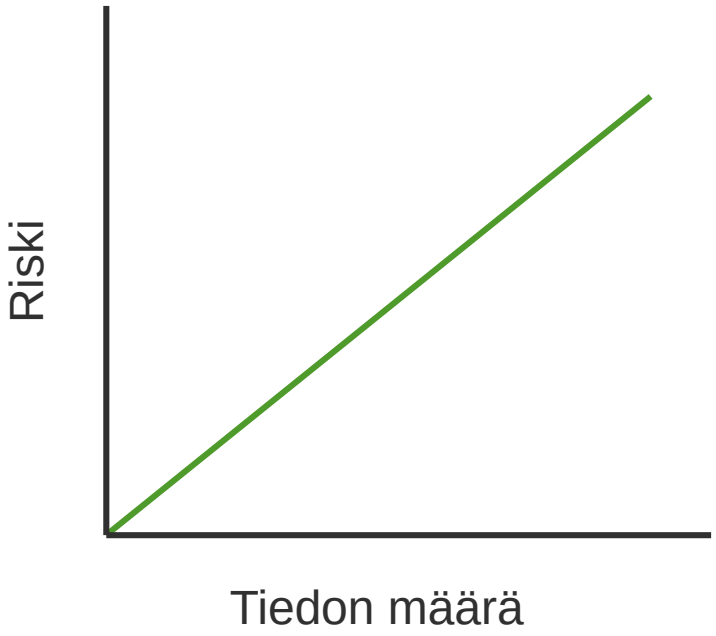
Kommenttipuheenvuoro tietosuojaan 7.5.2018

Christian Jämsén

THL

TERVEYDEN JA HYVINVOINNIN LAITOS

http://howtowriteabusinessplan.com/2016/08/business-plan-template/
Reynermedia CC By 2.0


Riski
Tiedon määrä

Aineistoihin kohdistuvat riskit

- Yksittäiset ja satunnaiset tietomurrot. Ulkopuoliset yksittäiset toimijat, jotka satunnaisesti kokeilevat järjestelmässä olevia heikkouksia ja tilaisuuden ilmaantuessa toteuttavat tietomurron
- Tutkimusryhmän omat jäsenet ja aineiston väärinkäyttö hankkeen sisällä
- Aineiston päätyminen muun ulkopuolisen käsiin ja sen väärinkäyttäminen
- Aineiston rikastaminen muulla tiedolla tunnistettavuuden lisäämiseksi

Insider and privilege misuse

Insider and privilege misuse accounted for 20% of the security incidents suffered by healthcare organizations — an increase of five percentage points since last year.

of incidents in the healthcare sector were attributed to insider and privilege misuse.

This type of incident covers situations where employees and business partners use their legitimate access rights to take confidential

Miscellaneous errors

Miscellaneous errors — security incidents that were the result of accidental actions rather than malicious intent — accounted for almost a fifth (19%) of data breaches last year.

of security incidents in the healthcare sector were due to human error.

The three most common causes of incidents covered by this pattern were:

- Misdelivery (37%) — data delivered to the wrong recipients

THL

Verizon 2015 Data Breach Investigations Report – Healthcare sector

- A specific pitfall is to consider pseudonymised data to be equivalent to anonymised data. The Technical Analysis section will explain that pseudonymised data cannot be equated to anonymised information as they continue to allow an individual data subject to be singled out and linkable across different data sets. Pseudonymity is likely to allow for identifiability, and therefore stays inside the scope of the legal regime of data protection. This is especially relevant in the context of scientific, statistical or historical research.[10]

EXAMPLE:
A typical instance of the misconceptions surrounding pseudonymisation is provided by the well-known "AOL (America On Line) incident". In 2006, a database containing twenty million search keywords for over 650,000 users over a 3-month period was publically released, with the only privacy preserving measure consisting in replacing AOL user ID by a numerical attribute. This led to the public identification and location of some of them. Pseudonymised search engine query strings, especially if coupled with other attributes, such as IP addresses or other client configuration parameters, possess a very high power of identification.

ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN WP216; http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

THL

b) ZIP, Sex, Birth Date

Recall from the Introduction the study by Latanya Sweeney, professor of computer science, who crunched 1990 census data and discovered that 87.1% of people in the United States were uniquely identified by their combined five-digit ZIP code, birth date (including year), and sex.[79] According to her study, even less specific information can often reveal identity, as 53% of American citizens are uniquely identified by their *city*, birth date, and sex, and 18% by their *county*, birth date, and sex.[80]

Ohm, Paul, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization (August 13, 2009). UCLA Law Review, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. Available at SSRN: https://ssrn.com/abstract=1450006

At the time GIC released the data, William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers.[84] In response, then-graduate student Sweeney started hunting for the Governor's hospital records in the GIC data.[85] She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date, only three of them men, and of them, only he lived in his ZIP code.[86] In a theatrical flourish, Dr. Sweeney sent the Governor's health records (which included diagnoses and prescriptions) to his office.[87]

Ohm, Paul, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization (August 13, 2009). UCLA Law Review, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. Available at SSRN: https://ssrn.com/abstract=1450006

THL

1. The Trade-off Dilemma

There is a fundamental tension at the heart of every statistical agency's mission. Each is charged with collecting high quality data to inform national policy and enable statistical research. This goal necessitates dissemination of both summary data and microdata. Each is also charged with protecting the confidentiality of survey respondents—not only because of legal and ethical mandates, but because public trust and perceptions of that trust are important contributors to data quality and response rates.

Protecting confidentiality necessitates perturbing or summarizing the data in some fashion so that the individual respondent cannot be identified. Greater protection of confidentiality means that the data, which cost so much to collect and produce, are likely to become less valuable. The resulting trade-off dilemma, which could well be stated as protecting confidentiality (avoiding disclosure) but maximizing data quality and data access, has become more complex as both technological advances and public perceptions have changed in this Information Age. In sum, while statistical agencies go to great lengths to collect high quality data, the necessity of protecting confidentiality results in some data quality compromises. This book describes new theoretical, practical, and technological responses to the challenges that statistical agencies face.

Doyle et al (2001): Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies

Korkea luottamus ja mahdollisimman suuri hyöty
–
miten se saavutetaan?