Privacy and security
Myths and fallacies of “Personally identifiable information”

Section A03

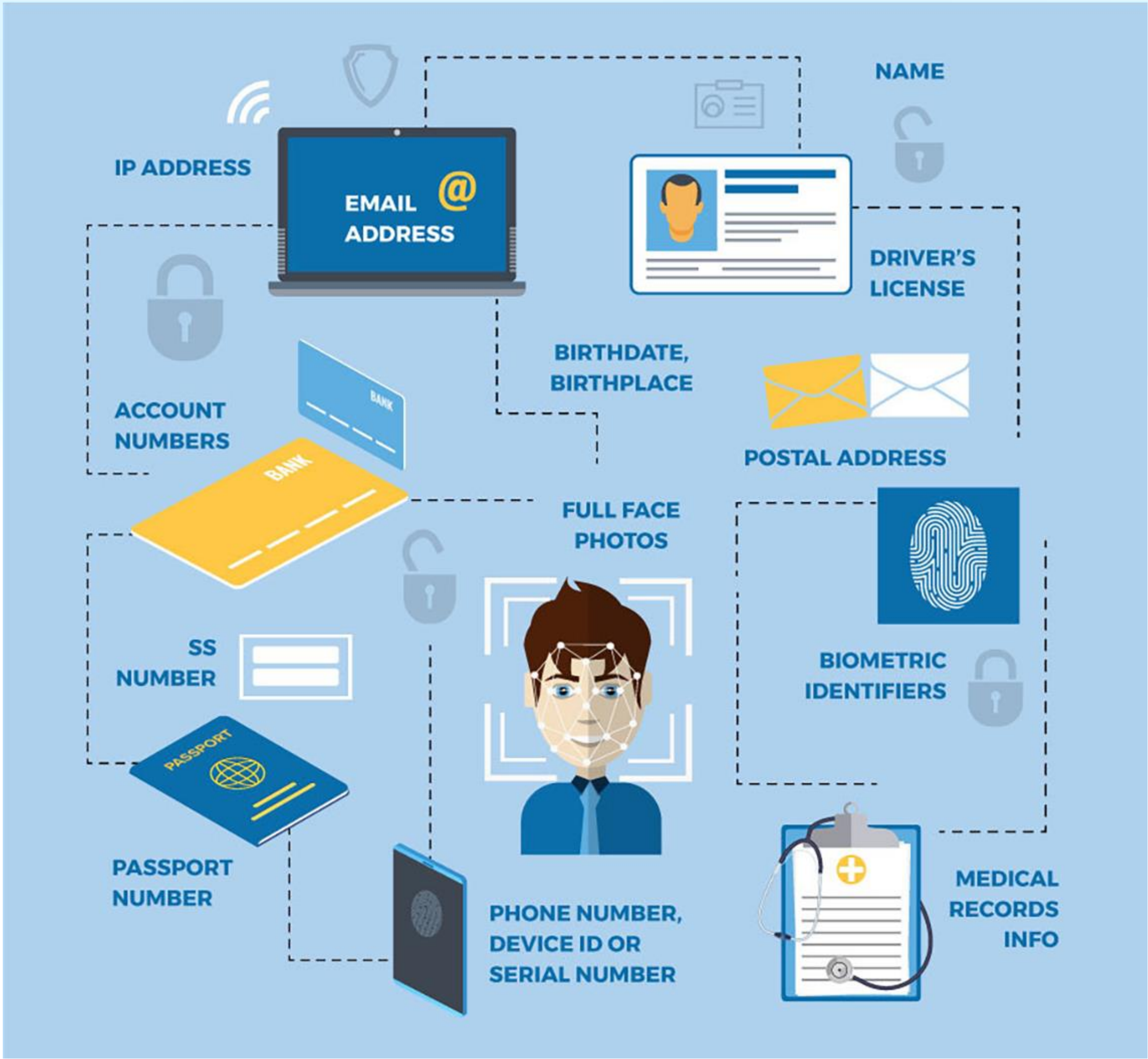
What is PII?

- Personally identifiable information

- Any information that distinguishes one person from another can be used for re-identifying data.

TYPES OF PII

(PERSONALLY IDENTIFIABLE INFORMATION)



Two legal context

California Senate Bill 1386

- Included
  - Social Security numbers
  - Driver’s license numbers
  - Financial accounts
- Not Included
  - Email addresses
  - Telephone numbers
- Focus on the types of data that are commonly used for authenticating an individual

Privacy Act of 1974

- Regulates the collection of personal information by government agencies
- No overarching federal law regulating Private entities

State-level & Worldwide

- California’s Online Privacy Protection Act of 2003
- Personal Information Protection and Electronic Documents Act (PIPEDA, Canada)
- Data Protection Directive (EU)

Privacy laws defined PII in a broader way

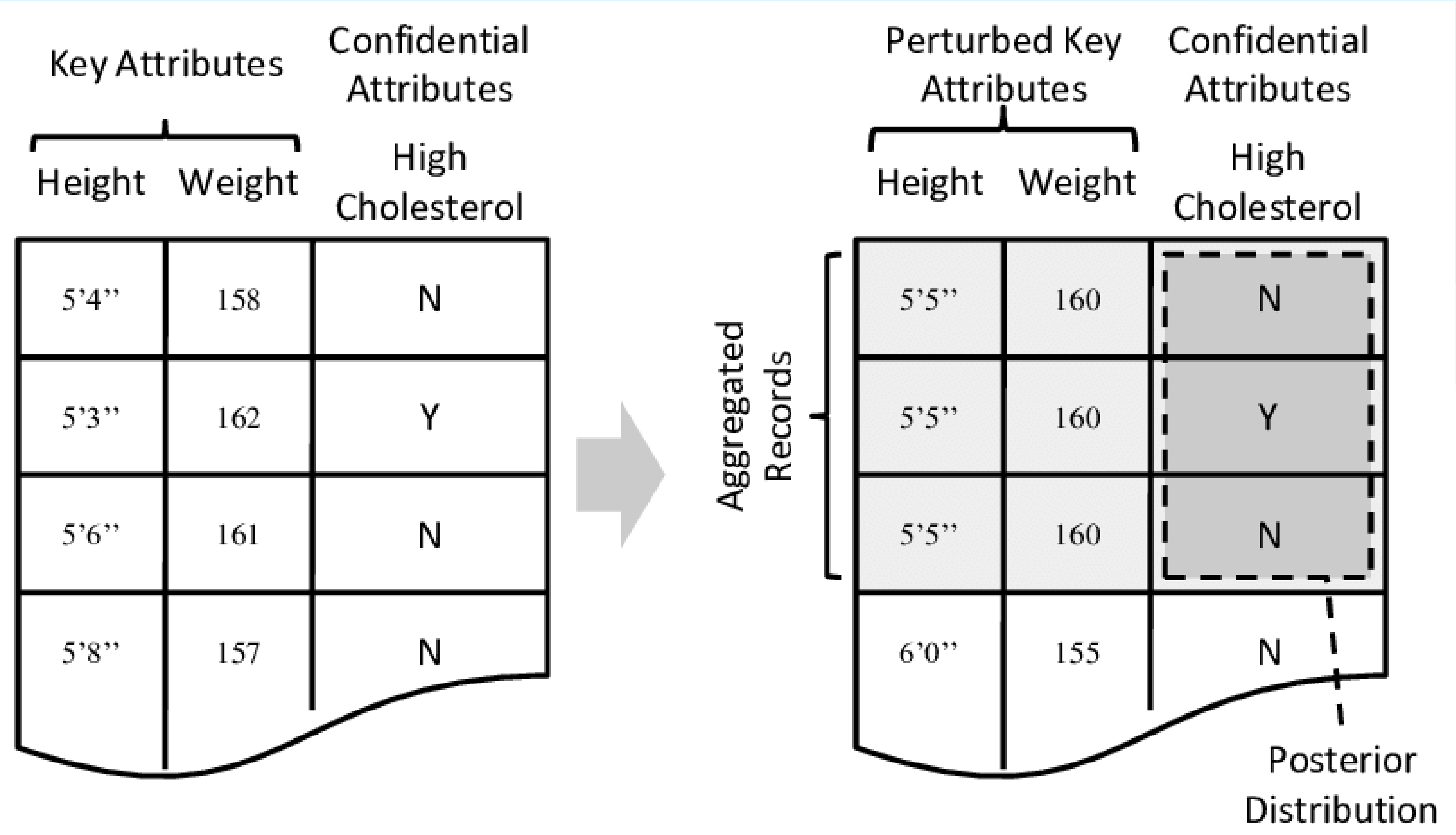
> “any information relating to an [...] natural person [...] who can be identified, directly or indirectly, in particular by reference [...] to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity.”

Data Protection Directive

How to protect privacy?

K-anonymity

- A data is said to have k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k−1 individuals whose information also appear in the release.
- Height and weight are considered as quasi-identifiers here.
- Cholesterol data is considered as non-identifying.



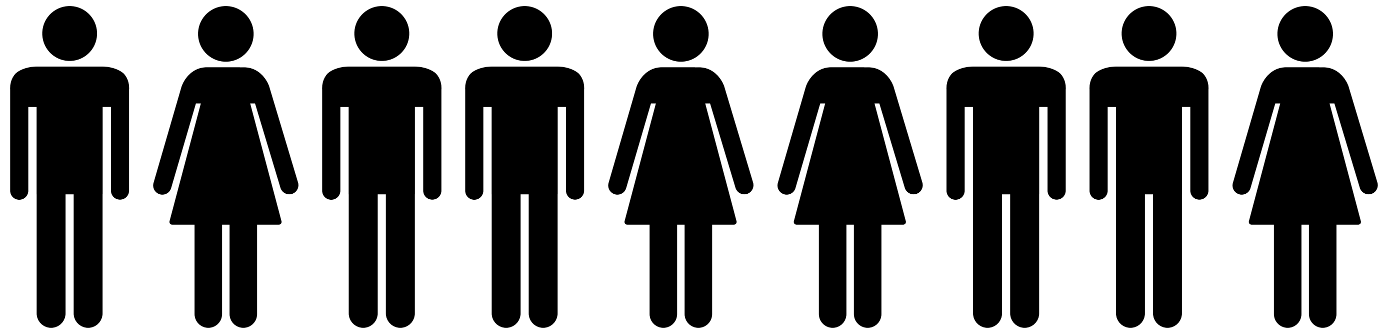
Key Attributes		Confidential Attributes
Height	Weight	High Cholesterol
5’4’’	158	N
5’3’’	162	Y
5’6’’	161	N
5’8’’	157	N

Perturbed Key Attributes		Confidential Attributes
Height	Weight	High Cholesterol
5’5’’	160	N
5’5’’	160	Y
5’5’’	160	N
6’0’’	155	N

Re-identification

- Behavioral or transactional profile
- Location information and stylometry
- Consumption preferences,
- Commercial transactions
- Web browsing
- Search histories

- Reasonably stable across time and contexts
- Corresponding data attributes are sufficiently numerous and fine-grained that no two people are similar



The next step

Privacy Protection

- Privacy protection has to be built and reasoned about on a case-by-case basis
- Interactive, query-based approach is generally superior from the privacy perspective to the “release-and-forget” approach
- Strong access control mechanisms
- Non-technological protection methods such as informed consent and contracts specifying acceptable uses of data