

# white paper

## Data retention: avoiding the traps

The complexities and the costs of complying with data retention rules and regulations must be carefully investigated by Communication Service Providers before they make any investment decision. Many common views on data retention systems are only partly correct and may generate misleading business perceptions. This paper analyzes some common statements and shows their weaknesses.

# Contents

<b>1</b>	<b>Executive summary</b>	<b>3</b>
<b>2</b>	<b>Data retention background</b>	<b>4</b>
2.1	Why a data retention solution?	4
<b>3</b>	<b>The DR pitfalls, and how to avoid them</b>	<b>5</b>
	The “billing systems” pitfall	5
	The “existing data” pitfall	6
	The “over reliance on existing systems” pitfall	7
	The “too expensive” pitfall	7
	The “install and forget” pitfall	9
<b>4</b>	<b>Conclusion</b>	<b>10</b>
<b>5</b>	<b>Glossary</b>	<b>11</b>
<b>6</b>	<b>References</b>	<b>12</b>

# 1 Executive summary

Telecommunications data retention (DR) refers to legal requirements to store telecommunications-related data for the purpose of combating serious crimes. European Union (EU) Directive 2006/24/EC has been issued to harmonize these requirements across the EU, but such regulations also exist in many countries outside the EU.

A data retention solution is not usually a revenue generating investment, yet selecting the most suitable solution can make a difference in terms of costs, ease of use, fulfillment of regulations, protection of citizens' right to privacy and operator brand value.

There are many ideas concerning data retention that seem well-founded from the point of view of a service provider. Yet, subject to deeper scrutiny, they lose much of their value.

Some of those statements include the following:

- ❖ “All data needed for data retention is already in our billing systems. We just need to store it for a little longer.”
- ❖ “All the necessary data is somewhere in one of our many systems, so we can always collect the required info if need be.”
- ❖ “Our existing systems are secure enough for data retention purposes.”
- ❖ “A specialized data retention solution is expensive, and we do not have a business case for it.”
- ❖ “Once our data retention solution is put into operation then we can forget about it because we have complied with the law.”

Only with a careful analysis of all the factors addressed above can a Communication Service Provider (CSP) fulfill data retention regulations at a minimum cost.

## 2 Data retention background

### 2.1 Why a data retention solution?

Many countries have issued rules and regulations concerning the collection and retention of telecommunications traffic-related data for investigative purposes.

A data retention solution provides the CSP with the means to fulfill these obligations.

All countries within the EU must comply with their respective national data retention laws based on EU Directive 2006/24/EC issued on March 15, 2006. Most countries have implemented this directive in their national legislation, but many have not yet issued detailed implementation regulations.

The EU Directive aims to harmonize a CSP's obligations to provide communication data to relevant national authorities for the purpose of investigation, detection and prosecution of serious crimes. Likewise the directive requires the data to be handled with full respect for the fundamental right to privacy of all citizens.

The communications data include the traffic data, the location data and the related data necessary to identify the legal entity or person using a publicly available communications service. Six categories of data must be retained:

- the source of a communication
- the destination (including supplementary services) of a communication
- the date, time and duration of a communication
- the type of a communication
- the user's communication equipment
- the location of mobile communication equipment (including geographic location of cells).

The data must be retained also in relation to unsuccessful communication attempts, e.g. calls with no answer (due to a busy line or if

nobody answers), call diversion, e-mail rejection or failed sending of an SMS.

The directive does not require retention of any content of communication.

The directive defines the constraints of the retention period: not less than six months and not more than two years from the date of the communication. Data must be destroyed at the end of the retention period. The EU Directive does also allow Member States to specify their own time periods overriding the constraints.

Principles of security and protection of retained data are stated in the directive. Appropriate technical and organizational measures must be taken to ensure that data can be accessed by authorized personnel only and to protect data against accidental or unlawful destruction, accidental loss or alteration, unauthorized or unlawful storage, processing, access or disclosure.

The European Telecommunications Standards Institute (ETSI) has worked on defining technical specifications for data retention in order to provide a standard framework for the solution suppliers. The ETSI TS 102 656 collects requirements for a DR solution from various authorities both inside and outside the EU, while the ETSI TS 102 657 specifies the electronic interface to interrogate the retained data and receive results from a DR solution. Moreover ETSI has also elaborated a Technical Report, ETSI TR 102 661, on the security requirements for a DR solution.

Some countries outside the EU already have DR laws in place. In some cases those are very much aligned with the EU Directive, as in Mexico, for example, while the rules in other cases are very country specific.

# 3 The DR pitfalls, and how to avoid them

Data retention regulations do not usually mandate any specific technical approach to CSPs for the realization of a data retention solution.

CSPs are thus free to explore different alternatives for a suitable solution. Sometimes their subsequent analysis is based on questionable assumptions that can lead to flawed conclusions and consequently larger than necessary investments.

These questionable assumptions, or “pitfalls,” look like solid ground, but then often give way when someone tests them. This paper examines some common statements and provides material that can help CSPs make the right decision – and avoid the pitfalls.

---

## The “billing systems” pitfall

***“All data needed for data retention is already in our billing systems. We just need to store it for a little longer.”***

This assertion is only partly true as data is available only when there is a genuine business need to collect it.

A mobile CSP will collect data about the origins of voice calls, SMS and MMS, as these services are usually charged per unit. Likewise, records about mobile PS sessions will also be collected as mobile CSPs often use volume or time-based charging for data services.

More rarely, records will be collected on terminating calls, for the purpose of reverse charging, prepaid credit reloading schema, or for evaluating the quality of service.

Records for terminating SMS and MMS are seldom collected, likewise data about unsuccessful call attempts.

The gap between billing data and the data required by data retention laws arises from a fundamental difference in purpose. The billing data is collected to bring revenues; hence events that do not lead to charging are of little interest and can even be illegal to store for other purposes than data retention.

Data retention, on the other hand, is interested in traces of communications, locations of possible suspects, and connections and relationships between suspects. A busy call attempt can be irrelevant for billing, yet it might be the only

available trace of a relationship between two people. An unanswered call is again irrelevant in most countries for billing, yet it can be the only link between a person and an event, like the explosion of a bomb detonated via a mobile phone.

The EU Directive recognizes this difference and requires CSPs to provide data for services provided by the CSP to users, including data about terminating calls and communications (clearly a service provided to users) and unsuccessful calls, albeit if available.

It is possible to modify billing systems in order to collect and store this additional wealth of data on voice calls, SMS and MMS, but this is not easy due to the extended retention period.

A lot of data is already collected for billing, and indeed, Call Detail Records (CDRs) contain many parameters that could be useful for DR. But these records are kept for a relatively short time period, usually three to six months.

Data needed for DR is more limited in details, and many CDR parameters do not need to be retained, but the retention period is often in the range of one to two years, at least according to many national regulations.

There are two possible solutions to this issue, both with many drawbacks. Either a

different retention period is implemented for different CDR parameters or an extended retention for the complete CDR is performed.

In the first case, the erasure procedure for the expired parameter becomes very cumbersome, or a restructuring of the data model is required.

In the second case, the amount of storage needed will grow far beyond the absolute minimum needed for DR, as much useless data will be stored for a prolonged time.

A different question is the collection and storage of IP services. Data on e-mails is never collected in billing systems, as e-mails are never charged per unit. Moreover the number of e-mail events can be much higher compared to the services discussed above.

Another example concerns the network address translation (NAT) functionality usually implemented at the border between an operator network and the public internet. Such function assigns a temporary public IP address to a user to replace the private IP address obtained when the user has been

granted connection to the IP network.

Clearly these are irrelevant events from a charging point of view, but unfortunately, keeping track of these translations is the only way to know which subscriber used a public IP address in the operator-assigned range in a certain moment. These events occur in at least two orders of magnitude more frequently than calls or SMS. A billing system adapted to collect and store this type of records would quickly choke on all the data.

### Conclusion

Billing systems are designed for a completely different purpose than complying with DR rules and regulations. Adapting them to serve DR needs is both expensive and unrealistic for some type of data, like IP-related information. Investments in adaptations of billing systems must be carefully evaluated and compared with investment in a specialized data retention solution that collects and stores only the necessary data.

---

## The “existing data” pitfall

***“All the necessary data is somewhere in one of our many systems, so we can always collect the required info if need be.”***

At first glance this approach seems interesting. It requires no additional investment for a specialized DR solution, and, with some adaptations, it is possible to retrieve the required data when an interrogation request comes from the relevant authorities.

Unfortunately this approach also has many shortcomings from security, operational and maintenance points of view.

Searching for data in many systems is expensive in terms of opex, as the same manual or semi-automatic activity must be executed a number of times. Moreover results have to be consolidated in a coherent report before delivery.

This process of search and consolidation also has an impact on response time, making the investigations more cumbersome. Failure to quickly reply to urgent investigations may strain a CSP’s relationship with the authority

in question and increase the pressure to adapt the solution and improve its performances.

Another aspect to be considered is that log data in many systems is kept only for a short period of time. Log files from firewalls or e-mail servers rotate rather frequently (maybe weekly) since they are used mainly for troubleshooting. The implication of storing these logs for the retention period can be far fetched, in terms of storage cost and system performance.

### Conclusion

A specialized DR solution is much more cost effective when compared to a distributed DR solution assembled by reusing the existing network nodes. The investment and the opex for a distributed approach needs to be carefully evaluated and compared with a centralized DR solution.

## The “over reliance on existing systems” pitfall

### ***“Our existing systems are secure enough for data retention purposes.”***

The data retention directive does not establish a higher standard of security than Directive 2002/58/EC on “privacy and electronic communications.”

Yet, considering the nature of the data to be retained, its purpose and the retention period, there is a higher level of risk connected to this data when compared to personal data acquired and processed for commercial or operational purposes only.

Many systems used in a CSP network do not have strict requirements on integrity and availability of logged information. For example a firewall performing NAT will function perfectly even if someone alters some rows in the log file, and there’s no real business damage caused by tampering with the log file. Yet such change may erase important information for an investigation.

Keeping data on the network nodes either requires the implementation of strict security measures to protect the integrity of data, or greatly diminishes the legal value of the data itself. It will be difficult to bring evidence to court based on data that cannot be guaranteed to be a true representation of what has happened in the network.

Integrity of data is not the only security

issue. Systems used for traffic are accessed by a lot of people for different reasons, like maintenance, administration and trouble shooting. It is very likely that those people are able to trace data searches for judicial purposes. For example the system access made by the user responsible for the legal DR interrogations could be subject to stealth logging. Information about ongoing investigations could be easily collected and is very precious. Both organized crime and white collar criminals have enough money to tempt someone with access to the information.

Separating the data retention system from billing systems and traffic systems ensures the right level of protection for retained data. This is indeed the approach taken for example by the Italian Garante per la Protezione dei Dati Personali (2008) and by ETSI (2008).

### **Conclusion**

The investment needed to implement strong security requirements in existing systems should be considered when evaluating a specialized DR solution.

---

## The “too expensive” pitfall

### ***“A specialized data retention solution is expensive, and we don’t have a business case for it.”***

A specialized data retention solution requires a significant investment for a system that does not then contribute directly to revenues. A logical first reaction is to consider such an investment too expensive, no matter how big or small it is.

Yet the opex and capex involved in assembling a DR solution by keeping data spread over the existing different systems can vastly offset the cost of an effective specialized solution.

The business case for a data retention system must not be approached from the

point of view of cash flows and returns over time, since it is primarily costs involved, not revenues. Therefore the best framework to evaluate different options for data retention is the business case based on the total cost of ownership (TCO) concept.

The TCO method is a technique that is used to ensure that all associated costs over a given time period are considered when acquiring an asset. TCO does not only reflect the costs of purchase. It also includes all other costs during use and maintenance of the asset.

The TCO method works well in clarifying the cost structure of a dedicated data retention solution.

Let us say that the alternative of reusing an existing set of systems for data retention purposes is evaluated. Then a TCO cost structure needs to be created in order to compare it with the specialized data retention TCO.

Capex costs should cover at least:

- technical data retention requirements for adapting each existing system (for example, the cost of collecting the whole set of data to be retained and the implementation of technical security requirements)
- non-technical data retention requirements for each existing system (for example, costs associated with organizational security measures to be adopted)
- evolution and adaptations due to the introduction of new services or new nodes.

Opex costs should cover at least:

- handling of DR functionality
- erasure of data past the retention period
- handling of data interrogations from authorities
- security audits.

These opex costs could be sensibly high if the number of systems is not small and the amount of data and related queries is high.

A DR solution can also present some opportunities for generating some revenues that might partially compensate the costs, or in some cases even provide a genuine profit.

For example, a specialized data retention solution is the best place to generate anonymous usage and QoS statistics for the whole network and all services, as data for every communication is collected. This has a positive impact on cost efficiency because savings in other systems are possible by concentrating reporting activity in the data retention solution.

Likewise a DR solution with support for hosting can open the possibility for a medium-to large-sized CSP to provide DR services to small CSPs.

Finally, in some countries reimbursement is discussed for both opex (see reference [7]) and capex (see reference [8]).

### Conclusion

A specialized data retention solution is not expensive compared with a solution that reuses the existing network infrastructure. An operator business case will prove that a specialized DR solution minimizes the TCO over a given period.

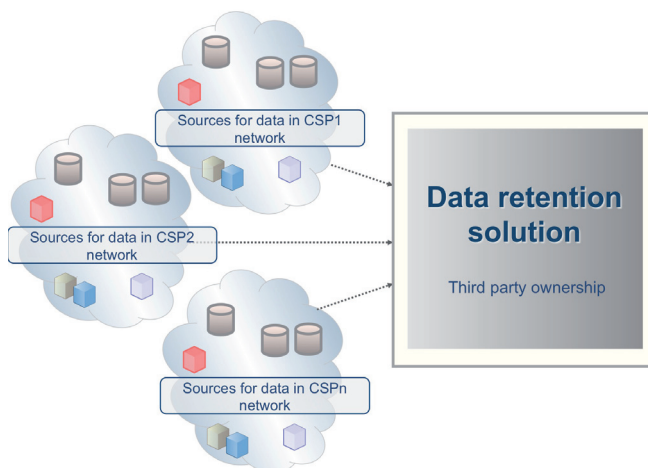


Figure 1: More CSPs are supported by the same data retention solution

## The “install and forget” pitfall

***“Once our data retention solution is put in operations, we can forget about it because we have complied with the law.”***

This assertion is very appealing but hardly close to the truth. On the contrary, the go live of a data retention solution is only the start of a journey for which the CSP must be well prepared.

In the normal course of business, CSP’s are constantly looking for new services and applications that can differentiate them from their competitors. Moreover they are also constantly looking at optimizing their network infrastructure in order to reduce costs and increase efficiency.

So it is normal for a CSP to introduce new services (with old ones improved) and new nodes (with old ones upgraded).

Such changes can have large impacts on a data retention solution.

New nodes may deliver the same data as old nodes but in a complete different format. Deployment of IMS-based services may imply a completely different way of reporting events from the network nodes, and new call scenarios might appear.

Changes in regulations, like extension of the retention period, might also drive a DR system past its maximum capacity of storage.

It is essential to understand and consider the impacts, either direct or as side effects, on the DR solution of any change performed in the network. The risk of not doing so is that the deployment of new services or nodes might be delayed because of a lack of compliance with the DR regulations.

Another aspect to be carefully considered is the cost of evolving the DR solution. The best approach for a CSP is to know beforehand the cost for updating the complete solution.

This implies that a CSP will know the costs in advance for:

- ❖ software release upgrades
- ❖ support for new services
- ❖ adaptations to existing services support
- ❖ increased capacity both in terms of SW licenses and HW.

### **Conclusion**

A telecom-centric approach to data retention is the enabling factor for properly assessing the impact of network evolution, and for evolving the DR solution to keep in compliance with DR regulations.

# 4 Conclusion

The EU Directive 2006/24/EC on data retention has been in force since 2006, and its transposition into national laws across the EU is inevitable.

Communication Service Providers have a myriad of technical choices to ensure compliance with the national laws, but not all choices are equally good. Providers must consider all data retention factors and related

costs when comparing solutions. Wrong or incomplete assumptions concerning data retention could become pitfalls and lead to sub-optimized and costly solutions.

A specialized data retention solution is an effective and efficient investment that assures full compliance without financial surprises for years to come.

# 5 Glossary

<b>capex</b>	capital expenditure
<b>CDR</b>	Call Detail Record
<b>CSP</b>	Communication Services Provider
<b>DR</b>	data retention
<b>ETSI</b>	European Telecommunications Standards Institute
<b>IMS</b>	IP Multimedia Subsystem
<b>IP</b>	internet protocol
<b>MMS</b>	Multimedia Messaging Service
<b>NAT</b>	network address translation
<b>opex</b>	operational expenditure
<b>QoS</b>	quality of service
<b>PS</b>	packet switched
<b>SMS</b>	Short Message Service
<b>TCO</b>	total cost of ownership
<b>TR</b>	Technical Report
<b>TS</b>	Technical Specification

## 6 References

- [1] ETSI, 2008. ETSI TS 102 656 Requirements of Law Enforcement Agencies for handling Retained Data.
- [2] ETSI, 2009. *ETSI TS 102 657 Retained data handling; handover interface for the request and delivery of retained data*
- [3] ETSI, 2008. *ETSI TR 102 661 Security framework in Lawful Interception and Retained Data environment.*
- [4] European Parliament and Council, 2002. *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)*. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:EN:PDF> [Accessed September 15 2009].
- [5] European Parliament and Council. 2006. *EU Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC*. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF> [Accessed on September 14 2009].
- [6] Garante per la Protezione dei Dati Personali, 2008. *Secure Retention of Telephone and Internet Traffic Data*. Available at: <http://www.garanteprivacy.it/garante/doc.jsp?ID=1542849> [Accessed on September 14 2009].
- [7] Netherlands Secretary of State for Economic Affairs, September 5 2008/Nr. ET/TM 8126895. *Indexering tarieven Regeling kosten aftappen en gegevensverstrekking*. Available at: [http://www.justitie.nl/images/Indexering%20tarieven%20Regeling%20kosten%20aftappen%20en%20gegevensverstrekking%201%20juni%202008-%201%20juni%202009\\_tcm34-141544.pdf](http://www.justitie.nl/images/Indexering%20tarieven%20Regeling%20kosten%20aftappen%20en%20gegevensverstrekking%201%20juni%202008-%201%20juni%202009_tcm34-141544.pdf) [Accessed on September 14 2009].
- [8] UK Secretary of State, Statutory Instruments 2007 No. 2199. *The Data Retention (EC Directive) regulations 2007*. Available at: [http://www.opsi.gov.uk/si/si2007/pdf/uksi\\_20072199\\_en.pdf](http://www.opsi.gov.uk/si/si2007/pdf/uksi_20072199_en.pdf) [Accessed on September 14 2009].