Advanced Survey Statistics: Disclosure Control

Part 3: Rechtliches

Matthias Templ

Institut für Datenanalyse und Prozessdesign
School of Engineering
Zürcher Hochschule für Angewandte Wissenschaften

FU-Berlin, 2019

Zürcher Hochschule
für Angewandte Wissenschaften

zh aw School of Engineering
IDP Institut für Datenanalyse und Prozessdesign

Überblick

Im Folgenden steht die rechtliche Sitation im Fokus. Dabei ist bei folgenden Themen eher die praktische Auslegung als die juristische Feinheit im Visier.

- Bearbeitungsgrundsätze
- DSGVO
- Datenschutz-Folgeabschätzung
- Natürliche vs. juristische Personen
- Wird Datenschutz auch umgesetzt?
- Begriffsbestimmungen (Bestimmbarkeit, De-facto Anonymität)
- Sensibilität
- Datensicherungsmassnahmen
- Was bedeuted der Begriff Anonymisierung

Ist eine Pseudo-Anonymisierung ausreichend?

Oft wird Pseudo-Anonymisierung mit Anonymisierung verwechselt.

- Die Massachusetts Group Insurance Commission (GIC) hatte Mitte der neunziger Jahre eine exzellente Idee - sie beschloss, „anonymisierte“ Daten über staatliche Angestellte zu veröffentlichen, die jeden einzelnen Krankenhausbesuch zeigten.
- Zu dem Zeitpunkt als GIC die Daten veröffentlichte versicherte William Weld - der damalige Gouverneur von Massachusetts - der Öffentlichkeit, dass die GIC die Privatsphäre der Patienten absolut geschützt hat, indem sie Kennungen (direkte Indentifizierungsvariablen) gelöscht habe.

Ist eine Pseudo-Anonymisierung ausreichend?

- Eine Studentin suchte in GIC-Daten nach den Krankenhausakten des Gouverneurs. Sie wusste, dass Gouverneur Weld in Cambridge, Massachusetts, einer Stadt mit 54.000 Einwohnern und sieben Postleitzahlen, residierte.
- Für zwanzig Dollar kaufte sie die vollständigen Wählerlisten aus der Stadt Cambridge, eine Datenbank, die unter anderem den Namen, die Adresse, die Postleitzahl, das Geburtsdatum und das Geschlecht jedes Wählers enthielt.
- Durch einfaches Zusammenführen dieser Daten mit den GIC-Daten fand Sweeney Gouverneur Weld mit Leichtigkeit obwohl der Datensatz pseudoanonymisiert wurde.
- Die Studentin schickte die Gesundheitsakten des Gouverneurs (die Diagnosen und Rezepte enthielten) filmreif in sein Büro.

Folge: seit über 20 Jahren weiss der Gesetzgeber Bescheid.

Bearbeitungsgrundsätze (1/3)

Unter anderem hat der Datenhalter auch zu achten, dass die technischen (Datenbanken, IT-Sicherheit) und organisatorischen (wer hat Zugriff auf welche Daten und wie erfolgt der Zugriff) Massnahmen (bereits) umgesetzt sind.

Die generellen Prinzipien (Bearbeitungsgrundsätze) des Datenschutzes - so wie er zur Zeit gelebt wird - lauten:

1. Grundsatz der Rechtmässigkeit. Dies bedeutet, dass sie nicht durch Drohung oder Täuschung oder ohne das Wissen der Betroffenen beschafft werden dürfen.
2. Grundsatz von Treu und Glauben (loyales und vertrauenswürdiges Verhalten im Rechtsverkehr)

Bearbeitungsgrundsätze (2/3)

3. Grundsatz der Verhältnismässigkeit. Im einzelnen Fall zwar so viele Daten wie nötig erheben/bearbeiten, gleichzeitig aber so wenige wie möglich zu bearbeiten sind $\rightarrow$ Grundsatz der Datenminimierung leitet sich davon ab (Artikel 5 der DSGVO)
4. Grundsatz der Zweckbindung, d.h. Zweck der Bearbeitung der Daten müssen für die betroffene Person erkennbar sein.
5. Grundsatz der Transparenz bzw. der Erkennbarkeit. Für betroffene/erhobene Personen muss unter normalen Umständen erkennbar sein muss, dass Daten, die sie betreffen, beschafft wurden oder möglicherweise beschafft werden.

Bearbeitungsgrundsätze (3/3)

6. Grundsatz der Datenrichtigkeit und der Datensicherheit (angemessene technische und organisatorische Massnahmen gegen unbefugtes Bearbeiten). Dies bedeutet zum Beispiel, dass nur berechtigte Personen auf eine Datenbank Zugriff haben dürfen.

Der Datenhalter hat deshalb Personen zu informieren, was mit den Daten geschieht und begründen warum und welche Daten sie sammelt.
Zentral für diese Vorlesung sind aber nur indirekt Teile des Punktes 6.

U.S. Health Insurance Portability and Accountability Act

Datensätze gelten als anonym, wenn die folgenden Identifyer gelöscht oder aggregiert werden:

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: (i) The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people. 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 >89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a category of age 90+

U.S. Health Insurance Portability and Accountability Act

4. Telephone numbers and
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.

U.S. Health Insurance Portability and Accountability Act

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.

Punkte 1-17 sind trivial bzgl. Datenschutz: direkte Identifizierungsvariablen löschen oder pseudoanonymisieren.

Punkt 18 enthält eine wesentliche Erweiterung, die für Daten und Geheimhaltungskonzept von Daten ganz entscheidend ist: *"unless otherwise permitted by the Privacy Rule for re-identification"*.

$\rightarrow$ Theorie und Methoden der Statistischen Geheimhaltung sind zu berücksichtigen und anzuwenden.

Die neue Datenschutzgrundverordnung (DSGVO) der EU

- Trat am 25.05.2018 in Kraft und ist in allen Mitgliedstaaten der EU direkt anwendbares Recht. Nationale Gesetze dürfen in der EU dem DSGVO daher nicht widersprechen.
- Sie bringt bringt keine neuen Gesichtspunkte gegenüber der vormals geltenden Richtlinie **im Hinblick auf** den Begriff der **Anonymisierung** und sie ist auch für die gesamte EU richtungsgebend
- Der freie Datenverkehr wird ausdrücklich unterstützt, d.h. ausschliesslich dem Schutz von natürlichen (und zZ noch nach Schweizer Recht auch juristischen) Personen.

Zentrale Ziele der DSGVO

- Schutz von Individuen (Informationelle Selbstbestimmung)
- Freier Datenverkehr
- Ausgleich konfligierender Rechte: Pressefreiheit, Forschungsfreiheit etc. gegenüber Informationeller Selbstbestimmung

Dem freien Datenverkehr steht also sowohl innerhalb der Helsana als auch für „extern“ nichts im Wege, d.h. ausschliesslich dem Schutz von natürlichen und juristischen Personen.

Wesentliche Neuerungen in der DSGVO

- *Privacy by design* (Artikel 25 und 32 auch für Auftragsdatenverarbeiter). Dies beinhaltet den **Grundsatz der Datenvermeidung und Datensparsamkeit**, ergo so wenig Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen.

Dies bedeutet, dass ein Datenhalter sämtliche Verfahren zu überprüfen hat, ob Anwendungen z.B. nicht mit weniger datenschutzrechtlicher „Reibung“ funktionieren könnten.

Insbesondere auch ob durch Datenanonymisierung Anwendungen (z.B. Datenanalysen) trotzdem noch im gleichen Aussmass möglich sind.

Wesentliche Neuerungen in der DSGVO

- Rechenschaftspflicht durch Verzeichnis der Verarbeitungstätigkeiten. Hier sei einem Datenhalter empfohlen, alle Datenbanken intern an zentraler Stelle aufzulisten, als auch festzuhalten, welche Abteilung/Stelle und Person welche Verarbeitung von welchen Daten durchführt. Ebenfalls Inhalt der DSGVO sind
- Data Protection Impact Assessments / Folgeabschätzung
- Mandatory Security Breach Reporting. Dies erfordert, dass Helsana, bei einem Datenleck seine Kunden und andere Parteien informiert und Schritte zur Behebung unternimmt.
- Bussgelder in Artikel 83 (bis zu 20 Mio oder bis zu 4% des weltweiten Jahresumsatzes)
- Betroffenenrechte: Recht auf Datenübertragbarkeit, Recht auf Vergessen-Werden, Recht auf Auskunft der Datenverarbeitung.

Folgeabschätzung

1. eine systematische Beschreibung der geplanten Verarbeitungsvorgänge und die Zwecke der Verarbeitung ...
2. Zwecke. Eine Beurteilung der Notwendigkeit und Verhältnismässigkeit der Verarbeitung
3. eine Bewertung der **Risiken** bzbl. der Daten; Herkunft, Art, Besonderheit und Schwere solcher Risiken und
4. die geplanten **Massnahmen** zur Bewältigung der Risiken, einschließlich Sicherheitsmaßnahmen und Mechanismen, um den persönlichen Schutz zu gewährleisten Daten.

Der 3. und 4. Punkt ist ganz zentral in diesem Projekt. In diesem Projekt werden Risiken quantifiziert und Massnahmen getroffen (Anonymisierungen von Daten).

Natürliche oder juristische Personen

Laut jetziger Rechtssprechung bzgl DSVGO sind juristische Personen vom Datenschutz ausgenommen.

Ausnahmen:

- juristische Person ist Einpersonenunternehmen
- es darf keinen Rückschluss (Re-Identifizierbarkeit) auf einzelne Beschäftige in einem Unternehmen geben.

Wir wissen noch nicht ob und wie weit dieser Punkt Ende 2019 von der Schweiz übernommen wird.
(Deshalb auch zusätzlich eine Anonymisierung der ambulanten und stationäre Leistungserbringer in diesem Projekt)

Datenschutz: Beispiel Umsetzung in Schweizer Unternehmen

- Eine am 3.10.2018 erschienen Studie zufolge, beurteilen der Grossteil von 265 befragten Unternehmen den Datenschutz als wichtig oder eher wichtig.
- Die aktuellen Gesetze sind dieser Studie zufolge etwa der Hälfte der Unternehmer bekannt, im Gegensatz zur neuen DSGVO, die noch wenig bekannt ist.
- Nur rund ein Viertel der Unternehmen gehen davon aus von der Verordnung der EU betroffen zu sein, was im klaren Widerspruch zur Einschätzung von Experten, Anwälten und Beratern steht.
- Die für den Datenschutz eingesetzten Ressourcen in den Unternehmen sind sehr beschränkt und zumeist Null.
- Selten sind auch formalisierte Abläufe und durchgeführte Schulungen.

Nicht erwähnte Punkte in der DSVGO

- Nebenwirkungen der Anonymisierung ist das Recht des Daten"spenders", seine Einwilligung zu widerrufen. Dies wird durch anonymisierte Daten im Wesentlichen vereitelt.
- Bei erfolgreicher Anonymisierung der Daten kann kein Personenbezug per Definition mehr hergestellt werden. Somit kann ein Datenspender für anonymisierte Datensätze seine Einwilligung gar nicht mehr widerrufen, da seine personenbezogene Information eben nicht mehr (re-)identifizierbar ist.

Bemerkungen

- Anonymisierte Daten bieten oft die falsche Vorstellung, dass Daten ohne Sicherheitsmassnahmen ausgetauscht werden können. Es gilt jedoch: je mehr Daten geteilt und verknüpft werden, desto gröser das Risiko einer Identifizierung
- Grundsätzlich kritisch zu sehen sind Fälle von enger Verzahnung von Behandlung und Forschung in der Biostatistik und Life-Sciences (z.B. Krebsbehandlung, „personalised medicine“).
- Bsp. Therapie und Merkmale des Versicherten sind im Datensatz zu finden, wenn auch möglicherweise nicht nur einem Patienten zuzuordnen, sondern mehreren (Verweis: *l*-diversity, später)

Personenbezogene Daten

Art. 4 DSGVO : "personenbezogene Daten“ sind alle Informationen

- über eine identifizierte (bestimmte)
- oder bestimmbare natürliche Person („betroffene Person“)
- als bestimmbar wird eine Person angesehen, die direkt oder indirekt identifiziert werden kann, insbesondere durch Zuordnung zu einer Kennnummer oder zu einem oder mehreren spezifischen Elementen, die Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind Gegenbegriff: anonyme Daten (= Daten ohne Personenbezug)

Zur Auslegung des Begriffs „Bestimmbarkeit“:

- „. . . bei der Entscheidung, ob eine Person bestimmbar ist, sollten alle Mittel berücksichtigt werden, die vernünftigerweise entweder von dem Verantwortlichen für die Verarbeitung oder von einem Dritten eingesetzt werden könnten, um die betreffende Person zu bestimmen.“ (Rec. 26 Deutsche Stiftung für Recht und Informatik (DSRI))
- 2 mögliche Interpretationen: Absolute Anonymität (ein für alle mal und unumkehrbar)
- Die Artikel-29-Datenschutzgruppe, engl. Article 29 Data Protection Working Party, teilw. abgekürzt (G29) Opinion on anonymisation techniques: „. . . outcome of anonymisation as a technique applied to personal data should be as permanent as erasure.“

De-facto Anonymität

- Wenn der **Aufwand höher** ist Daten zu re-identifizieren **als der Nutzen** wird von de-facto Anonymität gesprochen.

Von der G29 an anderer Stelle und ganz überwiegend wird folgendes vertreten (wie auch im Schweizer DSG):

- alle verfügbaren Mittel sind zu berücksichtigen, d.h. alle möglichen oder vorstellbaren Angriffe müssen in Betracht gezogen werden
- die vom Verantwortlichen oder von Dritten ausgehen können, d.h. Sicherheitsmaßnahmen müssen die Möglichkeit der Re-identifizierung sowohl eigener Angestellter als auch Externer im Blick haben
- Mittel müssen vernünftigerweise eingesetzt werden können. Welche Kriterien sind zu berücksichtigen? (s. WP29 „Opinion on concept of personal data“)

De-facto Anonymität, Bemerkung

- Eine (rigide) Anonymisierung kann Daten für die Analyse nutzlos machen. Es muss immer darauf geachtet werden, dass die Datenqualität durch Anonymisierung der Daten nicht zu sehr leidet.
- Der Begriff der Anonymität ist nicht statisch: dieselben Daten können in einem Kontext anonym sein und in einem anderen personenbezogen!
  - Beispiel: Voller Name (z.B. Urs Meier) im Census versus voller Name innerhalb einer Uni-Abteilung.
- Ziele, Sensibilität der Daten und technische Massnahmen: einfache Maßnahmen mögen ausreichen, um eher uninteressante Daten zu anonymisieren, während sie für hochsensible Daten unzureichend sind.

Sensibilität der Daten berücksichtigen (DSG, 1. Abschnitt) und vertragliche Verbote

Organisatorische Sicherheitsmaßnahmen: üblich sind inzwischen Zugangsbeschränkungen zu Forschungsdaten, auch wenn sie derzeit als (noch) nicht re-identifizierbar eingestuft werden können:

- Research Data Alliance https://rd-alliance.org/,
- EC policy on open science data http://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/Aboutus/Policy/Spotlight-issues/Data-sharing/,
- MRC Data Sharing Policy http://www.mrc.ac.uk/research/research-policy-ethics/datasharing/policy/

Sensibilität der Daten berücksichtigen (DSG, 1. Abschnitt) und vertragliche Verbote

Gesetzgeber:

- Vertragliche Verbote der Re-identifizierung sind nur ein schwaches Mittel und allein unzureichend, zumal wenn die Einhaltung nicht wirksam kontrolliert wird/werden kann.

Persönliche Erfahrung:

- Vertragliche Nutzungsbedingungen sind oft sehr wirksam, vor allem wenn z.B. Forschungsinstitutionen langfristig Daten beziehen. Würden sie nicht penibel die Verträge erfüllen, würden sie keine Daten mehr bekommen.

Abgrenzung zur Pseudonymisierung

- Bei der Pseudonymisierung wird der Name oder ein anderes Identifikationsmerkmal durch ein Pseudonym (zumeist eine mehrstellige Buchstaben- oder Zahlenkombination, auch (*hash*)Code genannt) ersetzt, um die Identifizierung des Betroffenen auszuschließen oder wesentlich zu erschweren. Methodisch: kommt später.
- Bei Pseudoanonymisierung kann der "Code" zur De-identifizierung an sicherer Stelle aufbewahrt werden, d.h. getrennt von den Daten aufbewahrt. Technischen und organisatorische Maßnahmen müssen getroffen werden.
- Aber: Pseudoanonymisierung ist trivial und löst das Datenschutzproblem nicht. Dies wurde bereits am Beispiel des Governeurs William Weld aufgezeigt.
- Trotz dieser Beispiele fehlt oft das Bewusstsein dafür.

Bemerkungen (1/2)

Häufig werden Methoden und Sicherheitsbegriffe durcheinander gebracht:

- Die Datenbank ist anonymisiert vs. es wurden Namen gestrichen
- Die Verbindung ist verschlüsselt vs. die Verbindung ist „sicher“

Eine Anonymisierung ist immer anwendungsspezifisch (im Ggs. zu Verschlüsselung). Anonymisierungsverfahren sind anhand eines Algorithmus / einer Regel auszuführen, um Transparenz und Nachvollziehbarkeit zu garantieren

Bemerkungen (2/2)

- Weitere relevante Punkte betreffen die wiederholte Freigabe des (fast) gleichen Datensatzes oder Methoden für hochdimensionale Daten. - Datenanonymisierung ist ein komplexes Problem und keines der Verfahren für sich genommen garantiert die Entfernung des Personenbezugs im rechtlichen Sinne.
- Deshalb sind neben der Anonymisierung zusätzliche technische und organisatorisch Massnahmen zu betrachten.

Gerade das neue Recht „to be forgotten“ in der DSVGO erfordert dazu mehr Anstrengungen und Safeguards (data protection by design and default, Article 23) als bis dato.

Einige relevante Datensicherheitsmassnahmen

Einige IT- und Vertrags-lastige Datensicherheitsmassnahmen werden oft aufgelistet.

- Verschlüsselte Speicherung, Kommunikation und Suche
- Privacy Preserving Computation (Mining as well)
- Policy-Enforcement (Nutzen von Data Use Agreements, die Linkage, Sharing und Re-Identifizierungsversuche verbieten)
- Data Usage Control

Vorteile von Data Use Agreements: Nutzungsverträge auch mit externen Datenempfängern sollen explizit darauf hinweisen, dass die Daten nur für den Zweck der Untersuchung verwendet werden und die Zusammenführung mit anderen Datenbeständen - sofern nicht vorab geregelt - untersagt ist.

Anonymisierung

ISO/TS 25237:2008 on anonymization: „Process that removes the association between the identifying data set and the data subject.“

ISO 29100:2011: „Anonymization is the process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party.“

Eine Anonymisierung beinhaltet die Verwendung komplexer Methoden der statistischen Geheimhaltung.

Heuristiken z.B. HIPAA-Liste zum Entfernen von QIDs, einfache Strukturbedingungen wie z.B. k-anonymity und komplexere statistische Modelle: hier wird der Attacker explizit modelliert und das aktuelle Re-Identifikations-Risiko anhand parametrisierter Wahrscheinlichkeiten gemessen. Mehr dazu später.

Bemerkung

- Das Anwendungs- und Forschungsgebiet welche Anonymisierung, Risikomessung, Beurteilung der Datenqualität anonymisierter Daten, und alle Methoden welche der Anonymisierung der Daten dienlich sind behinhaltet, wird oft als **Statistical Disclosure Control** (SDC) oder zu Deutsch **Statistische Geheimhaltung** bezeichnet.
- Im Amerikanischen Raum gebräuchlich: **Statistical Disclosure Limitation**.
- Oft fällt der Begriff auch unter dem breiteren Gebiet der **Privacy** oder **Data Privacy**. Auch der Begriff **Confidentiality** ist geläufig.

Warum Anonymisieren?

- Recital 26 EU DSGVO:

  „The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.“

Wenn also der Personenbezug durch Anonymisierung auch wirklich nicht mehr besteht, ist dies die Befreiung von allen Datenschutzbeschränkungen.

Warum Anonymisieren?

- Anonymisierte Daten = nicht-personen-zuordenbare Daten. Ein Datenhalter muss alles tun um Herauszufinden, ob einzelnen Personen in einem Datensatz identifiziert werden können.
- Dabei spielt wiederum die Facto-Anonuität (im Artikel 4 der DSGVO) eine Rolle.
- „Absolute Anonymität“ ist nicht möglich und wird von der DSGVO nicht gefordert. Anonymität ist kein statisches Konzept, sondern hängt vom Kontext ab. Der Name „Harry Smith“ wird jemanden in einem Klassenzimmer identifizieren, aber nicht global.

Public-Use vs Scientific-Use

- Public-use-files:
  - Datensatz wird open-access gestellt (oder mittels einfacher Registrierung)
  - Qualität variiert nach Anwendung (Teaching, Research, Testdaten)
  - Generell starke Eingriffe in die Daten
- Scientific-Use-Files
  - Nur für Forscher oder Forschungsinstitution inklusive Nutzungsvertrag
  - Weniger starke Eingriffe in die Daten als bei Public-Use-Files
  - Üblicherweise müssen Forscher genau erklären wofür sie die Daten nutzen, und mit welchen anderen Daten verknüpft wird.
  - Üblicherweise sollen Forscher auch die Ergebnisse zur Verfügung stellen. Etc.

Scientific-Use cont.

Typische Vertragspartner/Datenempfänger für Scientific-Use-Files:

- Vertragspartner aus der Industrie und Wirtschaft
- öffentliche Forschungsinstitute und Forscher

Bei externen Datenemfänger ist es also nicht bekannt, welche Sicherheitsvorkehrungen bzgl IT-Sicherheit, Protokollen und Loggen von Aktivitäten der Usern getroffen wurden oder welche Analysen er im Detail machen wird.

Es wird vorausgesetzt, dass der externe Vertragspartner einen Geheimhaltungsvertrag/Nutzungsvertrag unterschreibt und Kenntnis über die Pflichten und Rechte als auch technischen Voraussetzungen bzgl. Datenschutz kennt.

Alternativen zur Anonymisierung? (1/4)

Privacy Preserving Computation ist ein umstrittenes Konzept.

- Hier kann zumeist nicht direkt auf Daten eingesehen werden, aber Auswertungen mit bestimmten zur Verfügung gestellten Funktionen/Prozeduren werden unterstützt.
- Diese Prozeduren sind derart gestrickt, dass z.B. keine Einzelfälle ausgewiesen werden. Wird zum Beispiel die Verteilung einer Variablen visualisiert, werden die Einzeldaten zuvor gruppiert und Ausreisser womöglich überhaupt gelöscht. Gezeigt wird daher nur mehr eine Informationsverdichtung, die keinen Rückschluss auf Einzelpersonen erlaubt.
- In der Praxis ist dies sehr schwierig, da nie völlig sicher gestellt werden kann, dass der Rückschluss auf Einzelpersonen nicht möglich ist und andererseits bedeutet es einen sehr grossen Aufwand solch ein System zu etablieren.

Besser: Die Daten werden vorab anonymisiert, sodass jede Berechnung erlaubt ist und nicht nur vorab definierte.

Alternativen zur Anonymisierung? (2/4)

Das **kontrollierte Lab** findet kaum Zustimmung bei Forschern.

- Hier wird ein abgeschotteter PC mit allen Sicherheitsmassnahmen (kein USB, kein Internet, etc.) zur Verfügung gestellt.
- Der Forscher muss also zum Datenhalter kommen.
- Üblicherweise werden nur direkte Identifizierungsvariablen gelöscht.
- Ergebnisse werden vom Datenhalter überprüft und nur wenn sie die Anonymität nicht verletzen dem Forscher gegegeben.
- Das ist somit auch aufwändig für den Datenhalter

Alternativen zur Anonymisierung? (3/4)

Mit **Remote Execution** ist wohl die meiste Arbeit des Datenhalters von Nöten.

- Strukturdaten müssen dem Forscher zur Verfügung gestellt werden, sodass er Code entwickeln kann.
- Der Code wird vom Datenhalter auf den Originaldaten laufen gelassen.
- Ergebnisse werden vom Datenhalter überprüft und nur wenn sie die Anonymität nicht verletzen dem Forscher zur Verfügung gestellt.
- Oft beginnt hier eine (fast) Endlosschleife. Der Forscher sieht seine Ergebnisse, passt darauf hin seine Analyse an $\rightarrow$ der neue Code/Ergebnisse wird wieder vom Datenhalter überprüft...
- Darum nicht sehr brauchbar in der Praxis.

Alternativen zur Anonymisierung? (4/4)

Vom Forscher wird oft **Remote Access** bevorzugt, aber es ist mit rechtlichen Problemen verbunden.

- Mit einer sicheren Verbindung verbindet sich der Forscher an einem Server des Datenhalters. Dort ist alle benötigte Software installiert (theoretisch).
- Dort hat er Zugriff auf die Originaldaten (minus direkte Identifizierungsvariablen)
- Die Resultate bleiben am Server
- Final checkt der Datenhalter (Code und) Ergebnisse und gibt die sicheren Resultate frei.
- Rechtlich umstritten, da man nicht weiss, ob der Datenhalter Daten fotografiert und Fotos mittels OGR wieder in Daten umwandelt. . .
- Praxis in den Niederlanden, nicht vorgesehen in D, AUT, CH, . . .