Information Commissioner's Office

# ICO response to the Nuffield Council on Bioethics consultation on

# The linking and use of biological and health data

8 January 2014

ico.

Information Commissioner's Office

# Contents

## About the ICO

**The ICO's mission is to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.**

The ICO is the UK's independent public authority set up to uphold information rights. We do this by promoting good practice, ruling on complaints providing information to individuals and organisations and taking appropriate action where the law is broken.

The ICO enforces and oversees the Freedom of Information Act, the Environmental Information Regulations, the Data Protection Act and the Privacy and Electronic Communication Regulations.

# Introduction

1. The Information Commissioner’s Office welcomes the opportunity to respond to the Nuffield Council on Bioethics’ consultation on the linking and use of biological and health data.

2. We approach this consultation from the perspective of the regulator of the Data Protection Act 1998 (DPA). That is to say, our primary concern in the linking and use of biological and health data is whether there are any implications for compliance with the requirements for the processing of personal data, and sensitive personal data, under the DPA. For example, how it may impact on fairness and privacy rights. We do not approach it with any expertise on the ethical issues that are raised, although there will often be a clear link between them and the data protection considerations that come into play.

3. This is recognised in ‘Background to the consultation’, for example, the reference to the legal implications of the developments in the areas of health-related research, clinical practice and governmental activities and that the ethical issues “are often articulated in relation to the concepts of privacy and the public interest.”

4. One of the key issues for us is that there is a basic distinction between information that identifies individual people (and so comprises personal data for the purposes of the DPA) and information that does not. However, we do recognise that the distinction between personal data and non-personal data is not always straightforward. We recommend that, wherever possible, personal information should be anonymised, in particular in circumstances where information from a number of different sources is being linked and there is an increased possibility of individuals being identified with the associated risks to their privacy. In such circumstances, there must be a strong commitment to anonymisation by default. Further detail on this is provided in our Anonymisation Code of Practice.

# Specific questions

## Question one: Do biomedical data have special significance?

1. From the perspective of the DPA, biomedical data that relates to a living individual is especially significant in the sense that it describes something deeply personal about somebody's health and family relations. It has the potential to reveal some of the most private information about an individual and his/her family. The natural reaction of society is that this is very sensitive information with significant implications for privacy rights if it is processed inappropriately. For example, people will be aware that such information has the potential to be used in a predictive and/or discriminatory way to the detriment of individuals. This is also reflected in the DPA which recognises that "sensitive personal data", of which biomedical data relating to an individual is one example, needs to be treated with greater care than other personal data.

2. This should be borne in mind when undertaking any linking of biomedical and health data, both when the information is sensitive personal data under the DPA and also when individual items of data are initially anonymised but the act of linkage increases the risk of re-identification of individuals. This emphasises that, in view of the numerous possibilities, and demand, for the data created through such linkage, there are special risks that need to be considered before proceeding. It also serves as a reminder of the importance of retaining anonymity, and organisations should take the necessary steps to ensure that individuals cannot be identified as a result of the linkage of information. This requires careful consideration of whether, on release of anonymised information, it is reasonably likely that an individual could be identified from those data together with other information that is available.

3. In the case of genomic datasets there will be circumstances where data potentially relates to more than one family member. In terms of the DPA this means that there will be more than one data subject and so data controllers would have to be alert to this and ensure that the DPA is complied with in respect of all of them. Clearly, there are particular privacy issues where the collection of biomedical data about one person reveals personal data about another – although 'ordinary' health data can also do this.

## Question two: What are the new privacy issues?

1. In broad terms new technologies and big data science do not present new privacy issues in the sense that the fundamental requirements for transparency – being as open as possible about how an individual's data will be used - and no inappropriate use or disclosure remain. Moreover, the algorithms used to analyse a high volume of data, and combine data, increase the risk of anonymised personal data being re-identified such that the data becomes associated with specific individuals and regains the characteristics of personal data. So, whilst no new privacy issues are raised per se, it is the new technology and science which means that a large amount of data has the potential to be subject to privacy laws where it wasn't before.

2. The scale of the data, and the increasingly wide range of uses to which it can be put, does raise certain concerns in terms of compliance with the DPA and privacy issues more broadly. For example, due to the large volume of data that is analysed in big data science, together with the nature of the analysis, a degree of unpredictability is introduced in terms of the uses to which the data will be put. As a result it will be more difficult for people to give meaningful consent to the processing of their data at the time it is collected as the future purposes of the processing will not have been determined. Nevertheless, we would recommend that individuals should be advised at the outset, by means of a well-drafted privacy notice, that new risks might arise with the advent of new technologies. This can also be done as part of the normal patient / research subject's relationship with the healthcare professional / researcher.

3. It is a well-established principle that for consent to be meaningful it must be given in the context of a specific purpose. However, with big data it is very likely that data collected for one purpose will be used for another at some future point. Not only does this raise issues of fairness in terms of the first data protection principle, as individuals are not told of future uses to which their data will be put, but compliance with the second principle is also unlikely if such future purposes are incompatible with the original purpose. This can also be described in terms of an individual's control over his personal data. As information becomes linked with other databases and it becomes multiplied and shared widely, an individual is more likely to lose control of his information. This being the case, it will be necessary to consider whether the outcomes of the data linkage and big data science offer some sort of trade-off for this loss of control, for example in terms of the wider benefits to society. The key issue here is therefore to put the procedures in place to ensure that individuals know what they are agreeing to and that their personal data is used for a defined and limited set of purposes. (The use of properly anonymised information in big data type contexts is far less of an issue.)

4. As well as considering the data protection principles, it is important also to take into account the common law duty of confidence when personal information is used for other purposes. Indeed, this may prove to be a more powerful barrier to the linking of personal data and its use for future unspecified purposes than the data protection principles. (Where a confidence is breached this is also unlawful processing for the purposes of the DPA – although we suspect more people working in this field look to duties of medical confidentiality prior to DP compliance.)

5. The issue of control is also relevant in the context of big data being used to make decisions about individuals. As mentioned in Part One of the consultation documentation, developments in the use of data, algorithms and predictive analytics are informing health interventions and other health-related decisions in respect of individuals. The potential consequences for individuals in this regard, together with the advances in scientific techniques and the related ethical issues mean that data controllers need to be innovative in terms how they explain to individuals the uses to which their information will, or may, be put. As an example, the complexity of the algorithms and analytics may also make it difficult for data controllers to fulfil their duty to explain to a data subject the basis for an automated decision made about them.

6. In terms of harm caused by the use of biomedical data, it may be difficult to articulate what this may be in data protection terms, in particular as future use of data following data linkage and the application of big data science may not be clearly anticipated. However, we can be clear in terms of what we are trying to avoid, namely open, matchable and vulnerable databases of weakly anonymised data with their inherent risk to privacy, for example in view of the increased possibility of re-identification. Such databases are likely to evidence a lack of clarity of purpose in their creation. It is worth referring in general terms to the potential personal and societal benefits that may be achieved and how these should be balanced against any specific or hypothetical harm that may be caused to the individual. This also relates to the issue of the relationship between fairness to an individual in data protection terms and the wider public interest. This is a difficult balance to assess – there is no easy formula to calculate where the balance lies – but it does at least indicate that it will always be necessary to consider the impact on individuals of any processing of their personal information. Even if such processing results in uses of data that will not affect individuals, if this involves the processing of their personal data, the DPA still requires those individuals to be made aware of it.

7. As to whether it would be helpful to treat biomedical data as ‘property’, this is something that does not fall within the Commissioner’s regulatory jurisdiction. As regulator of the DPA our concern is much more with the responsibilities that the legislation imposes on data controllers as enshrined in the eight data protection principles. However, we are aware that the concept of personal information as property has become a topic for discussion, and the issue of patient “ownership” of their data

is something that is perhaps worthy of further consideration. For example, as with other property rights, the data should only be used subject to limits imposed by the 'owner' of the personal biomedical data and cannot be transferred or assigned without the consent of the individual. In the same way, though, a balance can be struck in the sense that in return for giving up 'property' rights an individual receives some value – for example, improved health outcomes – in return. It may be possible to argue that the treatment of biomedical data as 'property' reinforces data protection objectives such as the encouragement of privacy enhancing technology and data security.

## Question six: What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

1. This question raises more data protection issues, for example in terms of fairness and reasonable further use of personal biomedical data. As we have seen, the first two principles of the DPA require the processing of personal data to be fair and lawful and that what organisations do with the information is in line with the reasonable expectations of the individuals concerned. This means that if an organisation wishes to use or disclose the personal data for any purpose that is additional to or different from the originally specified purpose, it must ensure that the new use or disclosure is fair. In practice this will mean that organisations should contact patients to, at the very least, inform them of the new use, but more likely will require them to obtain consent. However, there will be circumstances where personal information which has been collected and stored for the purposes of either health care provision or biomedical research can be lawfully used for other purposes without the need to comply with the first data protection principle and with no breach of the DPA, for example for the prevention and detection of crime.

2. In terms of compatibility under the DPA, we would say that, when deciding whether disclosing personal data is compatible with the purpose for which it was obtained, an organisation should bear in mind the purposes for which the information is intended to be used by any person to whom it is disclosed. However, recognising that it can be difficult to distinguish clearly between purposes that are compatible and those that are not, we would say that the focus should be on whether the intended use complies with the fair processing requirements of the DPA. In other words, if that use is 'fair' then it will also be compatible. Note that section 33 of the DPA provides an important exemption in relation to research, in that the further processing of personal data only for research purposes is not considered to be incompatible with the purpose for which it was obtained.

3. In general, though, if an organisation wishes to use or disclose personal data for a purpose that was not contemplated at the time of collection (and therefore not specified in a privacy notice), it has to consider whether this will be fair. If using or disclosing the information would be unfair because it would be outside what the individual concerned would reasonably expect, or would have an unjustified adverse effect on them, then you should regard the use or disclosure as incompatible with the purpose you obtained the information for. In practice this will often require the organisation to get prior consent to use or disclose personal data for a purpose that is additional to, or different from, the purpose it originally obtained it for.

4. Clearly, if data were to be used, say, for insurance purposes or for marketing, this would be incompatible with the original purpose. In order for such uses to be compliant with the DPA, the individual would have to be informed and his consent obtained for use for the new purpose. However, it is accepted that this is unlikely to happen in the context of information that was originally obtained for biomedical purposes. This is more likely to occur, certainly in the case of insurance, in the context of a direct relationship between the data subject and the insurance provider, i.e. where the individual is applying for insurance cover.

5. However, when new uses/purposes are being proposed for health data, although the data protection principles have an important role to play, it is likely to be the case that the common law duty of confidentiality that will act as a stronger barrier. (Although, as mentioned above, a breach of confidence will comprise unlawful processing and so be a breach of the first principle under the DPA.) This also links to the important issue of patient control of their data, in that strict limits on the use of biomedical data as a result of confidentiality mean that if further uses of the data are proposed then the patient must be advised of this and his consent obtained before such uses can proceed. Organisations need to consider as a matter of policy how much control patients should have over the information they supply, and be open with them regarding this. The degree of control that patients have over the data need not prevent that data being put to other uses. As we have mentioned, there are mechanisms within the DPA that legitimise alternative uses of patient data, such as consent.

6. As we have indicated, we are not in a position to comment on ethical issues, but the use of predictive analytic techniques is likely to have data protection implications. For example, if such techniques are adopted using an individual's biomedical and other health-related data to predict the likelihood of future health, or even behavioural, issues, this could have an impact in the employment context if a data controller made employment decisions on the basis of such predictions. From a data protection perspective, this would also raise issues of compliance with the fairness aspect of the first data protection principle. Similarly, if

insurance companies were to make decisions on the basis of such predictive analysis, it would also be necessary to consider whether this was fair to an individual applicant. As predictive technologies become more reliable there is an argument to suggest that it is reasonable to make decisions on this basis and that as long as an individual is made aware that the predictive techniques are being used fairness is ensured. However, it is questionable whether transparency is sufficient in this context as issues such as social exclusion may well have an impact on whether or not such processing of personal data is in compliance with the DPA. As well as fairness and principle one, it may also be necessary to consider whether there are any issues relating to compliance with principle three, which states that "personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed."

## Question seven: What legal and governance mechanisms might support the ethical linking of biomedical data?

1. We would recommend adopting a governance structure which addresses the practical issues arising from the need to anonymise the biomedical data in order to avoid, as far as possible, the re-identification of individuals when such data is linked. In our Anonymisation Code of Practice there is a chapter on governance which provides advice on the procedures and mechanisms that can be put in place in this regard. These include senior responsibility, staff training, knowledge management, sectorial co-ordination and more specific arrangements such as the adoption of privacy impact assessments, transparency measures and regular reviews of the consequences of the anonymisation programme. Techniques such as the use of trusted third parties and re-identification testing can also be built into the governance structure.

2. As regards transparency, even though anonymisation has no direct impact on individuals, it is still important for individuals to be told about the approach of an organisation to anonymisation. For example, if there are any risks associated with anonymisation these should be described, and the safeguards that may be in place, such as disclosure to a limited number of recipients, should also be explained. Transparency and patient engagement can work well together. For example, if the fairness requirements of the DPA are complied with, the uses to which the data are put are reasonable and the patient retains a reasonable amount of control, patients can be effective partners in the research that is undertaken. In such ways, positive results can be achieved both in terms of ethics and legal compliance.

3. An adequate level of IT security is clearly necessary as part of an organisation's overall mechanisms for supporting the linkage of biomedical data. Should there be weaknesses in IT security, this may

compromise the anonymisation techniques that have been adopted and so, for example, may increase the risk of re-identification.

4. In terms of consent, please again refer to our Anonymisation Code of Practice where we discuss whether consent is needed to produce or disclose anonymised data. We make the point that it is a safer proposition to publish the data in anonymised form in view of the difficulties where consent originally given for the processing of personal data is withdrawn, as there may be circumstances where this has no effect. . In order to comply with the DPA, the processing of personal data, which includes its anonymisation, must be legitimised. Although this can be achieved by consent, there are other alternatives and, in any event, it is our view that the DPA will not prevent the use of a privacy-enhancing technology such as anonymisation.

# Summary

In summary, the issues we would particularly like to emphasise are:

- Researchers and others should be encouraged to find alternatives to using personal data wherever possible. Producing individual-level, linkable but non-personally identifiable data for use in research and other contexts can be challenging.

- Organisations collecting personal data from patients and others need to find accessible, innovative ways of explaining the purposes for which the data will be used – and the consequences of this – to individuals. This becomes more difficult as information systems become more complex, with more data-linkage and sharing.

- Those collecting personal data need to strike the right balance between being so clear and specific about how data will be used, whilst leaving the door open to future new uses of the data.

- Organisations need to decide how much control individuals should have over their personal data. Consent gives individuals control but obtaining consent is not always feasible. This is a policy call for organisations and the DPA provides alternatives to consent.

The issue of the ‘ownership’ of personal data is coming more to the fore. If, for example, a pharmaceutical company benefits from the analysis of a patient’s personal data, how should the patient benefit from this? How do we deal with patients who may not want their personal data – or even anonymised data derived from it – to be used in a particular way?

## Nuffield Council on Bioethics: The linking and use of biological and health data

For further information on this submission, please contact Policy Delivery, ICO, casework@ico.org.uk

If you would like to contact us please call 0303 123 1113.

www.ico.gov.uk

Information Commissioner's Office
Wycliffe House, Water Lane
Wilmslow, Cheshire, SK9 5AF

ico.
Information Commissioner's Office