Consideration on de-identification guideline of Personal Information for data protection & safe use of data

Apr. 24. 2017

Kibyoung Kim

Director
Global e-government division
Ministry of the Interior

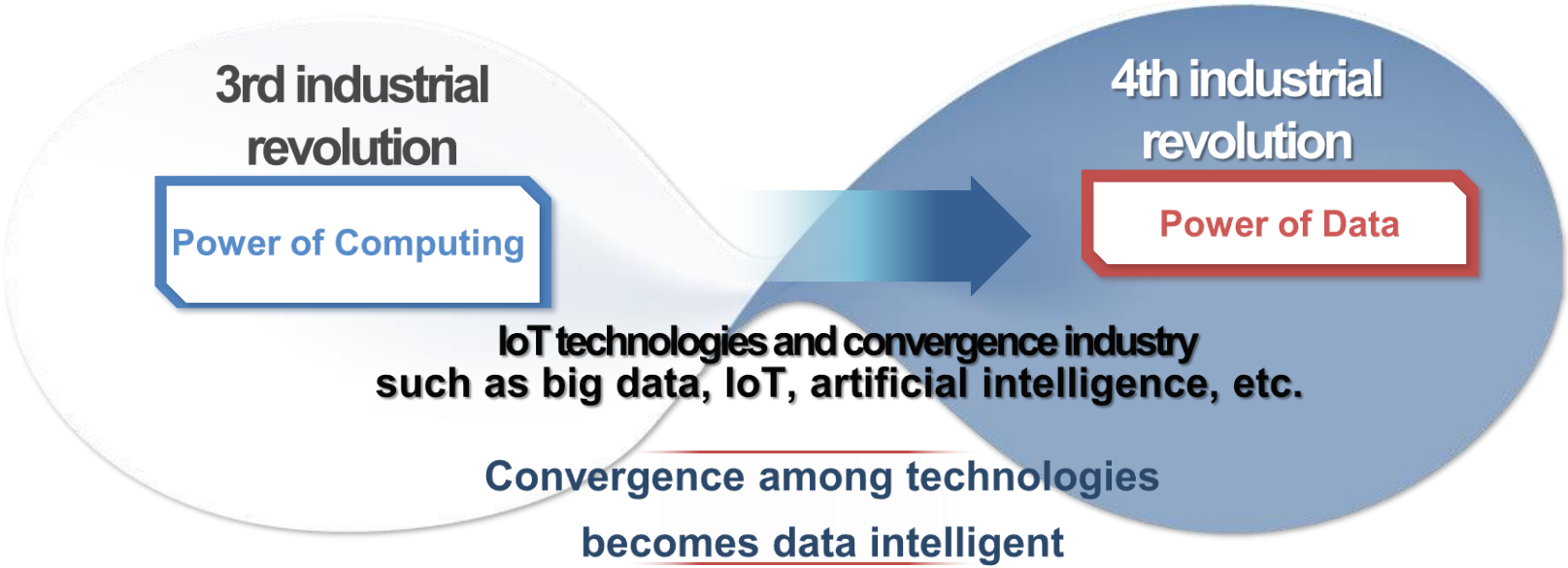
Ministry of the Interior

Republic of Korea

Contents

Background

In many area, data usage becomes violating GDPR, and, we need to find how to avoid violation of GDPR while increasing the value of data, big data analysis



Personal data protection *vs.* **Utilization of data – Big data, IOT, AI**

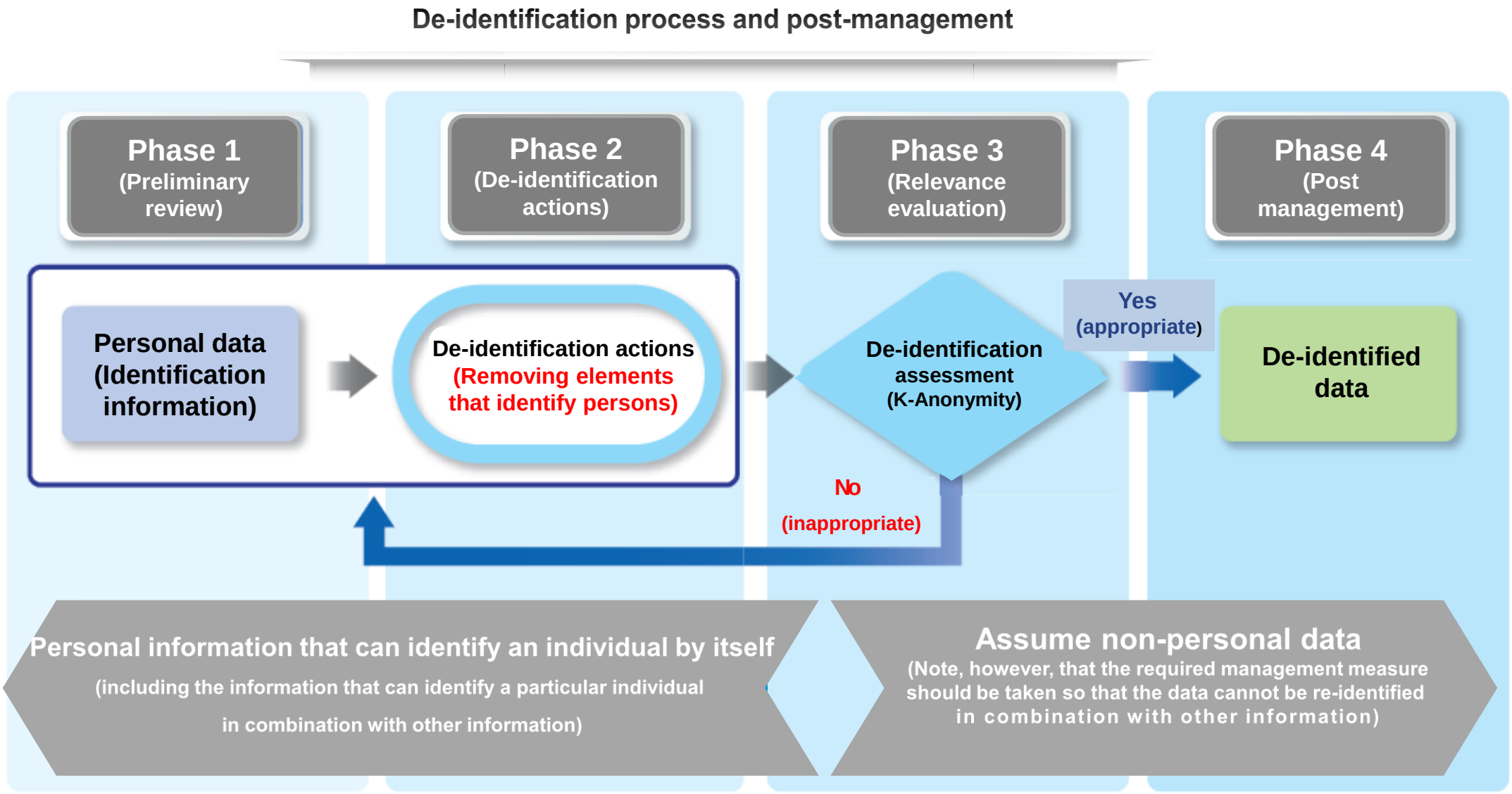
* Human right, Consumer's right
* Transfer data to other countries? or localization?

- De-identification
- Adequacy evaluation
- Join de-identified data

We need to prepare environment to use big data without violating GDPR in the new 4th industrial revolution era

Approach

Data can be used without violating GDPR if there's no personal information in the data, or, if the processed data meet the guideline of de-identification guideline



De-identification methods

Personal data can be de-identified using one of various methods such as replace, aggregation, reduction, categorization, and masking or together

Methods	Examples	Technologies
Replace	○ James Hong, 35 years old, living in Seoul, student of Hanguk Univ. → Kevin Lee, 30s, living in Seoul, student of Korea University	① Heuristic replacement of name ② Encryption ③ Exchange method
Aggregation	○ Kevin Lee - 180cm, James Hong - 170cm, Paul Kim – 150cm Edwin Park - 160cm → Total height of physics department students: 660cm, average height - 165cm	④ Sum ⑤ Partial sum ⑥ Rounding up/down ⑦ Rearrangement
Suppression (Reduction)	○ Resident registration number: 901206-1234567 born in the 1990s, male → born in the 1990s, male	⑧ Identifier removal ⑨ Partial identifier removal ⑩ Record removal ⑪ All identifiers removal
Generalization	○ James Hong - 35 years old, → Mr. Hong, age of 30s ~ 40s	⑫ Hiding ⑬ Random rounding up/down ⑭ Scoping ⑮ Control rounding
Masking	○ James Hong, 35 years old, living in Seoul, student of Hanguk Univ. → OO Hong, 35 years old, living in Seoul, student of OO University	⑯ Adding random noise ⑰ Replacing with blank/character

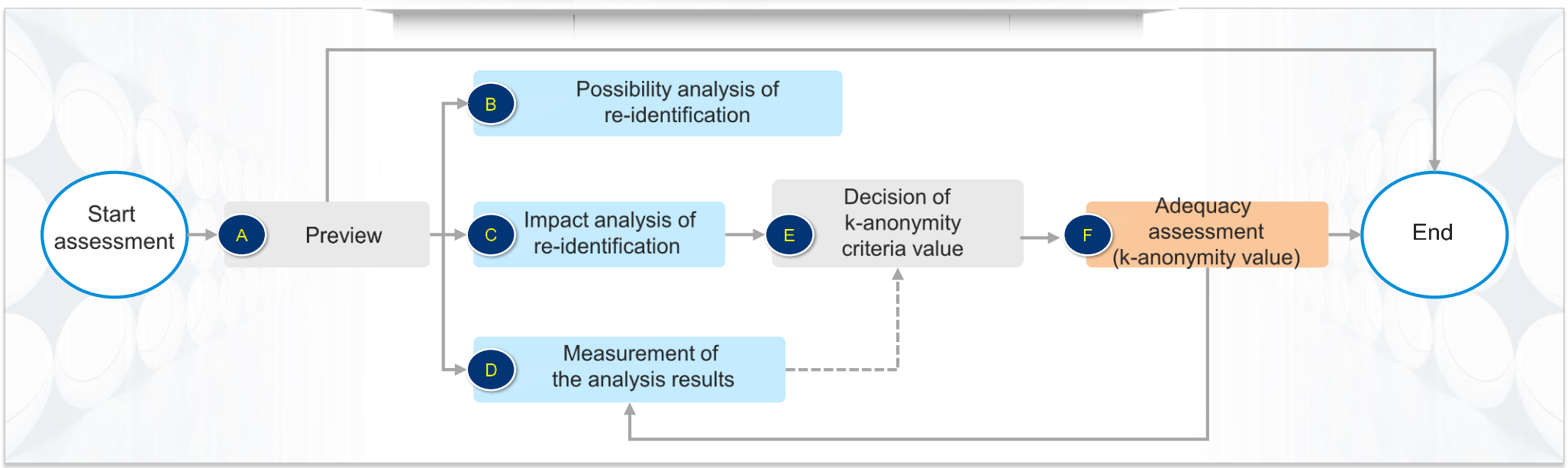
Adequacy evaluation

We can check whether de-identified data contain personal information or not using adequacy evaluation based on k-anonymity model

Evaluation committee checks the adquacy of the de-identification level using the sample data provided by the personal data handler and K-anonymity model.

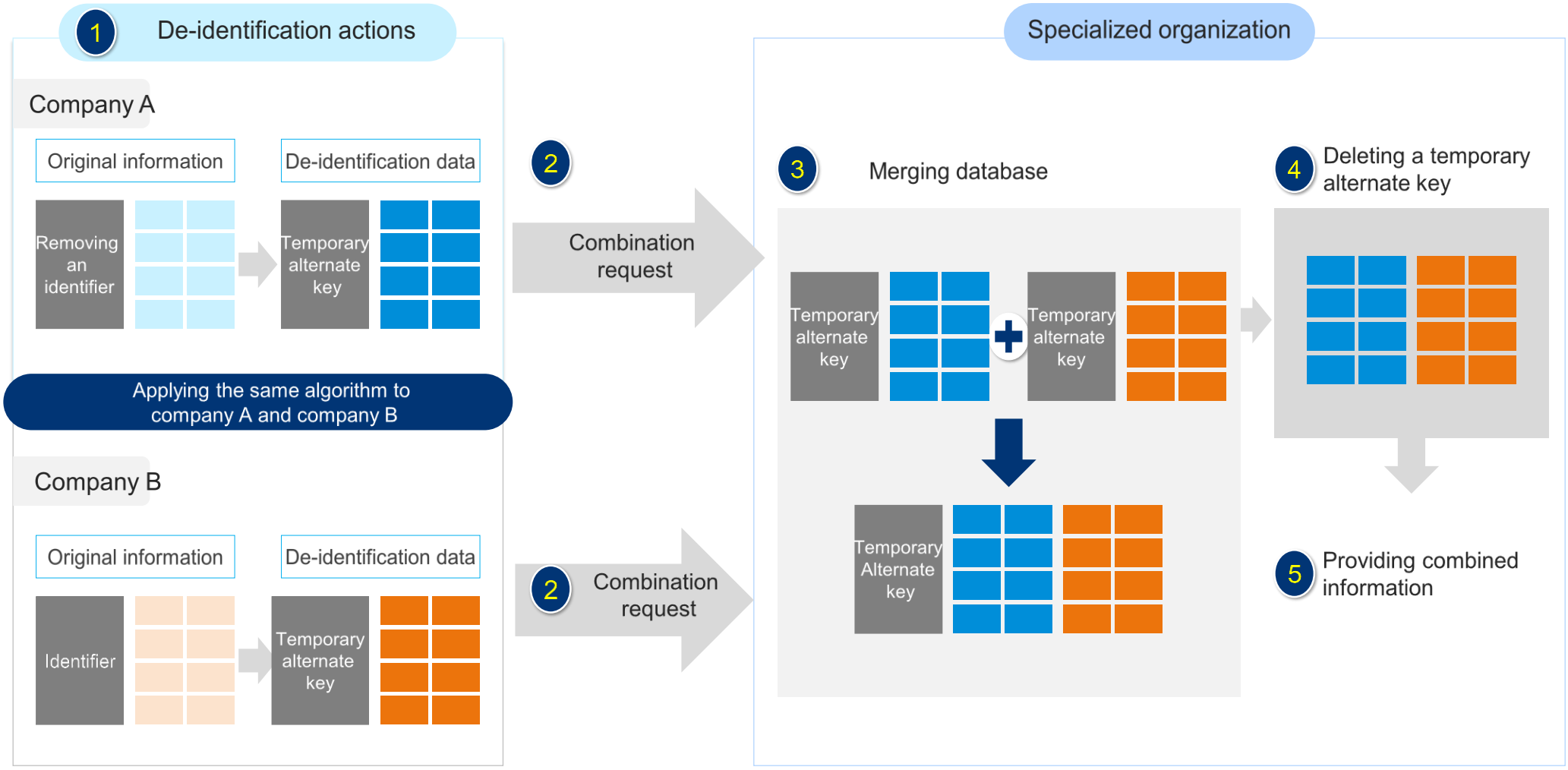
- Assess as “appropriate” if the k-anonymity value is 4 and the assessment criteria value is 3. (>)
- Assess as “inappropriate” if the k-anonymity value is 4 and the assessment criteria value is 6. (<)

Adequacy Assessment Process



Combination of data

In order to merge/join two or more data sets, we can use temporary key to join data sets, and, then, de-identify the results without violating k-anonymity rule



Summary

It’s good time for considering the guideline both for data protection & safe use of data.
And, it’s a guideline, not act/law itself. It needs more discussion and enhancement.

1. **De-identification process**
2. **Adequacy evaluation**
3. **Join/combine data sets without personal information**

Is it OK to use personal Information?

K-Anonymity rule is not a silver bullet to solve the problem of big data analysis In data protection point of view

but, it’s a good start for discussion