Share Analysis. Not Data.

PRANA-DATA

Highlights GDPR for PRANA-DATA (scientific research/ datasharing)

Project	PRANA-DATA
Project leader	Wessel Kraaij (TNO)
Work package	
Deliverable number	
Authors	Marie José Bonthuis (University Medical Hospital Groningen, UMCG)
Reviewers	Wessel Kraaij, Jessica Doorn
Date	31-05-2017
Version	1
Access Rights	Public
Status	Final

PRANA-DATA Partners:
Portavita, TNO, Radboud Universiteit Nijmegen, Maastricht UMC+, UMCG

COMMIT/ COMMIT is a public-private research community solving grand challenges in information and communication science shaping tomorrow's society

COMMIT/ A public- private research community

This publication was supported by the Dutch national program COMMIT.

Introduction

The General Data Protection Regulation (hereafter: GDPR), that will be implemented on 25 May 2018 in all member states of the EU, adopts new specific provisions to ensure adapted data protection in the field of scientific research. This field remains widely regulated at national level, in particular, regarding the application of research participants' rights (consent/ objection and notice). However, the GDPR set up clearer rules that will positively serve research practices for reusing health related personal data for another purpose, assessing the risks of data processing in the context of Data Protection Impact Assessment (DPIA), adopting accountable management systems of processing operations and building or reinforcing internal data protection competencies with the Data Protection Officer (DPO). In this deliverable the special provisions for Privacy Respecting ANAlysis of distributed patient health DATA[1] (PRANA-DATA) -a Swallow project-, will be presented.

PRANA-DATA

The Privacy Respecting ANAlysis of distributed patient health DATA (PRANA-DATA) explored decentralized (distributed) privacy preserving analytics models and possibilities to combine data from various distributed sources, without revealing the data itself, by new advanced cryptographic techniques like secure multi-party computation and/or homomorphic encryption. PRANA-DATA brings the algorithms to the data and combines the models resulting from the analysis at various repositories in a clever way. PRANA-DATA opens new possibilities for going forward in the structuring of data sharing in scientific research with measures encouraging self-regulation development. Especially the provisions that contribute to self-regulation development are described in this deliverable.

Legal framework

The legal framework that is applicable for PRANA-DATA is the GDPR, but also the Dutch implementing law (draft) of the GDPR and the regulation at national level: 'Dutch Medical Treatment Act (*Wet geneeskundige behandelingsovereenkomst: Wgbo*)'. The GDPR will provide for new rights and obligations that need to be practically implemented, whereby technology alone is not enough. PRANA-DATA embeds privacy enhancing technologies in the analytics design which is an integral part of avoiding personal data breaches and rebuilding trust between users and service providers, both individuals' privacy and big data quality of results.

The corresponding provisions that are described in this deliverable are:

i. Consent, the right to object to the re-use of personal data for scientific research and notice requirements;
ii. Personal data; anonymization and pseudonymization;
iii. New legal definitions of special categories of personal data;
iv. Privacy by design;
v. Data Protection Impact Assesment for scientific research.

[1] The aim of PRANA-DATA is to explore the **potential approaches** for collecting, storing, combining and analyzing health related **distributed** personal data in a unified secure and **privacy respecting** system architecture that supports the interests of stakeholders of **different domains**, i.e. patients, LSH and medical researchers, health professionals, health policy makers and companies (i.e. pharma, food, insurance).

Using PRANA-DATA-outcomes to process personal data for research purposes can avoid restrictions on secondary processing and on processing sensitive categories of data (Article 6(4); Recital 50 GDPR), whereby controllers remain responsible for the implementation of the entire legal framework.

i. Consent, the right to object and notice reguirements

Controllers (organizations/ natural persons responsible for the personal data) are not required to obtain consent from the patient/ participant for all processing for research purposes. However, article 12(1) GDPR requires controllers to take appropriate measures to inform data subjects of the nature of the processing activities and the rights available to them. Controllers are required to provide this information in all circumstances, regardless of whether consent is the basis for processing, "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". Notice should be provided at the time when the data is *first collected* and it must include the controller's identity and contact information, the intended purposes of the processing activities, as well as notice of "the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period." PRANA-DATA supports the interests of data subjects of different domains, by not revealing the data itself. That is why the controllers are still required to obtain consent and/ or take appropriate measures to inform data subjects about the privacy respecting way of their processing activities. However, PRANA-DATA combines the securing of data with other measures that will provide for informed consent based transparency about how the data is used in order to safeguard the vital interests of the data subjects, contribute to a common goal (e.g. public health/ improved logistics) while protecting privacy.

ii. Personal data; anonymization, pseudonymization

Article 89(1) GDPR provides that one way for a controller to comply with the mandate for technical and organizational measures is through deployment of pseudonymization. This means that the data no longer can 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 organizational measures to ensure non-attribution to an identified or identifiable individual. Even if a researcher no longer has the ability to re-identify a data set, such data set may still be regulated under the GDPR if it could be re-identified with reasonable effort. We can speak of 'unreasonable effort' when that much time, cost and manpower (as an excessive effort) is required, that in fact a risk of identification is insignificant. By new advanced cryptographic techniques like secure multi-party computation, homomorphic encryption, the inputs to the analysis remain confidential, and only the output of the analysis becomes available. PRANA-DATA brings the algorithms to the data to combine the models resulting from the analysis at various repositories in a clever way, whereby additional information is kept separately, whereby it could not be re-identified with reasonable effort.

iii. New legal definitions of special categories of personal data

The GDPR introduces some new definitions of certain special categories of personal data:
-data concerning health,
-genetic data and
-biometric data.
Also for PRANA-DATA the processing of these categories of personal data is exceptionally admitted for research or archiving purposes in the public interest in the respect of Articles 9 and 89 of the GDPR, and genetic data only with consent.

iv. Privacy by design

PRANA-DATA can be considered as 'privacy by design' as mentioned in article 25 GDPR. Privacy by design opts for solutions that minimize processing of personal data to perform effectively. PRANA-DATA only processes personal data that is necessary for the specific scientific research and minimizes the amount of personal data, the extent to which data is processed, the period for which they are stored and the accessibility. PRANA-DATA has built necessary safeguards in the processing to comply with the requirements of the GDPR to protect the rights of those involved whereby only the results of scientific research will be shared. The architecture that PRANA-DATA provides is a form of data protection and *processing by design*.

v. Dataprotection Impact Assessment (DPIA) for scientific research

With the GDPR controllers will have to practice a DPIA according to Article 35 of the GDPR. The DPIA is an entirely new self-assessment exercise which somewhat prolongs the requirements of most of the funding agencies requiring, as an integrated part of the ethics assessment of a research proposal, to describe how personal data will used and responsibly managed in the research (e.g. in H2020 or ERC programs). The DPIA concretizes the risk-based approach of the GDPR. The aim of the DPIA is to assess the likelihood and severity of the risk regarding data subjects' rights and freedoms before undertaking the processing. The DPIA serves not only to know the state of the art of data protection means in a certain context, to plan and manage the necessary enhancements to ensure compliance of the system, but also to determine if a prior consultation of the supervisory authority is necessary. According to Article 35(7) of the GDPR, 'the assessment shall contain at least: a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller; an assessment of the necessity and proportionality of the processing operations in relation to the purposes; an assessment of the risks to the rights and freedoms of data subjects […]; and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with the GDPR considering the rights and legitimate interests of data subjects and other persons concerned.' One of the main deliverables of PRANA-DATA is a survey report describing the listed critical assessments and promising approaches towards privacy respecting data analysis in distributed repositories.

Conclusion

The GDPR allow researchers (organizations) to further process personal data beyond the purposes for which they were first collected. Research may furnish a legitimate basis for processing without a data subject's consent. By using PRANA-DATA the inputs to the analysis remain confidential, and only the output of the analysis becomes available. To benefit from these exemptions, researchers must also implement (other) appropriate safeguards, in keeping with recognized ethical standards, that lower the risks of research for the rights of individuals. However, PRANA-DATA has explored a privacy friendly way of datasharing, because it minimizes processing of personal data by:

- using decentralized privacy preserving analytics models (distributed);
- combining data from various distributed sources, without revealing the data itself;
- new advanced cryptographic techniques like secure multi-party computation, homomorphic encryption;
- bringing the algorithms to the data and
- combining the models resulting from the analysis at various repositories in a clever way.

In this way, with PRANA-DATA the inputs to the analysis remain confidential, and only the output of the analysis becomes available. Because the storage of the data remains at the source, limitation of data collection or the reduction of data the amount of data, by not storing the data collectively, are important data protection measures, which are also mentioned in the recommended (Dutch) Privacy Impact Assessment-Framework of NOREA. Another topic mentioned in this framework and 'caught' by PRANA-DATA, is that it uses mathematical methods without retrieving and registering the underlying data. On behalf of the issue of data security PRANA-DATA provides the use of encryption and logical access protection.

For more information, see www.pranadata.nl

Sources:
https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5243137/
https://www.norea.nl/download/?id=522