DATA PROTECTION: WHAT EVERY LAB MANAGER NEEDS TO KNOW

A new European law will have far-reaching consequences for laboratory information. Erik Vollebregt and Sofie van der Meulen discuss the potential impact of the proposed General Data Protection Regulation

Changes to key concepts in data protection law – personal data, anonymisation and the processing of personal data for research purposes – could impact data collected and processed in the laboratory. While the General Data Protection Regulation (GDPR) has not yet been formally adopted, it is clear that this new European law will have consequences for the use of personal data in laboratories.

Currently, the European Data Protection Directive (DPD) regulates the protection of personal data within the European Union.[1] Although it has been transposed into the national laws of all 28 member states, this legal framework – which dates back to 1995, is considered fragmented, outdated, and unclear. The European Commission therefore proposed the GDPR in 2012.[2] The aim was to update data protection rules and harmonise divergent approaches across the EU member states. The fact that the GDPR is a 'regulation' instead of a 'directive' means it will be directly applicable to all EU member states without the need for national implementing legislation. As of June 2015, a general approach to the GDPR has been agreed by the Council of Ministers of the European Union, creating a compromised position between the European Commission's and the European Parliament's draft of the GDPR. The final outcome of the current tripartite negotiations is expected by the end of 2015. The GDPR will likely enter into force two years after the date of publication.

SCOPE OF THE OLD DPD: PERSONAL DATA

Collecting and processing[3] of (personal) data may give rise to obligations under data protection law. According to recital 26 of the DPD (the existing directive), the principles of protection must apply to any information concerning an identified or identifiable person. Personal data is defined as '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'.[4] To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller[5] or by any other person to identify the said person.[6]

The DPD does not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.[7] Anonymised data would therefore be data that previously referred to an identifiable person, but for which identification has become reasonably impossible. This concept evolves over time, because what is reasonably impossible depends on the state of the art of decryption technology. When the data that is processed does not fall within the concept of 'personal data', the consequence is that the DPD does not apply, pursuant to Article 3. The DPD also has a separate category of 'sensitive personal data'. This is personal data that is given extra protection under the DPD, such as data relating to racial or ethnic origin, political opinions, and health or sex life.

RELEVANCE FOR LABORATORY DATA

Data generated in an analytical laboratory will often include personally identifiable information, because the results need to be linked to an individual. For example, laboratory informatics systems may well deal with samples taken from human subjects. Depending on the measures taken, the data may fall outside the scope of data protection law as soon as the data is anonymised. Nevertheless, prior to anonymisation, the data still qualifies as personal data covered by data protection law. Software used for the management of clinical trials and biobanks will also almost always process personal data.

SCOPE OF THE NEW GDPR: PERSONAL DATA

The draft text of the proposed new GDPR introduces additional definitions for 'genetic data',[8] 'biometric data'[9] and 'data concerning health'[10] apart from a revised definition of

'personal data'.[11] The definition of genetic data is of particular relevance for the laboratory environment, as it encompasses all personal data resulting from an analysis of a biological sample[12] from the individual in question within the scope of data protection law. Furthermore, the definition of 'data concerning health' raises the question, whether it is only intended to apply to personal data within the category 'health' or to all data related to health as the words 'personal data' are missing in this definition. Although the definition does refer to the physical or mental health of an individual, further clarification is necessary to understand the scope.[13]

Both 'genetic data' and 'data concerning health' are treated as sensitive personal data under the proposed GDPR[14], which means the processing of these data is prohibited unless an exemption is applicable. The most common exemptions are explicit informed consent by the patient, and processing in the context of treatment under a healthcare professional's duty of confidentiality.

ANONYMISATION: THE CURRENT STATE

Anonymisation is a technique applied to personal data in order to achieve irreversible de-identification. Therefore, the starting assumption is that the personal data must have been collected and processed (in order to anonymise it).[15] In this context, the anonymisation process, meaning the processing of such personal data to achieve its anonymisation, is an instance of 'further processing'. As such, this processing must comply with the data protection law, such as informed consent for processing.[16]

For data not to be considered as personal data within the scope of the DPD, it must be rendered anonymous in such a way that identification of the data subject is no longer possible.[17] The DPD itself does not provide further guidance on the concept of anonymisation, but the Article 29 Data Protection Working Party[18] adopted an opinion on anonymisation techniques on 10 April 2014.[19] The main anonymisation techniques, namely randomisation and generalisation, are described in this opinion. In particular, the opinion discusses noise addition, permutation, differential privacy, aggregation, k-anonymity, l-diversity and t-closeness. The opinion helps to choose how to design an adequate anonymisation process in a given context, and furthermore elaborates on the robustness of each technique based on three criteria:

- is it still possible to single out an individual?
- is it still possible to link records relating to an individual?
- can information concerning an individual be inferred?

Relating to these three criteria, an overview of anonymisation techniques is provided in the opinion – see panel below.

Pseudonymisation is also addressed; not a method of anonymisation, it merely reduces the linkability of a dataset to the original identity of a data subject. Accordingly, it is a useful security measure to reduce risk in relation to a set of personal data, but it is not a method for anonymisation of personal data.

The outcome of anonymisation as a technique applied to personal data should, in the current state of technology, be as permanent as erasure of the personal data. It should make processing of personal data impossible. The optimal solution for anonymisation should be decided on a case-by-case basis, possibly by using a combination of different techniques. Furthermore, anonymisation should not be regarded as a one-off exercise as even anonymised data – like statistics, – may be used to enrich existing profiles of individuals. A dataset considered to be anonymous may be combined with another dataset in such a way that one or more individuals can be identified, thus creating new data protection issues.

Currently, the use of personal data concerning health cannot be legally justified on the basis of conducting research only

The following example is described in Opinion 05/2014: *'Genetic data profiles are an example of personal data that can be at risk of identification if the sole technique used is the removal of the identity of the donor due to the unique nature of certain profiles. It has already been shown in the literature[20] that the combination of publically available genetic resources (e.g. genealogy registers, obituary, results of search engine queries) and the metadata about DNA donors (time of donation, age, place of residence) can reveal the identity of certain individuals even if that DNA was donated "anonymously".'*[21]

ANONYMISATION UNDER THE GDPR

According to recital 23 of the GDPR anonymous data remain outside the scope of the GDPR: *'The principles of data protection should therefore not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable. This Regulation does therefore not concern the processing of such anonymous information, including for statistical and research purposes.'* As tools and computational power evolve, it is neither possible nor useful to provide an exhaustive enumeration of circumstances when identification is no longer possible. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into account consideration of both available technology at the time of the processing and technological development.[22]

The Council defines 'pseudonymisation'[23] as: *'the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person (…).'* This definition introduces a pseudo-category of personal data, leaving it uncertain what standard applies.[24]

PROCESSING OF DATA FOR RESEARCH PURPOSES

Currently, the use of personal data concerning health cannot be legally justified on the basis of conducting research only. That is, use of personal data for research must be legally ➤

Anonymisation techniques

	(1) Is singling out still a risk?	(2) Is linkability still a risk?	(3) Is inference still a risk?
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenisation	Yes	Yes	May not

justified on the basis of (explicit) consent or another legal ground specified in Article 7 of the DPD. It should be noted that consent to inclusion in a clinical trial is not equal to the consent often required for the (further) processing of (sensitive) personal data[25], e.g. for inclusion in an aggregated dataset that is used for other research.

The GDPR introduces a specific legal ground[26] for processing personal data, which is necessary for archiving purposes in the public interest, or for historical, statistical or scientific purposes. The processing of such data is lawful if the conditions and safeguards under Article 83 GDPR are also met. If research relies upon the reuse of existing data sets, data controllers would need to demonstrate that the further processing of the data for research is compatible with the original purposes for collection of the data.[27] In this regard the Council's Article 5(1)(b) of the GDPR is helpful, as it provides that further processing of personal data for scientific, statistical or historical purposes that is in accordance with Article 83 – the reuse of data for research – would automatically be considered 'compatible' and in compliance with the principle of purpose limitation. The problem here is that the scope of consent initially obtained from the patient does not usually permit further processing, because no consent was obtained for anything else than a specific test. If compliance with Article 83 is not possible, another legal ground for processing needs to be satisfied. In practice, this legal ground may be found in obtaining consent[28] from the data subject. In recital 25 of the Council's text, the difficulty of identifying all scientific purposes at the time of data collection is acknowledged. '*Therefore data subjects can give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research.*' This seems to be a valuable recognition of a broad consent in the context of research, which is crucial for longitudinal studies and the application of big data analytics in research.

RECOMMENDATIONS

This article has discussed the impact of just a few key concepts of data protection law on laboratory informatics. In the light of the upcoming changes under the GDPR, we recommend laboratories to revisit what personal data is collected and processed and to determine whether it is caught by the personal data requirements.

The extended scope of sensitive data – including genetic data – attracts a greater protection under the GDPR. As unlawful processing of personal data may give rise to penalties up to two per cent to five per cent of worldwide turnover, the risk of non-compliance under the GDPR has to be taken seriously. To ensure compliance with data protection law in the future, the appointment of a data protection officer becomes mandatory under the GDPR for data controllers and processors that employ 250 persons or more, or that process the personal data of 5,000 or more people. Also, privacy impact assessments become mandatory for processing of personal data concerning health.

Erik Vollebregt and Sofie van der Meulen are with Axon Lawyers, Amsterdam

References

1. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
2. 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, COM(2012) 11 final.
3. Processing of personal data covers any operation or set of operations performed on personal data such as: collecting, capturing, ordering, saving, modifying, looking up, use, sending to, spreading by means of making accessible, bringing together, linking, hiding or destroying (partially) personal data. The DPD applies to both automated and manual data processing that is entered in a file or intended to be entered therein. The processing must be limited to only those activities, are necessary to fulfill the identified purposes for which the data were collected.
4. Article 2(a) DPD. See also Opinion 04/2007 of Article 29 Data Protection Working Party on the concept of personal data.
5. Definition of 'controller' in Article 2(d) DPD: 'the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.'
6. This means that a mere hypothetical possibility to single out the individual is not enough to consider the person as 'identifiable'. If, taking into account 'all the means likely reasonably to be used by the controller or any other person', that possibility does not exist or is negligible, the person should not be considered as 'identifiable', and the information would not be considered as 'personal data'. The criterion of 'all the means likely reasonably to be used either by the controller or by any other person' should in particular take into account all the factors at stake.
7. Recital 26 DPD
8. Definition (Council Preparation of a General Approach 15 June 2015): 'Genetic data' means all personal data relating to the genetic characteristics of an individual that have been inherited or acquired, (…) which give unique information about the physiology or the health of that individual, resulting in particular from an analysis of a biological sample from the individual in question.
9. Definition (Council Preparation of a General Approach 15 June 2015): 'Biometric data' means any personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of an individual that allows or confirms the unique identification of that individual, such as facial images, or dactyloscopic data.
10. Definition (Council Preparation of a General Approach 15 June 2015): 'data concerning health' means data related to the physical or mental health of an individual, which reveal information about his or her health status.
11. Personal data is defined as: *'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 identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.'(Council position 15 June 2015)*
12. Recital 25a (Council Preparation of a General Approach 15 June 2015) mentions: 'in particular by chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis or analysis of any other element enabling equivalent information to be obtained.'
13. See Recital 26 GDPR (Council Preparation of a General Approach 15 June 2015): 'Personal data concerning health should include (…) data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health of the data subject; including information about the registration of the individual for the provision of health services (…); a number, symbol or particular assigned to an individual to uniquely identify the individual for health purposes; (…) information derived from the testing or examination of a body part or bodily substance, including genetic data and biological samples; (…) or any information on for example a disease, disability, disease risk, medical history, clinical treatment, or the actual physiological or biomedical state of the data subject independent of its source, such as for example from a physician or other health professional, a hospital, a medical device, or an in vitro diagnostic test.'
14. Article 9 GDPR (Council Preparation of a General Approach 15 June 2015).
15. Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, WP216.
16. For example: further processing has to be compliant with the principle of purpose limitation. See Article 29 Data Protection Working Party Opinion 03/2013 on purpose limitation.
17. Recital 26 DPD
18. The Article 29 Data Protection Working Party was set up under the DPD. It has advisory status and acts independently. (See: http://ec.europa.eu/justice/data-protection/article-29/index_en.htm)
19. Article 29 Data Protection Working Party, Opion 05/2014 on Anonymisation Techniques, WP216.
20. See John Bohannon, Genealogy Databases Enable Naming of Anonymous DNA Donors, Science, Vol. 339, No. 6117 (18 January 2013), p. 262
21. Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, p. 10
22. Recital 23 GDPR (Council Preparation of a General Approach 15 June 2015).
23. Article 4(3)(b) (Council Preparation of a General Approach 15 June 2015).
24. The Article 29 Data Protection Working Party 29 urges the European Commission not to define pseudonymous data as a new subset of personal data allowing for derogations from obligations under the GDPR. (Letter from the Article 29 Data Protection Working Party on Trilogue to Ms Ver Jourova, Commissioner for Justice, Consumers and Gender Equality of the European Commission, 17 June 2015.)
25. If consent is given in a written document, and that document also concerns other matters, the consent for the use of personal data must be presented in a form that is clearly distinguishable from the remaining contents of that document.
26. Article 6(2) GDPR (Council Preparation of a General Approach 15 June 2015).
27. According to the principle of purpose limitation.
28. According to Council's Article (1)(a) GDPR unambiguous consent to the processing of personal data for one or more specific purposes is required. Furthermore, consent should be freely-given, informed and specific (for specific purposes).