

# Global Data Leakage Report

H1 2010



2010

## Table of Contents

Executive Summary.....	2
Accidental and intentional leaks .....	3
Leak sources .....	4
What data leaks most .....	6
Leak channels and technologies .....	7
Leak distribution per country .....	10
Largest leaks .....	11
Summary .....	11

## Executive Summary

InfoWatch presents the latest issue of our regular analytical study of reported incidents in the field of confidential data leaks. This issue covers the incidents reported during H1 2010 and is based upon the daily-updated leak database maintained by InfoWatch analytical center since 2004. The database includes global data on any leaks reported by media, blogs, web forums, and any other public sources worldwide.

Total number of registered data leaks in H1 2010 (181 day) equaled to **382**, in average **2.1** leaks/day. This is slighter less as compared with H1 2009 (**2.3** leaks/day). This minor decrease can be attributed to the statistic fluctuation.

However these figures testify that the mass media and expert attention to the issue worldwide remains high.

Total amount of compromised data entries (compromised personal data is meant) equals to **539+ million**, i.e. about 3 million entries/day. Considering the previous statistics, data latency and ageing, we come to the conclusion that in developed countries at least one personal data entry on almost every citizen has been once compromised. Fortunately, only a part of leaked data can be financially misused.

Speaking of being prepared, organizations mostly remain sluggish in detecting and responding to incidents.

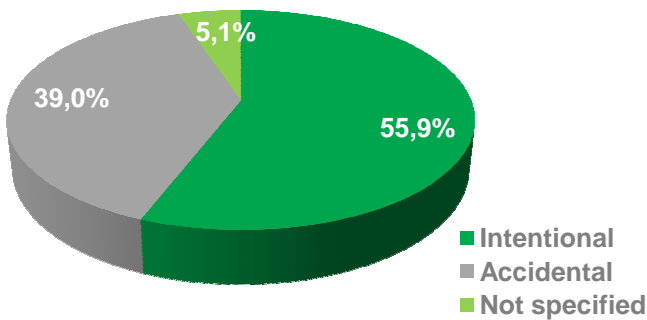
## Accidental and intentional leaks

Leaks are divided into accidental (these leaks happen because of negligence) and intentional. As the countermeasures for these two leak types differ greatly, the table below includes separate statistics on intentional and accidental leaks.

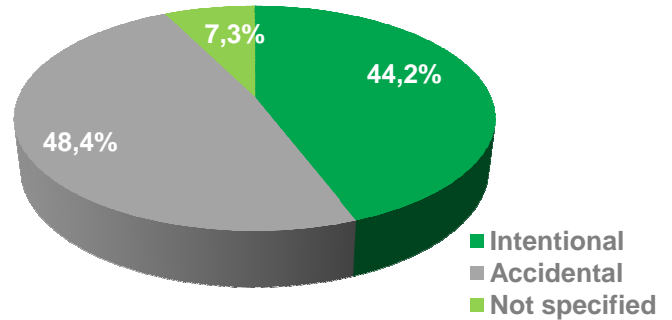
**Table 1: Leak distribution by intent**

Leak type	H1 2009		H1 2010	
	Amount	%	Amount	%
Intentional	231	55.9%	169	44.2%
Accidental	161	39.0%	185	48.4%
Unspecified	21	5.1%	28	7.3%

**H1 2009**



**H1 2010**



Three years ago the percentage of accidental leaks was drastically higher (up to 75%), then it started decreasing. This decrease is easily explained by introducing and implementing DLP-solutions and other protection measures that are most efficient in preventing accidental leaks. A well-designed DLP-solution monitors all communication channels and is nearly 100% efficient in preventing accidental leaks. By intentional leaks the efficiency of protection measures to a great extent depends on the ratio malefactor's / security team skills. The same principle is valid for organizational measures: they are more successful in preventing accidental leaks than intentional ones.

During 2009 the percentage of accidental leaks continued decreasing. However in H1 2010 we experience a slight increase in the amount of these incidents.

InfoWatch analytical team still considers the decrease of accidental leaks a long-term trend. The distribution of leaks by intent in H1 2010 is influenced by the global credit crunch after-effects: lack of budget slowed down the massive implementation of complicated and expensive DLP-solutions, while personnel layoffs had negative impact on organizational protection measures. This caused a slight rise in the amount of accidental leaks, what however could be also attributed to a statistic fluctuation.

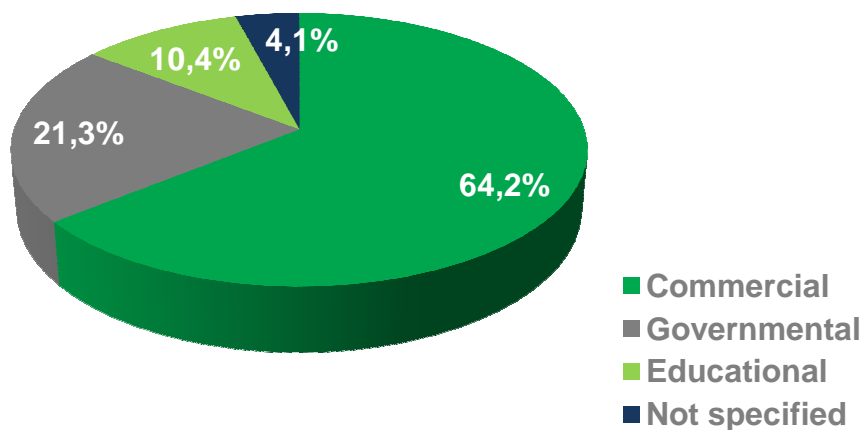
## Leak sources

All enterprises and organizations that encounter public leaks are classified by InfoWatch experts into three categories: governmental agencies, commercial enterprises, and educational institutions plus public non-profit organizations. Educational institutions can be either commercial or non-profit, but yet InfoWatch experts separate them into a specific category as procedures for processing the students' personal data significantly differ from the same procedures in traditional commercial businesses, such as banks, clinics, supermarkets, etc., while data protection procedures in governmental institutions and educational organizations are mostly similar.

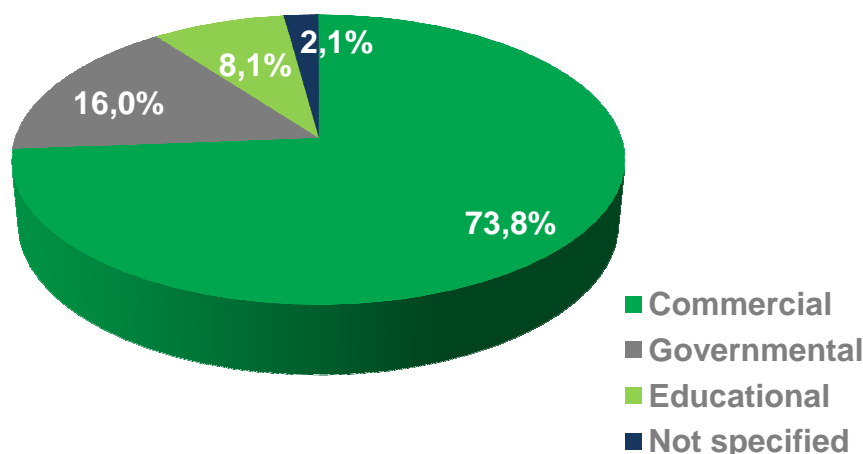
**Table 2a: Leak distribution per organization type**

Organization type	H1 2009		H1 2010	
	Amount	%	Amount	%
Commercial	265	64.2%	296	73.8%
Governmental	88	21.3%	61	16.0%
Educational / non-profit	43	10.4%	127	8.1%
Unspecified	17	4.1%	8	2.1%

### H1 2009



### H1 2010



During the last year no significant changes in leak distribution by organization type were discovered.

A slight increase in commercial sector leaks can be attributed to budget cuts, caused by financial crisis after-effects. This reason is less relevant for governmental and non-profit organizations. Governmental and commercial sectors differ greatly in both motivation and data protection approach.

Legal requirements and higher authority resolutions are considered first while implementing data protection procedures in governmental institutions. The efficiency of these procedures is rated by the total number of data leakage incidents – successful and prevented. This approach is also valid for non-profit organizations. However in commercial sector the situation is different. As the main task of every commercial organization is profitability, the necessity of data protection procedures implementation is also considered from this point of view. Of course legal requirements do influence company decisions, but they are of a less value.

From the profitability perspective, some leaks are tolerable. Businesses can take this risk and reserve financing for incident clean up, as the clean-up costs can be lower than investment in data protection procedures implementation. Another option for commercial organizations is dispose of the risk by outsourcing confidential information processing (for example, outsourcing personal data processing). Governmental institutions generally cannot accept this approach. They are to process the data as required by the legislation.

These differences influence leak statistics in the three organization types. However implementation of data protection solutions has similar impact in all organization types: intentional incidents are partially and accidental – almost totally prevented.

Now we will examine leak distribution per three sectors, separately for accidental and intentional leaks.

**Table 2b: Intentional and accidental leaks distribution per organization type**

Type of organization	Accidental		Intentional	
	Amount	% <sup>1</sup>	Amount	%
Commercial	137	74.1%	126	74.6%
Governmental	28	15.1%	27	16.0%
Educational / non-profit	18	9.7%	10	5.9%
Unspecified	2	1.0%	6	3.6%

No significant changes in these statistics are expected in the nearest future. Obligatory implementation of DLP-solutions in governmental institutions or in several market segments is hardly expectable, though InfoWatch experts assume the process of introducing technical protection measures to continue. This will result in gradual reduction of total leak amount and of the accidental leak percentage.

<sup>1</sup> Percentage of total leaks of selected type (intentional or accidental)

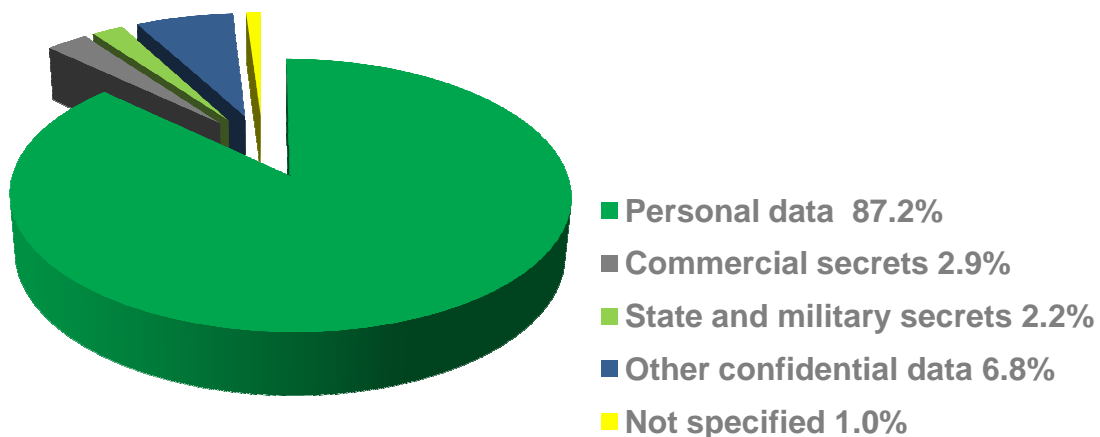
## What data leaks most

Starting the leak database InfoWatch experts divided all incidents into 3 categories: incidents with personal data, incidents with state and incidents with commercial secrets. All leaks included into the database were attributed to one of these categories. Total majority of incidents (90-98 per cent) during the whole period of record contain personal data.

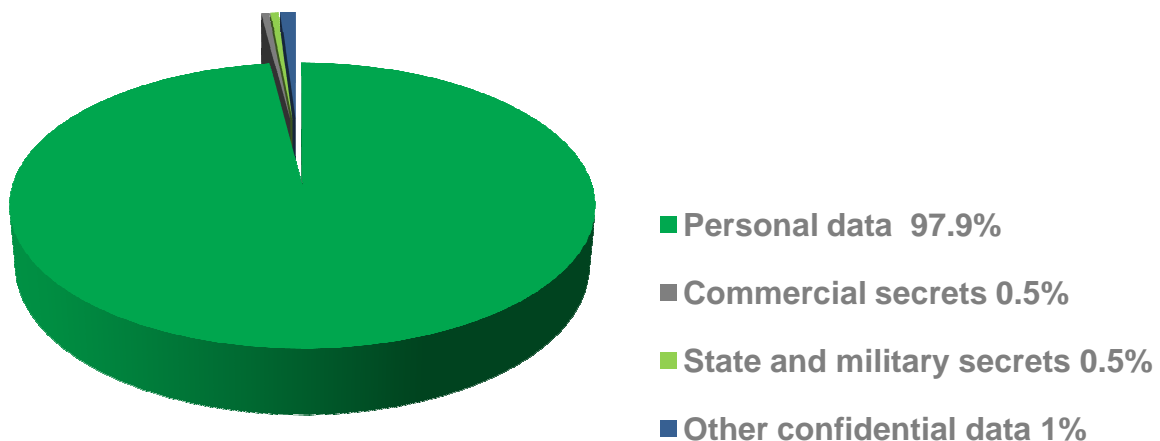
**Table 3: Leak sources, distributed per confidential data type**

Type of confidential data	H1, 2009		H1, 2010	
	Amount	%	Amount	%
Personal data	360	87.2%	374	97.9%
Commercial secrets, know-how	12	2.9%	2	0.5%
State and military secrets	9	2.2%	2	0.5%
Other confidential data	28	6.8%	4	1.0%
Unspecified	4	1.0%	0	0%

### H1 2009



### H1 2010



In H1 2010 personal data incidents are still in the lead. In H1 2009 the percentage of personal data incidents decreased slightly (87.2% in H1 2009, 89.8% in 2009 generally), but in H1 2010 it regained the previous almost 100 per cent majority.

This absolute domination is self-evident. There's an overwhelming mass of personal data worldwide, and they are in constant use. Thousands of organizations have to deal with personal data processing, while commercial secrets are much less common and state secrets are even rarer.

Moreover the validity of commercial data is limited: commercial secrets can be used only by direct competitors and only under specific circumstances. On the contrary the validity of some personal data is undeniable: there is an established market for lost/stolen personal data. That's why such data theft or improper use of accidentally lost personal data is a widespread phenomenon. Even a person without malicious intent sometimes cannot withstand the temptation. Other types of confidential data are usually stolen only by a specific request.

In the coming years the situation is likely to worsen. A possible solution to the problem could be transferring some categories of personal data from confidential into non-confidential, which is unfortunately mostly unlikely. Companies widely use personal data for customer authentication providing remote services. In this case non-confidential data become confidential that needs to be protected. Government initiatives mostly embrace the concept of mandatory protection of all types of personal data. Taking this into account InfoWatch experts expect further growth in the personal data category.

## Leak channels and technologies

Analysis of data carriers that were used to transfer the classified data over the secured perimeter or to deliver it to a non-authorized user allows predicting further leaks via this carrier and estimating risk reduction resulting from implementation of various security measures.

DLP-solutions and other counter-leak measures cover specific channels that can be used to transfer information over the security perimeter, such as network connections via various protocols, mobile data carriers, mobile computers, printers, etc. Companies not always have enough financing and technical skills to cover all possible data leak channels. Uncontrolled channels usually are responsible for the majority of accidental leaks.

The situation is different in regard of intentional leaks. A malicious insider with the knowledge about controlled channels chooses other – unprotected – channels to get the information out of the security perimeter. Probability of intentional leaks is hardly influenced by implementation of DLP-solutions covering only selected data transfer channels. To efficiently prevent both accidental and intentional leaks the data protection solution (supported by organizational measures) should provide comprehensive control over all channels and information carries.

**Table 4a: Major data leak channels**

Leak channel	H1 2009		H1 2010	
	Amount	%	Amount	%
Mobile computers (laptops, PDAs)	49	11.9%	40	10.5%
Mobile data carriers (flash drives, CD, DVD, etc)	23	5.6%	32	8.4%
Desktop computers, servers, HDD	41	9.9%	90	23.6%
Internet (incl. e-mail)	97	23.5%	82	21.4%
Paper document	84	20.3%	78	20.4%
Archived media	48	11.6%	6	1.6%
Other	36	8.7%	25	6.5%
Unspecified	35	8.5%	29	7.6%

As compared with the previous year, the amount of leaks associated with desktop computers, servers and HDDs has grown significantly.

Leaks via mobile computers and mobile data carriers were extremely popular 2-3 years ago. Last year the percentage of data leaks via mobile computers and mobile data carriers reduced. This year's growth is insignificant and can be attributed to statistical discrepancies. This reduction can be explained by implementation of encryption solutions. Encrypted carriers are not calculated within these statistics in case of their loss or theft.

Unfortunately encryption implementation has slowed down. Encryption has not become mandatory for corporate laptops, USB-sticks and CDs. As the statistics show, only responsible employees and departments under their management have fully embraced the encryption concept. Mostly employees and companies neglect this protection measure, considering the leak via unencrypted laptop unlikely. The situation is similar as with the usage of safety-belts while driving. Though everyone is aware that a safety-belt can significantly reduce the risk of injury in a car-accident, the car-accident itself seems highly unlikely. One of the major reasons for safety-belts usage is external enforcement (including fines and penalties).

Today almost everyone is aware of consequences caused by corporate laptop loss or theft. Encryption could successfully mitigate this risk, but as the incident probability is low, encryption goes underestimated. We believe mass-implementation of encryption solutions can be facilitated only by legal enforcement, for example, by introducing penalties for non-encrypting all mobile data carriers. Only a minority of companies have taken this approach. Surprisingly even governmental organizations do not pay enough attention to encryption, though their laptops can contain information about state and military secrets.

InfoWatch experts assume mobile data carriers encryption to continue spreading slowly. Voluntary actions have already been taken, further implementation can be facilitated with administrative support, at corporate, industry or government level.

From another point of view, the amount of used mobile data carriers grows steadily. This could give rise to the amount of leaks via laptops, flash-drives, etc.

The percentage of data leaks via paper documents remains steadily high. As mentioned in the previous report, the cause for this is self-evident. Efficient technical measures can block accidental leaks via electronic channels, but not all DLP-solutions feature control over document printing. Once out of printer, hard copies can be controlled only via personal security, which is more complicated and less efficient.

A typical "paper" leak is a system failure at customer letters printing. The address at the envelope is also printed automatically, sometimes the letters are also sealed automatically. In case of a small failure customers receive letters with other people's personal data.

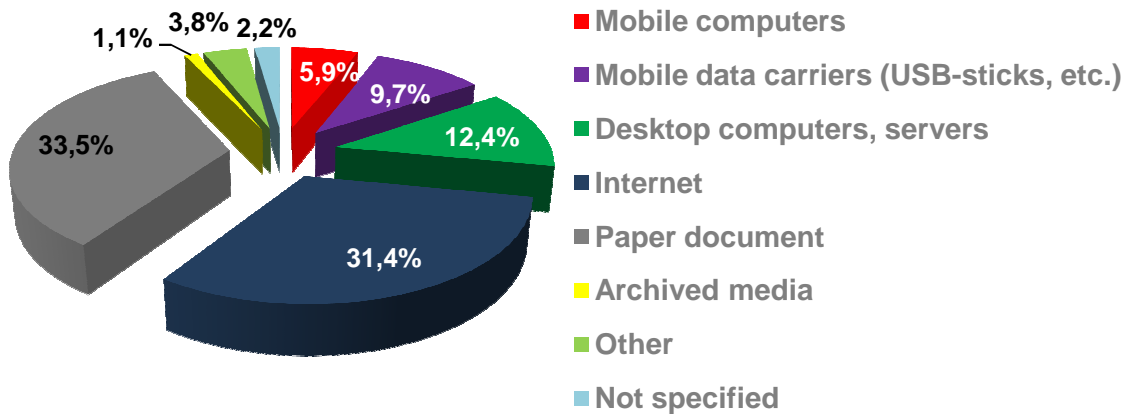
To reduce the amount of "paper" leaks a pool of organizational and technical measures is required: technically – a DLP-solution, featuring active control over document printing (including blocking) and a special utility to check the match between the addressee and the address, organizationally – a procedure for printed confidential document tracking. These measures are relatively expensive, especially as compared with printer costs, so the number of "paper" leaks is to decrease slowly, mostly because of reduced usage of printed documents.

The table below demonstrates leak distribution by channel, separately for intentional and accidental leaks.

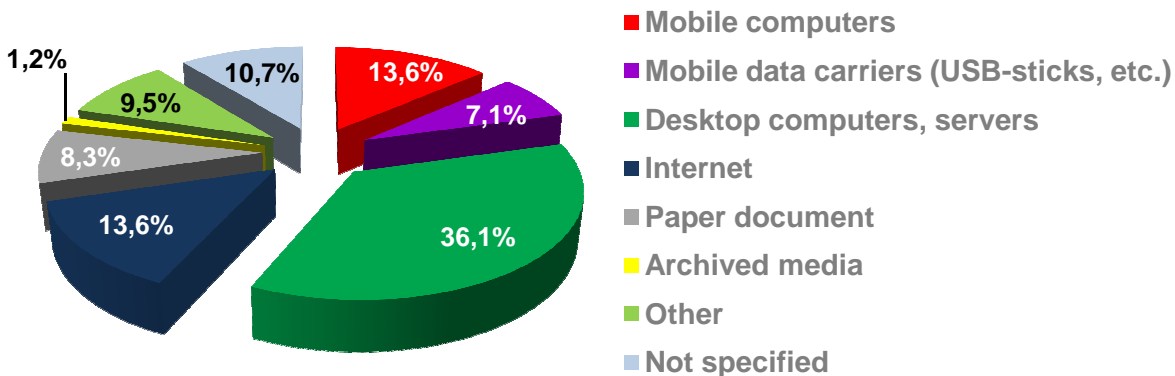
**Table 4b: Major channels for intentional and accidental data leaks**

Leak channel	Accidental		Intentional	
	Amount	%	Amount	%
Mobile computers (laptops, PDAs)	11	5.9%	23	13.6%
Mobile data carriers (flash drives, CD, DVD, etc)	18	9.7%	12	7.1%
Desktop computers, servers, HDD	23	12.4%	61	36.1%
Internet (incl. e-mail)	<b>58</b>	<b>31.4%</b>	<b>23</b>	<b>13.6%</b>
Paper document	<b>62</b>	<b>33.5%</b>	<b>14</b>	<b>8.3%</b>
Archived media	2	1.1%	2	1.2%
Other	7	3.8%	16	9.5%
Unspecified	4	2.2%	18	10.7

## Accidental leaks



## Intentional leaks



The most drastic differences between accidental and intentional incidents can be observed in the categories Internet and paper document.

Despite the common opinion, external violators intrude into corporate networks not so often. The majority of network data breaches as associates with irresponsible insider actions (compare the Verizon 2010 Data Breach Investigations Report: The percentage of breaches attributed to external agents slid 9%, insiders more than doubled). A typical example of an insider-related data breach is accidental copying of a confidential document into a catalog used by a web-server as data source. As a result the confidential document is published in the net.

Paper documents are also more often stolen by own employees, than by intruders. Moreover, increased public interest in the issue of data leaks leads to general concernment over improper utilization of confidential documents. Utilized documents with confidential data often are retrieved from the disposal bins and delivered to mass media. Finding confidential documents in garbage bins requires significantly less skills than finding web-site vulnerabilities. To the popularity of paper leak topic contributes also general people attitude to personal data: for example, when the customer receives a letter with other people data (as a result of automatic letter printing system failure), he/she is more likely to draw media attention to the fact.

The above diagrams indicate that a DLP-solution should be first of all deployed at the gateway between protected corporate and public network, at printers (print-servers) and by data copying to portable devices and carriers (USB-sticks). These channels are responsible for the majority of leaks. As portable devices and mobile data carriers increase productivity, blocking of them would be unreasonable. To secure these channels and control information outside physical corporate perimeter (office) encryption should be implemented.

As opposed to DLP-solutions, encryption tools are mostly price-efficient. Strength of encryption algorithms has no defining value, as the potential violator has no required mathematic and analytical skills to brute-force the encryption key. Though cryptography is considered insufficient for state secrets protection, it is perfectly suitable to protect the data on a lost laptop or USB-stick from unauthorized access and usage.

## Leak distribution per country

The table below illustrates leak distribution by countries. Because of incident latency, this table is not absolutely relevant to demonstrate geographic leak distribution, but it is highly relevant to illustrate incident latency – how often an incident gets concealed in a country. The index LPC is calculated as the ratio between the amount of leaks in the country and this country population (in millions). Countries with low latency, such as USA and Great Britain are considered as reference points. Legislation in USA and Great Britain binds the companies to notify the government about all leaks. Comparing US and Britain’s figures with the figures of other countries, we can conclude what amount of leaks stays publicly unknown.

**Table 5: Leak distribution per country**

(CC) Country	Amount	%	LPC
AU Australia	2	0.56%	0.100
CA Canada	11	2.88%	0.338
CH Switzerland	2	0.56%	0.257
CN China	1	0.26%	0.001
DE Germany	5	1.31%	0.061
ES Spain	1	0.26%	0.025
GB Great Britain	36	9.42%	0.597
GR Greece	1	0.26%	0.090
IE Ireland	3	0.56%	0.333
IL Israel	1	0.26%	0.164
IN India	3	0.84%	0.003
IT Italy	1	0.26%	0.017
JP Japan	1	0.26%	0.008
MX Mexico	1	0.26%	0.010
NL Netherlands	5	1.31%	0.304
NO Norway	1	0.26%	0.208
NZ New Zealand	1	0.26%	0.232
PK Pakistan	1	0.26%	0.006
RU Russia	15	3.93%	0.104
UA Ukraine	1	0.26%	0.021
US USA	284	74.4%	0.969
Several countries	1	0.26%	
Not specified	5	1.31%	

USA and Great Britain were in the lead in the “LPC” index last year. In 2008, when Great Britain had no legal regulation about mandatory leak notification, this country had a significantly lower “LPC” index.

Following the example of its southern neighbor, Canada has performed much better in data protection this year. Its “LPC” index has experienced drastic increase, so the country ranks number 3 in LPC in H1 2010.

InfoWatch experts assume that the total amount of leaks in developed countries is about the same and comprises about 2 leaks per year per million people. The differences in public statistics are explained by different notification regulations. Undoubtedly, USA has the lowest latency.

## Largest leaks

The table below includes top incidents for H1 2010.

**Table 6: Largest leaks**

Date	Description	Link
18.01.10	A vulnerability in a Britain database, containing personal data of <b>11 million</b> children, was detected.	<a href="http://www.telegraph.co.uk/news/newstopics/politics/lawandorder/6836911/ContactPoint-database-of-11-million-children-suffers-security-breaches-in-trials.html">http://www.telegraph.co.uk/news/newstopics/politics/lawandorder/6836911/ContactPoint-database-of-11-million-children-suffers-security-breaches-in-trials.html</a> ),
16.02.10	A vulnerability in Microsoft authentication system can influence up to <b>460 million</b> users.	<a href="http://www.bloomberg.com/apps/news?pid=20601087&amp;sid=affldPDlbclA&amp;pos=7">http://www.bloomberg.com/apps/news?pid=20601087&amp;sid=affldPDlbclA&amp;pos=7</a>
27.03.10	In USA the malefactors stole the data on <b>3.3 million</b> students from a company engaged in educational loans.	<a href="http://updatednews.ca/?p=10521">http://updatednews.ca/?p=10521</a>
01.05.10	A tax payers database with <b>25 million</b> records was suspected to be stolen and sold to spammers in Great Britain.	<a href="http://www.telegraph.co.uk/news/uknews/7665782/Tax-records-sold-to-junk-mail-firms.html">http://www.telegraph.co.uk/news/uknews/7665782/Tax-records-sold-to-junk-mail-firms.html</a>

## Summary

- Total amount of reported data leaks remains stable and equals about 2 incidents/day. This is about 10% lower than in the same period of 2009.
- The topic of data leaks (especially of personal data leaks) is popular throughout the world.
- The percentage of accidental leaks, tending to significant decrease in the previous years, stopped decreasing and has risen slightly.
- Governmental, commercial, educational and non-profit organizations should apply similar data leakage prevention tools and procedures.
- The amount of leaks because of unencrypted mobile computers and mobile data carriers loss or theft, stays steadily high.
- The amount of network leaks lowered slightly.
- Paper documents are still responsible for a majority of incidents, as control over printers and printed documents often goes overlooked.
- Leak latency is high in all countries, excluding USA and Great Britain, where obligatory leak notification regulations exist. The total amount of leaks in developed countries equals about 2 leaks per year per million people.