HITRUST

Health Information Trust Alliance

De-Identification Framework

Agenda

- Introduction
- What is de-identification?
- Re-identification attacks and other risks to de-identified data
- De-identification methodologies and experts
- Conclusion

HITRUST De-identification Working Group

- The Health Information Trust Alliance (HITRUST) formed the HITRUST De-Identification Working Group in 2012 to propose standards for health data de-identification.
- The intent of the Working Group was to establish a uniform and practical approach to data de-identification that balances the risks and benefits of using the data, while taking into account the advancement of healthcare innovation, increased access to healthcare, and the protection of individual patient privacy.
- The working group suggested qualifications for the professionals who can certify de-identification methods and de-identified data sets.
- The Working Group recommended changes to the Common Security Framework (CSF) to ensure consistency with these recommendations.

Working Group Phases

- **Phase one** defined the multiple levels of anonymization and recommended specific use cases for each variant.
- **Phase two** developed criteria for:
  - Evaluating de-identification methodologies;
  - Estimating re-identification likelihood; and
  - Certifying expertise in these methodologies.
- **Phase three** created a framework for mitigating the risks associated with the use, storage, and maintenance of a data. The controls create a baseline security framework for de-identified data and include controls to mitigate re-identification risks.
- **Phase four** established control requirements based on the HITRUST Common Security Framework for assessing risk related to de-identified data.

WHAT IS DE-IDENTIFICATION

The Anonymization Spectrum

Identifiable

- Privacy Board approved studies
- Public health purposes
- Medical and health plan services

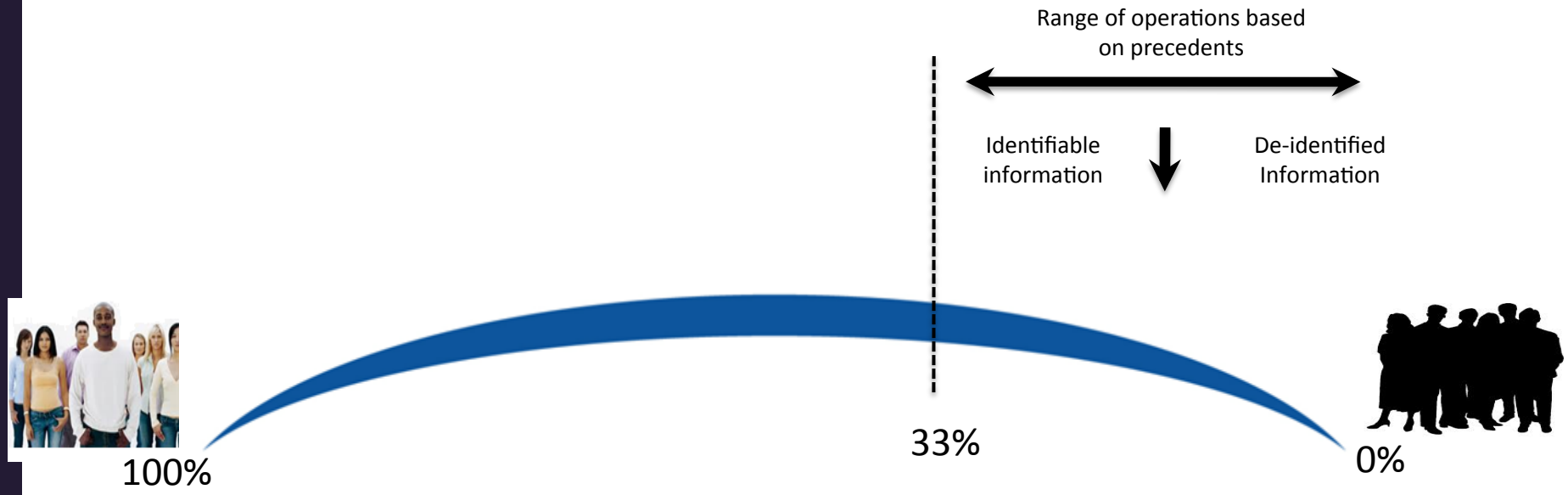
Anonymized

- De-identified data
- Longitudinal de-identified data
- Examples
  - Comparative Effectiveness Research (CER)
  - Health Economics and Outcomes Research (HEOR)
  - Early disease outbreaks detection and geographic tendencies
  - Quality and outcomes research analysis
  - Reduction of medical errors and improved patient safety

Non-identifiable data

- Healthcare cost management/Provider cost comparison data
- De-identified, aggregated analytics regarding member/patient use of health care services
- FDA uses of large population database records to identify risk factors

Spectrum of Identifiability



Identifiability Spectrum

RE-IDENTIFICATION ATTACKS AND OTHER RISK TO DE-IDENTIFIED DATA

Re-identification of Personal Data

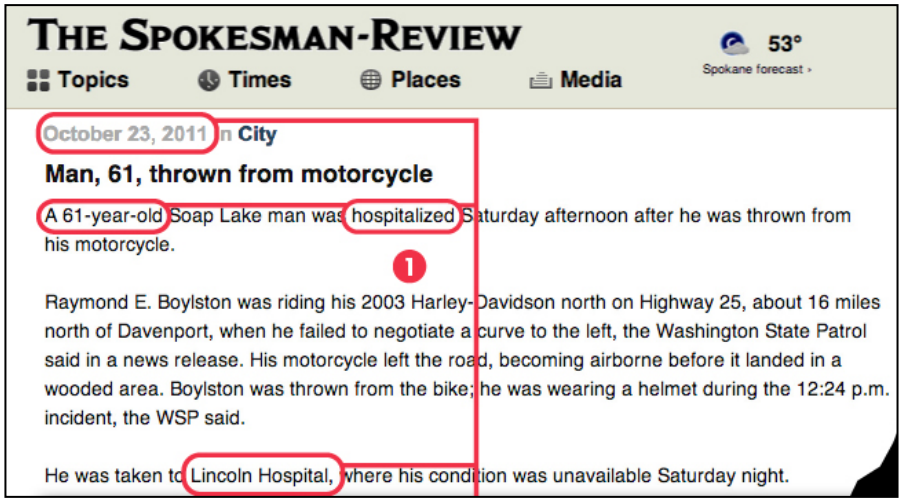
- There are some claims that health data is easy to re-identify
- Increasingly more information exists in public sphere to enable re-identification
- Real-world examples are often used to support that argument

Re-identification Attacks

- A whole discipline is emerging that is focused on re-identification attacks
- The main focus is in demonstrating that either certain data sets have not been anonymized properly or to demonstrate a new technique
- Almost exclusively conducted by academics and the media

State Discharge Database Re-identification

- Information about medical incidents that were published *in newspapers were matched with the White Pages* and the publicly available state hospital discharge database



THE SPOKESMAN-REVIEW

53°
Spokane forecast ›

Topics Times Places Media

October 23, 2011 n City

Man, 61, thrown from motorcycle

A 61-year-old Soap Lake man was hospitalized Saturday afternoon after he was thrown from his motorcycle.

1

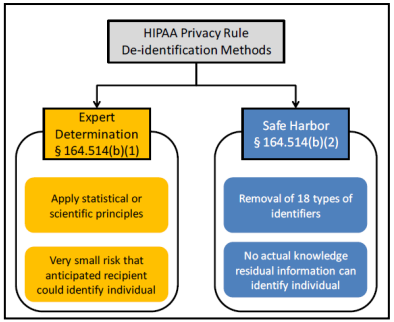
Raymond E. Boylston was riding his 2003 Harley-Davidson north on Highway 25, about 16 miles north of Davenport, when he failed to negotiate a curve to the left, the Washington State Patrol said in a news release. His motorcycle left the road, becoming airborne before it landed in a wooded area. Boylston was thrown from the bike; he was wearing a helmet during the 12:24 p.m. incident, the WSP said.

He was taken to Lincoln Hospital, where his condition was unavailable Saturday night.

DE-IDENTIFICATION METHODOLOGIES AND EXPERTS

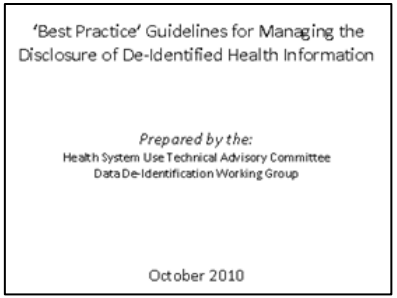
Standards

USA (HIPAA)



UK

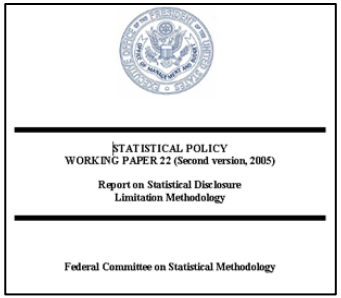
CANADA


'Best Practice' Guidelines for Managing the Disclosure of De-Identified Health Information

Prepared by the:
Health System Use Technical Advisory Committee
Data De-Identification Working Group

October 2010

USA


STATISTICAL POLICY
WORKING PAPER 22 (Second version, 2005)

Report on Statistical Disclosure
Limitation Methodology

Federal Committee on Statistical Methodology

FTC De-identification

- The company must take reasonable steps to ensure that the data is de-identified, that is, the data cannot reasonably be used to infer information about or link to a particular consumer, computer, or device.
- The company publicly commits not to re-identify the data.
- The company requires downstream users of the data to keep it in de-identified form.

Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers, FTC Report March 2012 at 21-22.

HIPAA Safe Harbor Method

Safe Harbor Direct Identifiers and Quasi-identifiers

1. Names
2. ZIP Codes (except first three)
3. All elements of dates (except year)
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images;
18. Any other unique identifying number, characteristic, or code

Expert Determination (Statistical) Method

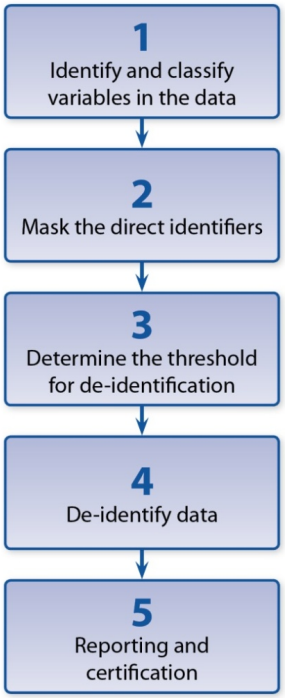
A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

- Applying such principles and methods; determines that the risk is “very small” that the information could be used, alone or in combination with other reasonably available information by an anticipated recipient to identify an individual who is a subject of the information; and
- Documents the methods and results of the analysis that justify such determination

Who is an Expert?

- No specific professional degree or certification program required.
- Recommended experts include statisticians, mathematicians, or other professions with a scientific background.
- An expert should have:
  - A well defined de-identification methodology
  - Been trained on risk management
  - Been educated on de-identification methods
  - Relevant professional and academic experience
  - The ability to measure or quantify the risk of re-identification
  - Actual experience of the expert using health information de-identification technologies
- Determinations should be unbiased and based on objective standards – consider using an external expert.

Anonymization Process

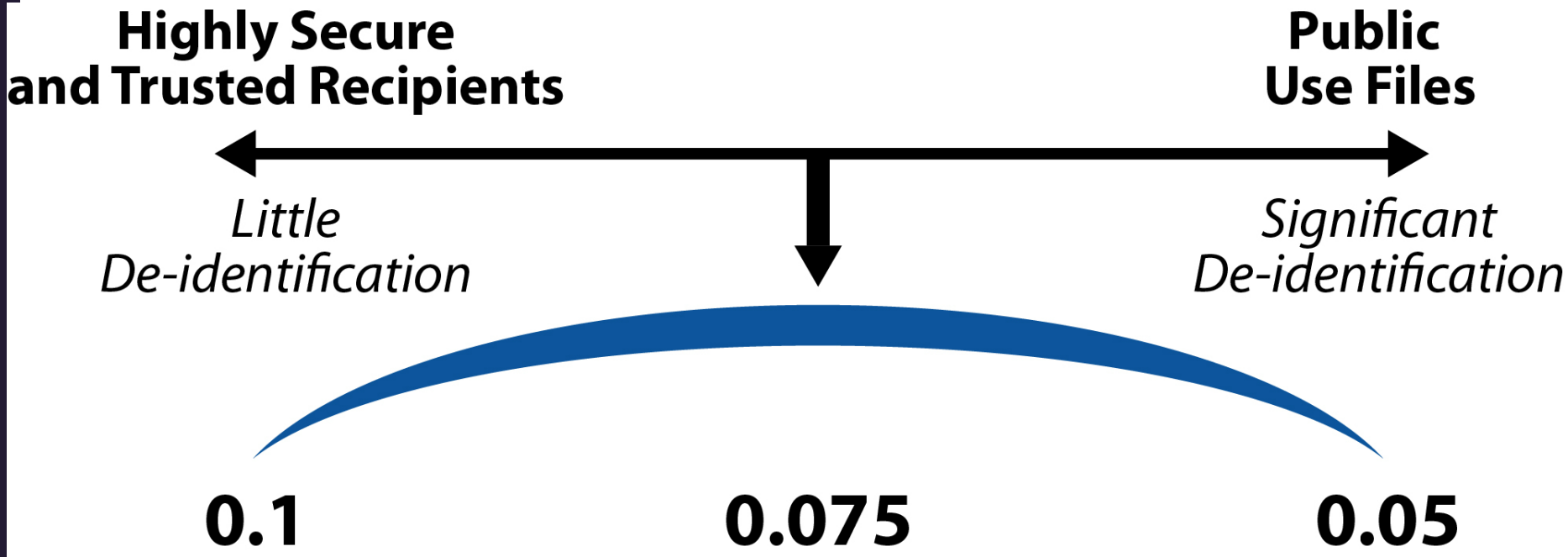


Direct and Indirect Identifiers

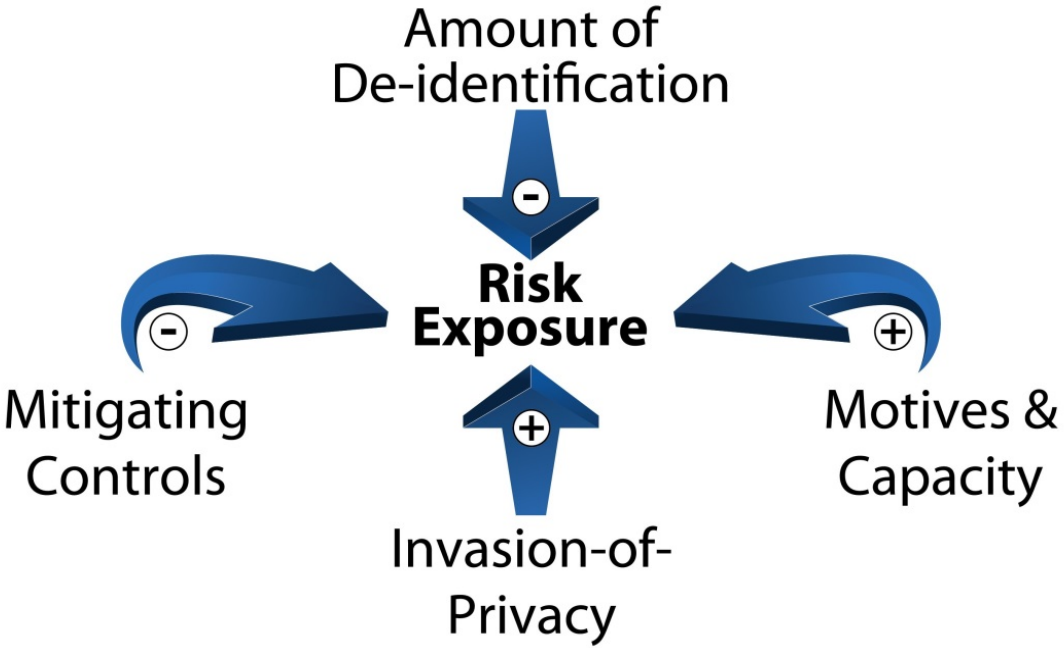
Examples of direct identifiers: Name, address, telephone number, fax number, MRN, health card number, health plan beneficiary number, license plate number, email address, photograph, biometrics, SSN, device number, clinical trial record number

Examples of quasi identifiers: sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, total years of schooling, marital status, criminal history, total income, visible minority status, profession, event dates, number of children, high level diagnoses and procedures

Re-identification Risk Spectrum



Managing Re-identification Risk



Recommendations

- Taking into account the size, complexity, and capabilities of your company and the amount of de-identification activities, establish a de-identification program, including:
  - Governance
  - Documentation
  - Explicit identification of the data custodian and recipients
  - External or independent scrutiny
  - Explicit agreement by recipients that re-identification will not be attempted.

Recommendations II

- Establish a de-identification methodology, that takes into account:
  - Re-Identification Risk Thresholds
  - Measurement Of Actual Re-Identification Risks
  - Identification And Management Of Direct Identifiers And Quasi-Identifiers;
  - Identification Of Plausible Adversaries And Attacks;
  - Identification Of Specific Data Transformation Methods And How They Reduce The Risks
  - Process And Template For The Implementation Of Re-Identification Risk Assessment And De-Identification
  - Mitigating Controls To Manage Residual Risk
  - Data Utility

CONCLUSION

HITRUST

Health Information Trust Alliance

Visit www.HITRUSTAlliance.net for more information

To view our latest documents, visit the Content Spotlight