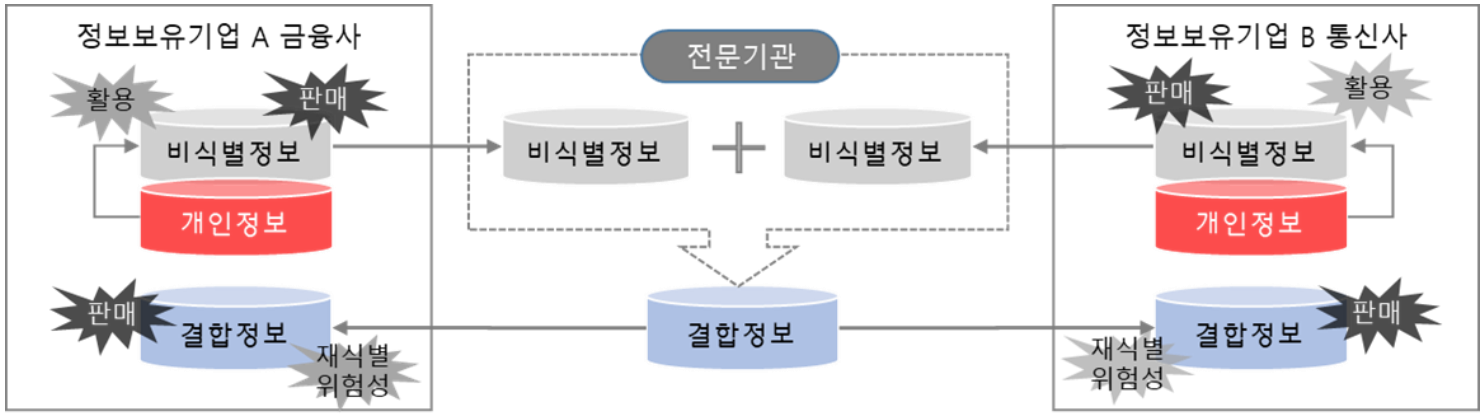
※'가명처리' 기법만 단독 활용된 경우는 충분한 비식별조치로 보기 어려움

이것으로 모든 것이 충분한가?

- 식별가능성의 판단을 활용주체가 판단
- 정보결합시만 적정성 평가
- 조건부 제3자 제공 가능(계약 맺은 제3자)
  ※ 불특정 다수에 공개 금지

→

- 식별가능성 객관성 결여
- 비식별화 정보의 상품화 및 집중화
- 부정이용의 무방비



3. 가이드의 문제점

개인정보의 비식별화 시사점

개인정보의 비식별화

- De-identification Data의 적정성 평가제 필요
- 법률적으로 개인정보의 수집이 필수가 아닌 경우 익명(Anonymous)으로 수집

  예) 사물인터넷(IoT) 영역은 개인정보와 직접 관련이 없으므로 익명(Anonymous) 수집 도입

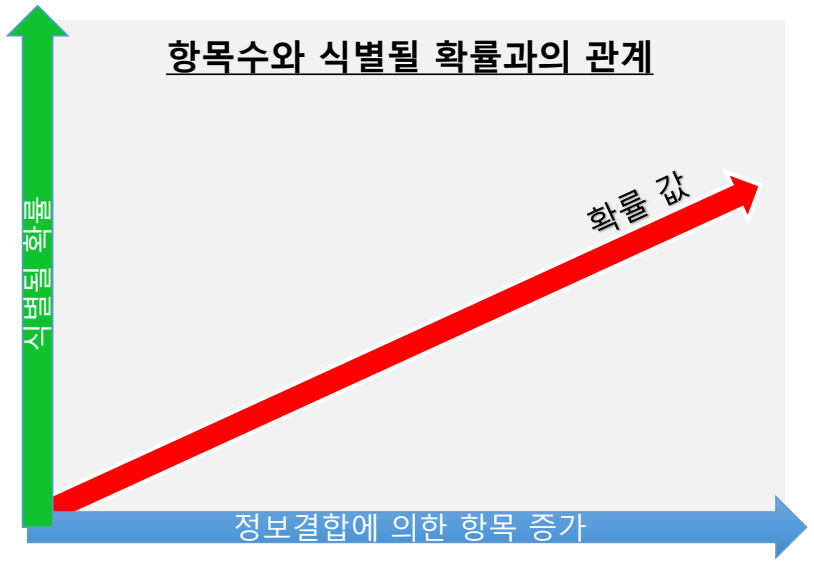
Pseudonymous data ≠ Pseudonymisation data
Anonymous data ≠ Anonymisation data

제3의 기관(Honest Broker)의 관리 필요

- 적정성 평가 모델 개발, 평가로 익명성 보장을 통한 재식별의 위험성 제거
- 합리적인 공유모델 수립을 통해 정보 유통 및 활용의 투명성 확보

정보 결합은 추론결합만 허용 필요

- 결합으로 발생하는 속성 증가에 따른 재식별 위험성 제거
- 추론 기술 연구 개발을 통한 비식별화 기술 경쟁력 강화



4. 비식별화 방향

4. 비식별화 방향

개인정보의 비식별화

개인정보 정보주체가 이해하는 활용

- 개인정보의 활용 실태 공표
- 비식별화 외부 감시 가능
- 개방을 통한 투명한 활용

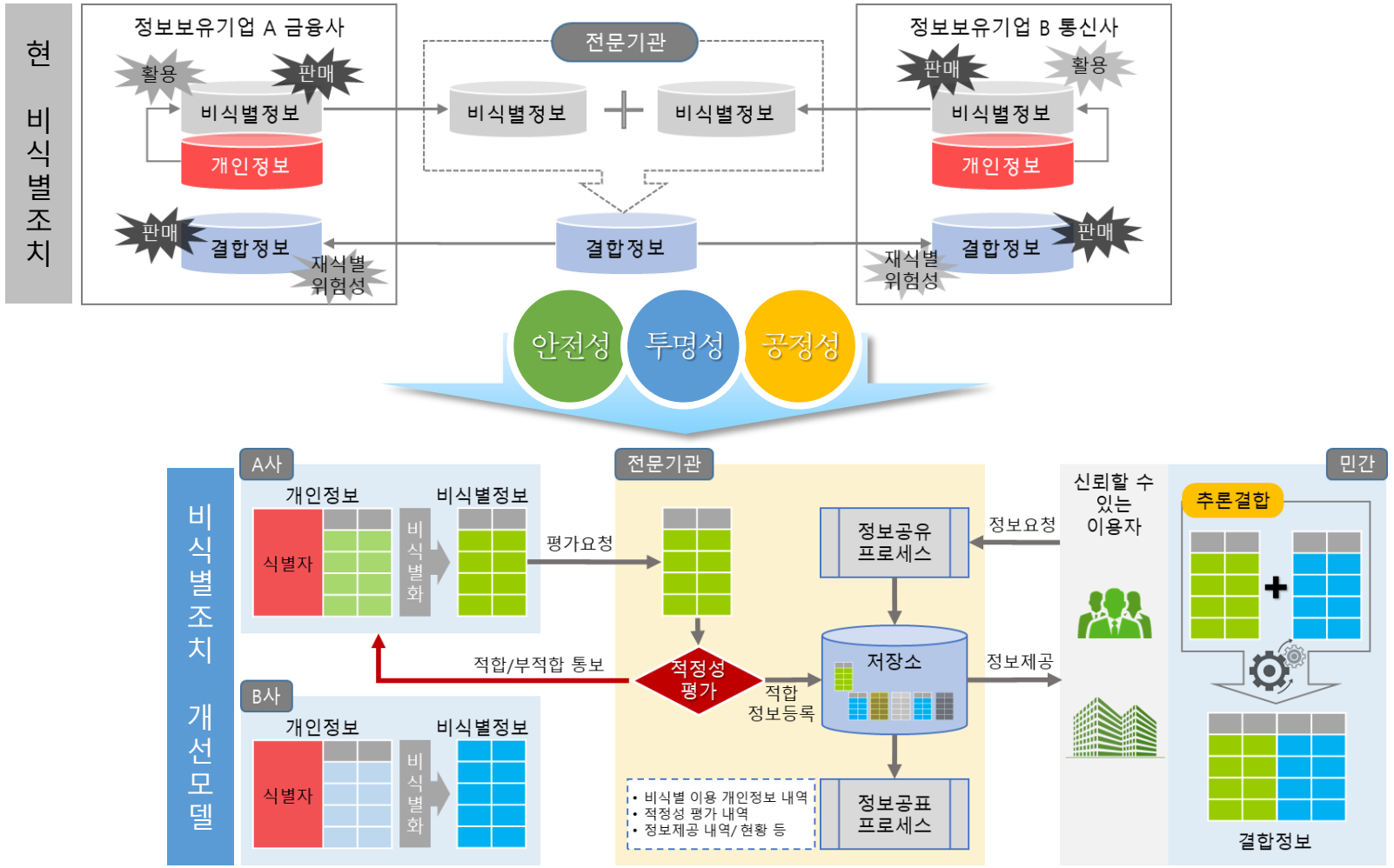
- 비식별화 정보의 적정성 확보
- 추론결합을 통한 익명성 강화

- 정보의 집중화 및 상품화 방지
- 비식별화 정보 공정 활용
- 기술 개발을 통한 경쟁력 강화

정보보유자	전문기관	비 정보보유자
• 보유정보 비식별화 • 비식별화 정보 확보 • 추론 결합 • 분석 • 분석결과 활용 및 유통	• 비식별화 정보 적정성 심사 • 비식별화 정보 공표 • 공유 유통 모델 수립, 운영 • 정보 생명주기 수립, 운영	• 비식별화 정보 확보 • 추론 결합 • 분석 • 분석결과 활용 및 유통

4. 비식별화 방향

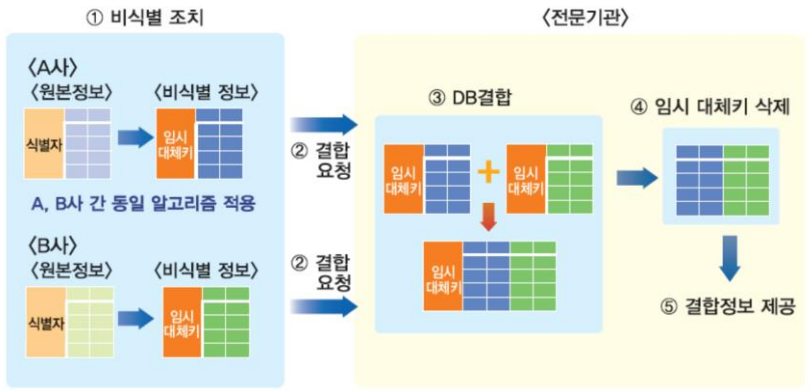
비식별화 조치 비교



4. 비식별화 방향

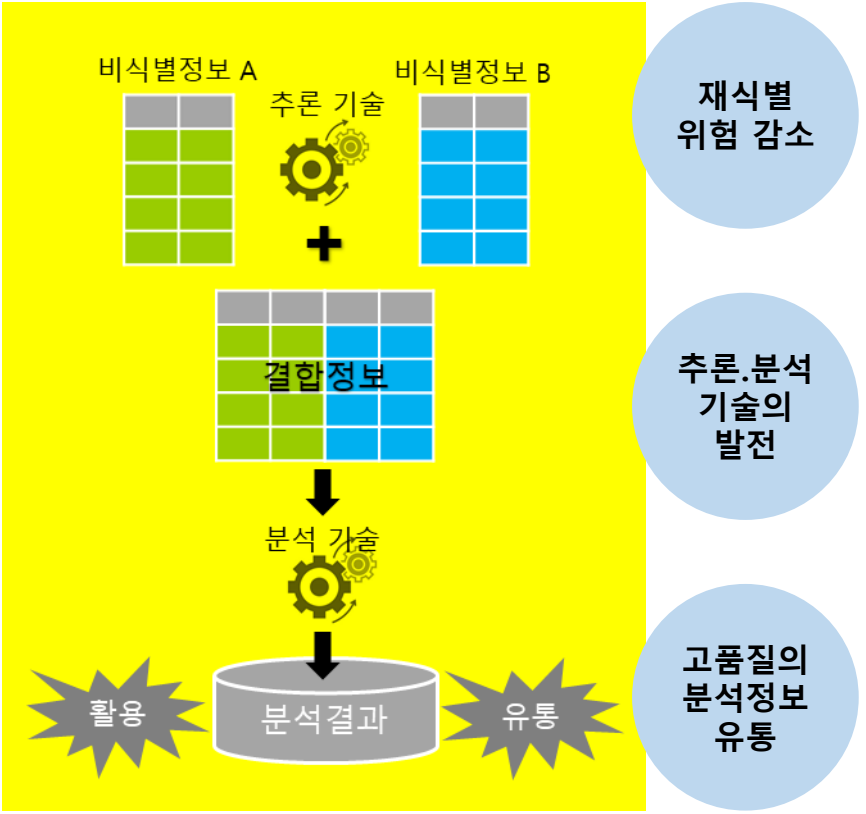
비식별 정보 결합방식 비교

직접 결합 방식



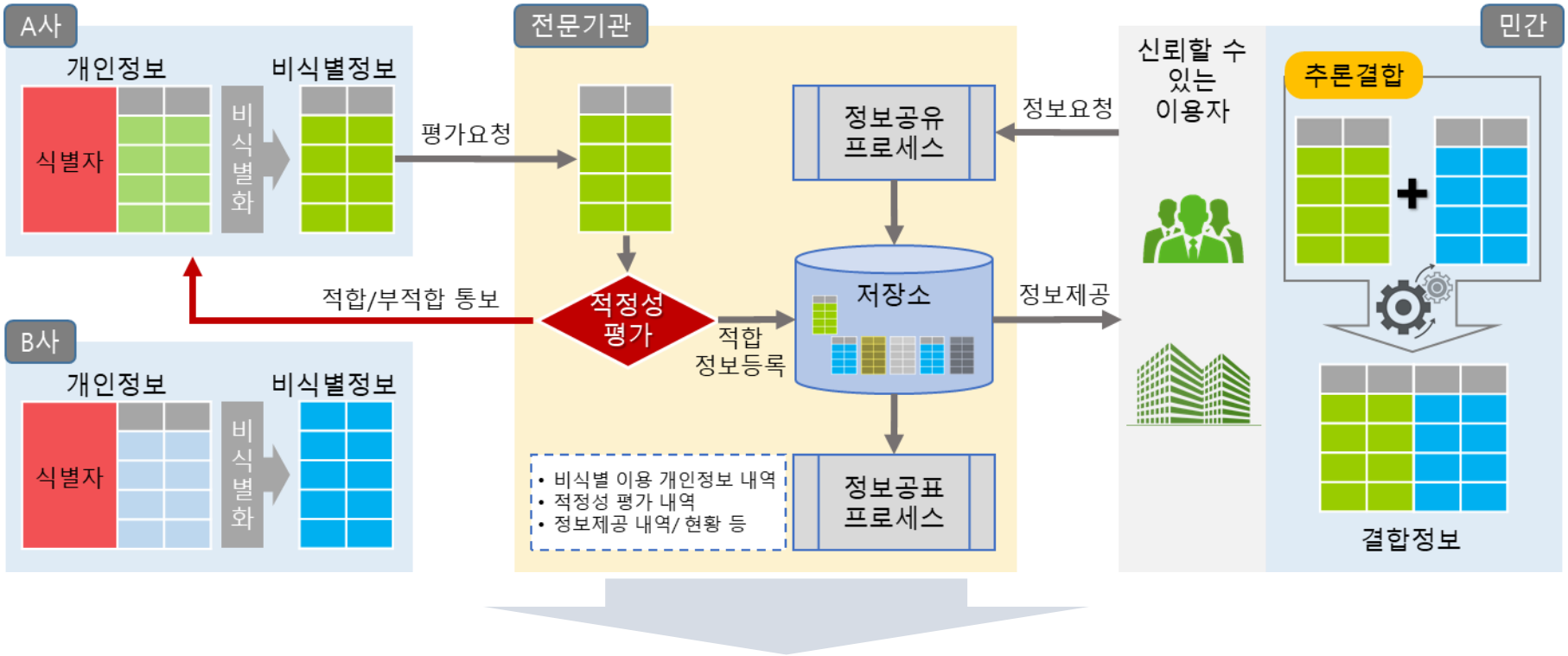
- 결합항목 증가로 인한 재식별 위험 증가
- 대량집단간 이해관계에 의한 결합으로 정보 집중의 가속화
- 개인정보의 단순 가공을 통한 수익화

추론 결합 방식



4. 비식별화 방향

비식별화 모델



개인정보 → 안전하고 투명한 비식별정보 활용
추론 기술 및 분석 기술에 의한 빅데이터 산업 육성 → 비식별화

4. 비식별화 방향

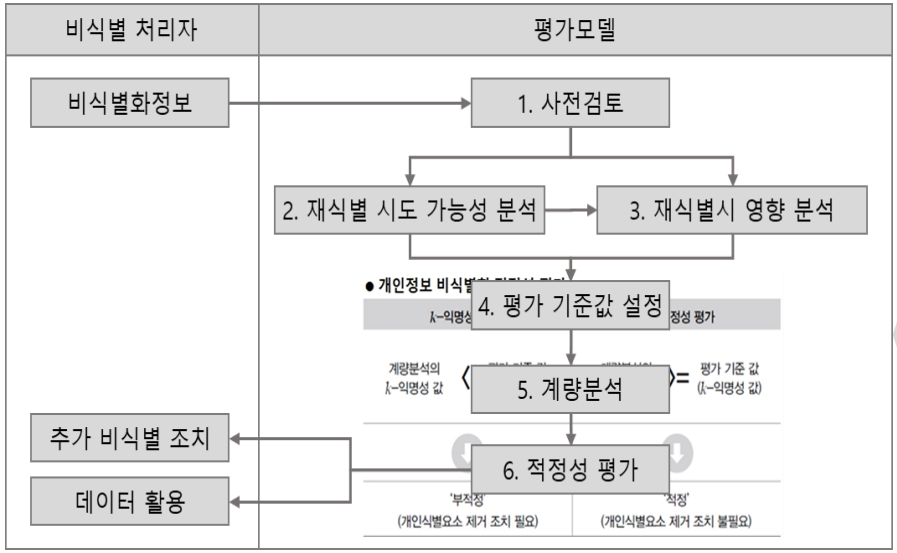
중재인(Honest Broker)

중재인(Honest Broker)의 필요성

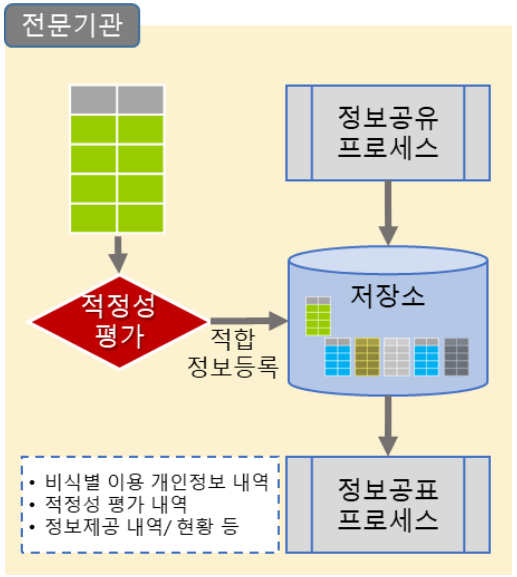
- 비식별화 정보의 안전성 확보
- 비식별화 정보 활용의 공정성 확보
- 비식별화 정보의 효율적 관리
- 개인정보 활용의 투명성 확보
- 개인정보의 합리적 관리
- 사전단계 익명정보(Anonymous Data) 도입

중재인(Honest Broker)의 역활

- 비식별화 정보의 적정성 평가모델 수립 및 평가
- 비식별화 정보의 공유모델 수립 및 제공
- 비식별화에 활용된 개인정보 실태 분석
- 사전단계 Anonymous Data의 모델 수립
- 비식별화 정보 이용 실태 분석
- 각 현황의 작성 및 공표

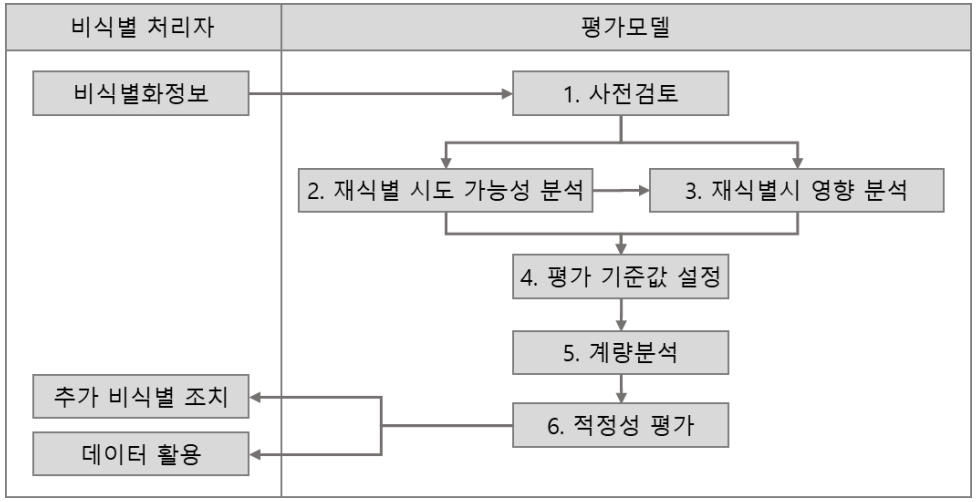


- 합리적 익명성의 평가 모델 수립
- 적정/부적정 평가 통보



4. 비식별화 방향

비식별 처리의 평가



● **개인정보 비식별화 적정성 평가**

k-익명성 값을 이용한 개인정보 비식별화에 대한 적정성 평가

계량분석의 k-익명성 값 < 평가 기준 값 (k-익명성 값)	계량분석의 k-익명성 값 >= 평가 기준 값 (k-익명성 값)
'부적정' (개인식별요소 제거 조치 필요)	'적정' (개인식별요소 제거 조치 불필요)

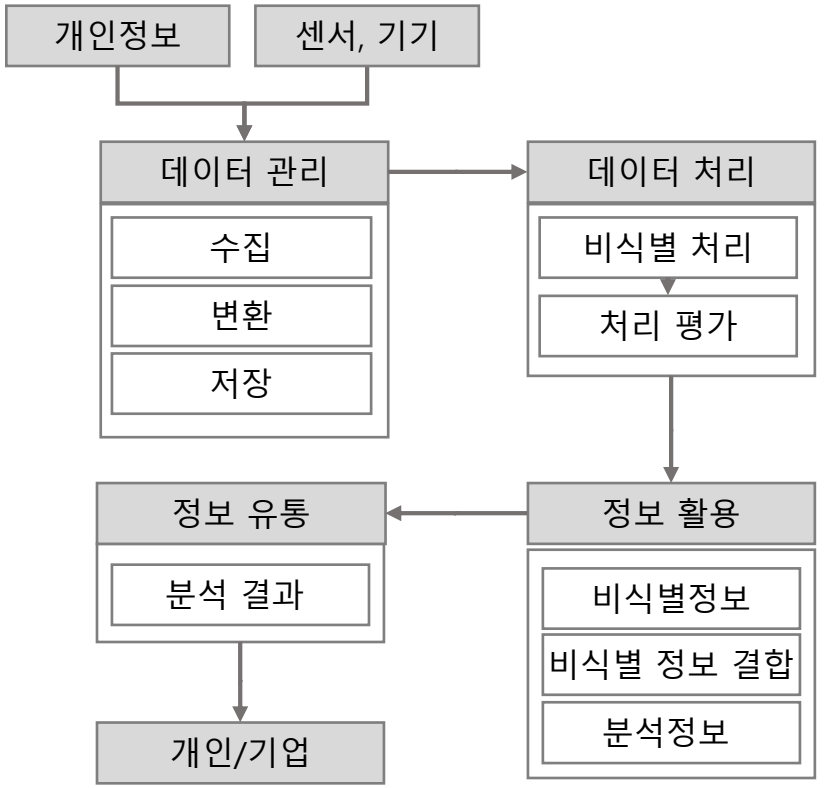
비식별 처리의 평가

1. 사전검토 : 신청기관의 기조차료와 비식별자료를 검토하여 데이터 특성 및 비식별 조치 확인
2. 재식별 시도 가능성 분석 : 데이터 관리수준을 검토하여 재식별 가능성 분석
3. 재식별 시 영향 분석 : 데이터가 재식별 가능할 경우의 영향을 평가
4. 평가 기준 값 설정 : 데이터 특성을 고려하여 분석 기법을 선택하여 적정 수치를 설정
   - K-익명성(필수) 외 l-다양성, t-근접성 등 평가 지표 개발
5. 계량 분석 : 현재 적정성 수준의 정확한 파악을 위해 관련 수치 도출
6. 적정성 평가 : 평가 기준 값과 계량분석결과를 비교하여 "적정" 또는 "부적정"으로 평가
   - 부적정 평가 시 추가 비식별 조치 시행 후 재평가
7. 데이터 활용 : 신청기관은 비식별 조치 적정성 평가 결과가 '적정'일 경우에만 해당 비식별 정보의 활용이 가능

4. 비식별화 방향

비식별 처리를 위한 생명주기

빅데이터 환경에서 비식별 처리를 위한 생명주기



■ 빅데이터 분석 과정에서 데이터 생명주기

빅데이터 처리 과정에서 생명주기는 비식별화 정보에서 발생 가능한 프라이버시 위협과 취약성을 예측하여 비식별 처리가 의도된 목적과 용도에 따라 사용될 수 있도록 함

■ 빅데이터 분석 과정에서 비식별 처리

✓데이터 관리 단계

- 불필요한 개인정보의 최소 수집 원칙
- 개인정보의 안전한 관리

✓데이터 처리 단계

- 데이터 변환 방법은 요구되는 개인정보 노출 위협을 방지하기 위한 적절한 기법의 고려
- 관리 기관은 비식별화 데이터를 활용할 때 해당 데이터와 관련된 다양한 이해관계자를 포함한 평가단을 통하여 비식별정보의 안전성 검증

✓정보활용 단계

- 서로 독자 기술 및 분석을 통해 경쟁

✓정보유통 단계

- 분석결과의 자유로운 유통

4. 비식별화 방향

정보 공유 모델

정보의 성격별 공유 모델

데이터 성격	접근 정책	데이터 용도
비 개인적인/기밀이 아닌 non-personal/non-confidential	 공개 접근	• 추세 분석 집계 • 시장 분석 및 요약 • 환경 및 물리적 자산 현황
확인되지 않은 De-identified	 공개 접근/ 신뢰할 수 있는 이용자	• 시나리오 개발 및 테스트 • 프로그램 / 정책 분석 및 평가 • 광범위한 서비스 기능 및 분석 • 마켓팅 성과 및 관리
식별될 수 있는 Pseudonymization leval	 신뢰할 수 있는 이용자	• 타겟 / 통합 제품 및 서비스 • 프로그램 / 규정 시행 및 준수
식별 할 수 있는 identifiable	 신뢰할 수 있는 이용자	• 타겟 / 통합 제품 및 서비스 • 프로그램 / 규정 시행 및 준수
비밀/보호받는 confidential/proteced	 특정개인	• 소비자 선택 및 의사 결정 • 개인의 활동, 행동 또는 결과를 모니터링

데이터 관리자가 공개를 결정한 정보

4. 비식별화 방향

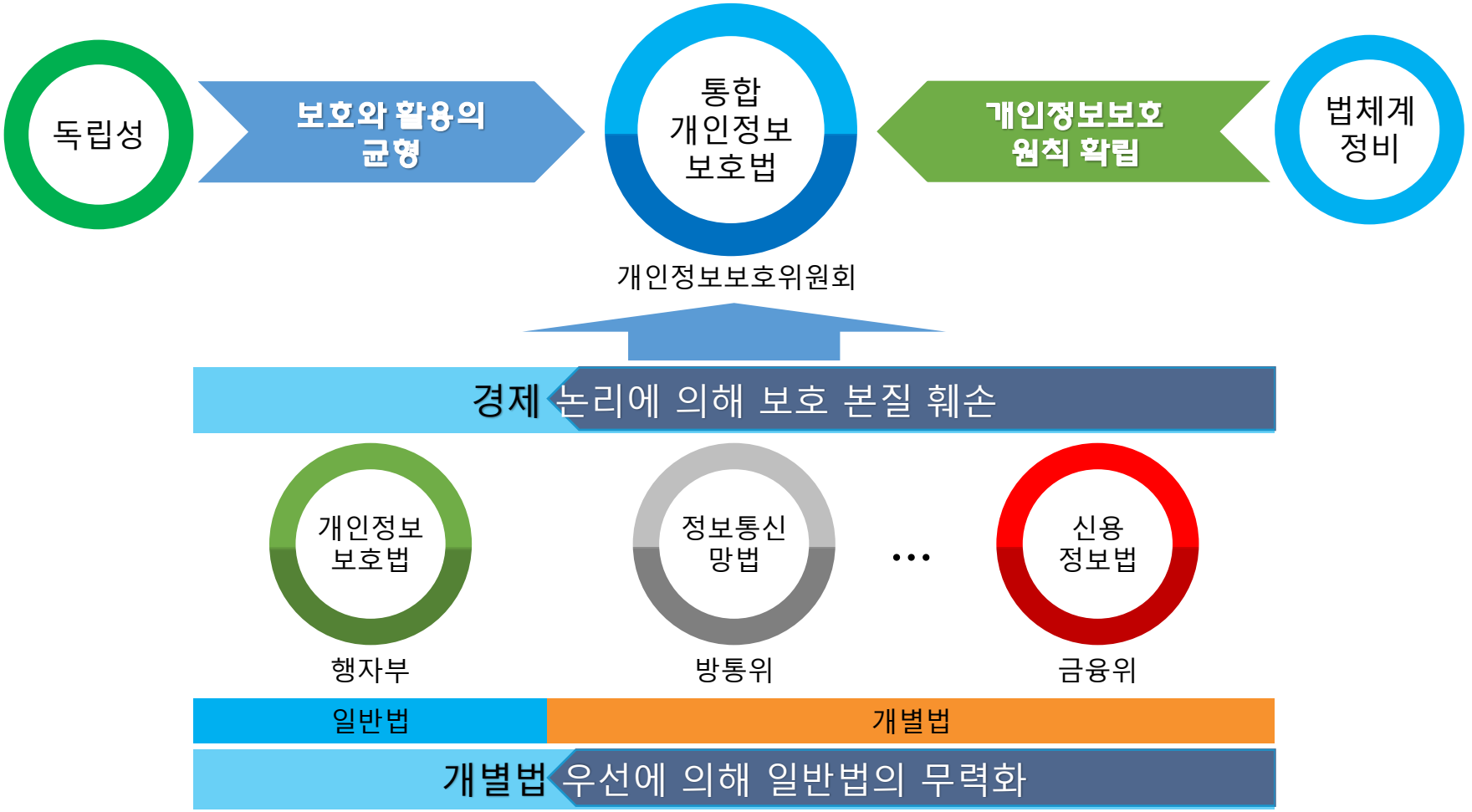
정보 공유 모델

정보 공유 모델의 비교

항목		공개적 활용 모델	반 공개적 활용 모델	비공개적 활용 모델
일반	접근성	모든 사람	**신뢰할 수 있는 이용자**	특정 대상
	활용 사례	• 웹으로 공공 데이터 공개	**• 요청에 의한 정보 전달**	• 이해관계자간에 정보 전달
비식별 정보	정보의 활용성	높음	**높음**	특정 집단에 국한
	익명화 신뢰성	높음	**보통**	낮음
	재식별 위험성	높음	**보통**	낮음
	독점화 및 상품화	낮음	**낮음**	높음
	부정이용 투명성	높음	**높음**	낮음

4. 비식별화 방향

개인정보보호 제도 개선



4. 비식별화 방향

개인정보 보호 제도 이것만은 개선하자

동의만능

동의만 받으면 무엇이든 수집 가능

- ✓ 최소수집원칙 무용론
- ✓ 포괄적 목적 정보 요구
- ✓ 권리는 스스로 찾아라

활용우선

솜방망이 처벌의 보호조치

- ✓ 빈번한 유출 사고
- ✓ 형식적인 보호책임자
- ✓ 기본조차 안 지키는 법

선택권 없는 주체

유명무실한 정보주체 권리

- ✓ 사소한것도 정보 요구
- ✓ 사규가 법보다 우선
- ✓ 싫으면 이용하지마

산업우선

산업 우선의 개별법에 일반법 속수무책

- ✓ 근거 없는 비식별 조치
- ✓ 보호 없는 개별법 개정
- ✓ 개인정보의 판매 허용

4. 비식별화 방향

개인정보 보호 제도 개선 방향

동의 만능

동의만 받으면 무엇이든 수집 가능

- 최소수집원칙 무용론
- 포괄적 목적 정보 요구
- 권리는 스스로 찾아라

활용 우선

솜방망이 처벌의 보호조치

- 빈번한 유출 사고
- 형식적인 보호책임자
- 기본조차 안 지키는 법

선택권 없는 주체

유명무실한 정보주체 권리

- 사소한 것도 정보 요구
- 사규가 법보다 우선
- 싫으면 이용 하지마

산업 우선

산업 우선의 개별법에 일반법 속수무책

- 근거 없는 비식별 조치
- 보호 없는 개별법 개정
- 개인정보의 판매 허용

감사합니다