

Ministry of Justice: Call for Evidence on EU Data Protection Proposals

Response by the Wellcome Trust

March 2012

KEY POINTS

- It is essential that Article 83 and associated derogations are maintained as the Regulation moves through the legislative process.
- Amendments to clarify and strengthen research provisions would be beneficial to ensure these achieve their intended purpose and do not inhibit important health research.
- Amendments are needed to ensure that the use of pseudonymised data in health research is regulated proportionately and to ensure clarity in the scope of the Regulation.

INTRODUCTION

1. Information from patient records provides the foundation for much health research, and offers significant potential to answer questions about the factors that influence health and disease. Information from patient records can be used for epidemiological research; to understand more about the causes of disease; to detect outbreaks of infectious diseases; to monitor the safety and efficacy of drugs and medical devices; and to study the effectiveness of treatments and interventions. Identifiable data is also used to identify participants for research studies. Researchers may wish to approach individuals in order to gain their consent to participating in a particular piece of research, for example the trial of a new treatment for a particular disease.
2. We welcome the opportunity to respond to this call for evidence since it is vital that the EU and UK can establish a regulatory framework that balances the rights and interests of individuals with the societal benefits of research using patient information.
3. This response has been informed by a legal analysis of the potential impacts of the Directive commissioned by the Trust and conducted by Lawford Davies Denoon. This detailed response supplements a joint statement from the Trust and other health research organisations setting out the impacts of the data protection proposals for the sector and including a number of case studies. Without further information on the intention and detail of requirements set out in the Regulation, it has not been possible to provide information on costs.

HEALTH RESEARCH IMPACTS

Research derogations

4. The Regulation appears to provide a number of derogations from particular requirements for the use of 'personal data' for scientific research, providing that personal data is processed in accordance with the conditions set out in Article 83. These derogations do not exempt research studies from all the requirements set out in the Regulation. The Wellcome Trust warmly welcomes this approach since it provides a framework that balances the facilitation of research with the protection of the interests of research participants. The Government must prioritise the protection of Article 83 and ensure the associated derogations for research are protected as the Regulation moves through the legislative process.
5. There are a number of issues around Article 83 and the associated derogations that would benefit from clarification to better reflect the intent of the clauses. The lack of clarity in the current UK Data Protection Act has contributed to a risk-averse culture among those sharing and using data for research. This has led to delays to research that would be in the public interest.
6. In order to avoid replicating these difficulties, it is essential that any lack of clarity is rectified in the new Regulation. The following clarifications are needed and suggested amendments to achieve these are set out in detail in box 1:
 - Clarification of Article 6.4 and Recital 40 to ensure that the processing of personal data for other purposes intends scientific research to be viewed as a compatible purpose in itself (amendment A).
 - Clarification that the reference to Article 83 (processing for historical, statistical and scientific research purposes) within Article 81 (processing of personal data concerning health) is intended to link the two sections, rather than to impose an additional restriction on research (amendment B).
7. A number of aspects of the research requirements and derogations rely on demonstrating 'necessity'.¹ While this approach is reasonable in principle, it will be important that an appropriate and consistent definition of 'necessity' can be applied in this context to ensure clarity in implementation.

¹ For example Articles 6.2; 9.29(i); 17.3(c); 83.1(a); and 83.2(c).

Box 1

Deletions are ~~struck through~~
Insertions are underlined

Amendment A

Clarify that scientific research is a compatible purpose *per se*:

Article 6.4: revise by appending the following sentence: "Processing necessary for historical, statistical or scientific research purposes shall always be deemed compatible processing, provided it is conducted with the rules and conditions laid down in Article 83."

Recital 40: The processing of personal data for other purposes should be only allowed where the processing is compatible with those purposes for which the data have been initially collected, such as ~~in particular~~ where the further processing is necessary for historical, statistical or scientific research purposes. Where the other purpose is not compatible with the initial one for which the data are collected, the controller should obtain the consent of the data subject for this other purpose or should base the processing on another legitimate ground for lawful processing, in particular where provided by Union law or the law of the Member State to which the controller is subject. In any case, the application of the principles set out by this Regulation ~~and in particular the information of the data subject on these other purposes~~ should be ensured.

Add a new Article to allow Member States to enact delegated legislation regarding processing for scientific research being a legitimate interest.

Amendment B

Clarify that the reference to Article 83 in Article 81.2 is not intended to impose an additional restriction on processing for research purposes, but rather is intended to connect the two sections. This could be achieved by moving Article 81.2 to 81.1(d).

Scope of the Regulation

8. The scope of the Regulation is 'personal data' that identifies a natural person, or from which a natural person can be identified.² It is important that the research community is clear about how 'personal data' relates to the different types of data used in research – anonymised data; key-coded or pseudonymised data; and identifiable data (see Annex A) – since the scope determines which research studies are brought within its remit and therefore must comply with its requirements. Clarity in the scope is essential so that those sharing and using patient data in research are fully aware of their responsibilities, but do not implement beyond the requirements that are necessary in law.
9. The Regulation is not explicit on whether pseudonymised data are intended to be included within its scope. Pseudonymised or key-coded data underpin a substantial amount of research, for example studies at the Wellcome Trust Sanger Institute and the UK Biobank research resource. Inclusion of pseudonymised data within the scope will therefore dramatically increase the regulatory burden on research.

² Articles 3 and 4

10. The use of pseudonymised data in health research is well-established and operates within a system designed to reduce the possibility of re-identification of participants. It is important that the use of pseudonymised data in research is handled within a proportionate regulatory framework that takes into account the actual likelihood of re-identification under current conditions, not just the technical possibility of re-identification. Conditions that will reduce the actual likelihood of re-identification could include the use of 'safe havens', such as England's new Clinical Practice Research Datalink and comparable services in the devolved nations, and contractual data sharing agreements and professional standards for researchers that prohibit re-identification. In many instances the identifying code will not be held at the research site where the pseudonymised data are used in research, but at a hospital or by a safe haven. Article 23 should be amended to provide greater clarity on this issue for research by noting that conditions could be established in a Member State that preclude re-identification, therefore ensuring that re-identification would not be considered "reasonably likely" (amendment C).

Amendment C

Clarify that the likelihood of identification of the individual (Recital 23) should take into account the conditions of access to data so that if the conditions of access for research purposes preclude re-identification then re-identification would not be "reasonably likely".

11. Anonymous data falls outside of the scope of the Regulation. However, the act of removing identifiers to ensure that data are no longer personal – anonymisation – could fall within the definition of processing (Article 4). This would mean that the process of anonymisation itself would have to comply with the requirements of the Regulation to be lawful. We suggest that the Regulation should be revised to expressly permit anonymisation, while prohibiting re-identification for data that has been anonymised (amendment D).

Amendment D

Expressly permit processing for the purpose of anonymisation (new definition of anonymisation and new Articles 6.1(g) and 9.2(k)) together with a new prohibition of re-identification for data that has been anonymised:

'anonymisation' means any steps taken towards processing personal data in such a manner that it can subsequently no longer be considered personal data;

6.1(g)/9.2(k): "processing is conducted for the purpose of anonymisation"

"the processing of data which has undergone a process of anonymisation with the objective of re-identifying a data subject or otherwise rendering such data personal data shall be prohibited except where any of the conditions in Article 9.2(a), (c), (e), (g), (h) or (i) apply."

12. Clarification is needed around 'genetic data' and 'data concerning health' to ensure that these definitions are only intended to apply to personal data that falls within these categories, rather than all related data (amendment E).

Amendment E

Deletions are ~~struck through~~

Insertions are underlined

Clarify that the Regulation does not capture non-personal data in the definitions of *data concerning health* (Article 4.12) and *genetic data* (Article 4.10):

'data concerning health' means personal data ~~any information~~ which relates to the physical or mental health of an individual, or to the provision of health services to the individual;

'genetic data' means ~~all~~ personal data, ~~of whatever type,~~ concerning the hereditary genetic characteristics of an individual which are inherited or acquired during early prenatal development;

13. Furthermore, the definition of 'data concerning health' must be consistent with the related Recital. Recital 26 includes "information derived from the testing or examination of a body part or bodily substance, including biological samples" in its description of personal data relating to health. However, no reference is made to biological samples in the definition at Article 4.12. This inconsistency should be rectified to clarify that data concerning health does not include biological samples *per se*, but rather to personal data obtained from testing such material (amendment F).

Amendment F

Deletions are ~~struck through~~

Insertions are underlined

Amend Recital 26 to clarify that the Regulation does not capture biological samples *per se* within its scope, but personal data derived from these samples. For example:

Draft Recital 26: Personal data relating to health should include in particular all personal data pertaining to the health status of a data subject;...; ~~information~~ personal data derived from the testing or examination of a body part or bodily substance, including biological samples;...

Increases in the regulatory burden for health research

14. Beyond potential increases in scope, the Regulation increases the regulatory burden for health research compared to the current Data Protection Directive and the corresponding UK Data Protection Act (1998). If implemented, these additional burdens will make it increasingly difficult to conduct this important research in the UK and EU. The following

issues present particular concerns and potential amendments to address these are presented in Box 3:

- Article 5(e) on **data storage** provides a welcome derogation that enables data to be held for extended, potentially indefinite, periods for research purposes. However, this derogation imposes a requirement to undertake periodic review to assess the necessity to continue storage. These reviews would be impractical since data are routinely held over long periods and it can be difficult to predict future uses or need for the data. Furthermore, these reviews would create a substantial burden for research institutions that currently hold valuable data and research resources, which may not be sustainable for the sector. We therefore recommend amending Article 5 to remove the need for such review (amendment G).
- The **right of the data subject to information** (Article 14) could be problematic for research in situations where notifying participants would create a disproportionate burden that could prevent the research from proceeding. The Regulation includes a 'disproportionate effort'³ provision, but this only applies where the data are not collected from the data subject. It would be helpful to clarify the situation for research by amending this Article to create a specific 'disproportionate effort' provision for research, in line with the current Data Protection Directive (amendment H).
- The **right to rectification** (Article 16) is inherently problematic for health research since researchers routinely hold data generated through their studies that cannot be guaranteed to be accurate. For example, data generated by genetic sequencing in the laboratory environment will rarely meet diagnostic standards used in a clinical setting. As a result, such data cannot be considered analytically accurate. In addition, a person's health status changes over time, for example pregnancy. The Regulation does not contain any guidance as regards practical means for researchers to assess or rectify such "inaccuracies". The Regulation should be amended to take this reality of health research into account (amendment I).

Box 3

Deletions are ~~struck through~~
Insertions are underlined

Amendment G

Amend Article 5.3 to remove the need for periodic review to resolve practical issues and the burden this creates for research. For example:

Draft Recital 5.3(e): [Personal data may be] kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data will be processed solely for historical, statistical or scientific research purposes in accordance with the rules and conditions of Article 83 and ~~if a periodic review is carried out to assess the necessity to continue the storage~~ until it becomes apparent that continued storage is no longer necessary.

³ Article 14.5(b)

Amendment H

Amend Article 14 to ensure that there is express 'disproportionate effort' provision from notification for research. For example:

Draft Recital 14.5 (xx): the data are processed for historical, statistical or scientific research purposes and the provision of such information proves impossible or would involve a disproportionate effort.

Amendment I

Amend Article 16 or provide a Recital to reflect the fact that data generated in research cannot be guaranteed to be accurate and to propose limits as regards the steps that researchers should be required to take to assess or to rectify any potential inaccuracies.

The Wellcome Trust is a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health. We are independent of both political and commercial interests

ANNEX A: THE TYPES OF PATIENT DATA USED ON HEALTH RESEARCH

Health data can be accessed by researchers in the following forms:

- **Identifiable data** – these include information in patient records such as patients' names, addresses, dates of birth and NHS numbers. There are also aspects of health data that could become identifying when they relate to a diagnosis of a rare condition or when combined with other data. Identifiable data are needed when future contact is needed with the participant, for example to contact them to take part in a study, or to link information across different data sets.
- **Key-coded or pseudonymised data** – these cannot directly identify an individual, but are provided with an identifier that enables the patient's identity to be re-connected to the data by reference to a separate database containing the identifiers and identifiable data. Pseudonymised data can often be used in place of identifiable data.
- **Anonymised data** – these data cannot be connected to the original patient record. Anonymised data are suitable when no contact is needed with the participant or where the data does not need to be linked to any other data sources.