IT Solutions for Privacy Protection in Biobanking

J. Eder^a H. Gottweis^{b, d} K. Zatloukal^c

^aDepartment of Informatics Systems, University of Klagenfurt, Klagenfurt, ^bDepartment of Political Science, Life-Science-Governance Research Platform, University of Vienna, and ^cInstitute of Pathology, Medical University of Graz, Graz, Austria; ^dDepartment of Sociology, Kyung Hee University, Seoul, South Korea

Key Words

Biobank donors · Biobanks · Personal data · Privacy concerns · Research needs

Abstract

Biobanks containing human biological samples and associated data are key resources for the advancement of medical research. Efficient access to samples and data increases competitiveness in medical research, reduces effort and time for achieving scientific results and promotes scientific progress. In order to address upcoming health challenges, there is increasing need for transnational collaboration. This requires innovative solutions improving interoperability of biobanks in fields such as sample and data management as well as governance including ethical and legal frameworks. In this context, rights and expectations of donors to determine the usage of their biological material and data and to ensure their privacy have to be observed. We discuss the benefits of biobanks, the needs to support medical research and the societal demands and regulations, in particular, securing the rights of donors and present IT solutions that allow both to maintain the security of personal data and to increase the efficiency of access to data in biobanks. Disclosure filters are discussed as a strategy to combine European public expectations concerning informed consent with the requirements of biobank research.

Copyright © 2012 S. Karger AG, Basel

Introduction

Most of our current knowledge of diseases and of diagnostic and therapeutic opportunities is based on our systematic investigation of human biological samples from healthy populations or diseased individuals and arrived at by linking data on clinical and molecular disease phenotypes with therapeutic interventions and diseases outcomes. Biobanks contain collections of a variety of human biological samples, such as blood, cells, tissues, and DNA as well as associated information on the sample and the sample donor. As such, they provide a well-defined framework enabling medical research in a quality-controlled and ethically and legally compliant manner. Biobanks are typically established in the context of prospective cohort studies involving individuals of a healthy population or as clinical biobanks established within health care containing samples and data from patients with certain diseases or disease groups.

In order to address the global health challenges, such as providing sustainable health care to aging populations, there must be international collaborative research that delivers innovative diagnostic, preventive and therapeutic solutions. Currently, however, transnational or international collaborative medical research is severely hampered by the lack of common quality standards, nonharmonized data management systems, and heterogeneity of ethical

and legal requirements from country to country. To improve the interoperability of biobanks, major efforts will be required, which would include reducing the time it currently takes from the formulation of a specific medical research question to the time when the researcher gains access to the human biological samples and data in order to actually perform the research project. The current lack of standardization and interoperability also prevents the integration of and the reuse of data generated in different research projects, resulting in significant duplication of effort and even the limited reproducibility of research. All of these factors conspire to create an urgent need for researchers to have better access to, and the integration of, both biological samples and their associated data that are stored in biobanks in various institutions and countries.

These needs were recognized as a strategic priority in the ESFRI (European Strategy Forum for Research Infrastructures) roadmap for research infrastructures that listed a research infrastructure for biobanks and biomolecular resources (BBMRI). ESFRI research infrastructures are implemented by EU member states and are intended to provide a single pan-European and sustainable (30 years or longer) solution for providing access to key technologies and resources for a given research field. According to these requirements, BBMRI was designed as a distributed infrastructure with operational sites in almost all EU member states. This infrastructure integrates existing and newly established human biological sample and data collections, resources, technologies, and expertise to facilitate high-quality medical research [1]. The planning of BBMRI was finalized in January 2011 and has involved more than 270 institutions in 33 countries. The concept developed not only had to provide solutions to improve interoperability of biobanks across Europe and internationally, but also had to balance research needs with societal demands, such as improved health and health care on one hand, and protection of individuals contributing to medical research by making their samples and data accessible through biobanks, on the other hand.

In the context of medical research on biological samples and associated data (not clinical trials), the protection of privacy is the most sensitive issue. This is also reflected in international conventions as well as European and national legislations (for examples see: <http://www.wma.net/en/30publications/10policies/b3/index.html>; <http://conventions.coe.int/Treaty/en/Treaties/Html/164.htm>; and <http://www.oecd.org/dataoecd/41/47/44054609.pdf>). The general principle is that research on identifiable health-related data requires specific informed consent that covers the type and implications of planned research

and possible risks for the participant. Since research on anonymous data is considered as presenting no privacy risk for participants, it therefore does not require informed consent. However, it is an inherent feature of medical research that research on biological samples has to be correlated with long-term disease outcomes, which requires that samples and data collected at different times can be integrated over several years, which in turn does not allow anonymization. Furthermore, informed consent obtained in the context of the collection of a sample several years ago cannot properly reflect the current state of the art of medical research, particularly in the field of rapidly developing genomic technologies. Therefore, a 'specific' informed consent is almost impossible to be applied in modern medical research. At the same time, anonymization is not easily achieved since detailed medical data or genetic information (e.g. as required for personalized medicine) can, at least in theory, be attributed to an individual (because there is, for instance, only one individual with a certain pattern of laboratory parameters or a certain gene sequence). The limited applicability of specific informed consent and the limited opportunities of using anonymized data have caused much-discussed ethical and legal dilemmas for medical research.

In this article, we discuss current European public opinion on privacy issues and concerning informed consent, and then propose IT solutions that help to harmonize research needs and societal demands. The strategy to create disclosure filters to deal with privacy issues in biobanking is proposed as an approach to reconcile societal expectations with research requirement.

European Publics, Biobanks and the Question of Privacy

Protecting privacy and data security are key concerns when it comes to the public perception issues in cohort studies and related biobanks. Also, recent focus group research shows that there are clear concerns about data protection in the European public, but if certain appropriate measures are taken and there are trustworthy actors in matters involving data, people seem to be less concerned about data flows and related privacy and data protection issues (see Snell et al. and Hobbs et al. in this issue).

Figure 1 shows that, overall, there is a substantial level of concern about different institutions in European countries holding data. Clearly, there are substantial differences of concern between countries, such as between Germany and Iceland, but the data indicate clearly that, across

Fig. 1. Concerns about data privacy by the organization holding personal data. Question asked: ‘Different private and public organizations keep personal information about people. Are you concerned or not that your personal information is being protected by these organizations?’ Source: European Commission 2008/7/ EU. DK = Don’t know; NA = no answer.

Fig. 2. Willingness to provide information about oneself in a biobank. Question asked: ‘Would you be willing to provide information about yourself to a biobank?’ Source: European Commission 2010/160/EU. DK = Don’t know.

Europe, data protection is an issue of concern for a good part of the society. These general concerns are reflected in people’s answers to the question of whether they are willing to provide data about themselves to biobank research (fig. 2). Again, while in Europe there seems to be good overall readiness to participate in biobank research, there are stark contrasts between countries. For example, in Iceland 92% of the population are definitely or probably willing to provide information to a biobank, whereas that figure drops to 24% in Turkey. At the same time, there seems to be a consistent level of concern about giving genetic data to biobanks (fig. 3). These patterns of concern seem to correspond with a strong preference among European publics to prefer narrow, individualized forms of consent over broad forms of consent. At the same time, it must be considered that there is clear indication that people in Europe lack knowledge of and information about biobanks, and that there is a strong correlation between their knowledge of biobanks and their readiness to participate in biobank research, but also with respect to questions of informed consent. The more people know about biobank research, the more they seem to be prepared to participate in such research and also to be more open toward giving broad forms of informed consent [2–4]. This constellation poses a considerable challenge for European biobank research; a clear public expectation seems to exist that biobank research not only should protect the privacy of research participants, but also adhere to an informed consent model in which the given consent is specific rather than broad and general (fig. 4). However, it must be pointed out that public attitudes toward certain issues such as informed consent or privacy should not be seen as set in stone, but are always fluid and contextual. It is therefore a good strategy for biobanks to reach out to the public and to engage in a discussion of how people’s preferences for particular forms of informed consent can be reconciled with the daily requirements of biobank research. There is a need for innovative solutions at the interface between biobanks and society. In the following section, we present such an innovative solution to negotiate public preferences with requirements of biobank research in the form of disclosure filters opening up the possibility for individualized informed consent.

Biobanks, Privacy and Consent: IT Support for Privacy Maintenance in Searchable Biobanks

Biobanks are collections of biological material (items) and data. They contain data *describing* the collected items, such as the type of the material and details of

Fig. 3. Concern about the collection of types of data or materials. Question asked: 'In order to understand the causes of disease researchers need as much information as possible about people in the biobank. Would you personally be concerned or reluctant about the collection of any of the following types of data and ma-

terials from you?' (BL = Blood samples; TI = tissue collected during medical operations; GE = your genetic profile; MR = medical record from your doctor; LS = lifestyle). Source: European Commission 2010/161/EU.

the harvesting, the preservation procedure and storage conditions. In addition, they store data *derived* from the items, such as pathological diagnostic findings, gene expression profiles, biochemical analysis, and gene sequencing. And they store or link to data *associated* with the items. These last data are in particular data about the donor, the medical case history, and the therapies as well as lifestyle data typically supplied by the donor on questionnaires. There are several architectural layouts for the implementation of information systems for biobanks [5]. In any case, information systems have to integrate data coming from various sources, such as hospital information systems, lab data, official registries, and research data sets. Data might be stored in an integrated form in biobank databases (data warehouse architecture), or they might be integrated as part of the query processing (federated database architecture). Biobanks can be organized in networks [1, 6] that enable users to search a vast collection from a single entry point. For the envisioned international research infra-

structure, only an architecture based on federated databases is viable to recognize the various legal and institutional regulations for the sharing data and material from biobanks.

In this architecture, a user has potential access to hundreds of biobanks through a single query interface.

Data stored in biobank databases consist of personal donor data, which are highly sensitive and must therefore be protected to avoid misuse. Consequently, access and use of these data are strictly restricted to the defined purposes for the use of the material as laid down in general laws and bylaws and, in particular, in the informed consent of the donor.

As a first measure, security mechanisms must ensure that no unauthorized parties can access the data. For this purpose, standard security mechanisms for databases and information systems can be applied. Data in biobanks are at least as secure as all health-related data collections as, for example, in hospital information systems or in the databases of health insurance companies, al-

Fig. 4. Informed consent. Question asked: ‘In a hospital doctors ask the patient to sign a form giving permission to carry out an operation – this is called “informed” consent and it is also required of medical researchers who do research involving members of the public. When a scientist does research on data in a biobank, what do you think about the need for this kind of permission? Researchers should ...’. Source: European Commission 2010/157/EU. DK = Don’t know.

though in biobanks, typically fewer individuals have access privileges to sensitive data than in hospitals.

However, a main purpose of biobanks is for researchers to access and to use these data and the biomaterials for research. The challenge is how researchers can access the data in biobank information systems without compromising the anonymity of the donors. Or, in other words: which kind of interface can we offer researchers so that their research is not prohibited, but donors’ personal data remain protected?

We analyze the different phases in preparing and performing a medical study and discuss the appropriate privacy measures for each phase.

We follow here the scheme proposed by BBMRI, which in principle consists of 3 phases:

(1) In a first phase of acquiring material and data, a researcher investigates which of the biobanks participating in a consortium store data that may be relevant for the

proposed study. In this phase, the researcher has no access to data detailed enough to violate anonymity requirements and has no access whatsoever to biomaterials.

(2) Based on the results of this research, the researcher in stage 2 formulates a project proposal and submits it to the institution’s ethics committee for approval. Additionally, the proposal is reviewed by the biobank, so it can manage access to rare resources. Part of the approval process is the analysis of whether all the data requested is indeed necessary for the project (‘need to know’ principle).

(3) Only after approval from the ethics committee is the researcher granted, in stage 3, access to detailed data and material. In this phase, the actual research project is performed, and the researcher is bound to observe all necessary precautions to maintain the security of the data and the confidentiality of the donors.

As indicated in stage 1 above, the protocol of the first phase is *intended* to safeguard the anonymity of donors and their personal data. A researcher might, for example, ask ‘which biobanks store cryopreserved tissue of patients below 50 years of age, diagnosed with liver cancer, with a BMI below 20?’ The system responds with a table containing the biobanks and the amount of matching cases (if this number is above a threshold). In this phase, the system must avoid only that personal data can be inferred from this very aggregated information, that is, from the number of items satisfying the restrictions in the query. This is more subtle than is obvious at first glance. Even if only a single number is returned as a result for a query, a prying researcher can still deduce information about an individual by making a specific sequence of queries. For example: researcher A has enough information to uniquely identify the record of a male person in the database and intends to violate the male’s privacy rights by trying to determine whether he has been diagnosed with HIV. If the number of results for the query (restrictions to identify the person and HIV diagnosis) returns 1, then researcher A has the desired information. This kind of unethical snooping can be avoided by establishing a lower limit for the results. But even with such a limit, researcher A might ask how many females are diagnosed with HIV and then submit the following query: number of persons with diagnosis HIV who are either female or satisfy the restrictions identifying the person of interest. This technique is called tracking [7].

There are 3 possible technical strategies to prohibit such prying: tracker detection, the anonymization of the data set and the statistical blurring of query results returned to requesters.

The first strategy implies that the database installs mechanisms to identify trackers and stops query answering, if a tracker is detected [8]. Mechanisms for discovering trackers are costly, however. In principle, this strategy requires that series of queries are analyzed to determine whether the results allow the inference of sensitive information. Adam and Wortmann [9] give an overview of tracker detection methods.

The second strategy, anonymization, is to allow queries to be performed not on the original database, but on a derived database containing a version of the original data that is prepared in such a way that no data of any individual can be inferred. Such a database can be constructed by removing data that might compromise anonymity (personal identifiers such as Social Security Numbers), replacing absolute data with relative data (e.g. age instead of birth date) and generalizing data to make them more coarse-grained (e.g. age groups in 5-year intervals). K-anonymity [10] and l-diversity [11] provide the formal basis and algorithms for the construction of such databases. A data set is k-anonymous: if for any combination of search attributes there are at least k records satisfying the query. A data set is also l-diverse: if in any query result there are at least l different values for each non-search attribute. As such a database version does not contain information detailed enough to allow any proposition about an individual, the results of any query on such a database do not admit inferencing sensitive personal data.

The third technique is to add statistical noise to the answers, which blurs the results in such a way that no individual data can be deduced with certainty [12, 13]. For example, adding a small random number between -5 and 5 to the query result in the attack sketched in the example above would make the results useless for an attacker, but the results of such blurred queries would still be valuable for a researcher legitimately searching for relevant cases. In practice, the statistical blurring would have to be sophisticated enough to prevent the reduction of the statistical noise through repeated queries by an attacker who is attempting to violate the privacy of the donor.

To conclude: the threat that individual personal data are unethically obtained by an attacker posing statistical queries to biobank databases, presumably in stage 1 of their research, must not be overlooked, although there are strategies to repel such attackers.

In phase 2, a researcher formulates a project and an application for the use of material and data. Part of this application is the exact specification of which data and material are needed and a justification for this demand.

A researcher also has to describe how data received from the biobank will be protected against misuse. Also part of phase 2 is an assessment by an ethics committee as to whether the proposed study observes all ethical regulations and whether the envisioned results justify the use of the material and of personal data. Included in this assessment is the evaluation of the data protection concept of the applicant and the justification for the data and material requested. In the third phase, the researcher gains access to sensitive data to perform the planned study. For this phase, the need to know principle should be applied: a researcher should receive only the data necessary for the project and only as detailed as necessary in order to perform the study approved by the ethics committee in stage 2. Whenever possible, the data set relayed to the researcher should be made k-anonymous and l-diverse, so that it is not possible to infer sensitive personal data from the dataset. Since anonymization is accompanied by information loss, it must be carried out carefully, taking into consideration the particular information requirements of the project. Priority-driven anonymization procedures [14, 15] reduce the information in loss for data which needs to be processed in detail at the cost of higher information loss in other data. Nevertheless, we take note that anonymized data are not always sufficient for carrying out medical studies. For studies that require the researcher to have access to detailed data, securing these data against misuse must rely on legal means.

Disclosure Model and Disclosure Filter: Toward Individualized Consent

In the previous section, we analyzed how the right of donors to maintain their privacy rights can be harmonized with the necessity of researchers to search for useful material and data. In this section, we introduce a novel concept, disclosure filters, that reconciles 2 objectives: individualization of informed consents on one hand, and increasing efficiency of the quest for material and data for research, on the other. Disclosure filters allow donors to individualize the informed consent, that is, to constrain the purposes for which their material and data can be used. Additionally, disclosure filters allow researchers to concentrate their search on material and data for which there is a realistic possibility that they can be actually used in the planned project.

Individualization of consents means that donors are offered the possibility of stipulating constraints on the use of their data and material. These restrictions can be

Fig. 5. Generic disclosure model.

specified in several dimensions, such as the field of research (cancer, metabolics, etc.), type of research (basic research, drug development, etc.), type of research lab (university, company, etc.), location of the research (country, federation), permitted data (case history, lifestyle data, lab data, family data, etc.), qualifications of the research organization or researcher, or accreditation of a project by some organization or ombudsman. Individualized informed consents raise the complexity of managing the content of biobanks and the access to it considerably. Nevertheless, as we discuss below, suitable IT support makes individualized consents viable.

A disclosure filter is a software component that helps the biobank hosts to answer the following multilayered question: *who* is allowed to receive *what* from *whom* under *which* circumstances and *how*?

A disclosure filter is built upon a disclosure model, capturing all the information needed to answer the question outlined above. The disclosure model is based on laws, contracts between participants, policies of participants, rulings by courts and ethics boards, and, last but not least, by restrictions specified by the donors in their informed consents. Figure 5 shows the basic elements of a disclosure model: requester, provider, item, and circumstances. It captures information of the *requester* of material and data. At an individual level it contains the qualifications of the requester, both on an individual and on an institutional level. These qualifications can be grouped in classes (e.g. recognized lab for basic research or private company). Additionally, the disclosure model contains information about the location and other details of the requester. Then, the model contains information about

the *provider*, for example, the institution hosting the biobank, and technical and organizational information on the biobank, including its policies and procedures. Providers may be affiliated with other providers (e.g. forming a network of biobanks with an integrated interface for searching material). Information about *items* consists of the type of item (material or informational item) and available information about the item. This information is usually stored in biobank database systems. The relationship between requester and donor represents which properties a requester (or his or her study) has to fulfill to gain access to the item. This relationship, in particular, contains the restrictions established in the informed consent of the donor of the item. Additionally, restrictions might come from law, rulings or policies of the provider.

These requester-item-provider relationship is further detailed with *circumstances* that are additional conditions that either must be fulfilled by the provider as a precondition (e.g. the requester cannot be a for-profit company) or must be satisfied as additional constraints (e.g. samples and data may be relayed if a project is formulated and approved by the ethics board of the provider biobank).

Disclosure models and disclosure filters are used in the following ways:

- **Result filtering:** The biobank query interfaces removes data from query answers that is not supposed to be seen by requesters. In this role, the disclosure filter supports a policy-based access control scheme. For example, the answer set for a request might be restricted to those set of cases for which access was granted by a prior proposal of the project.

- *Technical optimization:* Query system optimizes query processing using disclosure information. The disclosure model adds additional restrictions to the query posed by the requester and thus can be used to reduce the technical effort for computing the result set of the query. In particular, in networks of biobanks this technical optimization may reduce the number of biobanks to which the query is forwarded and thus significantly reduces both query response time and total system effort toward answering the query. For an example, biobanks that do not allow exporting data and material to the country of the requester do not need to be contacted.
- *Query restriction:* The requester might restrict the query to consider only biobanks and items that might grant access to their material and data given the profile of the requester and the particularities of the planned research project. With this restriction, the requester can avoid vain efforts to identify and select items and data, and take steps to apply for access, if the attempt is hopeless a priori. For example, if the project is in the area of metabolic diseases, cases that might only be used for ontological research are not retrieved for the requester.
- *Access information:* The disclosure model can be queried itself for information on the access policies of a biobank. In particular, the conditions that need to be satisfied by the requester and his or her intended research project and the procedures for gaining access to material can be retrieved from the disclosure model. For example, the requester might ask about the procedures for using material for a cancer research project by a private company in a member state of the European Union.

What is needed for applying disclosure models and disclosure filters? First of all, it will be necessary to develop standards for the description of requesters, providers and items. Only if description protocols are comparable across biobanks, a disclosure model can work beyond individual biobanks. The description of the requester can be general, typically in the form of requester classes. However, for the advanced functionality of disclosure filters, the description of the requester – both the person and the institution – should be detailed and individual. The collection of the necessary data is organized by enrolling and registering participants before they can search the biobank databases for material and data. In order for requesters, projects and regulations to be matched correctly with material and data, descriptions of the research projects, the materials and the circumstanc-

es must be standardized. The standardization needs must also be considered when formulating the informed consent, as the disclosure filter has to analyze whether a particular requester project is included in the allowed usage in the informed consent and thus must be able to match restrictions in informed consents and project descriptions. The development of an ontology for all these descriptions has to be part of mentioned endeavors to create an international research infrastructure.

Conclusions

We have argued that during the last decade European biobank research has expanded considerably and that cooperation between European biobanks is significantly increasing as well as increasing in importance to researchers. At the same time, it is clear that the European publics expect that the data they give to biobanks are well protected and that this privacy protection in terms of given consent is specific rather than broad. We have argued that these expectations, however, are in development and fluid and that there is a need for biobank research to propose innovative models for how to reconcile biobank research with public expectations. We have shown that the intensification of biobank research and the operation of individualized consent strategies are not mutually exclusive. Implementing disclosure filters have been proposed as a key element in this strategy.

The disclosure filter demonstrates that it is possible to harmonize seemingly conflicting requirements by research needs on one hand and societal demands on the other hand and in this way helps both researchers and donors. Researchers avoid sunk costs in chasing material and data for their projects that they ultimately cannot obtain. This alone makes the search for material more efficient and also reduces the time before the project can start by avoiding lengthy detours. Disclosure filters also admit that donors can execute their rights to restrict the use of their data and material without rendering it unusable for research. The integration of the disclosure model in the query processing makes individualized informed consents possible without jeopardizing efficiency.

Research needs and societal demands are often seen as conflicting requirements for the organization of research infrastructures such as biobanks for biomedical research. We claim that frequently these seemingly conflicting requirements can be harmonized when suitable organizations and IT solutions are employed. We discussed and analyzed 2 pairs of conflicts: maintaining privacy versus

access to data by researcher, and search efficiency versus individualized informed consents. For both cases, we presented IT solutions that support our claim. The discussion of solutions for privacy maintaining search for data and material show that the practice of overly restrictive access is not necessary, and it demonstrates that suitable interfaces to biobanks support researchers without risking the privacy of the donors. The concept of disclosure filter shows that individualized informed consents do not render donated data and material useless if the donor demands restrictions on the purposes and circumstances in which her or his data and material can legitimately be used. Additionally, integration of the access restrictions into disclosure filters not only enforces these restrictions, but also speeds up the researchers' quest for usable material and data. Developing and applying such IT solutions raises the value of biobanks for research, making them better and more quickly accessible, and it

encourages donations because donors can understand the measures taken to protect their privacy and to respect their intentions. And this, last but not least, meets the societal demand for better health care through improved therapies as a consequence of medical research.

Acknowledgments

This work was supported by the EU FP7 Grant Agreement number: 212111 (Project BBMRI) and the Austrian Genome Programme GEN-AU (Project GATIB II). Herbert Gottweis's work was also supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-330-B00169) and the project 'PRIVATE Gen' (Privacy Regimes Investigated: Variations, Adaptations and Transformations in the age of (Post-)Genomics) which is supported by the German Federal Ministry of Education and Research (BMBF), DLR, Academy of Finland and FFG.

References

- 1 Yuille M, van Ommen GJ, Bréchet C, Cambon-Thomsen A, Dagher G, Landegren U, Litton JE, Pasterk M, Peltonen L, Taussig M, Wichmann HE, Zatloukal K: Biobanking for Europe. *Brief Bioinform* 2008;9:14–24.
- 2 Gaskell G, Gottweis H: Biobanks need publicity. *Nature* 2011;471:159–160.
- 3 Gaskell G, et al: Europeans and biotechnology in 2010. Winds of change? European Commission. http://ec.europa.eu/public_opinion/archives/ebs/ebs_341_winds_en.pdf.
- 4 Gottweis H, Gaskell G, Starkbaum J: Connecting the public with biobank research: reciprocity matters. *Nat Rev Genet* 2011;12:738–739.
- 5 Eder J, Dabringer C, Schicho M, Stark K: Information systems for federated biobanks. *TLDKS* 2009;1:156–190.
- 6 Wichmann HE, Kuhn K, Waldenberger M, Schmelcher D, Schuffenhauer S, Meitinger T, Wurst S, Lamla G, Fortier I, Burton P, Peltonen L, Perola M, Metspalu A, Riegman P, Landegren U, Taussig M, Litton JE, Fransson M, Eder J, Cambon-Thomsen A, Bovenberg J, Dagher G, Van Ommen GJ, Griffith M, Yuille M, Zatloukal K: Comprehensive catalog of European biobanks. *Nat Biotechnol* 2011;29:795–797.
- 7 Denning DE, Denning PJ, Schwartz MD: The tracker. A threat to statistical database security. *ACM T Database Syst* 1979;4:75–95.
- 8 Denning DE, Schlörner J: A fast procedure for finding a tracker in a statistical database. *ACM T Database Syst* 1980;5:88–102.
- 9 Adam NR, Wortmann JC: Security-control methods for statistical databases: a comparative study. *ACM Comput Surv* 1989;21:515–556.
- 10 Sweeney L: Achieving k-anonymity privacy protection using generalization and suppression. *Int J Uncertain Fuzz* 2002;10:571–588.
- 11 Machanavajjhala A, Kifer D, Gehrke J, Venkatasubramanian M: L-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Discov Data* DOI: [10.1145/1217299.1217302](https://doi.org/10.1145/1217299.1217302).
- 12 Matthews GJ, Harel O: Data confidentiality: a review of methods for statistical disclosure limitation and methods for assessing privacy. *Statist Surv* 2011;5:1–29.
- 13 Li F, Zou X, Liu P, Chen JY: New threats to health data privacy. *BMC Bioinformatics* 2011;12(suppl 12):S7.
- 14 Stark K, Eder J, Zatloukal K: Priority-based k-anonymity accomplished by weighted generalisation structures. *Springer LNCS* 2006; 4081:394–404.
- 15 Stark K, Eder J, Zatloukal K: Achieving k-anonymity in DataMarts used for gene expressions exploitation. *JIB* DOI: [10.2390/biecoll-jib-2007-58](https://doi.org/10.2390/biecoll-jib-2007-58).