NSHG-PM

Nordic Society of Human Genetics and Precision Medicine

21 December 2020

Comments on EDPB Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data

In recognition of EDPB Recommendations 01/2020, the legal working group of the Nordic Society of Human Genetics and Precision Medicine (NSHG-PM) respectfully submits comments for technical, organizational, and legal supplementary measures to achieve an essentially equivalent standard of protection when transferring personal data for medical scientific research to collaborators outside the EEA.

Our aim is to ensure that medical scientific research will still be feasible to conduct and that research participants' fundamental rights will be respected in the process. Regrettably, we do not believe that the Recommendations in the current version achieve this aim.

While a goal of the Recommendations may be to ensure that research participants' fundamental rights will be respected, in their current version, the Recommendations will hinder a large proportion of medical scientific research. Hence, we respectfully submit suggestions for a manner in which medical scientific research can be conducted while research participants' fundamental rights remain protected. We also include a case example.

We would welcome the opportunity to expand on these comments and to provide case studies/examples for further clarification.

On behalf of the NSHG-PM legal working group,

Heidi Beate Bentzen
University of Oslo
h.b.bentzen@medisin.uio.no
Phone +47 22 85 00 86

Comments on EDPB Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data

The Nordic countries have some of the world's most comprehensive public registry, health, and genetic databases combined with large biobanks. Building on our experience in processing and protecting such data, and our firm commitment to ensure that the data are used only for their intended purposes, we respectfully submit comments on EDPB Recommendations 01/2020 with suggestions for measures we believe are suitable not only for scientific research, but more generally to protect data transfers to non-EEA countries.

We note the current lack of *operational* appropriate safeguards according to Article 46 GDPR for many of the data transfers necessary for scientific research, but focus here on the situation in which a data transfer mechanism is in place, and on specific comments on the Recommendations.

Pseudonymization

Medical scientific research relies heavily on processing of vast amounts of genetic data, and on processing other large, high dimensional datasets, including large epidemiological, clinical or registry data. Such data cannot be anonymized because anonymization would sacrifice the utility of the data, and make it impossible to do follow-up studies, for example to see who has developed the disease in question or has died, or to combine findings from multiple studies on the same individuals. It would also in many instances be impossible to anonymize the data as the data would often be uniquely identifiable or too rich to exclude identification through linkage with other datasets. Anonymization furthermore makes it impossible to fulfill a duty of care where a finding is crucial to communicate to save a research participant's life, which is a legal obligation in some jurisdictions.

Therefore, data used for medical scientific research are instead pseudonymized in the manner pseudonymization is defined in Article 4(5) GDPR. Use case 2, paragraph 80-83 of the Recommendations, concerns transfer of pseudonymized data, for instance for scientific research purposes. However, *the criteria for pseudonymization in this section far exceed those of the pseudonymization criteria of the GDPR*, making it impossible to use pseudonymization as a supplementary measure for transfer of high dimensional or uniquely identifying data, such as whole exome or genome sequences.

Criteria 1 and 4 of use case 2 raise particular concern, and are phrased as anonymization rather than pseudonymization criteria. The first criterium, read in combination with paragraph 81, raises the question as to whether genetic and many other clinical and epidemiological data sets according to the Recommendations can rely on pseudonymization as a supplementary measure at all. The criterium requires that personal data can no longer be attributed to a specific data subject, nor be used to single out the data subject in a larger group, adding in paragraph 81 that genetic, physical, physiological, mental, and other factors may allow identification of the person even if identifiers are omitted. This goes beyond the GDPR definition of pseudonymization, where it is possible to pseudonymize also genetic and high dimensional data. Furthermore, in some instances, we must work with very small numbers. For example, cancer of the placenta is very rare, as is cancer of the heart muscle or hypopharyngeal cancer. These can go down to one case in a Nordic country per year. The one case in the country singles out and uniquely identifies one person in the world. If we are to follow survival rates of rare cancers across larger populations, then individual data on these cases must also be reported on an individual level to the World Health Organization International Agency for Research on Cancer. We suggest that criterium 1 is rewritten in line with Article 4(5) GDPR, specifically that the "nor be used to single out the data subject in a larger group" is omitted so that the criterium will read: "*a data exporter transfers personal data processed in such a manner that the personal data can no longer be attributed to a specific data subject, without the use of additional information*".

The fourth criterium in use case 2 is challenging; that the controller has to establish through thorough analysis that the personal data cannot be attributed to an identifiable individual, taking into account any information the public authorities of the recipient country may possess. The requirement is very broad, and it is impossible for individual controllers to determine what data the public authorities of a third country may possess. It also seems unreasonable to ask a controller to speculate on the strength of the investigative powers that public authorities possess. For genetic data, which by its nature is familial, recreational genealogy databases available publicly on the internet pose a particular concern in terms of reidentifiability in some populations. If all such sources must be considered, that would make the assessment substantial and likely make it impossible to fulfil the fourth criterium. By requiring *any information* to be taken into account, the requirement goes beyond the GDPR level of identifiability, as elaborated in Recital 26 GDPR, according to which account should be taken of all the means reasonably likely to be used. We therefore suggest the following rephrasing: "...taking into account any information that the public authorities of the recipient country *may reasonably likely possess and that could reasonably likely be used*, that the pseudonymized personal data cannot be attributed to an identified or identifiable natural person even if cross-referenced with such information".

If the EDPB against our advice should decide to keep the criteria of Use Case 2 as is, we strongly recommend not using the term 'pseudonymization', as that creates confusion when the criteria are not aligned with those of the GDPR. 'Pseudo*ano*nymization' is also not a good term. Perhaps 'strict relative anonymization' might work, but we would then encourage the EDPB to explain how C-582/14 *Breyer* and C-434/16 *Nowak* relate to the use case.

If keeping Use Case 2 as is, we respectfully ask the EDPB to explain how personal data can be transferred for medical scientific research purposes as the criteria of the Recommendations render much ongoing and future medical research impossible. Currently, data transfers in medical scientific research studies are stalled with grave consequences for medical scientific advancement globally.

The GDPR calls for a risk-based approach to data processing, which we believe should be reflected in the Recommendations.

A combination of supplementary measures as a solution

We suggest that a combination of supplementary measures can achieve an essentially equivalent standard of protection when transferring genetic and high dimensional data for medical scientific research to collaborators outside the EEA. We acknowledge that such transfers require particular vigilance, but medical scientific research depends on international collaboration to achieve necessary statistical power to draw valid conclusions. Scientific research is also one of the objectives of the European Union, as expressed in Article 179 of the Treaty on the Functioning of the European Union. Not establishing a solution for transfer of such data is therefore not an option.

We believe that the combination of supplementary measures we suggest below would respect research participants' fundamental rights while enabling the data transfers necessary for medical scientific research. We respectfully ask the EDPB to clarify whether a combination of supplementary measures along the lines we suggest achieves an essentially equivalent standard of protection.

A. Supplementary measures that should *always* be in place

Technical measures:

1. Pseudonymization, as defined in Article 4(5) GDPR.

2. Encryption when sending data out of the EEA and encrypted storage on safe servers in the non-EEA country.

Organizational measures:

3. Keycode for linking pseudonymized data with identity is kept separately from the researchers conducting the scientific research and not shared with non-EEA collaborators.

4. The encryption key should remain in the hands of the EEA data exporter.

Legal measures:

5. Contractual clause prohibiting personal data from being disclosed to non-EEA intelligence authorities or other third parties without a basis in law or a court order.

6. Contractual clause prohibiting attempts to identify the data subjects.

7. Contractual clause allowing contract termination if the non-EEA country introduces legislation that may jeopardize the protection of the data.

8. Contractual clause requiring encrypted storage on safe servers in the non-EEA country.

9. Contractual clause mandating details on technical measures in place to ensure data security.

B. Supplementary measures that *may be appropriate* depending on study design

For scientific research projects, study designs vary with the goals of the research, and not all measures listed below are suitable for all studies. Note in particular that provision of remote access or distributed/federated analyses are feasible for some, but not all, projects going forward. Projects in areas such as precision medicine rely on very large data sets where international collaborations are necessary in order to achieve stratification of subgroups. For such projects, it may be necessary to send data to non-EEA countries.

Technical measures:

10. Privacy-enhancing technologies such as those that allow for distributed/federated analyses (where "the analysis is brought to the data rather than the data to the analysis") are already in use and being further developed. However, the lack of statistical power associated with distributed analyses can impede discoveries in some contexts, limit the ability to return to data with other questions, and significantly slow the process in time-sensitive analyses.

11. Studies can often set up a harmonization plan with a metadata/codebook, to ensure that non-EEA collaborators only access the variables that are needed according to the analyses plan.

12. When sending data to non-EEA countries, ensure that the levels of detail of the epidemiological and demographic variables are in accordance with the analyses plan in order to reduce the risk of reidentification of individuals. For some studies, continuous variables such as age, weight, and height can be reported as categories, and the exact day can be omitted in a date.

13. Fuzzifying/noise adding/obfuscating: add noise to data to alter data while keeping the signal. This is only feasible in some instances when you know there is a signal and you know the noise distribution, but this is not often the case. Not useful or feasible on large datasets.

Organizational measures:

14. Provision of access to data on EEA computing infrastructure, with strict control: the EEA entity sets a time limit and gives access only to named investigators outside of the EEA to necessary data according to a detailed and approved analyses plan. Though provision of remote access is still considered data transfer, it allows for significantly increased control of technical and organizational measures, and can therefore also function as a useful supplementary measure. Such access must be combined with contractual clauses that researchers will not attempt to download individual data or take screen shots of individual data records. Remote access combined with the measures listed in section A above, including pseudonymization according to Article 4(5) GDPR, should yield sufficient protection even when data are high dimensional such as large clinical, epidemiological or genetic data. We respectfully ask the EDPB to confirm this. Note that precision medicine research requires large international data collaborations, and remote access is not always sufficient, for instance where the scientific research requires data generation or analysis in one location.

15. Ensure overview of who accesses the data on the non-EEA server through provision of logs.

16. Strict organizational measures for analyses of biological samples. Some laboratory analyses are only performed at sufficient quality at very few laboratories in the world, or it is important that the measures are done in the same laboratory to ensure compatibility across studies. Biological samples must then be shipped to the laboratory, and personal data will be generated from the samples. When shipping biological samples from which personal data will be generated by the non-EEA recipient, only sufficient material must be sent to run the specified analyses without significant surplus. The laboratory results should where possible be returned to EEA before analyses. In studies where it is possible and where provision of remote access to a secure non-EEA server is in place, other pseudonymized data should only be made available with the laboratory results on the server within the EEA.

17. Professional ethics norms may apply to the researchers and their conduct of the research.

18. Right to perform or request audits at the non-EEA sample processing site, including review of the technical tools used. Note that some non-EEA governmental institutions (e.g, governmental entities) will not be able to agree to audits by foreign entities. A third party in the non-EEA country, for example a governmental or accredited organization, could conduct the audit on behalf of the EEA data exporter.

Legal measures:

19. Contractual clause requiring notification of any disclosure requests received.

20. Contractual clause requiring data access log information from the non-EEA recipient.

21. Contractual clause with sanctions for violation of the contract provisions. Note that not all will be able to sign such clauses, for instance government agencies.

22. The non-EEA country may have privacy or sector-specific legislation in place that provides some protection. For instance, in the United States, this includes the Common Rule, the Health

Insurance Portability and Accountability Act (HIPAA), and Certificates of Confidentiality issued by the U.S. government.

Case example

Schizophrenia is a severe mental health disorder affecting close to 1% of the population. It results in substantial suffering, not only of patients but also their families. Due to its chronic nature and typical decline in cognitive abilities, it often results in chronic disability, shortened life span and lifelong suffering for both the patient and her/his family. Little is known about its basic mechanisms.

During the past few years genetics has shed new light to basic mechanisms of schizophrenia. For the first time there is solid evidence of a biological background of disease vulnerability. To discover this, very large samples that combine data from all over the world is needed. To accomplish this research, sharing of individual level genetic data is necessary. The last breakthrough is the SCHEMA consortium (https://schema.broadinstitute.org/) led by the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard. This study combines individual level DNA sequencing data from 24,248 schizophrenia cases and 97,322 controls from both Europe and the U.S. This study identifies a number of coding gene variants that are fundamental for potential new drug developments, in a disease where no new drugs have been developed in decades. Without a possibility to share individual level medical, genetic, and other biological data we hinder the development of new knowledge and methods to treat devastating diseases.

For data transfer in a study such as the one mentioned, we suggest applying supplementary measures 1 through 9 above that we believe should always be in place. Additionally, we suggest applying as many as possible of the measures that are appropriate according to the study design: 11, 12, (13 for some phenotype data), 15, 16, 17, 19, 20 and 22. For this study, this means that we would apply the following supplementary measures:

Technical measures:

- Pseudonymization, as defined in Article 4(5) GDPR.
- Encryption when sending data out of the EEA and encrypted storage on safe servers in the non-EEA country.
- A harmonization plan with a metadata/codebook, to ensure that non-EEA collaborators only access the variables that are needed according to the analyses plan.
- Ensure that the levels of detail of the epidemiological and demographic variables are in accordance with the analyses plan in order to reduce the risk of reidentification of individuals.
- For some phenotype data: Fuzzifying/noise adding/obfuscating: add noise to data to alter data while keeping the signal.

Organizational measures:

- Keycode for linking pseudonymized data with identity is kept separately from the researchers conducting the scientific research and not shared with non-EEA collaborators.
- The encryption key should remain in the hands of the EEA data exporter.
- Overview of who accesses the data on the non-EEA server through provision of logs.
- Strict organizational measures for analyses of biological samples. When shipping biological samples from which personal data will be generated by the non- EEA recipient, only sufficient material must be sent to run the specified analyses without significant surplus. The lab results should be returned to EEA before analyses.
- Professional ethics norms apply to the researchers and their conduct of the research.

Legal measures:

- Contractual clause prohibiting personal data from being disclosed to non-EEA intelligence authorities or other third parties without a basis in law or a court order.
- Contractual clause prohibiting attempts to identify the data subjects.

- Contractual clause allowing contract termination if the non-EEA country introduces legislation that may jeopardize the protection of the data.
- Contractual clause requiring encrypted storage on safe servers in the non-EEA country.
- Contractual clause mandating details on technical measures in place to ensure data security.
- Contractual clause requiring notification of any disclosure requests received.
- Contractual clause requiring data access log information from the non-EEA recipient.
- The non-EEA country has privacy or sector-specific legislation in place that provides some protection. In the United States, this includes the Common Rule, the Health Insurance Portability and Accountability Act (HIPAA), and Certificates of Confidentiality issued by the U.S. government.

In combination, we believe that this would result in protection of the research subjects' fundamental rights and provide an avenue for important medical progress.

***

We acknowledge the following contributors to these comments:

Kári Stefánsson, M.D., Chief Executive Officer, deCODE genetics, Iceland

Heidi Beate Bentzen, LL.M., Researcher, University of Oslo, Norway

Søren Brunak, Ph.D., Professor, University of Copenhagen, Denmark

Paul W. Franks, Ph.D., Professor, Lund University Diabetes Centre, Sweden

Hakon Heimer, Senior Research Advisor, University of Copenhagen, Denmark

Jóhann Hjartarson, LL.M., General Counsel, deCODE genetics, Iceland

Kristian Hveem, M.D., Professor, Chief Executive Officer, KG Jebsen Centre for Genetic Epidemiology, Norwegian University of Science and Technology, Norway

Tero Jyrhämä, LL.M., Data Protection Officer, FinnGen, Finland

Lili Milani, Ph.D., Professor, Head of Personalized Medicine, Estonian National Genome Center, University of Tartu, Estonia

Aarno Palotie, M.D., Ph.D., Professor, Research Director, Institute for Molecular Medicine FIMM, Finland

Sirpa Soini, LL.M., Director, National Institute for Health and Welfare Biobank, Finland

Giske Ursin, M.D., Ph.D., Director, Cancer Registry of Norway, Norway