Data Breaches in Europe:
Reported Breaches of Compromised Personal Records in Europe, 2005-2014

Philip N. Howard

CMDS Working Paper 2014.1

Center for Media, Data and Society
School of Public Policy
Central European University

October, 2014

Table of Contents

Tables

Figures

I. Executive Summary

A growing number of massive data breaches are degrading the personal privacy of people around the world. Data security and privacy policy are ongoing concerns in Europe. But it can be difficult to assess privacy breaches in Europe in particular, since many of the biggest incidents of compromised personal records involve people and organizations from around the world. This working paper offers early descriptive statistics and analysis of the first cross-national, systematized event log of data breaches in Europe. The data is available for download at http://cmds.ceu.hu/.

Methodology. The sample frame includes major media news reports on compromised personal records and is unique for:

- sampling 28 European Union member countries, plus Norway and Switzerland;
- sampling from 2005 through the third quarter of 2014;
- sampling credible news sources in national languages;
- high social science standards for event database construction, with multiple sourcing, inter-coder reliability tests, recoding, and specific exclusion criteria.

Findings. A data breach is defined as any incident involving the loss or exposure of digital personal records. Personal records are defined as a) data containing privileged information about an individual that cannot be readily obtained through other public means and b) this information only known by an individual or by an organization under the terms of a confidentiality agreement. Preliminary analysis reveals that over the last decade:

- Some 229 data breach incidents involved the personal records of people in Europe. Globally, all these incidents resulted in the loss of some 645 million records, though not all of these breaches exclusively involved people in Europe. Within Europe, we confirmed 200 cases involving people in Europe, and 227 million records lost in Europe-specific breaches.

- The total population of the countries covered in this study is 524 million, and the total population of internet users in these countries is 409 million. Expressed in ratios, this means that for every 100 people in the study countries, 43 personal records have been compromised. For every 100 internet users in the study countries, 56 records have been compromised.

- Fully 51 percent of all the breaches involved corporations and 89 percent of all the breached records were from compromised corporations. Among all the kinds of organizations from which personal records have been compromised, 41 percent of the incidents involved clear acts of theft by hackers, but 57 percent of the incidents involved organizational errors, insider abuse, or other internal mismanagement (2 percent unspecified).

- The level of sophistication and detail in journalism about issues of privacy and personal data has increased, but is largely driven by national “mandatory reporting” rules in particular countries. In other words, we know most about data leaks in countries where organizations are required to report that personal records have been compromised.

II. Reports of Data Breaches in Europe

1. Introduction

The internet, mobile phones, and a host of other new information technologies have allowed more and more people to conduct the business of their personal lives over digital media. And even people who are not heavy technology users are tracked, surveilled, and surveyed, making facts about their attitudes, behaviors, and other life details are tracked electronically. Many of those activities, such as banking, shopping, e-government, social networking, and emailing, require disclosure of a certain degree of personal data. The data citizens or companies store online ranges from email or postal addresses, login information or passwords to sensitive personal information, including bank and credit card account information. The more activities take place online, the more data is stored in servers. Such situations pose certain challenges to maintaining privacy and keeping data safe.

Almost everything we know about privacy violations in Europe comes from news reports on specific breaches. So what can we learn about the big picture assembling the highest quality reports—across countries and languages—about compromised personal records?

Privacy policy and data protection is a contemporary concern of policymakers in Europe. Countries such as Germany, the United Kingdom, and Ireland are implementing strict rules for information management and data protection. Due to software vulnerability, data mismanagement or simple human indiscretion data is stolen or lost quite frequently. There have been several attempts to measure the costs of such data loss. However, there have been few systematic efforts to measure the scale of data breaches across Europe. Given that Europe has unique values when it comes to privacy and surveillance, we seek to provide new knowledge on the scale and quantity of breached data.

Non-governmental organizations and data protection authorities acknowledge that there has been a steady increase in data breaches. Although Europe is moving towards the unified policy of data protection and requirements for reporting data breaches, there is a huge lack of information about exact cases and incidents. Not only are there few news accounts of big picture trends in data breaches, but public policy researchers have little comparative data to work with. Several years ago, another comparative event database about incidents of compromised records in the United States revealed that 1.9 billion records were compromised between 1985 and 2006. In 2006, this meant that for every hundred U.S. adults 875 personal records had been breached.[1]

For us, the lack of organized event records is both an empirical obstacle and an opportunity to generate new knowledge about data and privacy protection. This study investigates a decade of records to help assess both the changing volume and character of data breaches and the way in which those breaches are reported to the public in Europe.

2. Methodology

To understand the trends in data breaches, we built an original event database of incidents as reported in credible, multilingual news media in Europe. The database we built and on which our findings are based includes all the cases reported on the internet in which personal data of

[1] Erickson, Kris, and Philip N. Howard. (2007). "A case of mistaken identity? News accounts of hacker, consumer, and organizational responsibility for compromised digital records." *Journal of Computer-Mediated Communication* 12(4): 1229-1247.

European citizens—those of the 28 European Union Member States plus Norway and Switzerland—were compromised during the period 2005–2014, including data breaches, leaks, and identity thefts.[2] Because so few cases were reported prior to 2005, we limit our comparative event analysis to the incidents coded and catalogued over the last decade.

Since our study focuses on cases whose victims were European, instances in which European citizens compromised data of overseas citizens do not appear in our database, even though we found dozens of such cases. Although they are unquestionably examples of privacy violation, we also did not include cases of surveillance, as we were concerned with how already existing data is handled rather than with how data is collected. The database, moreover, excludes instances in which the compromised data were not personal in nature. Neither were we concerned with hacker attacks whose direct aim was financial gain—for example, those involving transferring money from one bank account to another—and otherwise did not compromise personal data. Incidents involving phishing or malware were excluded because their impact was impossible to estimate in the overwhelming majority of cases. We were not concerned with privacy breaches of paper-based records.

The compromised personal records had to concern citizens of the countries we were studying. There were four cases of breaches on embassies that were excluded from this analysis. These were the embassies of governments outside Europe that were physically located in Europe and were compromised. But the data lost probably involved nationals from outside Europe. If the news report included details about the kinds of documents compromised or the number of gigabytes of data put at risk, the case was coded but then excluded from analysis. Only cases with personally identifiable information of people living in the European Union, Norway and Switzerland. Given the diversity of reporting styles and variations in information completeness, we implemented the standard rubric for codifying and quantifying news reports. If the news reported “thousands” then the value 3,000 was recorded, if the news reported “tens of thousands” then the value 30,000 was recorded. Broad, nonspecific reports of phishing and malware attacks were excluded. Reports about the compromised personal records of European Union citizens by the national security agencies of other countries (such as the United States) were also excluded. Second hand reports and unsourced reports were excluded.

We looked for reports on personal data being compromised on specialized websites and in the LexisNexis Academic and the Google News databases. Appendix A identifies the coding variables, and Appendix B identifies the list of specialized websites used in the snowball sampling strategy. Appendix C presents the country by country breakdown of incidents. We also utilized the language skills present in the project team[3] by searching for relevant terms on Google and thus found related reports on national specialized and news websites. For the rest of the countries in our sample we did top level domain name searches using Google Translate. As much as possible, we relied on credible news reports rather than on user-generated content. We did not use sources that contained aggregated data because of the possibility of unidentifiable overlaps among them and their different conceptualizations of what constitutes a data breach. We also aimed to find more than one source for each of the cases. All in all, the

[2] The countries included in this case list include: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom.

[3] The language-specific searches include: Bulgarian, Croatian, Czech, Dutch, English, Estonian, Finnish, French, German, Hungarian, Italian, Latvian, Lithuanian, Polish, Slovakian, Slovenian, and Romanian.

research team of eleven people spent 450 hours identifying and evaluating data breach reports. As is common with event datasets, a standard set of incident descriptors was collected:

- when it was reported,
- when it occurred,
- which country and organization was affected by it,
- to which sector the affected organization belongs,
- how strong its impact was (measured as the total number of people, records, gigabytes, or emails compromised);
- what kind of data was compromised;
- whether the data was stolen, mismanaged, or lost; and exactly what kind of breach happened.

In case of hackings, we identify the country of the attacker, if known. Moreover, we classified instances of whistleblowing each time the source did so.

Every case had its own complications. An illustrative example is the case in which German tax authorities obtained records of German nationals from an employee of a Swiss bank: such an event may be seen as a case of whistleblowing from the German perspective, or as one involving illegally and illegitimately compromised data from the Swiss side. Assessing the impact of cases involving multiple countries also turned out to be a challenge. Through regular coder training sessions, inter-coder reliability scores, and searching for multiple sources, we were able to fact check incidents in a variety of ways. There were dozens of cases, in which malware is known (by security expert) to have compromised personal records. While we have been able to count the number of cases like this, we have no way to evaluate their impact on individual privacy.

In addition, there were many kinds of cases that are excluded from this analysis for methodological reasons. If too many details were unknown or the sourcing was of questionable quality the case was excluded. Several dramatic cases simply didn't qualify as incidents of compromised personal records. Bitcoin wallets, involving print files, browsing history, unspecified forms of data, laptop computers, USB sticks, contact lists, call histories, and photos were not included unless the reports specifically mentioned that personally identifiable information had been compromised. On one occasion the unreleased version of the Dutch government's annual budget was stolen. In another, an attack of the dating website www.beautifulpeople.com allowed a large group of "ugly people" to be admitted to the dating site. In one dramatic breach a large value of carbon credits were stolen—but no personally identifiable information was lost. While many of these cases are interesting examples of breaches, our sample frame and quality controls standards made it important to make consistent decisions about the inclusion and exclusion criteria.

3. Findings: Descriptive Statistics

Table 1 identifies some basic descriptive statistics that reveal both the nature of this event dataset and some important trends. All in all, there were 229 reported incidents in which the personal records of at least a few people in Europe were breached. Over all incidents around 641 million email addresses, names, passwords and other kinds of personally identifiable information was compromised, though most reports do not specifically disaggregate the

Table 1: Quick Fact Table	Values
Total Number of Breaches Involving European Targets	229
Total Volume of Breached Records Across All Incidents	641,979,541
Number of Times a Specific Country in Europe Was Identified as Target	267
Number of Global Breaches Involving European Targets	29
Volume of Records From Global Breaches that Impact People in Europe	415,012,618
Volume of Records From Europe-Specific Breaches	226,966,923
Total Number of Breaches Involving European Targets	229
Total Volume of Breached Records Across All Incidents	641,979,541
Number of Times a Specific Country in Europe Was Identified as Target	267
Number of Global Breaches Involving European Targets	29
Volume of Records From Global Breaches that Impact People in Europe	415,012,618
Volume of Records From Europe-Specific Breaches	226,966,923
Number of People Living in Study Countries	523,730,791
Number of Internet Users in Study Countries	408,583,658
Volume of Records from Europe-Specific Breaches per 100 People	43
Volume of Records from Europe-Specific Breaches per 100 Internet Users	56
Number of All Breaches in Which Attacker Was Unspecified	48
Number of All Breaches in Which There was No Attacker	116
Number of All Breaches in Which The Attacker Was Known and Specified	65
Number of All Breaches in Attack Originated Within the EU	45
Number of Cases in Which Attack Originated Within the UK	10
Number of All Breaches	274
Number of All Breaches in Which There Was an Attacker, The Attacker Was Specified, and the Attacker Was Not in Europe	20
Percent of All Breaches in Which Attacker Was Known and Specified As Being In Europe	69
Percent of Breaches Revealed by a Whistleblower	2
Percent of Breaches in Which Attack Originated Within the UK	15
Percent of Europe-Specific Breaches Known to Originate in the UK	24
Percent of All Breaches Involving Corporations	51
Percent of All Breaches Involving Hacker	41
Percent of All Breaches Involving Insider Abuse, Missing Hardware, Accidental Exposure Online, or Administrative Error	57

proportion of victims known to reside in Europe. Many of the reports, however, do list some of the countries in which there were known victims. European countries were specifically identified 267 times in the event dataset.

Reporting requirements make the problem more transparent. The stricter regulation the country has, the more cases can be identified and the more details about those breaches can be described. As organizations are obliged to report data breaches in some European countries, this leads to the greater number of cases.

Some breaches have an impact on people around the world. There were 29 of these uniquely global breaches, many of which involved major credit card companies or data mining firms that are incorporated in the United States but store data on people living in Europe. These global breaches accounted for 415 million personal records, though again news reports rarely disaggregated the proportion of this volume of breached data impacting people in Europe. This does mean, however, that 226 million records were compromised during security breaches for which a specific country in Europe was identified.

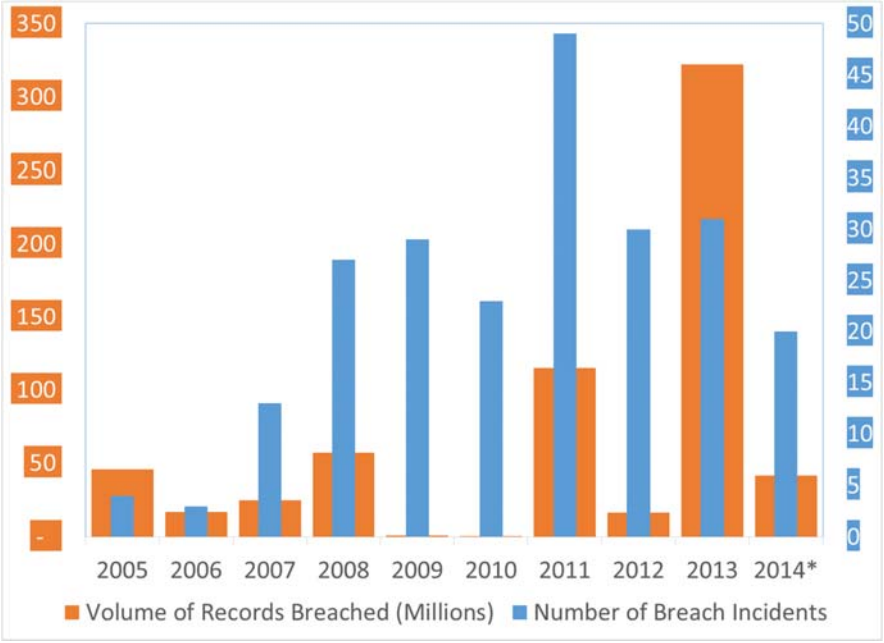
Since there are 524 million people living in the study countries and 227 million records have been compromised, the ratio of compromised records to people is 43 to 100. Since there are 409 million internet users living in the study countries the ratio of compromised records to internet users is 56 to 100. On the basis of news reports, however, we cannot say much about the real distribution of privacy violations across race, gender, class, or other social categories. We can, however, say something about the distribution of security breaches by country.

Some news reports provided additional information, and we coded for several additional trends across all reports. All the incidents in this event data set involve people targets residing in Europe. While news reports provide some comparable data and context, not all kinds of information about an incident are consistently reported. Many of the incidents involved organizational error. Other involved an external attack, but the identity and location of the attacker may not have been known. Among all the cases—global and Europe-specific—in which the breach of personal records involved an attack, 15 percent of the attacks were launched from the United Kingdom. Among the cases in which the targets resided in Europe, 24 percent were known to originate in the United Kingdom.

This study did not analyze incidents of privacy violations by government security agencies around the world, incidents we know about mostly because of whistleblowers like Chelsea Manning and Edward Snowden. We tracked reports that clearly identified a whistleblower but found that only 2 percent of the reported incidents clearly involved a whistleblower.

The last decade has seen a significant increase in the number of incidents and volume of records breached, and there are three reasons for this trend. While more and more people have been putting personal information online, journalists have become better at reporting on breach incidents and more and more governments are requiring that the victims of breaches be informed.

Figure 1: Volume and Number of Breach Incidents, 2005-2014



*Inclusive to Third Quarter 2014

Reporting requirements make the problem more transparent. The stricter regulation the country has, the more cases can be identified and the more details about those breaches can be described. As organizations are obliged to report data breaches in some European countries, this leads to the greater number of cases.

Which Countries in Europe Are Impacted the Most?

Table 2 identifies the countries in Europe with the most concerning patterns of personal information breaches over the last decade. Germany, Greece, Netherlands, Norway and the United Kingdom are the five countries with unusually high numbers of incidents and large volumes of records breached.

Table 2: Severity of Breach Patterns, Top 5 Country Targets	Compromised Records Per 100 People	Compromised Records Per 100 Internet Users
Germany	68	79
Greece	81	140
Netherlands	23	24
Norway	80	83
United Kingdom	220	245

According to the best publicly available data on incidents in Europe, organizations in these countries are doing a poor job managing personal information and are the biggest targets for cybercrime. Again, much of what we know is shaped by reporting requirements, and it is likely that the media's coverage of data breaches has improved over time but does not pick up all incidents. Nonetheless, the per capita trends across countries suggest that as a national average, the ratio of compromised records to people in the UK is over 2:1. The ratio of compromised records to internet users is more than 1:1 in Greece and 1:2 in the UK.

Several countries, including Croatia, Estonia, and Slovenia, had no reported incidents of citizens losing data. We expect that some citizens in these countries have been impacted by the large global breaches, but that their absence from the event data set can be explained by our sampling frame. Either journalists did not specifically mention these countries as being the targets of attack or the incident reports from these countries were of dubious quality.

a) How and Why Are Personal Records Compromised?

To understand how and why personal records were being compromised we came up with two coding schemes. The first was a simple, three category typology of breaches. Each case was coded for whether the data was stolen, lost, or mismanaged (exposed online or mismanaged in an organizational accident). **Table 3** demonstrates that by this typology, the some 57 percent of incidents involved theft, and 570 million records were stolen over the last 10 years.

Table 3: Type of Loss, Three Categories	By Number of Incidents	Percent	By Number of Records	Percent
Lost	20	9	34,980,276	5
Mismanagement	77	34	36,751,944	6
Stolen	131	57	570,079,321	89
TOTAL	228	100	641,811,541	100

Many of the reports of stolen data, upon further inspection, were cases in which a disgruntled employee or company insider stole the data. In these cases, a large part of the story was absent or incomplete security. So a more nuanced six category typology was developed, often using the same keywords used by technology reporters and the organizations revealing a

breach. The coding system in **Table 4** reveals that while 42 percent of the cases clearly involved external attack by criminal hackers, the majority of cases involved problems internal to the organization: insider abuse or theft, hardware that the organization either lost track of or lost to theft, and administrative errors. A common accident involved mistakenly putting personal records online.

Table 4: Type of Loss, Six Categories	By Number of Incidents	Percent	By Number of Records	Percent
Administrative Error	22	10	33,171,867	5
Exposed Online	49	22	2,381,386	0
Insider Abuse or Theft	25	11	12,150,489	2
Missing or Stolen Hardware	29	13	37,273,276	6
Stolen - Hacker	94	42	556,106,552	87
Unspecified	4	2	656,413	0
TOTAL	223	100	41,739,983	100

b) Which Organizations Are Breached Most Often?

However, different kinds of organizations are impacted by such breaches. Certainly not all are commercial firms—but most are. **Table 5** reveals that over half the organizations reporting on a breach were businesses, and they lost fully 89 percent—some 538 million—personal records. Government offices were the next largest kind of organizational target. Almost a quarter of the incidents involved public agencies, but these breaches tended to be much smaller in scale. All in all, only a few of the breaches impacting people living in Europe involved a data loss by a non-profit, the military, a medical facility, or educational organization.

Table 5: Data Breach by Type of Organization Compromised	By Number of Incidents	Percent	By Number of Records	Percent
Commercial	117	51	538,349,868	89
Educational	11	5	80,221	0
Government	55	24	59,173,346	10
Medical	18	8	9,337,197	2
Military	8	3	917,001	0
Non-profit	12	5	1,818,765	0
Unknown	8	3	32,303,143	5
TOTAL	229	100	06,940,632	100

4. Findings: Europe-wide Comparisons

Across the data set, we identified two transnational phenomena involving particular types of attackers and targets.

a) Attacks by Anonymous

Anonymous is an international hacktivist network. The group conducts disruptive online activism, in the form of cyberattacks on many kinds of individuals and organizations. Somewhere between an affiliation of clubs and a distributed network of temporary teams, groups of people collaborating under some country-specific Anonymous chapters have breached many collections of personal records in Europe. Anonymous-Spain attacked multiple organizations including Sony Corps, Spanish banks, governments and other actors. 77 million account details of the Sony network were stolen. All three hackers arrested for the crime were Spanish. Anonymous-Spain also leaked 5 GB of financial documents of the People's Party, Spain's largest right-wing political party for the past twenty years from 1991–2011.

Anonymous-Sweden took over the official website of Sweden's National Board of Health and Welfare after police raided the office of the Stockholm-based web-hosting company PRQ. Anonymous-United Kingdom revealed thousands of British email addresses and encrypted passwords, including those of defense, intelligence and police officials as well as politicians and NATO advisers, have been revealed on the internet following a security breach by hackers. Among the huge database of private information exposed by self-styled "hacktivists" are the details of 221 British military officials and 242 NATO staff.

Another incident involved Anonymous-Italy, a group that hacked the system of the National Anti-Crime Computer Centre for Critical Infrastructure Protection (CNAIPIC - Centro Nazionale Anticrimine Informatico per la Protezione delle Infrastrutture Critiche), stealing 8 gigabytes of confidential documents. Anonymous-Italy also exposed a public figure, Father Don Giacomo Ruggeri, by dumping his personal emails on the internet. Subsequently, he was accused of child abuse, suspended of his duties and arrested by the Italian police.

b) Breaches of Crypto-Currencies

One of the more recent phenomena regarding data leaks are hacks targeting crypto-currencies. These are usually open-source peer-to-peer digital currencies which make use of a public ledger to keep track of transactions and wallet balances. While the protocols themselves are usually very resilient against various types of attacks (thanks to the use of strong cryptography) the services associated with crypto-currencies might be a weak spot.

The primary goal of crypto-currency hacks is usually a theft of coins, but data leaks are inevitably associated with such events either directly or indirectly. Given that the transaction ledger is public, crypto-currencies are pseudonymous rather than anonymous. If someone gets hold of coins associated with a particular wallet he or she is then able to track the transactions back through the ledger. Many public addresses (wallets) can be tied to a known entity or individual and therefore analysis of the ledger may yield potentially interesting intelligence and in some cases it may even be possible to find the actual owner of the coins. This works both ways: the stolen money can be traced as long as the attacker does not use some money laundering service (which could prove difficult for large amounts of coins).

Because of the pseudonymous nature of crypto-currency networks and because significant part of the economy works underground in what is commonly called the "deep web", it is often difficult to figure out the geographic origin of the attacked service or the extent of the hack. Below you can find an overview of several notable cases related to bitcoin thefts in the past three years, even though not all of these are part of the data set that we worked with.

The first notable crypto-currency theft that can be tied to Europe is probably the Bitomat.pl hack in June 2011. The Polish exchange, which was the third largest at that time, lost around 17,000 bitcoins worth over €150,000. In April 2013, an unknown attacker managed to reset the password of French exchange Bitcoin Central through its hosting provider's web interface effectively locking it out of its own site. The attacker requested a server reboot in rescue mode and managed to get away with a few hundred bitcoins worth tens of thousands of euros at the time.

United Kingdom-based online bitcoin wallet Blockchain.info lost 50 bitcoins worth a few thousand euros in August 2013 when someone exploited vulnerability in the JavaScript random number generator. Online forum BitcoinTalk was hacked by a group calling themselves "The Hole Seekers" in October 2013. While the full extent of the attack is not known, it is

possible that the hackers were able to access a database containing users' information including password hashes.

During an attack on bitcoin payment processor BIPS in November 2013 bitcoins worth €750,000 were stolen. Poland-based Picostocks was hacked in late November 2013 and around 6,000 bitcoins worth approximately €4.5m were stolen. Around that time Czech exchange Bitcash.cz lost 480 bitcoins from around 4,000 wallet balances worth roughly €74,000. It is believed that the attack was performed through their web interface. Another Poland-based company—Bidextreme.pl—was hacked in the same month and undisclosed amount of bitcoins was stolen. In early February 2014 bitcoin hardware manufacturer Cointerra was hacked. Email conversations were stolen and the customers had to be warned to be wary of possible phishing attempts following the attack. Poland's leading bitcoin exchange Bitcurex was hacked in March 2014 but only small portion of their operational balance (so called "hot-wallet" balance) was stolen.

It is not very common that a deep web service can be tied to a particular location but in the case of Sheep Marketplace it is believed that its servers were running in the Czech Republic. This case is worth mentioning because it is one of the largest thefts in bitcoin history. At least 96,000 bitcoins were stolen, worth €165m. It is not clear whether the market was hacked or if it was an inside job.

5. Findings: Unusual Country Cases

There are some unusual examples of data breach, where the data was lost or published in a surprising way. One example is from Denmark, where personal information of HIV patients was included in a PowerPoint presentation. This in itself was an accidental leak, but only for the audience at the presentation, however, later the PPT was published online. Another incident happened in the United Kingdom when a staff member of an educational institution lost their camera that held sensitive information, namely photographs of job applicants' passports. Another case took place before the 2011 Bulgarian elections when the Ministry of Foreign Affairs accidentally published online the names as well as the addresses of the permanent residences of Bulgarian nationals living abroad. Although the information was available online for a few hours, it made these citizens an easy and open target for theft and burglary. In another incident, FC Manchester City opened investigations against a rival club which might have hacked confidential records of players' signings along with their personal records.

While there are a great many cases of compromised personal data over the years, there are also several unusual country cases that stand out as being particularly egregious or indicative of peculiar trends. Here are some of the country cases that we identified as being unusual, in a comparative context. Not every country in the sample has cases as unusual as these.

a) Belgium

Many identified cases coming from Belgium refer to hacking directed at foreign embassies from European and non-European countries and international organizations' private data.

b) Bulgaria

Not many specific incidents of data leaks were found in Bulgaria. However, there were general reports on widespread identity theft and misuse of personal data in the country. At the beginning of 2011, the Bulgarian head of the Computer Crimes and Intellectual Property Department of the Ministry of Interior's Chief Directorate for Combating Organized Crime

stated that these types of crimes have doubled in number between 2006 and 2010 compared to data from previous years. Internet penetration in the country has also increased during this time, meaning there have been more and more individuals using the internet in general.

c) Czech Republic

Given the widespread discrimination against Roma in the Czech Republic, this incident involving the Czech Ministry of Education was rather controversial and highly publicized. In November 2011, an administrative error exposed the data of 893 Roma recipients of a governmental stipend for their studies and vocational training. The dataset included the names of the students as well as the amount of the stipend they were awarded. The Ministry of Education was not only missing some €16,400, but it also received a Data Leak Award in 2013.

d) France

In 2012, the American Chamber of Commerce in France was hacked by a hacker collective known as DeleteSec. As a result, hundreds of email addresses and passwords were compromised. However, the hackers claimed that they warned the Chamber in a message about an SQL injection error, informing the organization of a possible security threat before actually breaking into the computer system. However, the Chamber ignored the warning and even sent a hateful response back, according to DeleteSec. The case illustrates an important lesson to be learned: overconfidence is not something one should practice as a virtue in the digital era.

e) Netherlands

Condoms are not supposed to leak… but what about the data of those who order them online? In 2009, a Dutch website where young people could order free condoms and have them mailed to their address turned out to have a serious case of leaking personal data. When customers requested condoms to be sent to them via the website, they were given a client number. By simply changing that number, it was possible to access other customers' data who have been ordering online: their name, address, zip code and city. The most embarrassing part is that the website was primarily intended for youths who were shy about buying condoms in a shop—the data leak affected about 10,000 of them.

f) Italy

The majority of cases found for Italy were data breaches caused by hackers. One incident involving personal data theft from Sony Italy was carried out by a Turkish hacker group called Turkish Ajan, who leaked confidential personal records of users to the public. The other four cases involving hacking were tied to the Italian wing of Anonymous. According to the hacker collective's public announcements, they wanted to demonstrate the weak digital security measures taken by government bodies. Therefore, in 2011, they attacked the Italian police on two occasions stealing police employees' login information to the police's computer system. Anonymous launched a larger operation in 2012 and stole about 3500 records from the state police including police reports, mobile phone numbers, personal emails, information on salaries, and even soft-porn pictures were found in the compromised dataset.

g) Slovak Republic

In 2003, Orange, the Slovak telecommunication company, had problems securing the privacy of its customers. Almost 900,000 personal records were exposed online, including phone numbers (even unlisted ones), names, addresses and birth identification numbers of

subscribers. This case is especially interesting given that their sister company, Orange France, already compromised over 2 million personal data records in the first half of 2014.

h) Ireland

The Ireland Department of Social and Family Affairs has had several data breaches: the department has lost at least 400,000 records between 1985 and 2014. Some of these incidents involved stolen laptop(s), while others were insider abuse cases. Usually the department lost personal data with sensitive social welfare information, such as social security numbers and other personal records.

i) United Kingdom

Due to strict legislation, the United Kingdom has an enormous number of data breach cases, both paper-based and digital ones. A large portion of these data breaches are a result of carelessness, either on the part of the owners or handlers of personal records. Most cases involve administrative errors or mismanagement, such as not wiping hard drives of old computers offered for re-sale.

Medical records are often treated with extra care, but there are still dramatic cases of data breaches involving confidential medical information. In London, a private clinic decided to contract another company to computerize the paper records they kept on their patients, which contained confidential details on the patients' conditions, names, home addresses and dates of birth. However, after scanning the documents this company sub-contracted other types of work on the files—such as compiling them into a database—to a company in India. The records were offered for sale there by the firm's local employees, primarily to insurance companies or marketing executives for health products. Hundreds of thousands of personal medical records of UK patients have been outsourced to Indian companies this way, even though under the United Kingdom Data Protection Act it is illegal to send such documents outside the European Union—unless appropriate security is guaranteed. The case is a good example of cross-sector breaches that speak to the difficulties of classifying the information that was gathered for this research.

Another example illustrating this is the Surrey and Borders Partnership NHS Foundation Trust, an organization that provides a comprehensive range of services, such as mental health services, drug and alcohol abuse related services, learning disability healthcare services, and so on. Consequently, it is difficult to define the sector the organization operates in, as it is both educational and medical. In this respect, the sector of the organization's activity was determined on a case-by-case basis. Clearly, most of the cases involve multiple levels of responsibility across many different kinds of organizations.

6. Conclusion: Moving Forward with Mandatory Reporting

Public policy oversight of personal records is evolving quickly. These preliminary findings demonstrate the size and complexity of the problem and the positive value of mandatory reporting, for both public awareness and policy evolution. In March 2014, the European Parliament voted to support a new General Data Protection Regulation that outlines a complicated and strict legal framework for processing personal data. The decision to back the renewed data protection plan was triggered by a number of high-profile incidents of personal data loss across Europe, which moved the question of secure handling of personal information to the fore.

In 2009, the European Union made a breach notification law as part of its Directive on Privacy and Electronic Communications (E-Privacy Directive). The directive was to be implemented in national laws by May 2011. A new regulation on mandatory personal data breach disclosures came into force in August 2013, building on the provisions laid out in the E-Privacy Directive. According to the regulation, European Union telecom operators and internet service providers are obliged to notify national authorities of "any theft, loss or unauthorized access to personal customer data, including emails, calling data and IP addresses. Details concerning any incident, including the timing and circumstances of the breach, nature and content of the data involved, and likely consequences of the breach, must be reported."

In addition, the report must be made within 24 hours of the detection of the incident and within three days of being alerted to the breach. Telecommunications firms and internet service providers must also report on the measures they took to address the breach. The company also has to inform directly the subscribers whose data might have been compromised. The aim of the relatively strict legislation is also to incentivize companies to better encrypt and secure personal data. Provided that companies comply with certain security measures and recommendations put forth by the European Commission, they might be exempt from having to report the data breaches they suffer.

The regulation has been criticized for the vast burden it would place on authorities due to the expected high number of breach notifications. Companies suggested that the regulation should also involve categorizing data breaches according to the level of security risk they pose, in order to avoid 'notification fatigue' both for clients and operators. Furthermore, operators wish to maintain control over the communication of data breaches to their clients as to avoid negative impacts on their brand as much as possible.

In January 2014, the European Union's Justice Commissioner called for much larger fines for companies that breach European data privacy laws. New proposals currently under debate in the EP involve the establishment of a single European Union regulator that would be able to issue fines.

The new European data protection law framework is still likely to go through several changes as three-way negotiations take place among the European Commission, European Parliament and the Council of Ministers. The full set of regulations involving many contested privacy regulations (limits on 'profiling', requirements, using clear and plain language in privacy policies, obtaining the explicit consent of data subjects on processing any form of their personal data) are not likely to come into effect before 2016.

Currently, the United Kingdom has a mandatory reporting requirement in place for "organizations who provide a service allowing members of the public to send electronic messages", such as telecommunications or internet service providers. They have to notify the Information Commissioner's Office, an independent authority dealing with information rights, within 24 hours of becoming aware of the data and/or security breach. Other companies are not required by law to report incidents at this point; however, the ICO has managed to establish reporting data breaches as a 'best practice'. The authority made it clear that if it learns about incidents that have been unreported (from the press or any other sources), it will take such cases more seriously.

The Netherlands is planning to pass such a mandatory reporting bill, however, the latest amendment to the proposal in April 2014 would only require data breaches to be reported

when the breach has seriously adverse consequences. Such wording is not only vague, it also erodes the purpose of a mandatory reporting requirement, which would be to prevent data breaches from happening through implementing more rigorous security measures.

According to the European Union's new data breach regulation that came into effect in August 2013, in all member states' providers of publicly available electronic communication services—including telecommunications firms and internet service providers—have an obligation to notify the competent national authorities of data breach cases within 24 hours of the incident. However, there are differences as to how seriously each member state takes this otherwise binding regulation. Most European countries have a consumer data protection office of some kind with a website of information on how individuals and organizations can protect themselves. But not every country has an archive of incidents affecting their citizens, or a form for reporting incidents.

The various cases of data breach demonstrate an extremely nuanced and complex array of very diverse scenarios, which very well illustrate the difficulty of making good policies that reflect the complicated nature of the issue without posing limitations—legal or otherwise—on our use of ICTs and curbing the advantages they offer for everyday life.

III. About the Project

1. Correspondence

Please direct correspondence to Philip N. Howard, Director, Center for Media, Data and Society, Central European University, Nador 9, Budapest, 1051, Hungary, howardp@ceu.hu, @pnhoward.

2. About the Authors

Philip N. Howard is director of the Center for Media, Data and Society and a professor in the School of Public Policy at Central European University. He is also a professor at the University of Washington and a fellow at the Tow Center for Digital Journalism at Columbia University. He is the author, most recently, of *Democracy's Fourth Wave? Digital Media and the Arab Spring*. Currently, he is writing *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up*, a book about the future of global information politics for Yale University Press. He blogs at http://philhoward.org and tweets from @pnhoward.

3. Research Team

This research was conducted by the Spring Media Practicum at the School of Public Policy of Central European University: Gulnara Alimbayeva, Roxana Damian, Tamilla Dauletbayeva, Orsolya Gulyas, Zintis Hermansons, Tautvydas Juskauskas, Attila Mester, Róbert Papp, Radka Pudilova, Marija Stojanovska Rupcic.

IV. Institutions and Funding

1. The Center for Media, Data and Society

The Center for Media, Data and Society is the leading center of research on media, communication, and information policy in Central and Eastern Europe. Based in the School of Public Policy at Central European University, CMDS produces scholarly and practice-oriented research addressing academic, policy and civil society needs. CMDS research and activities address media and communication policy, social media and free expression, civil society and participation, fundamental communication and informational rights, and the complexities of media and communication in transition.

2. The School of Public Policy

The School of Public Policy (SPP) at Central European University, in the words of its founder, George Soros, is a "new kind of global institution dealing with global problems" through multi-disciplinary study of public policy, innovative teaching and research, as well as meaningful engagement with policy practice.

3. Central European University

Central European University (CEU) is a graduate-level, English-language university accredited in the U.S. and Hungary and located in Budapest. The university offers degrees in the social sciences, humanities, law, public policy, business management, environmental science, and mathematics. CEU has more than 1,500 students from 100 countries and 300 faculty members from more than 30 countries.

V. Appendix A: Case and Variable Definitions

Variable Name	Definition
Administrative error	Accidentally disclosing private data, for example by misplacing hardware, or by selling hardware that had not been wiped of identifiable information.
Attacker country	Location from which the breach originated. Country-to-individual or individual-to-country cases have been taken into account; country-to-country and government-sponsored attacks on other governments are not included.
Compromised records	Collections of electronic personal records that have been breached by third parties through illegal or negligent acts. The cases where data is sold to third parties for marketing purposes without users' informed consent are not taken into consideration as compromised records.
Data exposed online	Personal records are made accessible either by publishing online, software error or accidental disclosure.
Electronic personal records	Data containing privileged information about an individual that cannot be readily obtained through other public means; this information is only known by an individual or by an organization under the terms of a confidentiality agreement. Examples include individual personal credit histories, credit card numbers, account numbers, medical records, social security numbers, grades earned in school.
Incident (list of incidents)	A case where one or more electronic personal records were compromised through negligence or theft.
Hacker	Intruder deemed responsible for compromising records.
Mismanagement	Exposing private records online, leaking data due to administrative error or using data for activities not related to the work of the organization.
Phishing	Cases where victims are deceived into voluntarily revealing their personal information.
Security breach in an organization	Accidental exposure of personal records online, inside abuse or theft, missing or stolen hardware, administrative error.
Target country	The country of residence for the people who had personal records compromised.
Unknown (reference to type of compromise of the records)	A case where an estimation of the compromise has not yet been made or it is impossible to be made.
Unspecified (reference to type of compromise of the records)	A case of unwillingness to disclose information about the type of compromise occurred.
Whistleblower	A person or network who disclose alleged wrongdoing or illegal activity occurring in an organization like law or rule violation, fraud, health and safety violations and corruption. The whistleblower attribute was reserved for cases where the source of the breach was described as serving the public interest.

VI. Appendix B: Sources

Specialized Databases
http://www.databreachtoday.eu/
http://datalossdb.org/
http://www.dataprotection.ie/
https://www.huntonprivacyblog.com/archives/
http://ico.org.uk/news/latest_news
http://www.infosecurity-magazine.com/
http://www.insideprivacy.com/data-security/data-breaches/
http://nakedsecurity.sophos.com/
http://www.pogowasright.org/
http://www.privacy-europe.com/blog/
http://www.scmagazineuk.com/
http://seclists.org/
http://thehackernews.com/

VII. Appendix C: Country-Specific Breaches

Country	Population	Internet users	Number of Breaches Involving Each Country	Volume of Breaches Exclusively Involving That Country	Records Per Person	Records Per Internet User	Breaches Originating In This Country
Austria	8,526,429	7,135,168	9	683,731	8.02	9.58	2
Belgium	11,144,420	9,441,116	4	9,700	0.09	0.10	1
Bulgaria	7,167,998	4,083,950	5	64,678	0.90	1.58	0
Croatia	4,272,044	2,780,534	0	-	0.00	0.00	0
Cyprus	1,153,058	726,663	1	-	0.00	0.00	0
Czech Republic	10,740,468	8,322,168	8	159,538	1.49	1.92	1
Denmark	5,640,184	5,419,113	6	32	0.00	0.00	1
Estonia	1,283,771	1,047,772	0	-	0.00	0.00	1
Finland	5,443,497	5,117,660	7	428,300	7.87	8.37	1
France	64,641,279	55,429,382	15	2,782,428	4.30	5.02	1
Germany	82,652,256	71,727,551	28	56,422,711	68.27	78.66	3
Greece	11,128,404	6,438,325	4	9,016,885	81.03	140.05	1
Hungary	9,933,173	7,388,776	2	55,146	0.56	0.75	1
Ireland	4,677,340	3,817,491	12	916,934	19.60	24.02	1
Italy	61,070,224	36,593,969	7	74,601	0.12	0.20	3
Latvia	2,041,111	1,560,452	2	3,500	0.17	0.22	0
Lithuania	3,008,287	2,113,393	3	107,475	3.57	5.09	0
Luxembourg	536,761	510,177	1	-	0.00	0.00	0
Malta	430,146	173,003	1	-	0.00	0.00	0
Netherlands	16,802,463	16,143,879	31	3,868,446	23.02	23.96	2
Norway	5,091,924	4,895,885	6	4,060,032	79.73	82.93	3
Poland	38,220,543	25,666,238	8	787,066	2.06	3.07	2
Portugal	10,610,304	7,015,519	3	657	0.01	0.01	0
Romania	21,640,168	11,178,477	3	5,000	0.02	0.04	2
Slovakia	5,454,154	4,507,849	6	2,401	0.04	0.05	0
Slovenia	2,075,592	1,501,039	0	-	0.00	0.00	0
Spain	47,066,402	35,010,273	9	15,444	0.03	0.04	3
Sweden	9,631,261	8,581,261	6	90,250	0.94	1.05	2
Switzerland	8,157,896	7,180,749	3	1,000	0.01	0.01	1
United Kingdom	63,489,234	57,075,826	77	139,666,768	219.98	244.70	10