Interní pravidla pro zacházení s osobními údaji při archivaci a sdílení výzkumných dat

Koščík, Michal
2017

10th Conference on Grey Literature and Repositories: proceedings [online]. Prague: National Library of Technology, 2017. ISSN 2336-5021. Available from: **http://nrgl.techlib.cz/conference/conference-proceedings/**.

INSTITUTIONAL RULES AND POLICIES FOR SHARING AND STORING RESEARCH DATA

Michal Koščík

michalkoscik@gmail.com

Institute of law and technology, Faculty of Law, Masaryk University

Department of public health, Faculty of Medicine, Masaryk University

Abstract

The paper aims to provide readers with a practical view on how to adapt the internal policies of research institutions to the upcoming General Data Protection Regulation. Since the Regulation enters force six months after the conference takes place, it can be expected that this issue of readjustment of internal processes to GDPR will be very important for majority of conference participants. With regard to the time and space limit, the paper will focus exclusively on the issues connected with archiving and sharing research data. Emphasis will be put on the rights of research subjects and the public interest in research as an entitlement to process of personal data without consent.

Key words

Privacy; Compliance; Data Protection; General Data Protection Regulation; GDPR

This article is a result of the research funded by the Czech Science Foundation as the project GA15-20763S Právni rámec sběru, zpracování, uchovávání a užívání výzkumných dat (*Legal Framework* for Collecting, Processing, Storing and Utilizing of Research Data).

Introduction

There is little point in an extensive introduction and description of the General Data Protection Regulation (hereinafter GDPR) which enters into force on 28 May 2018, as it has already been covered in numerous articles[1]. This article does not intend to provide readers with a detailed checklist of all data protection issues that need to be addressed, but rather provide practical guidance on where to begin and how to proceed in the process of adapting institutional rules and processes for operators that share research data through repositories. We presume that the majority of operators of repositories are research institutions (universities, academies or central authorities) that employ more than 250 employees or process special categories of data and are thus are subjected to the new obligation to keep records on processing activities under Article 30 of the GDPR, which in practice means a significant step towards the strong formalisation of the compliance procedures. Therefore, the main purpose of the article is to outline the steps that need to be taken in order to bring rules of institutions to the appropriate formal level.

The GDPR compliance procedure consists of two major steps. The first requires knowledgeto be gathered about the dataflows within the institution, and the second requires internal policies to be adopted or adjusted. Therefore, this article is divided into two chapters. The first chapter describes the steps that have to be taken in the process of collecting information, and the second chapter suggests ways in which the policies of institutions should be structured.

Gathering information and identifying personal data in repositories

The first step in compliance procedure is the identification of personal data in repositories and its life cycle. This means both the revision of existing data and its origin (source) as well as the identification of the channels where the data will flow in the future. It is necessary to identify storage spaces, departments or employees who are responsible for the administration of the data and the recipients of the data (i.e. the persons or entities that find the data useful).

Identification of personal data in all repositories

One has to keep in mind that the definition of “personal data” is very extensive and covers any information that can be directly or indirectly related to an individual. The data does not have to be structured in order to qualify as personal data. Any information in any media format, including photographs, audio and visual records, may meet the definition of personal data and thus make the repository of the institution subject to regulation.

It is important to point out that even pseudonymized information is to be considered personal information. The borderline between pseudonymized and anonymized information might not be exactly clear in many practical situations. Since the definition of personal data is very broad, it can be advised that even anonymized data be handled with great cautionn, if possible under the same standards as if the personal data were involved (see subchapter 1.3. of this article).

[1] See also previous articles of the author of the Article that covered the development in this area in recent years: KOŠČÍK, Michal. Privacy and anonymization in repositories of grey literature. In: Conference on Grey Literature and Repositories. 2015. p. 72. KOŠČÍK, Michal. The Impact of the General Data Protection Regulation on grey literature. Grey Journal (TGJ), 2017, 13. See also: WIPP EKMAN, Leon; BILLGREN, Petter. Compliance Challenges with the General Data Protection Regulation. 2017.

Identification of the purpose and activities related to data processing

Personal data processing is a daily activity in every public institution or business. The governance of personal data has to be based on the purpose served by the data being processed (i.e. its value to the organization) and on the activities (processes) that involve the particular data. After the personal data has been identified, it is necessary to attribute each set of records to a certain purpose (or purposes) for which they have been collected and processed. To put it simply, the institution has to seriously question each individual database record and answer the question "do we really need to keep this record and why?". Virtually no common purpose of processing is illegitimate per se[2]. After the purpose of processing is identified, it is possible to assess whether the processing is legitimate in this particular case and what steps need to be taken in order to keep the processing legitimate. Keeping personal data without a specific purpose[3] is equivalent to non-compliance with the regulation.

It is necessary to define the purpose of each set of data in order to determine whether or not the institution requires the consent of the data subject. The general regulatory principles of purpose limitation[4], data minimisation[5] and storage limitation[6] are directly related to the purpose of data processing. Hence, if the institution does not define the purpose of each particular set of personal data it processes, it cannot comply with these fundamental principles. The purpose of data processing is also crucial in dealing with requests for data erasure [7] or the right to restriction of processing[8].

Recital 39 of the GDPR states that the purpose needs to be determined at the time when the personal data is collected and that changing the purpose of processing after the data has been collected is limited by the GDPR and restricted to several explicitly defined cases[9]. Operators of repositories will benefit from the provisions of the second paragraph of Art. 9 of the GDPR, which enables so-called "further processing" or secondary use of data for archiving purposes in the public interest, for scientific or historical research purposes or for statistical purposes[10] even in cases where data in special data categories (sensitive data) is being proccesssed. Even if the repository operator intends to rely on the provisions of Art. 9 section 2, the purpose has to be defined. It is advised that the purpose be defined more specifically than by the mere declaration of public or scientific interest so that the proportionality between the public interest and the interest of the subject can be demonstrated.

[2] with the exception of clear excesses, usually well defined in criminal codes

[3] for example storing historical data collected during past activities just because someone failed to delete it or keeping data "just in case"

[4] Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes

[5] Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which it is processed

[6] Personal data shall be kept in a form allowing the identification of data subjects for no longer than is necessary

[7] See also COFONE, Ignacio N. Google v. Spain: A Right To Be Forgotten?. Browser Download This Paper, 2015.; ROSNAY, Melanie Dulong de; GUADAMUZ, Andres. Memory Hole or Right to Delist?. Implications of the Right to be Forgotten for Web Archiving. *RESET. Recherches en sciences sociales sur Internet*, 2016, 6. KOŠČÍK, Michal. The Impact of the General Data Protection Regulation on grey literature. Grey Journal (TGJ), 2017, 13.

[8] The data subject shall have the right to obtain from the controller restriction of processing the controller no longer needs the personal data for the purposes of the processing.

[9] One of the reasons may be protecting the vital interest of a data subject, or vital interest of another natural person or archiving. Here, it has to be noted that even the change in purpose of processing personal data collected for the public interest is limited.

[10] The national law could however specify requirements for link between those purposes and the purposes of the intended further processing - See recital 50 and

After the institution identifies the data and its purpose, it has to identify the activities in which it is necessary to process the particular personal data. Each activity in which the personal data needs to be processed shall have a delegated person who is responsible for compliance with internal rules and policies (see below). These persons are not necessarily (and most likely not) data protection officers, as the data protection officer is more the role of the internal auditor and not the person who will perform all the tasks associated with data protection.

Identification of the data sources

Every repository needs to identify sources from which it retrieves personal data, mainly for three compliance reasons:

A. Identifying whether the repository is a controller or processor. It should be noted that the repository is rarely established as a mere processing service without any interest of its operator in collecting data and determining what goes into the repository. We presume that the repository will be a controller of most of its data. In cases where the repository serves as a data processor, we strongly recommend that the the contractual framework be reviewed with the data contollers[11].

B. Determining whether the repository processes raw, pseudonymized or anonymized data

The first practical issue with anonymized data is the question of whether a repository storing data obtained from a third party which the repository's operator cannot himself attribute to an individual natural person is to be considered as anonymous or pseudonymous data if the subject that encrypted the data still keeps the key to its decryption. According to Article 4 of the GDPR, pseudonymized data is defined as "personal data that can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures for ensuring that the personal data is not attributed to an identified or identifiable natural person".

The first possible approach is to admit that the anonymity or pseudonymity of data is relative. Two (or more) subjects can process the same set of data, whereas one is unable to encrypt it and the other one is able to encrypt it. If we accept that the concept of data anonymity is relative, it would mean that the first subject can use and share data freely without any significant restrictions, whereas the person that possesses the encryption/ decryption key is restricted in handling the data. The second approach is to presume that the anonymity of the data is absolute. If the encryption key exists anywhere in the world or can be deciphered in any way, such data is not anonymous but only pseudonymous. One of the main problems of pseudonymized personal data is that (if shared) it can potentially be de-anonymized by a third party when merged with other data-sets[12], and basically any anonymized data can be de-anonymized by forensic methods. The CJEU addressed this issue in the judgment in Case C-582/14: Patrick Breyer v Bundesrepublik Deutschland, in which the CJEU ruled that the possibility to combine the data with this additional data must constitute a means of which it can

[11] Detailed formal requirements on a contract between controller and processor are described in the Article 28 GDPR.

[12] HARAŠTA, Jakub a Matěj MYŠKA. Secondary use of research data in the EU: Complex institutional approach. In Erich Schweighofer; Franz Kummer; Walter Hötzendorfer; Christoph Sorge. Trends and Communities of Legal Informatics IRIS 2017 Proceedings of the 20th International Legal Informatics Symposion. Wien: Oesterreichische Computer Gesellschaft, 2017. s. 539-542, 4 s. ISBN 978-3-903035-15-7.

reasonably be assumed that it will likely be used to identify the individual[13]. The interpretation of the Breyer case speaks in favour of the "relative approach". The data is not anonymous for a person that has a legal and material capacity to de-anonymize it. We can add that data may remain anonymous/anonymized for entities that lack legal and material capacities to de-cipher it. This approach is favourable for data repositories, since they can share anonymized research data with a certain degree of legal certainty.

C. Determining whether the source collects data in accordance with applicable rules. If the data processing requires the consent of the subject, the repository operator needs to make sure that the copy of the consent can be found at the source.

Adopting policies and internal rules

Only after the data and its sources and purpose have been defined is it possible to adapt the internal policies and formalize them into internal documents. Below, we have identified the key areas that have to be addressed by internal policies

General data protection policy addressing privacy by design and default

The repository should adopt a norm which will address risks, responsibilities and measures as they regard the security accessibility, pseudonymization and anonymization of data and the identification of processes, activities and involved employees[14]. If the repository shares data based on health information or other sensitive data (such as the ethnic origin or political stances of research subjects), it will likely be obliged to carry out a privacy impact assessment under Article 35 of the GDPR[15]. Even in institutions where the privacy impact assessment is not required, it is recommended to identify the major risks to the rights of data subjects and identify organizational units that are required to take measures to protect these rights.

The norm should also implement a notification policy for cases of personal data breaches. The GDPR requires response and notification of the data protection authority within the 72 hour time limit. It is, therefore, advisable to have defined responsibilities for notification of data breaches in advance.

We presume that the majority of institutional repositories will also fall under the obligation under Article 30 of the GDPR that obliges each controller to keep a record of processing activities for which he is responsible. The record must contain all of the following information:

[13] NIEMANN, Fabian a Lennart Schüßler CJEU decision on dynamic IP addresses touches fundamental DP law questions. Bird & Bird [online] [vid. 2017-10-09]. Available from: **https://www.twobirds.com/en/news/articles/2016/global/cjeu-decision-on-dynamic-ip-addresses-touches-fundamental-dp-law-questions** See also EL KHOURY, Alessandro. Dynamic IP Addresses Can be Personal Data, Sometimes. A Story of Binary Relations and Schrödinger's Cat. *European Journal of Risk Regulation*, 2017, 8.1: 191-197., POLČÁK, Radim. Stock Exchange Interconnections and Legal Issues in Data Exchange. *Masaryk University Journal of Law and Technology*, 2017, 11.2: 351-362.

[14] See also: GJERMUNDRØD, Harald; DIONYSIOU, Ioanna; COSTA, Kyriakos. privacyTracker: A Privacy-by-Design GDPR-Compliant Framework with Verifiable Data Traceability Controls. In: International Conference on Web Engineering. Springer International Publishing, 2016. p. 3-15.

[15] Accoriding to Art. 35 GDPR, the privacy impact assessment is required if the processing is „ likely to result in a high risk to the rights and freedoms of natural persons, see also BIEKER, Felix, et al. A process for data protection impact assessment under the European general data protection regulation. In: *Annual Privacy Forum*. Springer International Publishing, 2016. p. 21-37.

- the name and contact details of the controller
- the categories of data subjects and categories of personal data;
- the categories of recipients to whom the personal data has been or will be disclosed, including recipients in third countries or international organisations;
- transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and the documentation of suitable safeguards;
- the expected time limits for the erasure of the different data;
- description of the technical and organisational security measures referred to.

We strongly advise that obligations be formulated and record keeping be delegated to the respective departments and employees. Data protection by design and default should become a formal responsibility of every employee of an institution that has access or the right to upload content to a repository[16].

Privacy (transparency) policy

Section 2 of the GDPR, which deals with information and access to personal data, sets forth requirements for the information that has to be provided to the subject. It is recommended that a document containing the basic information that has to be provided to data subjects under Art. 13 to 15 of the GDPR be drafted and published. Among others, this information includes:

- the contact details of the controller and the controller's representative;
- the contact details of the data protection officer;
- the purposes for which the personal data is processed as well as the legal basis for the processing;
- the legitimate interests pursued by the controller or by a third party and the recipients or categories of recipients of the personal data;
- the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;
- the existence of the right to request from the controller access to and rectification or erasure of personal data or a restriction of processing concerning
- information regarding the existence of the right to withdraw consent[17]

It is evident that the recipients of the general policy (under Art. 2.1.) are the employees of the institution, whereas the privacy policy is a non-binding informative document for data subjects. The language and scope of detail should be adjusted accordingly. The privacy policy is supposed to be relatively short and approachable, whereas the general data protection policy will be more detailed and could be further specified by technical norms applicable to the respective divisions of an institution. Most academic institutions will likely have a general policy covering all the major data protection issues and might also adopt a separate policy governing data protection issues in a respective repository. This approach is recommended for larger institutions that operate several repositories which store data from distinctive research fields and serve different purposes.

[16] See also KOŠČÍK, Michal. Sharing Liability for a Repository Between Employer and Employee. In: *CONFERENCE ON GREY LITERATURE AND REPOSITORIES*. 2016. p. 69.

[17] The list is not exhaustive, the author has selected information that is most likely to be relevant for a repository

Conclusion

The article outlined two phases of the procedure for compliance with the GDPR at public institutions that operate a repository. We suggest that the institution needs to identify its data and processes and link the data to the processes (and thus define their purpose) before drafting new rules and documents. The institution needs to make it clear which set of data is processed in the role of "data controller" and which set of data is processed in the role of "data processor" (in cases where the institution is the data processor, it is also necessary to review the contractual framework with the controllers).

We presume (and also recommend) that most institutions will aim to draft at least two documents - one internal policy that will address most data involving processes in order to comply with the objective of "data protection by design and default" and one publicly available policy document that will provide information about the privacy standards of the institution operating the repository.

References

BIEKER, Felix, et al. A process for data protection impact assessment under the European general data protection regulation. In: *Annual Privacy Forum*. Springer International Publishing, 2016. p. 21-37.

COFONE, Ignacio N. Google v. Spain: A Right To Be Forgotten?. *Chicago-Kent Journal of International and Comparative Law* [online]. 2015, **15**(1). Available from: **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2548954**

GJERMUNDRØD, Harald, Ioanna DIONYSIOU a Kyriakos COSTA. PrivacyTracker: A Privacy-by-Design GDPR-Compliant Framework with Verifiable Data Traceability Controls. In: *Current Trends in Web Engineering*. Springer International Publishing, 2016, 3 - 15.

HARAŠTA, Jakub a Matěj MYŠKA. Secondary use of research data in the EU: Complex institutional approach. In: SCHWEIGHOFER, Erich, Franz KUMMER, Walter HÖTZENDORFER a Christoph SORGE. *Trends und Communities der Rechtsinformatik / Trends and Communities of Legal Informatics: Tagungsband des 20 Internationalen Rechtsinformatik Symposions IRIS 2017*. Wien: Oesterreichische Computer Gesellschaft, 2017, 539 - 542. ISBN 978-3-903035-15-7.

KOSCIK, Michal. Privacy Issues in Online Service Users' Details Disclosure in the Recent Case-Law: Analysis of Cases Youtube v. Viacom and Promusicae vs. Telefonica. *Masaryk University Journal of Law and Technology* [online]. 2009, 3, p. 259.

KOŠČÍK, Michal. Privacy and anonymization in repositories of grey literature. The Grey Journal (TGJ). 2015, **11**(special issue).

NIEMANN, Fabian and Lennart SCHÜßLER. CJEU decision on dynamic IP addresses touches fundamental DP law questions. In: *Bird & Bird* [online]. [2017-10-09]. Available from: **https://www.twobirds.com/en/news/articles/2016/global/cjeu-decision-on-dynamic-ip-addresses-touches-fundamental-dp-law-questions**

POLČAK, Radim. Getting European data protection off the ground. *International Data Privacy Law* [online]. 2014, **4**(4), 282-289 [cit. 2017-11-23]. DOI: 10.1093/idpl/ipu019. ISSN 2044-3994. Available from: **https://academic.oup.com/idpl/article-lookup/doi/10.1093/idpl/ipu019**

POLČÁK, Radim. Stock Exchange Interconnections and Legal Issues in Data Exchange. *Masaryk University Journal of Law and Technology* [online]. 2017, **11**(2), 351-362 [cit. 2017-11-23]. DOI: 10.5817/MUJLT2017-2-7. ISSN 18025943. Available from: **https://journals.muni.cz/mujlt/article/view/6681**

ROSNAY, Melanie Dulong de and Andres GUADAMUZ. Memory Hole or Right to Delist?: Implications of the Right to be Forgotten for Web Archiving. In: *RESET: Recherches en sciences sociales sur Internet*. 2017(6). ISSN 2264-6221. Available from: **https://reset.revues.org/807**

WIPP EKMAN, Leon a Petter BILLGREN. *Compliance Challenges with the General Data Protection Regula- tion* [online]. Lund, 2017 [cit. 2017-11-23]. Available from: **http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8911983&fileOId=8911995. Master thesis. Lund University**