Ochrana soukromí a anonymizace osobních údajů u repozitářů šedé literatury

Koščík, Michal
2014

Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-19]. Available from: <http://nrgl.techlib.cz/index.php/Proceedings>. ISSN 2336-5021.

Privacy and anonymization in repositories of grey literature

Michal Koščík

michalkoscik@gmail.com

Masaryk University, the Faculty of Law, the Faculty of Medicine, the Czech Republic

Abstract

Recent case law from the Court of justice of the European Union, such as the case of Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González indicates that the repositories of grey literature may be qualified as institutions processing the personal data of the subjects mentioned in the documents stored in repositories. These repositories may face requests for the anonymization or even the removal of their documents. The purpose of this paper is to outline the legal framework and suggest procedures to approach this issue in compliance with the EU legislation.

Keywords

Privacy, Personal Data, The Right to be Forgotten

Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-19]. Available from: <http://nrgl.techlib.cz/index.php/Proceedings>. ISSN 2336-5021.

Privacy and anonymization in grey literature repositories

Most grey literature repositories process documents containing random information which directly or indirectly relates to living or deceased individuals. Even if these repositories are not operated for the purpose of collecting specific information on the individuals who are mentioned in the archived documents, the recent case-law of the Court of Justice of the European Union (hereafter simply referred to as the CJEU), such as Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González[1] (hereafter simply referred to as Google Spain) indicates that such repositories may be considered to be data controllers and have to comply with data protection regulation. The case-law also suggests that EU regulation could also apply to repositories and search indexes operated from outside the EU, most notably from the USA.

The case of Google Spain and its consequences for grey literature

The case of Google Spain involved the "google search" service, whose primary purpose is to index information contained in the websites. The purpose of the "google search" is not to collect the personal data of any individuals coincidentally contained in the indexed websites. Nevertheless the CJEU was asked to decide, whether the Google was or was not the controller of personal data under the EU regulations. The facts of the Google Spain case can be summarized as follows:

"In 2010, a Spanish citizen lodged a complaint against a Spanish newspaper with the national Data Protection Agency and against Google Spain and Google Inc. The citizen complained that an auction notice of his repossessed home on Google's search results infringed his privacy rights because the proceedings concerning him had been fully resolved for a number of years and hence the reference to these was entirely irrelevant. He requested, firstly, that the newspaper be required either to remove or alter the pages in question so that the personal data relating to him no longer appeared; and secondly, that Google Spain or Google Inc. be required to remove the personal data relating to him, so that it no longer appeared in the search results.[2]"

The Spanish court which dealt with the case initiated a preliminary ruling procedure[3] at the CJEU. The CJEU was asked with question, whether the "google search[4]" internet service can be qualified as an activity that falls under the definition of personal data processing as per the

[1] See: Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.

[2] See: Factsheet on the right to be forgotten ruling: (C-131/12). EUROPEAN COMMISSION. Europa.eu [online]. [cit. 2014-10-25]. Available at: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

[3] The preliminary ruling procedure has been explained by the Court of Justice as follows: The national courts in each EU country are responsible for ensuring that EU law is properly applied in that country. But there is a risk that courts in different countries might interpret EU law in different ways. To prevent this happening, there is a 'preliminary ruling procedure'. If a national court is in doubt about the interpretation or validity of an EU law, it may – and sometimes must – ask the Court of Justice for advice. This advice is called a 'preliminary ruling'. See: Preliminary ruling procedure, EUROPEAN COUR OF JUSTICE, [cit. 2014-10-25]. Available at http://europa.eu/about-eu/institutions-bodies/court-justice/index_en.htm

[4] This activity was described as a provider of content, consisting of locating information published or included on the net by third parties, indexing it automatically, storing it temporarily and finally making it available to internet users according to a particular order of preference, when that information contains the personal data of third parties, see paragraph 20 of the Google Spain ruling

data protection directive[5]. A positive answer to this question would mean that Google is a "data controller" and as such has to comply with EU regulations and is obliged to grant certain individuals the "right to be forgotten[6]" i.e. to erase any processed information on request.

In a surprise to many, the CJEU eventually ruled that the activities of a search engine consisting of finding information published or placed on the internet by third parties, automatically indexing it, temporarily storing it and making it available to internet users according to a particular order of preference must be classified as 'personal data processing' in accordance with the Data Protection Directive when that information contains personal data and, secondly, the operator of the search engine must be regarded as the 'controller' with regard to said processing[7]. In order to comply with the rights laid down in the directive, the operator of a search engine is obliged to remove links to web pages published by third parties and containing information relating to a certain person from the list of results displayed following a search made on the basis of the given person's name,[8] if the information is inadequate, irrelevant or excessive in relation to the purpose of the processing.

The main practical consequence for grey literature repositories is that the activities which comprise the indexing and storing documents for the purpose of making them available on the internet are qualified as personal data processing, if the documents contain any personal information. Given that the repositories typically categorize and index their documents, enable full text searches within documents and make the documents available online, they share similarities to the functionality of the google search engine and the conclusions in Google Spain case would most likely apply to them in any eventual dispute with a data subject. Hence, the operators of the repositories have to acquaint themselves with the data protection rules and adapt their internal policies accordingly.

The basic legal framework

The general legal framework for data protection

The Charter of Fundamental Rights of the European Union guarantees every EU citizen and resident the "*right to respect for his or her private and family life home and communications*," and the right to the *"protection of personal data concerning him or her."* Personal data must be: "*processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.*

More detailed rules on the protection of personal data are contained in the secondary EU legislation. The European Union has harmonized its legal framework for data protection in

[5] DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, p.31) as amended by Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003

[6] For further reference on right to be forgotten see BENNETT, Steven C. Right to Be Forgotten: Reconciling EU and US Perspectives, The. *Berkeley J. Int'l L.*, 2012, 30: 161. ROSEN, Jeffrey. The right to be forgotten. *Stanford law review online*, 2012, 64: 88. Ausloos, Jef. "The 'Right to be Forgotten'–Worth remembering?" Computer Law & Security Review 28.2 (2012): 143-152.

[7] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.

[8] See Ibid.

the Data Protection Directive[9]. The rules outlined by the directive are subsequently implemented by individual member states within their respective national laws and are enforced by their national data protection agencies. Therefore, there is not a single European law on the protection of personal data, there are 28 national laws enforced by 28 national agencies. The national laws are based on the same concepts due to harmonization, but they are not identical among member countries.

The Data Protection Directive defines the rights and duties of the "controllers[10]" and "processors[11]" of personal data and the rights of the "data subject" i.e. the individuals to whom the processed personal data relates. The definition of personal data is very broad and comprises any information relating to identified or identifiable natural (i.e. not legal) person. The directive sets forth the conditions under which the processing of personal data is **legitimate**. Personal data can only be processed, if the data subject has unambiguously given his or her consent or under the exemptions provided by the directive, such as the performance of a contract, compliance with a legal obligation, a task carried out in the public interest or when pursuing other legitimate interests. Conversely, the data subjects have the right to withdraw their consent, to be informed of the extent of the processing of the data and the right to object to the processing of data or even to demand its erasure.

As was mentioned above, one of the main conclusions of the Google Spain case is that even the indexing of documents which contain personal data for the purpose of creating a search engine constitutes "data processing" and an entity which operates such a search engine or stores such documents is indeed a "data controller". The legal status of the "data controller" and the "data processor" is associated with specific duties prescribed by both European and national laws. These duties especially relate to the proper administration of the data, the legitimate use of the data and the relations towards the "data subjects"[12].

The duties related to the proper administration of the data would not be too burdensome for a repository which is operated in a professional manner. These include the duties related to data quality, such as the duty of specifying the purpose of the processing data or the duty of processing adequate, relevant, not excessive[13], accurate and, if possible, up-to-date data[14] and the duties related to data security (i.e. the appropriate organizational and technical measures). Repositories usually comply with these rules by complying with general professional standards without having specific data protection directives in mind.

Administering and monitoring the legitimacy of data use can be much more complex and challenging. The general rule, as set out in the Data Protection Directive is that personal data can be processed solely on the basis of data subject's consent or on the basis of the data controller's statutory right.

[9] DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, p.31) as amended by Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003

[10] The controller is the person who determines the purpose and means of processing of personal data

[11] The processor is the person who processes the personal data on behalf of the controller

[12] i.e. an identified or identifiable natural person See Art. 2 a) of the Data Protection Directive

[13] See Art. 6 par. 1. c) of the Data Protection Directive

[14] See Art. 6 par. 1. d) of the Data Protection Directive

The legal rules concerning the "right to be forgotten"

It has to be said that the "right to be forgotten" is a rather popular term which can be found in newspapers or scholarly articles more often than in the legislation of case-law. The Data Protection Directive does not mention the "right to be forgotten" explicitly. Even the Google Spain ruling (which is sometimes referred to as "the Right to be Forgotten Ruling[15]") uses this phrase only twice, and even then only when paraphrasing the complainant's argument.

From the perspective of the European Data Protection Directive, it is possible to define the right to be forgotten as the aggregate of several rights granted by the European Commission. Most notably the rights granted by: Article 12(b) - Every data subject has the right to obtain from the controller, *as appropriate, the rectification, erasure or blocking of data, the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data;* Article 14(b) – The data subject has right to object at *any time on compelling legitimate grounds relating to his particular situation to the processing of data relating to him. Where there is a justified objection, the processing instigated by the controller may no longer involve those data;*

Grey literature repositories may therefore face requests for the erasure of certain data or documents contained in their collections.

Possible approaches to the data protection

A rigorous pro-active approach

Considering the fact that repositories usually gather and collect their documents from third parties, the repositories might face a very complicated task in obtaining proof as to the fact that the uploaded documents have been created with the due consent of all the data subjects mentioned within the documents. The *most rigorous approach* to compliance with the personal data legislation would require:

1) the examination of all the documents contained in the repository,
2) the identification of which documents contain personal information,
3) the evaluation of which personal information may be processed without the consent of a data subject
4) the processing of the personal data identified in point 3
5) the selection of any personal data which can be processed with the consent of the data subject
6) the identification and contacting of the data subjects in order to receive their approval
7) processing the data for which consent has been obtained and erasing the data for which consent was not or could not be obtained

The aforementioned approach to the personal data is time consuming and expensive. The costs of this approach would not be justifiable for many repository operators, especially in cases where the repository stores large amounts of documents which contain only coincidental and random personal data.

[15] See: Factsheet on the right to be forgotten ruling: (C-131/12). EUROPEAN COMMISSION. Europa.eu [online]. [cit. 2014-10-25]. Available at: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

The repository operator might find itself in a paradoxical situation when it discovers that it stores or indexes documents with personal data towards which it does not have any legitimate entitlement. The natural reaction would be to anonymize or de-identify these documents or even to delete them completely. However, such actions would fall under the definition of data processing[16]. This leads to a formal loophole where the data controller is not entitled to either store the information or delete it. In these cases, we believe that the documents could (and should) still be de-identified[17], anonymized[18] or deleted, because the effort to terminate the unlawful status can be justified by Article 7 in the same directive which enables the processing of personal information in a way which is necessary for compliance with a legal obligation to which the controller is subject.

A reactive approach – a privacy policy

Grey literature repositories which make available large amounts of documents containing little, random and coincidental personal data produced by third parties are unlikely to have enough resources (neither financial nor personal) to adopt the rigorous approach described above. The necessary investments for a rigorous approach might not appear cost effective in relation to the actual risk of harming someone's rights.

In these cases, we recommend the formulation and publishing of a specific privacy policy that would transparently set out the process of how to indicate any potential infringements of the personal data rules and how the institution would proceed after being notified as a minimum standard. Such a policy should contain the exact identification of the institution that is responsible for operating the registry and indicate the contact person (or designated department) where the requests for the removal of displayed personal data can be forwarded to. We advise the definition of a specific request format or the provision of a form or electronic tool for users to report alleged data infringements. It is highly advised that the institution should indicate why it operates the repository (or search tool) and any public interest that lies behind making such documents available. This information will be useful when justifying the display of any such information, if the institution decides not to comply with a request to delete or anonymize any personal data contained in the repository.

The administration of the requests to be forgotten

Justified grounds for refusal

Neither the European personal data legislation nor the CJEU case-law can be interpreted as meaning that the repository is obliged to remove or anonymize any personal information contained in its repository. Whenever a repository operator receives a request to remove certain documents or personal information displayed online, the operator should consider whether or not it can justifiably refuse such request.

The justification for a refusal is not the same category as the legitimate basis for personal data processing. For example, in the case of Google Spain, the data controller processed data which

[16] Under the Article 2 of the Data Protection Directive, personal data processing ('processing') means any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;

[17] By de-identification, we mean the complete removal of a subject's personal information

[18] For the purposes of the article, anonymization is perceived to be a process of encrypting data in a form which enables the re-identification of data

it had gained lawfully, but it was still obliged to remove the personal information upon request, because it could not justify processing the personal information after receiving the data subject's request.

Therefore, dealing with a request to remove personal information contains two steps. **In the first step**, the operator decides, whether it is legally entitled to process such information. **In the second step**, the operator decides, whether the data subject's objection is justified or whether the data controller can justify a refusal of the request. The data controller considers whether the extent and manner of the processing of such information is truly relevant and proportionate (i.e. justifiable) to the purpose of the processing. The relevance and proportionality of this data may change over time, as the CJEU pointed out, *even the initially lawful processing of accurate data may, in the course of time, become incompatible with the directive where that data is no longer necessary in the light of the purposes for which it was collected or processed. That is particularly so where it appears to be inadequate, irrelevant or no longer relevant or excessive in relation to those purposes and in the light of the time that has elapsed.*[19]

The court also ruled that the denial of the erasure of personal data by the data controller cannot be justified merely by the data controller's economic interests or merely by the interests of the general public in having access to that information based on a search relating to the data subject's name[20], unless there is a particular reason for displaying such data because of the role of the individual in the public life. The question remains, what other grounds can justify a refusal to remove personal data. We believe that a grey literature repository has the right to display documents which contain specific personal information, if the processing of such personal information is necessary for the performance of a task carried out in the public interest[21,22], especially if it stores documents which have been created in the course of research into or the administration of public matters. We hold the opinion that the repositories cannot justify a refusal to grant a request with the argument that the data has been made available legally online by a third party. This argument might justify the processing of personal data under some jurisdictions, but it is not valid enough to justify the refusal of the data subjects' request to remove his or her personal information. The mere fact, that the information is available from other sources does not clarify whether the processing of such data serves a justifiable purpose.

Complying with the request

If the operator decides to comply with the request, it may remove the document completely, if the document is completely excess for the purposes of the repository. If the document itself is relevant, but the personal information contained therein is inaccurate or irrelevant for the purposes for which the document has been made available, it is possible to merely de-identify

[19] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González

[20] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González

[21] See TAN, Domingo R. Personal privacy in the information age: Comparison of internet data protection regulations in the United States and European Union. *Loy. LA Int'l & Comp. LJ*, 1999, 21: 661. NARAYANAN, Arvind, et al. A critical look at decentralized personal data architectures. *arXiv preprint arXiv:1202.4503*, 2012. MILLARD, Christopher; HON, W. Kuan. Defining 'personal data' in e-social science. *Information, Communication & Society*, 2012, 15.1: 66-84.

[22] See Art. 7 (e) of the Data protection directive

the document or anonymize certain information. If the institution operates a search engine, it may be enough to remove the information as a keyword from the search site so that the document still remains indexed, but cannot not found, if the user enters the name of the data subject in the query.

References

AUSLOOS, Jef. The 'Right to be Forgotten'–Worth remembering? *Computer Law & Security Review*, 2012, 28.2: 143-152. Factsheet on the right to be forgotten.

BENNETT, Steven C. Right to Be Forgotten: Reconciling EU and US Perspectives, The. *Berkeley J. Int'l L.*, 2012, 30: 161.

NARAYANAN, Arvind, et al. A critical look at decentralized personal data architectures. *arXiv preprint arXiv:1202.4503*, 2012.

HON, W. Kuan. Defining 'personal data' in e-social science. *Information, Communication & Society*, 2012, 15.1: 66-84.

ROSEN, Jeffrey. The right to be forgotten. *Stanford law review online*, 2012, 64: 88.

TAN, Domingo R. Personal privacy in the information age: Comparison of internet data protection regulations in the United States and European Union. *Loy. LA Int'l & Comp. LJ*, 1999, 21.

USTARAN, Eduardo, The wider effect of the 'right to be forgotten' case, *Privacy and data protection Journal*, vol. 14, Issue 8 (online) [cit. 20-11-2014]. Available from <http://www.hldataprotection.com/files/2014/09/The-wider-effect-of-the-right-to-be-forgotten-case-Eduardo-Ustaran-Hogan-Lovells.pdf>.

EUROPEAN COMMISSION. *Europa.eu* [online]. [cit. 2014-10-25]. Available from: <http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf>.

Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.