Data protection issues with regard to research in genetic data

Marian Arning, Nikolaus Forgó, Tina Krügel

Institute for Legal Informatics of the Leibniz University of Hannover

A person’s genetic data provides information about its descent, ethnical origin, and, with a certain probability, also about future diseases and possibly about their healing chances and much more. On one hand, for these reasons, the processing of highly sensitive genetic data requires strict compliance with data protection legislation. But on the other hand, data protection authorities are obliged not to restrict human-genetic medical research, but to show possible ways to promote genetic research in compliance with the existing legal framework.

This article deals with the characteristics of data protection legislation with regard to genetic data and especially the question of whether and of how genetic data can be rendered anonymous. It is motivated by the EU research project ACGT (Advancing Clinico-Genomic Trials on Cancer), which aims at the development of a transeuropean cancer gene bank to promote better and more efficient curability.¹

1 Genetic data

As stated above genetic data contains a huge amount of information about the person it refers to. Each individual’s genetic data is unique and can contain information even of yet unborn blood relatives. Therefore each person can be identified reliably by its genetic data.

Due to the amount of information they carry concerning an individual’s state of health, origins and descent, genetic data has to be classified as highly sensitive, so that it has to be protected in a strict way, as the unlawful processing would put the privacy of the data subject at high risk.

2 Special characteristics of genetic data

The characteristic features of genetic data are their uniqueness and the highly sensitive quality of the information they contain. Consequently, European data

¹ http://www.eu-acgt.org

* Marian Arning, LL.M., Prof. Dr. Nikolaus Forgó and Dr. Tina Krügel, LL.M. are staffs of the Institute for Legal Informatics of the Leibniz University of Hannover/Germany (http://www.iri.uni-hannover.de). In the framework of the EU research project ACGT (Advancing Clinico-Genomic Trials on Cancer) the authors are responsible for legal and especially for data protection issues.
protection legislation ranks data concerning health, such as genetic data\textsuperscript{2}, as data requiring special protection (see Art. 8 para. 1 Data Protection Directive 95/46/EC\textsuperscript{3}).

In general, the processing of genetic data is prohibited according to Art. 8 para 1 of the Directive, if it has to be qualified as personal data. According to Art. 2 lit.a of the Directive personal data shall mean any information relating to an identified or identifiable natural person (‘data subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

The processing of genetic data may only be lawful, if the data subject has given his explicit consent to the processing (Art. 8 para 2 lit a) or one of the exemptions stated in Art. 8 paras 2-5 applies.

However in the case of a scientific project it must be considered that to consent in advance to each individual operation performed upon the data is almost impossible as normally in the course of a project new research methods are developed which may demand other operations performed upon the data than the patient has consented to. Also the cooperation with other scientists may require the extension of the consent. It can be doubted whether a consent, which is worded too extensively, is valid.

Art. 8 para. 3 and 4 of the Directive state important exception for scientific research projects, permitting the processing of sensitive data. But first, genetic research does not fall under the scope of these exemptions, anyway.

Second, and more important, the Directive would still be applicable, if such an exemption would be used to legitimate the processing of genetic data. It would be much more practicable for a scientific researcher involved in a trans-European project like ACGT, if the data he uses wouldn’t fall under the scope of the Directive at all.

Therefore, it would be the best, if the researcher would use non-personal data for the research. But on the other hand, the identification of the data subject is needed, as the patient may benefit from the scientific research with his data.

In the following it is examined, whether genetic data can be anonymized at all and if yes under what conditions. Besides, it must always be taken into account that the data subject shall benefit from the research for example carried out in research projects like ACGT and must therefore be identifiable.

3 Anonymous genetic data

As soon as it’s data is rendered anonymous, the data subject requires no further protection, because re-identification is impossible. Anonymous data can be


\textsuperscript{3} In the following: Directive.
processed (collected, stored, published...) without restrictions, as it does not fall under the scope of the Directive.

Most of the times, for researchers it is not important to know the person to which the data, he examines, refer. In practice the data subject’s name etc. are often replaced with a label, in order to preclude identification of the data subject or to render such identification substantially difficult. The person can only be re-identified by using the appropriate key. The data is “pseudonymized”.

The question at this stage is, whether also pseudonymous genetic data can be regarded as anonymous data. Or does genetic data always have to be qualified as personal data because of it’s uniqueness?

The crucial point is, how to define the term “anonymous”. The Directive itself doesn’t contain an explicit definition of this term. Only Recital (26) of the Directive contains a definition of this term:

\[(26) \ldots \text{whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable} \ldots\]

According to the wording of Recital (26), data can only be classified as anonymous, if the re-identification of the data subject is impossible for everybody.

But a complete anonymization of genetic data is impossible. Take the example of a HIV study. In the course of the study, a sufficiently large gene sequence is published on the internet without personal details. If there’s already genetic information about the concerned person stored for a different purpose, e.g. because of a salivatestor as a compulsory requirement for a life insurance contract, an identification of the person concerned and his disease would be possible for all persons, who have access to these databases, by a matching-procedure.\(^4\)

This example shows that the unique quality of genetic data causes the problem, that despite comprehensive anonymization, a re-identification of the said person is possible, if relevant additional knowledge exists.

Nevertheless, on the basis of European legislation, too, the anonymization of genetic data seems to be possible, accepted and not objected. For example, the Article 29 Data Protection Working Party accepts the anonymization of genetic data as a means to limit the dangers of genetic research.\(^5\)

Indeed, there is a new view in coming: In the First Report on the implementation of the Data Protection Directive 95/46/EC,\(^6\) the Commission states, that the interpretation of the Directive must be sensible and flexible, and draws attention to an article of the European Privacy Officers Forum (EPOF),\(^7\) which

\(^4\) Huge databases containing genetic data are in use already e.g. by law enforcement agencies, insurance companies and in the USA even in labor relations (see Weichert, Der Schutz genetischer Informationen, in: DuD 2002, S.133).


emphasizes the practical orientation and exemplary function of the German definition of “anonymization”.

The German transposition of the Directive contains a broader definition. Section 3 para. 6 BDSG (Federal Data Protection Act) defines anonymization as the modification of personal data, so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual. In conclusion, the BDSG accepts also a second group of anonymous data: *de facto anonymous data*, because it can only be turned into personal data with a disproportionate amount of time, expense and labour.\(^8\)

According to that definition, genetic data can be regarded as anonymous data under certain conditions. But the question is, how to define the term disproportionate and especially for which person the amount of time, expense and labour has to be disproportionate to de-anonymize the data. In this context it is important, that the term “personal data” is relative. For each single data controller it has to be determined separately, whether he has the knowledge to identify the data subject or not, or whether additional knowledge has to be attributed to that particular data controller. In other words: it’s possible, that genetic data is anonymous for one researcher, while it’s personal for another. Therefore the central question is always, whether de-anonymization is possible for this particular data controller or not.

As stated above, pseudonymized data is significantly more useful in the framework of a medical research project. If the data controller does not have access to the pseudonymization key and if he has no access to other additional knowledge (e.g. a biobank) enabling him to identify the data subject, the data in question is anonymous for this particular data controller.\(^9\) Otherwise the genetic data has to be treated as personal data, with all restrictions of processing. It is beyond dispute, that additional knowledge, which the data controller actually has, is attributable to him, even if he doesn’t plan to use the additional knowledge to identify the data subject.\(^10\) As a second step the question arises, if and to what extent additional knowledge he does not have, but which could be obtained by him or any other person is attributable to the data controller.

---


\(^9\) See for example: Gola, Peter/ Schomerus, Rudolf: BDSG, Munich 2005, §3 marginal number 46.

\(^10\) Gola, Peter/ Schomerus, Rudolf: BDSG, Munich 2005, §3 marginal number 44.
4 Relevance of the character of data processing for the distinction between personal and anonymous data

Austria introduced a new category of data in addition to personal and non-personal data in the course of the transposition of the Data Protection Directive: indirectly personal data.\textsuperscript{11}

Data is indirectly personal “when the data relates to the subject in such a manner that the controller, processor or recipient of a transmission cannot establish the identity of the data subject by legal means”.\textsuperscript{12} The use of indirectly personal data is not considered an infringement of confidentiality interests requiring protection, even if sensitive data is processed.\textsuperscript{13} If data is only indirectly personal for a recipient (e.g. pseudonymized data), transborder transmission and committing of data do not require authorisation.\textsuperscript{14}

If data is only indirectly personal for the controller and will be used for scientific or statistical research purposes, whose goal is not to obtain results in a form referring to specific data subjects, the controller has the right to use indirectly personal data without having to comply with further requirements.\textsuperscript{15}
In this case he doesn’t need an informed consent of the concerned data subject to process its data lawfully.

In conclusion, according to the Austrian data protection legislation, additional knowledge is only attributable to the data processor, if it is accessible to him by legal means, e.g. by using Internet. The ability of third parties to re-establish the reference to the individual concerned remains unconsidered.

\textsuperscript{11} Sec. 4 no. 1 of the Austrian Federal Act Concerning the Protection of Personal Data (DSG 2000).
\textsuperscript{12} Sec. 4 No. 1, 2. clause DSG 2000.
\textsuperscript{13} Sec. 9 no. 2 DSG 2000.
\textsuperscript{14} Sec. 12 sub-para. 3 no. 2 DSG 2000.
\textsuperscript{15} Sec. 46 sub-para. 1 no. 3 DSG 2000.
However, the Austrian regulation regarding research in genetic data in accordance with data protection requirements cannot simply be applied to other Member States. A certain amount of freedom was given to the Member States to implement the Directive into national law. And the Directive doesn’t contain any regulation in certain areas, so that data protection legislation in the EU Member States still differs significantly.

First it is assumed, that knowledge the data controller doesn’t have himself and he could only get by using illegal means (e.g. by hacking into a biobank) is not attributable to him.

Second, the question arises, whose knowledge can be attributed to a data controller. Can only this kind of knowledge be attributed to a data controller, he himself actually has or could legally have access to? Or can also this kind of knowledge be attributed to a data controller, only a third person has access to?

---


17 But national transposition of the Directive 95/46/EC must not violate fundamental rights or principles like the principle of proportionality protected by Community Law, see: ECJ "Lindqvist" judgement of 06.11.2003, C-101/01: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62001J0101:EN:HTML, see summary 5 f. and holdings 87, 91 ff. The Member States may only take measures to ensure the protection of personal data that are consistent both with the provisions of Directive 95/46 and with its objective of maintaining a balance between freedom of movement of personal data and the protection of private life. However, nothing prevents a Member State from extending the scope of the national legislation implementing the provisions of Directive 95/46 to areas not included in the scope thereof provided that no other provision of Community law precludes it.
With regard to this question scientists in German legal literature predominantly hold the view, that only knowledge, which the data controller actually has or which is legally accessible for himself, can be attributed to him. In this respect, the German position is similar to the Austrian concept. The data controller would thus be free to deal with the data as he chooses, e.g. publish it on the Internet, as the Data Protection Directive would not be applicable in that case.

But this would enable for example law enforcement agencies or other third parties having a reference data set or link to re-establish the reference to an individual, e.g. by matching the data published on the Internet with data from their own database. The privacy of the data subjects, e.g. the patients taking part in ACGT, would be affected.

With regard to the Data Protection Directive, this opinion cannot convince. The question of whether certain additional knowledge is attributable to the data controller, and if, in consequence, a person is identifiable for the data controller, must be answered by statutory interpretation of the Directive.

Recital (26) states, that in order to determine, whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the data controller or by any other person to identify the said person. The interpretation of the wording of the Recital suggests, that also those means reasonably used by a third person can be attributed to the data controller. Without doubt, one of these means is the use of knowledge, which is legally accessible to this third person. The conclusion drawn from this directive-corresponding interpretation is that also knowledge, which is accessible only to a third person, can be attributed to the data controller.

A teleological interpretation of Art. 2 lit a) and Recital (26) of the Directive suggests, that our interpretation presented above is convincing. According to Art. 1 No. 1, the Directive aims to protect in particular their right to privacy with respect to the processing of personal data.

In the framework of a genetic research project, the data processor usually doesn’t have access to a reference data set to link his data used for research to a particular person. If only knowledge, the data controller actually has or could legally have access to, could be attributed to him, the data used for the research would have to be qualified as de facto anonymous data. The data processor could do with this genetic data whatever he wants, e.g. publish it or transmit it abroad.

As a result, third parties could access the data and re-establish the link to the said person, if they had a reference link to the person and an interest in the connected information. But this would be an infringement of the patient’s right

---


of privacy. The aim of data protection law and the Data Protection Directive 95/46/EC would be undermined.

For this reason it’s necessary, in accordance with the wording and the sense of the Directive, to attribute to the data controller also that kind of additional knowledge, to which only a third person has legal access. If a third party can legally access knowledge, which can be used to identify the said person, the genetic data concerned is personal data for the data processor as well, although the data processor himself cannot identify the person.

As consequence, the data processor would have to treat all genetic data as personal data in order to avoid responsibility, as he cannot know, whether there is a reference link to a person for a certain set of genetic data he uses available to a third party. Therefore the data controller would need an informed consent for each data processing operation as consequence of that opinion, since a legal basis is generally not available for this kind of data processing taking place within genetic research projects like ACGT.

On the one hand, the said person’s privacy would be effectively protected. But on the other hand, this interpretation would have the effect of a strong restriction on medical research, as an informed consent would be needed for each single data processing operation.

For this reason, the interpretation supported above must be applied restrictively. The privacy of the concerned data subject is not in danger, if, first, the data processor himself cannot legally access the additional knowledge of a third party and, secondly, the third party cannot access the data processor’s data used for the research. In these cases, when neither the data controller nor the third party can establish the link alone, the identification of the said person is not possible. Attributing additional knowledge of third parties to a data controller also in these cases would extend the scope of data protection legislation too far and would oppose the aim of data protection.

In conclusion, the attribution of additional knowledge of third parties depends on the data processing operation in question. If there’s a danger, that a third party can access the data processor’s data (e.g. following publication or transmission) and identify the said person, data protection legislation must provide effective protection of the individual’s privacy. Therefore, additional knowledge of a third party must be attributed to the data controller, if data processing causes any danger for the person’s privacy, e.g. in case of transmission or publication. In consequence this would mean, that for every transmission or publication of genetic data permission (by law or consent) is required, because the data processor cannot know, for which of the genetic data sets to be processed additional knowledge exists.

Data processing operations, which do not cause any danger for privacy, e.g. storage or use (research), do not require any consent or permission by law.
This opinion is also in accordance with Recital (26) of the Directive. Reasonably, a third person only uses means to identify the said person, if he can also access the data to be processed. If he cannot access the data, the third person does not reasonably use any means for identification, so that, following the directive-corresponding interpretation of Art. 1 No. 1 and recital (26) of the Directive, these means and also the third person’s knowledge cannot be attributed to the data controller with the result, that this data for him is de facto anonymous.

The above explained and supported opinion offers a solution for the attribution of additional knowledge, which not only corresponds with the Data Protection Directive and guarantees the privacy of the patients, but also has a practical orientation and promotes medical progress.