What do scientists, regulators and lawyers mean when they talk about de-identification? How does anonymous data differ from pseudonymous or de-identified information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability.

This is a primer on how to distinguish different categories of data.

**DEGREES OF IDENTIFIABILITY**
Information containing direct and indirect identifiers.

**DIRECT IDENTIFIERS**
Data that identifies a person without additional information or by linking to information in the public domain (e.g., name, SSN)

**INDIRECT IDENTIFIERS**
Data that identifies an individual indirectly. Helps connect pieces of information until an individual can be singled out (e.g., DOB, gender)

**SAFEGUARDS and CONTROLS**
Technical, organizational and legal controls preventing employees, researchers or other third parties from re-identifying individuals

**SELECTED EXAMPLES**
Name, address, phone number, SSN, government-issued ID (e.g., Jane Smith, 123 Main Street, 555-555-5555)
Unique device ID, license plate, medical record number, cookie, IP address (e.g., MAC address 68:4A:6D:35:65:03)
Same as Potentially Identifiable except data are also protected by safeguards and controls (e.g., hashed MAC addresses or legal representations)
Clinical or research datasets where only curator retains key (e.g., Jane Smith, diabetes, Hgb 15.1 g/dl = Csrk123)
Unique, artificial pseudonyms replace direct identifiers (e.g., HIPAA Limited Datasets, John Doe = 5L7T LX619Z (unique sequence not used anywhere else))

**PSEUDONYMOUS DATA**
Information from which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact.

**DE-IDENTIFIED DATA**
Direct and known indirect identifiers have been removed or manipulated to break the linkage to real world identities.

**ANONYMOUS DATA**
Direct and indirect identifiers have been removed or manipulated together with mathematical and technical guarantees to prevent re-identification.

**CONTROLS IN PLACE**
Limited or none in place

**LIMITED or NONE IN PLACE**
Not relevant due to nature of data

**NOT RELEVANT**
Due to high degree of data aggregation

**For example, noise is calibrated to a data set to hide whether an individual is present or not (differential privacy)**

**Very highly aggregated data (e.g., statistical data, census data, or population data that 52.6% of Washington, DC residents are women)**

**FPF.ORG**
Produced by FUTURE OF PRIVACY FORUM FPF.org