Integrated Citizen Centered Digital Health and Social Care
A. Värri et al. (Eds.)

doi:10.3233/SHTI200721

Description of Data Breaches Notifications in France and Lessons Learned for the Healthcare Stakeholders

Marie SIMON[a] and Vincent LOOTEN[b,1]
[a] *Université Paris-Est Créteil, Créteil, France*
[b] *UMRS 1138, Centre de Recherche des Cordeliers, Université de Paris, Paris, France*

Abstract. Although the consequences of the General Data Protection Regulation (GDPR) have been widely discussed, the violations have not been described in medical literature. In this study, we focus our analyses on the data breach notifications, in France, defined in the article 4 of GDPR as "a breach of security resulting, accidentally or unlawfully, in the destruction, loss, alteration, unauthorized disclosure of personal data transmitted, stored or otherwise processed, or unauthorized access to such data." Among 3,824 data breach notifications reported between May 2018 and February 2020, 244 (6.4%) is related to the health sector. Loss of confidentiality is the most important breach (80.7%) in this sector, followed by the loss of availability (27.5%). Malicious cause occurred in 58.2% of them. We hypothesized a phenomenon of underreported data breach incidents in health due to a mismatch between cybersecurity and data privacy issues.

Keywords. Policy, Data Privacy, Cybersecurity

1. Introduction

In 2017, the WannaCry cyberthreat affected more than 600 organizations as the National Health Service (NHS) in England; in 2018 the Singapore Health System reported a major breach of over one million of patient records: cybersecurity attacks are a growing threat to healthcare. Included in the General Data Protection Regulation (GDPR), cybersecurity in health is a major issue for the next decade. Although the consequences of GDPR have been widely discussed, the violations have not been described in medical literature. Since May 2018, the GDPR provides the mandatory legal framework for all data processing including European citizens' personal data[1]. National authorities across the European Union can sanction any company or administration performing non-conform data processing regarding to the GDPR. From the researcher's perspective, Peloquin *et al.*[2] exposed some technical challenges for data reuse: the anonymization or the pseudonymization of personal data, the management of consent, the cross-border transfers of personal data and the right limitations in the research context. Furthermore, Bernd Blobel and Pekka Ruotsalainen[3] proposed a model to implement data governance and data access management into a medical information systems. However, a description of the GDPR violations recorded by the national authorities in Europe could

[1] Corresponding Author, Dr Vincent Looten, Université de Paris, Paris, France; E-mail: lootenv@gmail.com

provide essential information about the legal practice of this regulation and the impact on its implementation. In this study, we focus our analyses on the data breach notifications defined in the article 4 of GDPR as “a breach of security resulting, accidentally or unlawfully, in the destruction, loss, alteration, unauthorized disclosure of personal data transmitted, stored or otherwise processed, or unauthorized access to such data”. The aim is to describe data breach notifications in France.

2. Methods

Definitions. The French national authority for data privacy is the CNIL (“*Commission nationale de l’informatique et des libertés*” in French). The GDPR have made mandatory to notify the CNIL of any personal data breach that poses a risk to the rights and freedoms of personal data. This notification to the CNIL must be made within 72 hours, by the responsible for processing or by its representative.

Data sources. We extracted data breach notifications reported to the CNIL from May 2018 to February 2020. The code and the data used in this study are available at www.github.com/vlooten/databreach, while more recent data can be downloaded from the open data governmental website (www.data.gouv.fr).

Outcomes. Three types of violation were described: the loss of confidentiality, the loss of integrity and the loss of availability. These categories are not exclusive. We described the number of people impacted by the breaches according to the same categories proposed in the original dataset. We described the cause of the breach (accidental, malicious or unknown) and the origin (Internal, external or unknown). Data breaches included individual identifiers are at higher risk regarding GDPR regulation. Thereby, we performed a focus in the health notification to compare data breaches included or not the national identification number (“*numéro d’inscription au répertoire national des personnes physiques*“ or *NIR* in French), which is a permanent identifier throughout the individual’s lifetime.

Statistical analyses. Data were expressed as numbers (%). Chi2 tests (for categorical data) was used to compare groups. All tests involved use of R 3.6.1(R Foundation, Vienna, Austria).

3. Results

Among 3,824 data breach notifications reported between May 2018 and February 2020, 675 (17.7%) are related to the administration, 660 (17.3%) to science and education activities, 485 (12.7%) to the financial and insurance activities, 326 (8.5%) to the Information and communication sectors and 244 (6.4%) to the Health sector. Table 1 presents a description of the whole dataset and a comparison between Health sector and the other activities.

Among the 503 notifications included the national identification number (NIR), 121 (24.1%) are related to the administration, 112 (22.3%) to the science and education, 87 (17.3%) to the commercial and industrial sectors, 71 (14.1%) to the financial and insurance activities, 36 (7.2%) to the health sector 28 (5.6%) to the information and communication sectors, and 48 (9.5%) to other sectors. Table 2 proposed a description of the data breach notification for the health sector and a comparison between notification included NIR and the others.

Table 1. Description of the data breach notification and comparison between Health sector and the other sectors

	All notifications (N=3824)	Health sector (N=244)	Other activity (N=3580)	p value
Year of accident				
2018	1170 (30.6%)	41 (16.8%)	1129 (31.5%)	<0.001
2019	2287 (59.8%)	174 (71.3%)	2113 (59.0%)	
2020	367 (9.6%)	29 (11.9%)	338 (9.44%)	
Type of violation				
Loss of confidentiality	3450 (90.2%)	197 (80.7%)	3253 (90.9%)	<0.001
Loss of integrity	406 (10.6%)	27 (11.1%)	379 (10.6%)	0.898
Loss of availability	659 (17.2%)	67 (27.5%)	592 (16.5%)	<0.001
Number of people impacted				0.009
<=5	919 (24.0%)	69 (28.3%)	850 (23.7%)	
[6-50]	652 (17.1%)	50 (20.5%)	602 (16.8%)	
[51-300]	746 (19.5%)	56 (23.0%)	690 (19.3%)	
[301-5000]	1010 (26.4%)	45 (18.4%)	965 (27.0%)	
>=5000	497 (13.0%)	24 (9.84%)	473 (13.2%)	
Cause of accident				0.516
Accidental	1151 (30.1%)	62 (25.4%)	1089 (30.4%)	
Malicious	2138 (55.9%)	142 (58.2%)	1996 (55.8%)	
Unknown	535 (14.0%)	40 (16.4%)	495 (13.8%)	
Origin of accident				0.200
Internal	1060 (27.7%)	64 (26.2%)	996 (27.8%)	
External	2229 (58.3%)	140 (57.4%)	2089 (58.4%)	
Unknown	535 (14.0%)	40 (16.4%)	495 (13.8%)	

Table 2. Comparison between data breach notifications with NIR and without NIR in the health sector

	Health sector (N=244)	Included NIR (N=36)	Without NIR (N=208)	p value
Year of accident				
2018	41 (16.8%)	5 (13.9%)	36 (17.3%)	0.809
2019	174 (71.3%)	26 (72.2%)	148 (71.2%)	
2020	29 (11.9%)	5 (13.9%)	24 (11.5%)	
Type of violation				
Loss of confidentiality	197 (80.7%)	29 (80.6%)	168 (80.8%)	1.000
Loss of integrity	27 (11.1%)	7 (19.4%)	20 (9.62%)	0.090
Loss of availability	67 (27.5%)	13 (36.1%)	54 (26.0%)	0.290
Number of people impacted				
<=5	69 (28.3%)	4 (11.1%)	65 (31.2%)	0.003
[6-50]	50 (20.5%)	3 (8.33%)	47 (22.6%)	
[51-300]	56 (23.0%)	14 (38.9%)	42 (20.2%)	
[301-5000]	45 (18.4%)	10 (27.8%)	35 (16.8%)	
>=5000	24 (9.84%)	5 (13.9%)	19 (9.13%)	

Cause of accident				
Accidental	62 (25.4%)	10 (27.8%)	52 (25.0%)	0.162
Malicious	142 (58.2%)	24 (66.7%)	118 (56.7%)	
Unknown	40 (16.4%)	2 (5.56%)	38 (18.3%)	
Origin of accident				
Internal	64 (26.2%)	9 (25.0%)	55 (26.4%)	0.127
External	140 (57.4%)	25 (69.4%)	115 (55.3%)	
Unknown	40 (16.4%)	2 (5.56%)	38 (18.3%)	

4. Discussion

Main results. Among 3,824 data breach notifications reported between May 2018 and February 2020, 244 (6.4%) is related to the health sector, increasing by a factor four between 2018 and 2019. Data breach characteristics of the health sector were similar to data breach characteristics of the other sectors. Loss of confidentiality is the most important breach (80.7%) in health sector, followed by the loss of availability (27.5%), some data breaches are mixed. 175 (71.7%) notifications reported fewer than 300 people impacted. Malicious cause occurred in 58.2% of them, accidental cause accounted for 25%.

Technical significance. Firstly, we didn't find important differences between data breach notifications in health and the other sectors but may lead to higher threat for citizens regarding to international experience [4]. Secondly, the French ministry of health and the French digital health agency have reported 327 incidents in 2018 and 392 in 2019, included respectively 276 and 333 hospitals. The rates of malicious incidents were 41% in 2018 and 43% in 2019[5]. Authorities hypothesized a phenomenon of underreported incidents: "*The total number of reports is still low compared to the number of structures concerned by the reporting obligation (more than 3,000) and the probability that at least half of the structures concerned have had to deal with an incident that has impacted its normal operation during the year*." Our results are similar with a lower amount of data breaches notification. Worldwide, healthcare lead in number of incidents (27%), as described in 2018 by the European Union Agency for Network and Information Security report [6], which is much more than the 6,4% notified in France based on our results. Thirdly, regarding the increase of data reuse for research purposes in France [7], the data processing included national identification numbers are regulated by the French law [8]. Nevertheless, only 36 (14.8%) notifications in the health sector included NIR, with 4 (11.1%) data breaches impacted 5 people or lower, which is non-realistic. Therefore, we hypothesized a phenomenon of underreported data breach incidents due to a mismatch between cybersecurity and data privacy issues. This underreported is likely a violation of GDPR. We could explain this underreporting by the distinction between data privacy and cybersecurity in the hospitals' organization in France. Data privacy is managed by the chief information officers with the data protection officer as advisors; they are focus on users' community and data processing purposes. Cybersecurity is leads by the chief information security officers focus on the data infrastructure integrity. We hypothesized that all data breaches cannot be detected by the chief information security officers (e.g. breaches with 5 people or less or internal breaches).

Perspectives and recommendations. Dean F. Sittig and Hardeep Singh[9] proposed a four steps socio-technical approach that organizations can undertake to secure an electronic health record system: (1) To ensure adequate system protection by correctly installing and configuring computers and networks (2) To ensure more reliable system defense by implementing user focused strategies (3) To ensure a comprehensive system monitoring of suspicious activities, and (4) To respond, to recover, to investigate, and to learn from ransomware attacks. For practical implementation, we recommend: (1) to plan seasonal assessments of information security management systems and to try to meet the international standards for information security with long-term and comprehensive perspectives as recommended by W.-S. Park at al[10], (2) to reduce the end point complexity (due to a technology saturated environment) and improving internal stakeholder alignment as recommended by M.S. Jalali, and J.P. Kaiser[11]. Finally, to improve the completeness of data breaches notification database, an electronic declaration system could be proposed to all users of the information system included physicians and patients.

Conclusion. We highlight a phenomenon of underreported data breach incidents in health possibly due to a mismatch between cybersecurity and data privacy issues.

References

[1] Demotes-Mainard J, Cornu C, Guérin A, et al. How the new European data protection regulation affects clinical research and recommendations? Therapies. 2019; 74: 31–42. doi:10.1016/j.therap.2018.12.004.

[2] Peloquin D, DiMaio M, Bierer B, and Barnes M. Disruptive and avoidable: GDPR challenges to secondary research uses of data. Eur. J. Hum. Genet. 2020. doi:10.1038/s41431-020-0596-x.

[3] Blobel B and Ruotsalainen P. How Does GDPR Support Healthcare Transformation to 5P Medicine?, Stud. Health Technol. Inform. 2019; 264: 1135–1139. doi:10.3233/SHTI190403.

[4] Ghafur S, Kristensen S, Honeyford K, Martin G, Darzi A, and Aylin P. A retrospective impact analysis of the WannaCry cyberattack on the NHS. Npj Digit. Med. 2019; 2: 98. doi:10.1038/s41746-019-0161-6.

[5] Observatoire des signalements d'incidents de sécurité des systèmes d'information pour le secteur santé. 2019. Ministère des solidarités et de la santé et Agence du Numérique en Santé. esante.gouv.fr/sites/default/files/media_entity/documents/ans_acss_rapport_public_observatoire_signalements_issis_2019_v0.10.pdf.

[6] ENISA Threat Landscape Report 2018 15 Top Cyberthreats and Trends. European Union Agency for Network and Information Security. doi:10.2824/622757

[7] Looten V and Simon M. Impact Analysis of the Policy for Access of Administrative Data in France: A Before-After Study. Stud. Health Technol. Inform. 2020; 270: 1133–1137. doi:10.3233/SHTI200339.

[8] Tout savoir sur le décret « cadre NIR » dans le champ de la protection sociale. Commission Nationale de l'Informatique et des Libertés, (2020). https://www.cnil.fr/fr/tout-savoir-sur-le-decret-cadre-nir-dans-le-champ-de-la-protection-sociale.

[9] Sittig D and Singh H. A Socio-technical Approach to Preventing, Mitigating, and Recovering from Ransomware Attacks. Appl. Clin. Inform. 2016; 7: 624–632. doi:10.4338/ACI-2016-04-SOA-0064.

[10] Park W-S, Seo S-W, Son S-S, Lee M-J, Kim S-H, Choi E-M, Bang J-E, Kim Y-E, and Kim O-N. Analysis of Information Security Management Systems at 5 Domestic Hospitals with More than 500 Beds. Healthc. Inform. Res. 2010; 16: 89. doi:10.4258/hir.2010.16.2.89.

[11] Jalali MS and Kaiser JP. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. J. Med. Internet Res. 2018; 20: e10059. doi:10.2196/10059.