

BBMRI-ERIC



We must urgently clarify data-sharing rules

Scientists have worked hard to ensure that Europe's new data laws do not harm science, but one last push is needed, says Jan-Eric Litton.

In a little over 12 months, the European Commission will roll out a new legal framework to govern the protection of personal data. There were many debates and discussions about the controversial regulations, which passed last year, and scientists and scientific bodies raised concerns over restrictions that the framework could have placed on the use of research data. We won several concessions, but the fight is not over yet. Scientists must now come together to work out a consistent way to implement the rules, and they must do so quickly.

Once in place, the European regulations will have to be interpreted by lawyers, administrative staff and others across a diverse patchwork of legal systems and cultures. To smooth their introduction and implementation, the European Commission is encouraging organizations that represent data users to prepare formal codes of conduct that would set out — in simple language — what can and cannot be done. Scientists need to help prepare these codes to ensure that the hard-won concessions for research are not lost in translation.

Legal texts are not easily accessible to non-lawyers. By developing codes of conduct that are as understandable as possible, we can help to guide researchers and administrative staff, reduce unnecessary fear about compliance and enhance data sharing for the sake of progress in research.

For example, the rules will allow data to be re-used for research, even when they were collected for another purpose, and enable personal data — that is, data about people who can be identified from those data — to be stored for longer periods of time for research than for other purposes. And as long as certain conditions are met, they allow researchers to use sensitive personal data collected for other reasons, such as health data, without seeking extra consent.

To benefit from these special rules, researchers must comply with safeguards to ensure that the use of personal data in research is proportionate — for example, they should make sure that anonymous data could not be used instead — and that individuals' data are used responsibly, safely and securely.

The problem is that many of these terms are ill-defined and leave too much room for interpretation by lawyers and others, especially across countries. Take the difference between anonymization and pseudonymization. With pseudonymization, data can be attributed to individuals using 'additional information' (such as a key or encryption code), whereas with anonymized data such information is not available. So far, so clear. However, in some countries, such as the United Kingdom, pseudonymized data (with the personal identifying tags removed) are widely considered anonymous once they have been passed along to researchers, who do not have access to those tags. This

is not the case in some other countries, including Germany.

These differences in emphasis and understanding could seed doubts when scientists and research groups ask to share others' data. As a consequence, it could take endless amounts of time to agree on the detailed conditions for sharing, and the costs of projects that require large, pooled data sets would rise.

It is thus not just scientists who would benefit from a code of conduct that examines and clarifies these possible grounds for confusion. Citizens taking part in research projects need and deserve a concise and understandable overview of how their sensitive data are handled in an appropriate and timely manner and protected against misuse. For this reason, we should be aiming for a code of conduct that is applicable to as many research projects as possible, to enhance transparency throughout research in Europe.

Next week, BBMRI-ERIC — which operates and is developing a pan-European distributed research infrastructure of biobanks and biomolecular resources — will hold an event to kick-start discussions on such a code. Framed as a discussion and consultation on a harmonized approach, the working meeting will bring together representatives from the European life-sciences research infrastructures, policymakers, medical associations, industry representatives, patient-advocacy groups and other interested stakeholders. We hope they will agree on a road map to develop a code, and commit to doing so.

There really is no time to lose. For the code to be in place when the regulation enters into force next year, we need to publish a draft for public consultation well before the end of this year.

A good example of what medical research using shared data could achieve, and how this might be under threat, can be found in the search for treatments for rare diseases. In the European Union, about 30 million people are affected by one of the 6,000 known rare diseases. It is clear that one country alone is unlikely to have enough cases to study any one rare disease. Linking with other data sets across research centres, countries and diseases is the only way to make progress.

Another important consideration is that the EU is planning to spend €6.7 billion (US\$7.2 billion) on a European Open Science Cloud initiative, which means that data sharing has to work across borders. If it does not, the cloud will be very cloudy indeed.

We invite the scientific community to become part of the important final step in this long-running saga. Stay tuned for the consultation process and contact BBMRI-ERIC to be kept informed. ■

Jan-Eric Litton is director-general of BBMRI-ERIC in Graz, Austria. e-mail: contact@bbmri-eric.eu

DIFFERENCES IN EMPHASIS AND UNDERSTANDING COULD SEED DOUBTS WHEN SCIENTISTS ASK TO SHARE OTHERS' DATA.