Contractual Approaches to Data Protection in Clinical Research Projects

EICAR, 24th Annual Conference
Nürnberg, October 2016

Dr. jur. Marc Stauch
Institute for Legal Informatics
Leibniz Universität Hannover

1. Data mining as a privacy challenge

- IT scientists mining data in ever more diverse forms and ever greater quantities. Increasing view that this places in question traditional attempts to protect the privacy of individuals (data subjects) by technical de-identification measures.
  - As Paul Ohm writes: "Computer scientists have recently undermined our faith in the privacy-protecting power of anonymization, the name for techniques for protecting the privacy of individuals in large databases… they can often 'reidentify' or 'deanonymize' individuals hidden in anonymized data with astonishing ease".

Data-mining as a privacy challenge (contd.)

- In response, attention has focused on radical data security conversions that seek to aggregate data or present it in a non-individuated form.
  - Presented as 'counts' rather than individual-level data-rows
  - Cleaning rows of outliers, perturbation, adding 'noise'
- Alternatively, technical security is applied externally as part of securely storing and controlling access to data, so only trusted persons with the clearest reasons for needing to see the data may do so.
  - Data 'siloed' in secure repositories; in some cases external users not allowed to access raw data, but only to pose predefined queries

Data-mining as a privacy challenge (contd.)

- Problem with both of these responses is that they potentially inhibit researchers from learning from and using the data in the optimal way.
  - If data is cleaned of many of the characteristics that individuate it and aggregated into 'counts', much of its value for research may be lost.
  - Placing the data in secure off-limit databases means most researchers will not see it in the first place.
  - In either case, slows down understanding of diseases, and hence of developing ways to prevent and treat them.

Data-mining as a privacy challenge (contd.)

- Also second problem, which is that defensive strategies of the above kind fail to build (or may even diminish) trust between the communities of would-be data researchers and the data subjects whose data is needed to do the research.
- Here proliferation of restrictive de-identification and access control techniques may reinforce public view that mining personal data for new knowledge is an inherently suspect and risky enterprise.

2. Data-sharing in clinical research

- Benefits of health data analysis for research
  - New means of mining (searching through and analysising) the data; sophisticated computer algorithms able to identify previously unsuspected associations/correlations in the data
  - Ever greater understanding of multi-factorial influences on development of diseases; encompasses past (how a condition arose), and future (ability to predict future health events)
- At same time potential risks to individual interests:
  - In privacy (easier for others to infer facts about a person – including about their health
  - In dignity/autonomy (others may make decisions that affect the person, accordingly, including without the person‘s knowledge)

Data-sharing in clinical research (contd.)

- Health data an especially sensitive and vulnerable category of data (and treated as such under the law: Art 8 DPD; Art 9 GDPR)
- So it is clear data protection law, backed up by technical data security, is *necessary* in order to minimise various risks, but in terms of promoting beneficial use of health data for research, not *sufficient*.
- Question is what can be added to achieve greater trust? (solution closely related to achievement of more transparency)

3. Contractual frameworks for individual clinical research projects

- Groups of researchers bound by contracts that create closed research communities
- Plausible solution in specific health data research projects, where a limited number of partners cooperate in sharing and mining data in order to pursue a pre-determined research goal.
  - Key scenario is when data originally stemming from individual patients are to be made centrally accessible via a central repository, or otherwise shared by multiple users.
  - AETIONOMY Project as example

Contractual frameworks for clinical research projects (contd.)

- AETIONOMY project funded by the Innovative Medicines Initiative (IMI)
- Aims to advance understanding of the causes of neuro-degenerative diseases (AD and PD).
  - [http://www.aetionomy.eu/en/vision.html].
- Large quantities of detailed clinical data shared by clinical and pharmaceutical partners. IT scientists carrying out stratification analysis, dividing patients into subgroups that reflect the discrete ways in which the diseases to develop.

Contractual frameworks for clinical research projects (contd.)

- For this purpose LUH has prepared a series of 'data provider' and 'data end user' contracts that the project partners have signed.
- Framework seeks to ensure that data users who access the data for project purposes handle it in legally and ethically compliant fashion.
  - Includes need for strict control of the data to prevent it from escaping into unforeseen interpretational contexts; entails state of art data security in accordance with the law (Art 17 DPD; Art 32 GDPR).

4. Contractual documents as a means of promoting transparency

- GDPR also stresses need for 'transparency':
  - Art 5(1) Need for personal data to be '(a) processed lawfully, fairly and in a transparent manner in relation to the data subject'.
- Here, particular feature of the contractual documents developed by LUH for projects like AETIONOMY:
- This is use of a so-called preamble (placed before the legally operative contractual provisions), i.e:
  - "[an] introductory written statement of facts or assumptions upon which a statute or contract is based".

Contractual documents as a means of promoting transparency (contd.)

- In the context of clinical data research projects the preamble in the data sharing agreements spells out the factual substratum – including the purposes of the project – against which the data processing in the project will take place.
  - For example, the preamble to the various AETIONOMY project data-sharing contracts states (in redacted and abridged form)….

4a. The AETIONOMY Preamble

- *"The AETIONOMY project is an IMI project that aims to analyse and structure different types of data (ranging from molecular data to information on symptoms) and apply this knowledge to construct a new classification of patient groups based on the underlying causes of their disease. This will facilitate the development of drugs that effectively target the specific disease mechanisms in different patient groups. At the same time, each of the project partners recognises as a priority the need to respect the fundamental interests and rights of patients, including the need to preserve the security and privacy of personal data processed in the project…*

AETIONOMY Preamble (contd.)

- *"The data-flow in the AETIONOMY project envisages that all individual (subject-level) data will ultimately be transferred to the [X] repository at [partner Y], where it will be made available to partners for mining and algorithm development as part of constructing the targeted disease taxonomies. In some instances data provider partners will send data directly to [XXX]; however two variant data flows are also planned. First, data may be sent initially to the AETIONOMY Knowledge base at [partner Z] for required data curation [by Z] before onward transfer to [X repository]. In the second place, in the case of image data, this may first be sent to [partner Q] for pre-processing, before inclusion in [X repository]….*

AETIONOMY Preamble (contd.)

- *"While the main obligations upon the data providers and users, respectively, are common to every agreement, individual data flows are covered by separate variant models: this allows the context of each data sharing transaction to be taken account of in specific additional provisions, so that parties' obligations reflect their specific position within the overall data flow* network.
- *The present model agreement is needed to state the obligations and conditions under which an AETIONOMY data provider will transfer data to [Y] for processing and storage in a data-cube in [Y's] X data repository, to which AETIONOMY data users will have access to carry out their tasks in the Project."*

5. Wider Application

- Use of using contractual preambles documents as a means of shining a light on data processing operations is potentially of wide application.
  - in most cases, extracting value from data will involve multi-party arrangements - transactions between separate legal entities, which are susceptible to contractual arrangements (usually wanted by the parties themselves).
- Thus in its Report on 'Big Data', the US PCAST Report (2014) distinguished three standard kinds of data-processing entity:
  - data collectors; data analysers; data users
  - (rare for all roles to be combined within one entity)

Wider Application (contd.)

- In this regard, data protection regulations could aim for a system, not dissimilar to that for present cases where agreements between different entities may raise legal or social concerns (such as in the area of competition law)
- Such contracts would need to be deposited with a central regulatory authority.
  - Admittedly, considerations of commercial confidentiality may sometimes require that the operative provisions of such contracts be kept from public view (albeit the regulator would be aware of them).

5. Wider Application

- However, in each case the parties could be required to include a preamble, in a standard prescribed format, setting out the main features of their arrangement
  - Would include the nature of the data to be shared; the intended purposes of processing it; and the source and final destination of the data.
- This would be available in public part of registry.
- Such an approach would promote transparency of processing operations using individual data, and thereby contribute to more public trust

Thank you for your attention!

Marc Stauch

stauch@iri.uni-hannover.de

• • •

Acknowledgement: the work on which this paper is based has received funding from the Innovative Medicines Initiative under grant agreement no 115568. The views expressed are those of the authors and do not necessarily reflect those of IMI.