Workshop

ANONYMISATION TOOLS AND THEIR PRACTICAL RELEVANCE

March 19, 2015

LEGAL BACKGROUND:

THE ROLE OF ANONYMISATION FOR DATA PROTECTION

Irene Schlünder, TMF

OUTLINE

1. LEGAL FRAMEWORK
2. LEGAL REASONS TO ANONYMISE
3. THE TERM „ANONYMISATION“ FROM THE LEGAL PERSPECTIVE
4. SELECTED KEY TOPICS

1. LEGAL FRAMEWORK

NO HIPAA List!

1. LEGAL FRAMEWORK

1. Names.
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census:
   a. The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people.
   b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers.
5. Facsimile 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 fingerprints and voiceprints.
17. Full-face photographic images and any comparable images.
18. **Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification**.

1. LEGAL FRAMEWORK

Removing all 18 identifiers leads to **de-identified data,** not to **anonymous** data!

The list makes only sense within the context of HIPAA and cannot be transferred into the European legal framework.

1. LEGAL FRAMEWORK

Common Europe-wide Legal Basis:

Directive 46/95 EC, Oct. 1995 "on the protection of individuals with regard to the processing of personal data and on the free movement of such data" (DPDir)

- Overall aim and objective:
  - protecting individuals (human right to privacy)
  - not hindering information exchange as such
  - (other) individuals might have conflicting rights: right to be informed, freedom of the press and of scientific research

- Impact on national legislation:
  - Leaves leeway for national legislation and interpretaion
  - Remains framework: national courts can seek advice from the European Court of Justice

1. LEGAL FRAMEWORK

The role of the Art. 29 Working Party

- The Art. 29 WP is composed of representatives of all national data protection supervising agencies (presidency France)

- Remit and most important activity: "…make recommendation on all matters relating to the protection of persons with regard to the processing of data within the Community"

- Documents on anonymisation: 2 basic "opinions":
  - "concept of personal data" (04/2007, WP 136)
    http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf
  - "anonymisation techniques" (05/2014, WP 216)
    http://www.cnpd.public.lu/de/publications/groupe-art29/wp216_en.pdf)

1. LEGAL FRAMEWORK

Relevance for the implementation of the DPDir:

- It is a general opinion, not deciding specific cases
- Due to the composition of the group the text is not free of contradictory statements
- The European Court of Justice is not bound by the opinion
- Nontheless, in absence of other sources of softlaw, national data protection supervisors and courts will take the opinions into account

2. LEGAL REASONS TO ANONYMISE

There are 2 legal reasons to anonymize:

- Dichotomy of data protection law: anonymous versus personal data:
  - Only personal data is protected by law
  - Anonymous data: no consent or other legal basis needed for processing (Rec. 26: "the principles of protection shall not apply to data rendered anonymous")

- Principle of data minimization (Art. 6 (c))
  - Personal data have to be de-identified as soon as possible to the extent that the research purpose is not defeated or considerably endangered
  - Anonymisation must be done by taking into account
    - the research purpose and
    - the donor's rights

2. LEGAL REASONS TO ANONYMISE

Not explicitly mentioned in the DPDir:

Side-effects (adverse events) of anonymisation

- Full (unlinked) anonymisation deprives the donor of the possibility to use their right to withdraw consent

  critical for biosamples

  or other cases of weak anonymisation

- It makes feeding back research results or incidental findings impossible

- It is not useful in cases where research is linked to treatment (oncology: personalised medicine)

3. ANONYMISATION AS LEGAL TERM

Definition of the term "anonymisation"

- Art. 2 (a) DPDir: personal data shall mean
  - any information
  - related to a natural person, who is
    - identified or
    - identifiable
      - directly (What is then the difference with regard to identified?) or
      - indirectly by reference to
        - Identification number
        - One or more factors specific to his physical, physiological, mental, economic, cultural or social identity

- counter term: anonymous data (= data that is not linkable to a natural person, so that the person is not identifiable)

3. ANONYMISATION AS LEGAL TERM

- "identifiable":

  "…to determine, whether a person is identifiable, account should be taken of **all the means likely reasonably to be used** either by the controller or by any other person to identify the said person" (Rec. 26 DPDir)

2 possible interpretations:

- WP29 opinion on anonymisation techniques:

  The underlying rationale of the DPDir is that „…the data must be stripped of sufficient elements such that the data subject can no longer be identified…An important factor is that the processing must be **irreversible**…the outcome of anonymisation as a technique applied to personal data should be **as permanent as erasure.**"

- Absolute anonymity for now and forever is not achievable

- Widely accepted approach: **De facto anonymity** is sufficient

  -> What does this mean?

  - All the means -> any possible/conceivable attacks
  - Used by the controller or by any other person
  - Likely reasonably to be used -> what criteria have to be taken into account? (see WP29 opinion concept of personal data)
    - Effort, money -> context of information, technical safeguards
    - expected advantage -> aim of potential attacker

  -> The term anonymity is **not static**:

3. ANONYMISATION AS LEGAL TERM

- the same information (data set) can be anonymous in one **context** and personal data in another!
  - Example: Even the full name (e.g. “Harry Smith”) might not be an identifier without additional information, whereas “Irene” is sufficient to be identified by my family or colleagues without adding my family name (the name itself may even not be necessary to identify an individual -> profiling!)

- Taking into account the **aim** of potential attackers: some safeguards may be appropriate to anonymise rather uninteresting data or data having a low impact on the privacy of persons involved, but completely insufficient regarding highly sensitive data
  - Example: counting people on the street versus counting famous patients with rare diseases

3. ANONYMISATION AS LEGAL TERM

- This leads to the question, if **access policies** as additional safeguards can have an impact on the status of anonymity
  - Research Data Alliance https://rd-alliance.org/,
  - EC policy on open science data ttp://ec.europa.eu/research/swafs/index.cfm?pg=policy&lib=science,
  - NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources http://grants.nih.gov/grants/sharing.htm,
  - Wellcome Trust Data Sharing Policy http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/,
  - MRC Data Sharing Policy http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/
- What about **contractual obligations** not to try to identify individuals?

4. SELECTED KEY TOPICS

Selected Key Topics:

"Relative" anonymity:

- in the UK "linked-anonymised" or "pseudonymised" data may be treated as "anonymous" where the data controller does not have access to the linkage key.
- The "Opinion on Anonymisation Techniques" of the Article 29 Data Protection Working Party leaves the question open.
- In Germany the Federal Court of Justice submitted the issue to the European Court of Justice in October 2014. (http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh& Art=pm& Datum=2014&Sort=3&nr=69184&pos=0&anz=152, in German).

4. SELECTED KEY TOPICS

Anonymisation of genetic data?

- DNA sequences alone do not disclose the identity of an individual
- But it can be enough information to single out a person
- Sharing detailed genetic data unique to one person (whole genome sequence data) increases the risk of re-identification.
- How much genetic information is sufficient to single out a person?
- How much additional information is needed to identify the donor?

Opinion on Anonymisation Techniques WP 29:

"Genetic data profiles are an example of personal data that can be at risk of identification if the sole technique used is the removal of the identity of the donor due to the unique nature of certain profiles. It has already been shown in the literature that the combination of publically available genetic resources (e.g. genealogy registers, obituary, results of search engine queries) and the metadata about DNA donors (time of donation, age, place of residence) can reveal the identity of certain individuals even if that DNA was donated 'anonymously'".

4. SELECTED KEY TOPICS

Biosamples "are" versus "contain" data?

Opinion WP29 concept of personal data is contradictory:

- Human tissue samples are themselves sources out of which biometric data are extracted, but they are not biometric data themselves
- Personal data includes information **available** in whatever form (including voice)

Next Generation sequencing

Dependent on context? -> available to whom?

Thank you very much for your attention!

irene.schluender@tmf-ev.de

3. ANONYMISATION AS LEGAL TERM

Anonymous versus anonymised?

Van't Noordende: Microdata: data or records that belong to a single individual, for example, a table with columns where each row conatins attributes that belong to an individual -> The background information that is available for re-identification is increasing over time.

Opinion on Anonymisation Techniques WP 29:

"…when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this data set (for example after removal or masking of identifiable data), the resulting dataset is still personal data. Only if the data controller would aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous."