Project Acronym: STORM CLOUDS

Grant Agreement number: 621089

Project Title: STORM CLOUDS – Surfing Towards the Opportunity of Real Migration to CLOUD-based public Services

Deliverable 4.1.1
Ethical issues and data protection report

Work Package: WP4
Version: 1.1
Date: 12/11/2015
Status: WP leader accepted
Nature: REPORT
Dissemination Level: PUBLIC

Editor: Alkiviadis Giannakoulias (European Dynamics SA)
Authors: Alkiviadis Giannakoulias (European Dynamics SA)
Adrian Slatcher (Manchester City Council)
Martine Tommis (Manchester City Council)
Reviewed by: Panagiotis Tsarchopoulos (Aristoteleio Panepistimio Thessalonikis – AUTH)

Version Control

Modified by	Date	Version	Comments
Alkiviadis Giannakoulias	26/5/2014	0.1	Initial version
Alkiviadis Giannakoulias	17/6/2014	0.2	Updated
Alkiviadis Giannakoulias	28/7/2014	0.3	City perspectives added, template updated
Alkiviadis Giannakoulias	30/7/2014	0.4	Removed IDS/IPS section
Alkiviadis Giannakoulias	14/12/2014	0.5	Edit by MCC
Alkiviadis Giannakoulias	14/12/2014	1.0	Final version submitted to the European Commission
Alkiviadis Giannakoulias	12/11/2015	1.1	Updated to meet review comments and new document template

Executive Summary

Public sector entities, such as government, education, and healthcare organisations are embracing clouds as a way to increase their operational efficiency and productivity, whilst at the same time maximizing investments and lowering costs. Whilst cloud offers the opportunity for and innovativion by consolidating, virtualising, and automating their ICT resources, public authorities must have confidence that the benefits can be achieved without compromising their core requirements. A pillar of these requirements is the need to consider the ethical and data frameworks which ensure full management and protection of the data they hold and manage.

Surfing Towards the Opportunity of Real Migration to Cloud-based public Services (STORM CLOUDS) is a project partially funded by the European Commission within the 7th Framework Program in the context of the CIP project (Grant Agreement No. 621089). The project aims to explore how the shift by public authorities to a cloud-based paradigm in service provisioning should be addressed, principally from the point of view of the end-users. In this context, the term "services" refers to applications available online. A definition of cloud computing is given in Section 2.

Section 2 focuses on the ethical and data protection issues that arise when public sector organisations consider transitioning to cloud computing. Data protection law was designed to be a fundamental and concrete dimension of the individual's right to privacy, the primary safeguard against misuse of personal information. City municipalities operate in a legal context – they are data controllers for a good deal of citizen focused data, much of which is sensitive, personal and highly regulated. They are also are trusted bodies, and citizens expect that their approach to data collection, retention, storage and sharing is in line with these responsibilities. Therefore the ethical dimension is a key part of the transition to cloud journey. Based on this premise, in Section 3 presents a roadmap for cloud computing adoption, outlining the main principals as well as steps to cloud implementation.

Section 3 focuses on data protection. Data protection is the level of availability or confidence in being able to access important data. There are many options to ensure data protection. There is still uncertainty in relation to cloud computing, when addressing the issues of data protection and other legal/ethical issues. Hence this document provides a level of detail and discussion on these areas, with a detailed analysis of the European Data Protection Directive.

The final section of the document addresses key recommendations and best practices including:

- Cloud computing must not lead to a lowering of data protection standards as compared with conventional data processing;
- Cloud service providers should offer greater transparency, security, accountability and trust in CC solutions in particular regarding information on potential data breaches and more balanced contractual clauses to promote data portability and data control by cloud users;
- Further efforts be put into third party certification, standardisation, privacy by design technologies and other related schemes in order to achieve a desired level of trust in CC; and
- Privacy and Data Protection Authorities continue to provide information to data controllers, cloud service providers and legislators on questions relating to privacy and data protection issues.

Table of Contents

List of Figures

Abbreviations

Acronym	Description
API	Application Programming Interface
CC	Cloud Computing
CDP	Continuous Data Protection
CLI	Command Line Interface
CSP	Cloud Service Provider
DBMS	Data Base Management System
DPA	Data Protection Authority
HMAC	Hash-based Message Authentication Code
IaaS	Infrastructure as a Service
IDS	Intrusion Detection Systems
IPS	Intrusion Prevention Systems
IT	Information Technology
NIST	National Institute of Standards and Technology
N/A	Not Available
PC	Personal Computer
PaaS	Platform as a Service
REST	REpresentational State Transfer
RPO	Recovery Point Objective
RSS	Really Simple Syndication
RTO	Recovery Time Objective
SaaS	Software as a Service
SSL	Secure Socket Layer
TCP-IP	Transmission Control Protocol – Internet Protocol (suite)
WLAN	Wireless Local Area Network

1 Cloud Computing

“Cloud computing is an evolving paradigm.” [2]

The National Institute of Standards and Technology (NIST) in September 2011 released a Special Publication SP 800-145, in which it defined cloud computing as:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models. [2]

Cloud computing intends to provide a baseline for discussion from what is cloud computing to how to best use cloud computing. [2]

Cloud computing is far more dynamic than traditional data processing. The location where data processing takes place can change dramatically. The current location of data and where it is processed can depend on a variety of factors to which end users and data controllers traditionally have given little thought, and into which they do not necessarily have the insight or ability to control. [17]

2 Ethical Issues

2.1 The City Context

Cities have a civic leadership role addressing major societal challenges, at the same time as they have legal and regulatory requirements that mean their resource focus and commitment is towards keeping the city working. Throughout Europe, cities are facing increasing challenges. Post-2008 their budgets have reduced, and at the same time, the expectations of their citizens have not decreased. The challenge is to do more for less. Cloud can offer a contribution to this.

The city's digital infrastructure is important. Cities often have many legacy systems but they may have limited resources or in-house skills. At the same time they must follow robust and complex procurement guidelines which restrict their direction and speed of travel. At the same time local government is responsible for a wide range of new directives from regional or national government, or from the EU. Many of these regulations aim to help citizens – e.g. through freedom of information legislation or directives on open government. At the same they are an additional workload.

New technologies, particularly where they are being created by private sector businesses, look to build on the advantages of innovation, often ahead of the ethical framework. The popularity of the "smart city" is growing as a route to city management. This needs to be developed through a partnership of stakeholders – city authorities, universities and other public sector bodies, large and small companies, as well as citizens themselves.

A key issue is that city municipalities operate in a legal context – they are data controllers for a good deal of citizen focused data, much of which is sensitive, personal and highly regulated. They are also are trusted bodies, and citizens expect that their approach to data collection, retention, storage and sharing is in line with these responsibilities. Data protection law was designed to be a fundamental and concrete dimension of the individual's right to privacy, the primary safeguard against misuse of personal information.Therefore the ethical dimension is key:

- What data is being stored?
- Who has access to that data?
- Where is that data being stored?
- Which applications can access?
- How is data shared?
- Is there a publicly available policy on data?
- Is there a mechanism for citizens to query compliance?

These issues must therefore be the fundamental starting point for any dialogue with suppliers, with clear responsibilities on ethical issues both within the procurement process and in deployment.

The context of this project is to enable cities to have the tools and methodologies to implement a robust system for deployment on the cloud. Much of this will be common across non-cloud services. However by formally implementing it within a municipality, this should ensure that all parties are aware of the ethical dimensions required at each stage- from citizen concerns on data

retention and use, through to procurement requirements and processes, as well as contractual arrangements with third party suppliers, engagement with the city's data controller (and other departments concerned with risk). The intention is that these guidelines should enable cloud computing implementations with robust processes in place.

2.2 Wider Ethical Context

A number of areas of ethics are impacted in the development of cloud computing. As cloud computing grows, it is starting to dominate some aspects of the internet. As such, it is important that we look at the ethical issue in cloud computing.

Whilst cloud computing is a technical and social reality, it is also still an emerging technology which is rapidly expanding. We do not yet know what it will be used for in the future and which social, ethical, or legal consequences these uses will have. Early recognition of ethical and related issues is essential. Timmermans, J et al, "The Ethics of Cloud Computing – A Conceptual Review" [23], three areas of ethical concern are raised:

- The shifting of control from technology users to the third parties
- The storage of data in multiple physical locations
- The interconnection of multiple services

In relation to control, the loss of (direct) control can become problematic if something goes wrong. Among risks associated with cloud computing are unauthorised access, data corruption, infrastructure failure, or unavailability/outing. The storage of data in multiple locations inherently means that the architecture creates a complex chain of events or systems, many people will have had a share in an action.

Handing over control to the cloud provider also raises the question of information self-determination. Informational self-determination refers to the right or ability of individuals to exercise personal control over the collection, use and disclosure of their personal data by others.

Also, say an important business hosts a file or a program on a cloud network. It may be possible for unauthorised people to access or copy that file or program. And if that cloud network goes down for a while, or permanently, then the company has lost that file or program until the network goes back up – perhaps permanently – unless they also have another copy stored somewhere else. And if the company stops using that cloud network, but have left files on it, then the problem of who has the rights over those files could be a major issue.

An important issue that needs to be considered is that cloud computing makes it possible for individuals and/or companies to access other people's information and personal details, without that person necessarily knowing that their information is being accessed. Timmermans et al, describe this as "function creep" where data collected for a specific purpose, over time may become used for other (unanticipated, unwanted) purposes. For example a database with biometric data of citizens may be designed for authentication purposes but may then turn out to be very helpful for crime investigations.

Particularly in the case of personal data stored in the cloud, vagueness about privacy can be potentially harmful. As data is no longer stored locally, control over the data is shifted to the service providers. Consumers then need to trust the cloud provider that certain personal information will not be exposed. In the cloud different services increasingly become intertwined: a

hosted application of one company for instance can be built on a development/deployment framework of another. Both reasons imply that to consumers it will not always be clear what they can expect from service providers in the cloud concerning privacy.

Jonathan Zittrain, Professor of Internet law at Harvard Law School, in his book "The Future of the Internet, and how to stop it" (Zittrain, 2008), illustrates how user's devices are devolving from autonomous systems into "tethered appliances" or "dumb terminals" whose functionalities are entirely dependent on the services proposed by the cloud operators (Zittrain, 2008; p. 41). Even previously decentralised applications based on open and decentralised protocols (such as SMTP for email, or IRC for live communication) are now turning into centralised cloud-based applications (such as Hotmail, Gmail, Gtalk or Facebook for synchronous and asynchronous communications). [23]

Cloud computing raises therefore a series of ethical issues concerning users' autonomy and control (Timmermans & al, 2010). In spite of the decentralised nature of the internet, the advent of cloud computing might, indeed, undermine the autonomy of users (De Filippi, 2013) who benefit from innovative and personalised online services at the costs of becoming increasingly dependent on them (Haeberlen, 2010).

Before the advent of cloud computing, even though hardware manufacturers could, to some extent, regulate users' behaviour by implementing specific features or technical constraints into particular devices, users were (at least theoretically) able to decide by themselves which applications to install and run on their own devices. Today, given that most applications are stored and run directly from the cloud, power is increasingly concentrated in the hands of a few large service providers, which have the ability to determine exactly what can or cannot be done on their platforms. In addition to the obvious concerns that this might entail in terms of data privacy and security (Nelson, 2009), relinquishing control over personal data or information can also undermine users' right to information self-determination (i.e. users' ability to determine, by themselves, how information can and will be used) – a right which Germany has recognised as one of the most important parts of the general right of personality (Allgemeines Persönlichkeitsrecht). [23]

Considering the boarder ethical picture, if the data owner is the city authority, the ethical responsibilities falls to them and must be translated in to actions. Indeed, without attention to this, it is likely they would find themselves in breach of data protection legislation as well as losing citizen confidence.

2.3 City Roadmap for Cloud Computing Adoption

Main Principles

1. Municipalities must abide with all national and European regulations relating to data protection.
2. Cloud computing is covered by the existing legislation. A broad definition of cloud computing is applied, and will include data stored on the cloud.
3. Identifying the data controller for a city's data is necessary to understand who is responsible for data protection on the cloud.
4. It is likely that more than one data controller may be involved when services are moved to the cloud and this responsibility will be retained by each of those data controllers.

5. “By processing data in the cloud an organisation may encounter risks to data protection that they were previously unaware of. It is important that data controllers take time to understand the data protection risks that cloud computing presents ”.
6. Selection of a cloud provider should take into account the responsibilities of the data controller (e.g. where the data is held, who has access to it, how the data is kept and for how long).
7. Moving data to the cloud should include an assessment of what data is being moved to the cloud and what the responsibilities of the data controller area.
8. Not only the location of the cloud services, but also any encryption is included. As all encryption can be vulnerable over a period of time the security of the access to that data must be considered when implementing a cloud solution.
9. Access to the data and by whom (including in certain circumstances the cloud provider) needs to comply with the relevant legislation.
10. As cloud-based data is unlikely to be just in one place, there must be a clear process, with reasonable timescales for deletion of that data.

How to Implement a Cloud Policy

1. Identify who in the Local Authority is the Data Controller and what mechanisms are already in place for data protection.
2. Ensure you have in place a clear process for the adoption of any cloud services.
3. Understand the difference between different types of data and the level of security that is required for each. (For example: decret; densitive; public data).
4. Develop security profiles for each type of data so that once it is classified you have clear guidance in place to implement.
5. Be aware of national and international legislation, and of relevant standards (e.g. around security).
6. Classify the data that you are looking to store against a security profile.
7. Identify the cloud provider that you are looking to use and get clarification that they comply with the classification.
8. Enter into a contract with the cloud provider to provide contractual compliance with your data protection needs.
9. Ensure you have a robust system (and clear audit trail) in place if you need to check the compliance with the cloud policy.
10. There are a wide range of cloud solutions and where restrictions might limit the adoption of cloud services, look at whether there are hybrid solutions that can still utilise the benefits of cloud storage.

3 Data Protection

There is still uncertainty in relation to cloud computing, when addressing the issues of data protection and other legal/ethical issues since:

- Data could be transferred to jurisdictions that do not provide adequate data protection;
- The controller accepts standard terms and conditions that give the cloud service provider too much leeway, including the possibility that the cloud service provider may process data in a way that contradicts the controller's instructions; [41]
- Cloud service providers or their subcontractors can use the controllers' data for their own purposes without the controllers' knowledge or permission; [41]
- The controller loses control of their data and data processing;
- The controller is unable to properly monitor the cloud service provider;
- Data protection authorities could be precluded from properly supervising the processing of personal data by the controller and the cloud service provider;
- The controller relies on unfounded trust in the absence of insight and monitoring, thereby potentially contravening the data protection legislation in force in the country of establishment;

3.1 Definition

Data protection is the level of availability or confidence in being able to access important data. There are many options to ensure data protection. These options often come in the form of RAID, data replication, and data archiving. Depending on the data itself, companies may replay on any number of these methods. Data that is changed frequently may be stored at certain intervals using SAN replays, a form of incremental backups.[2]

Old data is often stored as archive data so that it may be referred to if and when it is ever needed. It's heavily compressed using a data de-duplication algorithm. It may be stored on any available storage medium. [2]

3.2 European Data Protection Directive (95/46/EC)

In the following paragraphs we present how the Data Protection Directive (Directive 95/46/EC) determines the scope of what "personal data" means.

The Article 29 Working Party [9], acting within the confines of the EU's legal and regulatory environment and in concert with the European Commission, has confirmed that the Data Protection Directive applies to all instances where personal data is processed via cloud computing services. The spirit of the Directive extends and applies to clients, cloud service providers, communications providers, as well as infrastructure providers and others (European Commission 2012a, 8).

A key principle of the Data Protection Directive is, in order for users' personal data to be protected, the users must be informed about who processes their data as well as for what purposes their personal data is processed. Article 6 of the Data Protection Directive states that personal data must be "collected for specified, explicit and legitimate purposes and not further

processed in a way incompatible with those purposes" (Directive 95/46/EC, 1995, art. 6). Article 7 further states that data is to be processed only where the subjects have unambiguously given consent, or other requirements have been met (Directive 95/46/EC, 1995, art. 7).

3.2.1 Data Controller and Data Processor

Personal data[1] and sensitive personal data[2][3] are usually processed[4] in the cloud. In Europe, processing of personal data is mainly regulated by the Directive 95/46/EC. The Directive imposes quite stringent duties and obligations on the actors of such processing, mainly on the 'Controller'[5]

[1] 'Personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. Article 2 (a) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Directive 95/46/EC).

[2] 'Sensitive personal data' means personal data combined with any of the following:

a) the racial or ethnic origin of the data subject

b) his/her political opinions

c) his/her religious beliefs or other beliefs of a similar nature,

d) whether s/he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),

e) his/her physical or mental health or condition,

f) his/her sexual life,

g) the commission or alleged commission by him/her of any offence, or any proceedings for any offence committed or alleged to have been committed by him or her, the disposal of such proceedings or the sentence of any court in such proceedings.

[3] The concept of sensitive data carries different meanings in different legal cultures, cf. Art. 8 of Directive 95/46/EC, Art. 9 EU Draft General Data Protection Regulation and the FTC Report "Protecting Consumer Privacy in an Era of Rapid Change" (2012)

[4] 'Processing of personal data' ('processing') shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction. Article 2 (b) Directive 95/46/EC.

[5] 'Controller' shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law. Article 2 (d) Directive 95/46/EC.

but also on the 'Processor'[6]). Given the above, the fact that personal data can be rapidly transferred by the cloud service providers (CSP) from one datacenter to another and customer has usually no control or knowledge over the exact location of the provided resources (the 'location independence' concept), understandably stimulate customers' concerns on data protection and data security compliance.[4]

Applying such definitions to the cloud-computing environment is quite challenging. At first sight one may say that the customer is the Controller and the CSP the Processor.[5] Nevertheless, CSPs often determine the means and sometimes also the purposes of the processing – falling thus under the definition of Controller[7].

In a cloud-computing environment it remains quite unclear and such roles still need to be determined on a case-by-case basis, in the view of the nature of the cloud services.

3.3 Categories of Risk

Under the prism of cloud computing we can identify the following categories of risk:

Risk 1) Lack of control over the data in the cloud,

Risk 2) Lack of transparency related to the processing of data via cloud computing,

Risk 3) Inability to apply the EU data protection laws.

Regarding Risk 1) it is clear that usage of a cloud service means that users move personal data, into a virtualised system, where they do not have exclusive control of their data and full access to the sort of technical or organisational measures necessary to ensure the availability, integrity, confidentiality, transparency, isolation[8], intervenability and portability of the data.

The Working Party[9] identifies the following features of the cloud as causes that perpetuate this risk:

(1) use of proprietary technology resulting in lack of availability due to lack of interoperability (vendor lock-in), which complicates the data shift between different cloud-based systems

[6] 'Processor' shall mean a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller. Article 2 (e) Directive 95/46/EC.

[7] E.g., the CSP "of an IaaS, caring about the efficiency of its service, could automatically allocate processing and stocking capabilities between various facilities located worldwide. For instance, at a time "t", the most efficient could be to use a data center and processing capabilities located in Germany. But, due to the increasingly use of these facilities at a time "t+1", it could be more effective to have recourse to facilities located elsewhere in the world, for instance in India, in providing the service – which could involve a duplication of data, etc. In this respect, the technology at stake would automatically imply a transborder data flow the controller of whose is not necessarily easy to determine."[6]

[8] In Germany the broader concept of "unlinkability" has been introduced into legislation and is promoted by the Conference of Data Protection Commissioners

[9] Opinion 05/2012 on Cloud Computing

(data portability) as well as the exchange of data between entities using different cloud services;

(2) sharing of resources, resulting in lack of integrity. This is a result of a cloud being a shared system on infrastructure shared among clients;

(3) law enforcement requesting disclosure of information directly to a cloud provider, resulting in lack of confidentiality. If acting outside the EU this could result in a breach of EU data protection law;

(4) outsourcing by providers, resulting in lack of intervenability. The complexity and dynamics of the outsourcing chain can result in situations where services end up being facilitated by various providers without the client fully knowing who is looking after their contract;

(5) limited availability of necessary measures and tools, where a provider does not assist the controller to manage their data in terms of, e.g., access, deletion or correction of data, resulting in lack of intervenability;

(6) possible data leakage, resulting in lack of isolation. A cloud provider has multiple clients acting on its behalf (in an administrator role) are equipped with enough privileged access (high-risk roles) to adversely affect the security of individual clients.

Concerning Risk 2) it is apparent that insufficient information about a cloud service's processing operations poses a risk to controllers as well as to data subjects because they might not be aware of potential threats and risks and thus cannot take measures they deem appropriate.[32]

The Working Party identifies the following factors that affect matters of transparency:

(1) an improper or incomplete understanding of the chain of processing and whether multiple subcontractors may be involved; [25]

(2) lack of knowledge as to where data may actually be geographically located upon processing and throughout the duration of storage in the cloud; [25]

(3) unknown transfer of data to countries outside the European Economic Area (EEA[10]), which do not ensure an adequate level of protection; [25]

For Risk 3) the following factors affect whether a cloud provider is controlled by the EU data protection framework:

(1) cloud provider with a relevant establishment outside the EU;

(2) usage of equipment located outside the EU;

(3) cloud computing services offered to individuals as end users. Such services include storage of pictures, calendars and typically type of information that users usually keep at home and use for personal purposes. In this situation, there is some ambiguity whether the cloud

[10] Countries currently in EEA Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Iceland, Liechtenstein and Norway

provider would be covered by the EU data protection framework, and hence whether individuals' data would be properly protected.

Thus:

- A cloud provider established in the EU will in principle be 'caught' by EU law.
- A cloud provider which uses equipment (such as servers) in an EU Member State will also be caught.
- A cloud provider in other cases – even if it mainly and mostly targets European citizens – would not be caught by EU law.

3.4 Data Protection Principles

A commonly recognised data protection principle is that the processor must not process personal data to a greater extent than that which follows from the explicit instructions from the controller or by legislation (principle of purpose specification and limitation). For cloud computing, this implies that a cloud service provider cannot unilaterally make a decision or arrange for personal data (and its processing) to be transmitted more or less automatically to unknown cloud data centres. This is true whether the cloud service provider justifies such a transfer as a reduction of operating costs, management of peak loads (overflow), load balancing, copying to backup, etc. Nor may the cloud service provider use personal data for his own purposes without the knowledge of the controller. [17] In the latter case the cloud service provider should be seen as a co-controller and as such be held accountable for the unauthorised independent processing of data.

Encrypting data before they leave the controllers environment is another commonly recognised data protection principle. If the controller holds the encryption key, no one else will be able to easily decrypt and therefore read or use personal information. Whilst this is a good approach for a storage or archive service it will not be possible to share these files with anyone without also sharing the encryption key which can be difficult to manage.

The European Data Protection Directive offers some clear data protection principles:

a) Article 6(b), states that that personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes (principle of purpose specification and limitation). Moreover, it must be ensured that personal data are not (illegally) processed for further purposes by the cloud provider or one of his subcontractors.

b) Article 6(c) requires that personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.

c) Article 6(d) requires that personal data must be accurate and, where necessary, kept up to date. Personal data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, should be erased or rectified.

d) Article 6(e) requires that personal data must be kept in a form which permits the identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Personal data that are not

necessary any more must be erased or truly anonymised (principle of data erasure[11]). If this data cannot be erased due to legal retention rules (e.g., tax regulations), access to this personal data should be blocked.

According to [36] keeping personal data for too long may cause the following problems:

- Information can go out of date, and that outdated information will be used in error.
- Increased difficulty ensuring that information is accurate.
- Personal data even though no longer needed, still should be held securely.
- Holding more data than needed could creates problems with responding to subject access requests for their personal data.

e) Article 7 states that the controller should be in a position to give unambiguous consent for data processing, addressing the argument of diminished transparency due to unknown outsourcing of cloud services by one provider to another EU-based provider (Risk 2)).

f) Article 10 requires that the controller be provided with the identity of his processor(s) as well as information about any other recipients of the data (Risk 2)).

g) Article 12 states that a controller must be able to acquire information about any data being processed (Risk 1)). This information should come in a reasonable fashion without undue delay or expense.

h) Article 17 imposes the obligation upon data controllers and processors to apply technical and organisational measures to protect data against accidental or unlawful destruction loss disclosure, and other forms of unlawful processing. (Risk 1), Privacy by design)

i) Article 25 and 26, which explicitly establish that transfer of data outside the EEA, also known as transfer to "third countries", states that this is only permitted where the third country or the recipient has ensured an adequate level of protection for personal data.

j) Article 27 prevents the processor from processing data except on instruction from the controller or under applicable EU law.

k) Article 28 requires from the processor to adequately document its processing.

l) Article 29 requires the processor to co-operate with any relevant supervisory authority.

Moreover Article 14 of the European ePrivacy Directive, contains a similar provision: 'Where required, measures may be adopted to ensure that terminal equipment is constructed in a way that is compatible with the right of users to protect and control the use of their personal data, in accordance with Directive 1999/5/EC and Council Decision 87/95/EEC of 22 December 1986 on standardization in the field of information technology and communications. [8]

3.4.1 Guidance on Applying the Data Protection Principles

This section tries to give some practical examples to illustrate how the principles apply in practice.

[11] Erasure of data is an issue both throughout the duration of a cloud computing contract and upon it termination. It is also relevant in case of substitution or withdrawal of a subcontractor

3.4.1.1 Article 6(b)

The aim of the principle is to ensure that organisations are open about their reasons for obtaining personal data, and that what they do with the information is in line with the reasonable expectations of the individuals concerned.

According to [36] in practice this means that we should: [36]

- ✓ be clear from the beginning about why we are collecting personal data and what we intend to do with it;
- ✓ ensure that if we wish to use or disclose the personal data for any purpose that is additional to or different from the originally specified purpose, the new use or disclosure is fair

3.4.1.2 Article 6(c)

According to [36] in practice this means that we:

- ✓ should hold personal data about an individual that is sufficient for the purpose we are holding it for in relation to that individual;
- ✓ should not hold more information than needed for that purpose.

When it comes to sensitive personal data, we should make sure we maintain only the minimum amount of information needed.

If particular information regarding certain individuals only is needed, we should collect it ONLY for those individuals, since the information is likely to be excessive and irrelevant in relation to other people.

We should NOT hold personal data just in case that it might be useful in the future. However, it is permissible to hold information for a foreseeable event that may never occur, such as blood type groups, related to personnel in hazardous positions.

In order to access whether we are holding the right amount of personal data, we must first be clear about why we are holding and using it. If collected personal data is not adequate to process the purpose in question, we may collect more personal data than originally anticipated.

3.4.1.3 Article 6(d)

According to [36] in practice this means that we should:

- ✓ take reasonable steps to ensure the accuracy of any personal data obtained;
- ✓ ensure that the source of any personal data is clear;
- ✓ carefully consider any challenges to the accuracy of information; and
- ✓ consider whether it is necessary to update the information

However, how can we ensure data “accuracy” or “inaccuracy”? The Data Protection Act does not define the word “accurate”, but it does say that personal data is inaccurate if it is incorrect or misleading as to any matter of fact. This means that if for example a person has changed his/hers living address, from A to B a record showing that he/she currently lives in A is obviously inaccurate. But a record showing that he/she once lived in A remains accurate, even though he/she no longer lives there. Hence we must always be clear about what a record is intended to show.

Regarding keeping data up to date, it depends on what the information is used for. If the information is used for a purpose that relies on it remaining current, it should be kept up to date. For example, employee payroll records should be updated when there is a pay rise. Similarly, records should be updated for customers' changes of address so that goods are delivered to the correct location. In other circumstances, it will be equally obvious when information does not need to be updated.

Regarding the reasonable steps taken to ensure accuracy, we might have to get independent confirmation that the data is accurate. Again this is on a case-by-case basis and it depends on the nature of the personal data and what it will be used for. If data is to be used in making decisions that may significantly affect the individual concerned or others, in that case they will need to put more effort into ensuring accuracy.

Regarding the source of personal data, if it comes from a well-known organisation, then it will usually be reasonable to assume that they have given us accurate information. However, as already stated above, if inaccurate information could have serious consequences, then it makes sense to double-check.

Regarding challenges of individuals about the accuracy of information held about them then individuals should be able to provide convincing documentary evidence that, for example, a date of birth has been recorded incorrectly. Moreover, the information should be marked as being in dispute, avoiding any legal issues if indeed the information turns out to be inaccurate.

3.4.1.4 Article 6(e)

According to [36] in practice this means that we should: [36]

- ✓ review how long we keep personal data;
- ✓ consider the purpose or purposes we hold the information in order to decide whether (and for how long) to retain it;
- ✓ securely delete information that is no longer needed for this purpose or these purposes; and
- ✓ update, archive or securely delete information if it goes out of date

To avoid the problems arising from keeping personal data for too long it is good practice to regularly review personal data and delete anything no longer needed. Information that does not need to be accessed regularly, but which still needs to be retained, should be safely archived or put offline.

If the amount of personal data we hold is:

a) more than small, it is good practice to establish standard retention periods for different categories of information; a system for ensuring that we follow these retention periods in practice, and for documenting and reviewing the retention policy is also advisable;

b) modest, then we might not need a formal data retention policy; However, in order to comply with the protection principles we should conduct a regular audit and check through the records we hold to make sure that we are not holding onto personal data for too long, or deleting them too early.

According to [36] the appropriate personal data retention period is also likely to depend on:

1. The purpose for which it was obtained and its nature. If for example it is necessary to hold the data for reasons of performance of a public function or compliance with employment law, then we should retain it for as long as that reason applies. Similarly if personal data should be kept for historical, statistical or research purposes, then we may keep data indefinitely as long as it is not used in connection with decisions affecting particular individuals, or in a way that is likely to cause damage or distress. Data should be immediately by removed when it is no longer needed for these purposes.

   On the other hand, information with only a short-term value may have to be deleted within days.

2. The surrounding circumstances. If a user stops using a service, we must decide what personal data to retain and what to delete. We may need to keep some information so that we can confirm that the relationship existed – and that it has ended – as well as some of its details. For example contact details might be useful to keep so that we can deal with any complaints they might make about the provided services.

   In that sense, we may need to keep personal data so we can defend possible future legal claims. Unless there is some other reason for keeping it, personal data should be deleted when such a claim could no longer arise.

3. Any legal or regulatory requirements. If we need to keep personal data to comply with requirement such as information needed for income tax and audit purposes, or information on aspects of health and safety, in that case it is not considered that we kept the information longer than necessary; and

4. Agreed industry practices. For example, it is agreed that credit reference agencies are permitted to keep consumer credit data for six years.

At the end of the retention period, the platform can flag data entries for review or delete them after a pre-determined period, especially useful where many records of the same type are held.

However, there is a significant difference between permanently deleting a record and archiving it. [36] Archived data can reduce the risk of misuse or mistake at the expense of availability. Moreover, archived data must comply with the data protection principles and be readily accessed by their owner. If it is deemed necessary to delete data from the platform, it should also be deleted from any back-up made by the platform. Detailed data deletion guidance is presented on section 3.5.12

3.4.1.5 Article 17

According to [36] in practice this means that we need to: [36]

- ✓ have security measures following the current technological developments and regularly review our security arrangements as technology advances; However, we are also entitled to consider costs when deciding what security measures to take.
- ✓ review the personal data we hold and the way we use it to assess how valuable, sensitive or confidential it is, and what damage or distress could be caused to individuals if there were a security breach;

- ✓ make sure we have the right organisational security measures, such as information risk assessments, and be clear about who[12] is responsible for ensuring day-to-day information security. Accountability is important for such security measures;
- ✓ make sure we have the right physical (more on physical security is presented in section 3.5.4) and technical security, backed up by robust policies and procedures and reliable, well-trained staff. In that aspect we should ensure that:
  - ➢ only authorised people can access, alter, disclose or destroy personal data;
  - ➢ staff members only act within the scope of their authority;
  - ➢ staff members should be informed about the possibility that they may commit criminal offences if they deliberately try to access, or to disclose, information without authority;
  - ➢ staff members should be informed of the underlying dangers when trying to obtain personal data by deception or by persuading others to do so;
  - ➢ staff members should be trained on the right use of computers (to avoid, for example, virus infection or spam); and
  - ➢ if personal data is accidentally lost, altered or destroyed, it can be recovered to prevent any damage or distress to the individuals concerned.
- ✓ be ready to respond to any breach of security swiftly and effectively; to do so we must have a breach-management plan[13] dealing with information security breaches. This plan should have the following elements:
  1. Recovery plan and, where necessary, procedures for damage limitation (Containment)
  2. Risks assessment. In particular we should assess the potential adverse consequences for individuals; how serious or substantial these are; and how likely they are to happen.
  3. Notification of breaches. A plan on who needs to be notified and why is deemed appropriate.
  4. Evaluation and response, by investigating the causes of the breach and evaluate the effectiveness of our response to it.

Where a data processor is involved:

- ✓ we must choose a data processor that provides sufficient guarantees about its security measures to protect the processing they will do on our behalf;
- ✓ we must take reasonable steps to check that those security measures are being put into practice; and

[12] a person or department

[13] another example of an organisational security measure

- ✓ a written contract must be in place setting out what they are allowed to do with the personal data.

3.4.1.6 Article 25-26

If data is made anonymous, the information is not considered personal data, and the data protection principles will not apply, and we are free to transfer the information outside the EEA.

A transfer is not the same as the transit of personal data through a third country. This principle will only apply if the personal data moves to a third country, rather than passing through it on the way to its destination.

Putting personal data on a portal will often result in transfers to countries outside the EEA. The transfers will take place when someone outside the EEA accesses the portal. If we load information onto a server based in a country within the EEA so that it can be accessed through a portal, we should consider the likelihood that a transfer may take place and whether that would be fair for the individuals concerned. If we intend information on the portal to be accessed outside the EEA, then this is a transfer.

In order to decide whether there is adequate level of protection for the rights of individuals, in all the circumstances of the transfer, we should perform an assessment of adequacy. Hence we should look at: [36]

- the nature of the personal data being transferred;
- origin of the information in question;
- the country or territory of final destination of that information;
- how the data will be used and for how long; and
- the security measures to be taken in respect of the personal data in the country or territory where the data will be received.

If the assessment indicates that the transfer is 'high risk' (for example if the data is particularly sensitive), then we should perform a more comprehensive investigation of the legal adequacy criteria, considering the following:

- the extent to which the country has adopted data protection standards in its law;
- whether there are any enforceable codes or conduct or other rules, making sure that the standards are achieved in practice; and
- whether there is an effective procedure for individuals to enforce their rights or get compensation if things go wrong

If the assessment reveals that the risks are low the above considerations may not be necessary.

However, the level of protection is unlikely to be adequate if [36]:

- the transfer is to a processor in an unstable country; and
- the nature of the information means that it is at particular risk.

If it is not possible to make an assessment of adequacy it may be possible to put in place 'adequate safeguards' to ensure that the rights of individuals continue to be protected even after their data has been transferred outside the EEA. Examples of these safeguards are:

1. contracts based on the standard contractual clauses approved by the European Commission (EC model clauses); If these model clauses are used in their entirety in the contract, we will not have to make our own assessment of adequacy. However, if we rely solely on these clauses we cannot change them in any way, or remove parts or add other clauses to change the meaning.
2. other contracts we draw after a risk assessment to bring protection up to an adequate level.

One of the conditions for processing is that the individual has consented (see EU Data Protection Directive Article 7 above) to their personal data being collected and used in the manner and for the purposes in question. We can transfer therefore personal data overseas if we have the individual's consent, which should be given clearly and freely and may later be withdrawn by the individual.

We can transfer personal data overseas where it is necessary for reasons of substantial public interest. However, for the purposes of this project it is unlikely to be relevant.

3.5 Recommendations and Best Practices

3.5.1 General recommendations

The following general recommendations are foreseen [17]:

- Cloud computing must not lead to a lowering of data protection standards as compared with conventional data processing;
- Cloud service providers should offer greater transparency, security, accountability and trust in CC solutions in particular regarding information on potential data breaches and more balanced contractual clauses to promote data portability and data control by cloud users;
- Further efforts be put into third party certification, standardisation, privacy by design technologies and other related schemes in order to achieve a desired level of trust in CC; and
- Privacy and Data Protection Authorities continue to provide information to data controllers, cloud service providers and legislators on questions relating to privacy and data protection issues.

Due to strong regulator requirements, that require documentation of data destruction activities, we must:

- Establish procedures to sanitise tenant data when a program or project ends.
- Track the destruction of both the tenant data and metadata through ticketing in a CMDB.
- For Volume storage apart for issues related to the server physical location we should provide secure ephemeral instance storage, by implementing *qcow2* files on an encrypted filesystem.

Moreover, we must ensure that unauthorised users cannot access data either intentionally or accidentally (*Restrict access to sensitive information*).

We should also make sure that we only use supported software, i.e. software for which updates are still being provided.

3.5.2 Accountability

In 2010, the Article 29 Data Protection Working Party ('the Working Party') dedicated an entire Opinion to the principle of accountability [30].

Having assessed the potential benefits and 'possible overall legal architecture of accountability based mechanisms', Opinion 3/2010 advanced a concrete proposal for a general provision on accountability, which reads as follows:

"Article X – Implementation of data protection principles

(1) The controller shall implement appropriate and effective measures to ensure that the principles and obligations set out in the Directive are complied with.

(2) The controller shall demonstrate compliance with paragraph 1 to the supervisory authority on its request.

As far as the first component is concerned, it would appear that this provision would not bring any change to the existing framework. Under the current legal framework, controllers are already obliged to comply with the principles and obligations set forth by the Directive. [29]

The main innovation of the Working Party proposal is found in its second component. If adopted, the Directive would require Member States to provide data protection authorities with the power to request from controllers that they demonstrate the measures they have taken to ensure compliance. If a controller fails to demonstrate that it has implemented appropriate measures, this could be grounds for a separate cause of action, independently of an alleged violation of data protection principles[14].

If accountability shifts from a 'reactive' approach (under which actors are held accountable once their activities have resulted in a complaint), to a more 'proactive' approach, under which actors may called upon to demonstrate their compliance without being alleged of a violation of data protection principle, it can improve the current state-of-the-art in data protection regulation.

Accountability by itself is an amorphous concept, which has different meanings in different contexts. When discussing the role of accountability in the context of data protection, additional specification of its practical implications is necessary in order to give the concept further substance. Accountability as being defined[15] deals how the actors that are bound by these norms should demonstrate their compliance in practice.

3.5.3 Transborder Data Flow

Countries having national legislations have restricted flow of private data outside the national borders due to privacy laws and data protection regulations. However, personal information can be transferred between some countries, if either a model contract[16] is signed and approved by the country regulator, or if the owner of the data has given his free consent.

[14][30] paragraph 60.

[15] In Directive 95/46/EC and the Working Party

[16] Model contracts are agreements containing data protection commitments, liability requirements of the company and the liability to the concerned individuals

However, the cloud computing environment makes it difficult to understand which laws apply when the routes of information flow are not known, making transborder data transfer regulations in the cloud computing environment a very difficult process.

3.5.4 Infrastructure Design

The cloud's infrastructure, including servers, routers, storage devices, power supplies, and other components that support operations, should be physically secure.

The exterior perimeter of each anonymous building should be bullet resistant, have concrete vehicle barriers, closed-circuit television coverage, alarm systems, and manned guard stations, all of which help defend against non-entrance attack points. Inside each building, multiple biometric scans and guards should be used to limit access through interior doors and cages.

Some examples include:

- ✓ CCTV surveillance data should be maintained for at least 30 days.
- ✓ Datacenter access doors should be equipped with a local audible alarm.
- ✓ Use of computer-based controlled access system (CAS) with badge readers restricting access to only those with approval to enter controlled areas. All entries and exits to these areas should be logged.
- ✓ Biometric and card security should be present where appropriate.
- ✓ Use of anti-pass back (badge-out) function in order to prevent multiple users from using the same badge for datacenter entry.
- ✓ Secure disposal of paper waste.
- ✓ Secure storing of portable equipment.

3.5.5 Certification

Certification of cloud computing services allows cloud computing suppliers to show their customers that they meet certain standards, for example on network and information security.

Although certification is not a magic solution, we can see that certification can be of benefit to both cloud computing suppliers and users. Already, we can see existing solution available in the market for cloud computing. [18], [19], [20], [21]

3.5.6 Self-Assessment

This is available in the United Kingdom and certain other countries but not generally throughout the European Union. A customer could "self-assess", that the personal data once transferred is adequately protected. In many cloud situations, when 1) the data is not particularly sensitive, 2) a sensible security diligence has been undertaken, 3) proper contractual language is in place dealing with security, and 4) the cloud provider is a reputable company of substance, it will not be unreasonable for the customer to satisfy itself that there is adequate protection. [26]

3.5.7 Device Protection

Limit the number and types of devices that can access content with device pinning, or enforce the use of mobile device management (MDM-compatible) apps. Enforce application passcode locks and device encryption (on Android or via MDM), and report on device usage.

3.5.8 Protection of Sensitive Information in Transit and Storage

Encryption can and should be used to protect sensitive information in transit and storage.

For data in transit end-to-end encryption should be applied. It must be ensured that personal data in transit is protected against active (e.g. replays, traffic injection) and passive attacks (e.g. eavesdropping), thus ensuring data integrity[17].

Data must also be encrypted by the user, or by the provider when stored to the cloud. The data can be brought back through an encryption gateway for processing on secure servers. This makes encrypted data stored in the cloud a secure solution.

It is true however that ciphers can be broken, or the keys can be accessed. But solutions can be developed to make encryption as safe as can be. Once again the critical point of weakness is likely to be the human and procedural failings. Security authentication could for instance remain only in the hands of the data owner using the cloud. This would eliminate the risk that someone else can decipher the encryption keys, but would in most case require a reconfiguration of the typical data stack.[14]

The encryption keys should not be used by, or be accessible to anyone others than the controller and cloud service provider. The encryption keys should not be used by, or be accessible to other customers of the cloud service provider. Data should not be available in unencrypted form longer and more extensively than is absolutely necessary for the data processing process at hand. Methods rendering data unreadable to CC providers at any given time should also be explored. It could be useful to explore options by which the controller can effectively and quickly cut off the cloud service provider or its subcontractors from decrypting data (an emergency brake). [17]

Storage encryption adds an additional layer of protection that will continue protecting the data even if an attacker subverts the database access control layer. In that aspect we should:

1) **Store Sensitive Data That We Need. Never store unnecessary data.** The first thing we have to determine is which data is sensitive enough to require encryption. For example, passwords, credit cards, health records, and personal information should be encrypted.
2) **Use Strong Cryptographic Algorithms**. Use algorithms such as AES, RSA public key cryptography, and SHA-256 or better. Do not use weak algorithms, such as MD5 or SHA1.
3) **Ensure that Random Numbers Are Cryptographically Strong.** Ensure that all random numbers, random file names, random GUIDs, and random strings are generated in a cryptographically strong fashion. Also ensure that random algorithms are seeded with sufficient entropy.
4) **Use Widely Accepted Implementations of Cryptographic Algorithms.** Use widely accepted algorithms and widely accepted implementations. Ensure that the implementation has (at minimum) had some cryptography experts involved in its creation. If possible, use an implementation that is FIPS 140-2 certified.
5) **Ensure Data Integrity and Authenticity.** Encryption must be always combined with message integrity protection. Otherwise the ciphertext will be vulnerable to padding oracle attack

[17] Integrity may be defined as the property that data is authentic and has not been maliciously or accidentally altered during processing, storage or transmission.

and data manipulation, especially if it's being passed over untrusted channel (e.g. in an URL or cookie).

6) **Store the Hashed and Salted Value of Passwords.** Proper storage helps prevent theft, compromise, and malicious use of credentials. The cloudified service should not restrict the types of special characters (character set, or encoding) and the length (short or no length) of credentials. The cloudified service should use a cryptographically strong credential-specific salt[18]. We should "*design for failure*".

7) **Protect Secret Key from Unauthorised Access**. The lifecycle will specify when a key should no longer be used for encryption, when a key should no longer be used for decryption, and when a key should be removed from use all together.

   If the keys are stored with the data then any compromise of the data will easily compromise the keys as well. Hence we should store unencrypted keys on a different machine than where data is stored.

   Protect keys in a key vault.

   Document concrete procedures for managing keys through the lifecycle and train the key custodians.

   Change keys periodically. Key rotation is a must as all good keys do come to an end either through expiration or revocation.

8) **Generate keys offline and store private keys with extreme care**. Never transmit private keys over insecure channels.

9) Ensure offsite backups are encrypted, but the keys are managed and backed up separately.

3.5.8.1 Data Encryption and Data Tokenisation

Encryption is, so far, the best way to protect data. It uses algorithms to transform specified pieces of information so that they become unreadable until decrypted using cryptographic keys. Generally encryption works as follows: You have a file you want to move to a cloud, you use certain software with which you create a password for that file, you move that password-protected file to the cloud and no one is ever able to see the content of the file not knowing the password.

In general, trends in cloud encryption have remained relatively stable over the past three years.

Figure 3-1 summarises how encryption of data at rest in the cloud environment is applied. As shown, 44 percent and 40 percent of organisations using SaaS and IaaS/PaaS, respectively, apply encryption before data is sent to the cloud. Twenty-nine percent and 25 percent of SaaS and IaaS/PaaS encrypt data at rest in the cloud using tools placed in the cloud by the organization. Twenty-seven percent and 36 percent of data in SaaS and IaaS/PaaS environments rely on data at rest encryption that is applied directly in the cloud by the cloud provider.[12]

[18] A salt is fixed-length cryptographically-strong random value.

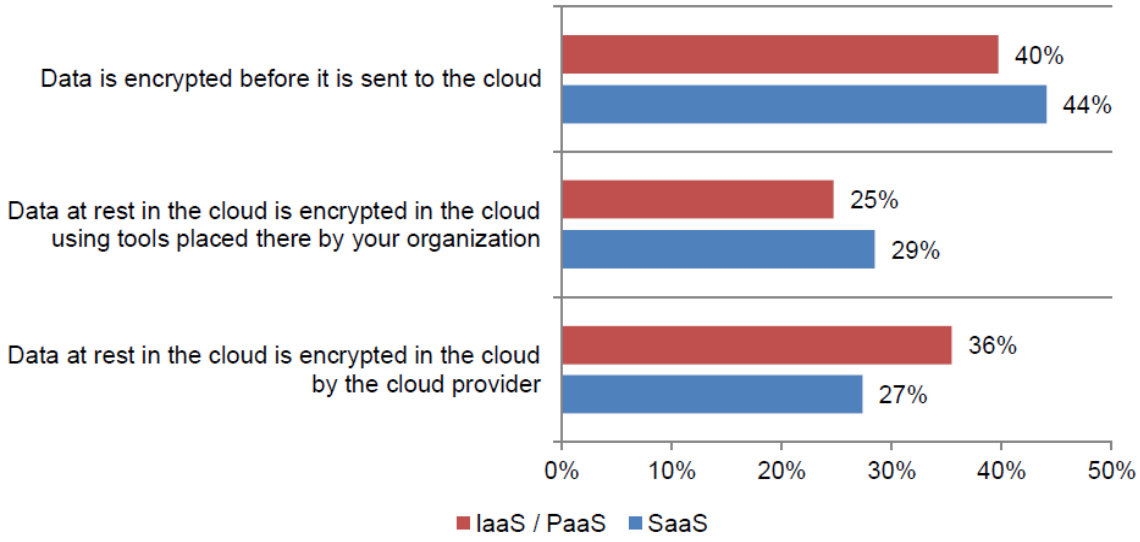

Figure 3-1 If data at rest in the cloud is encrypted, how is that protection applied?

Consolidated analysis for SaaS and IaaS / PaaS

Tokenisation, on the other hand protects data in a different fashion than encryption. Rather than using an algorithm to transform data, tokenisation replaces sensitive data with structurally similar but mathematically unrelated "tokens" before the data leaves our environment. The original information is not contained within the token, and thus the token cannot be reversed into true data. The real, sensitive information is stored in a different location entirely.

However, encryption has some drawbacks, especially when compared with tokenisation.

The first, and by far the biggest, problem with data encryption is that it's reversible. The strength of the encryption is based on the algorithm it uses to secure the data. However, all encryption is eventually breakable.

Another problem with encryption is that, because it's reversible, governing compliance entities still view encrypted data as sensitive data. In other words, it's data that must be protected, because it can be reversed back to the true information. This results in significant capital expenditure in purchasing solutions to protect encrypted data.

Tokenisation, on the other hand, has none of these problems. Since the original information is not contained within the token, the token cannot be reversed into true data. Even in the case that a hacker should manage to break into our environment and steal our tokens, they've really stolen nothing. Tokens cannot be used for fraudulent purposes. Furthermore, tokens can't be reversed independently of the secure platform or software by breaking an algorithm.

3.5.8.2 Open Source Data Encryption

Now that we know that corporations—or at least individuals in corporations—have worked with **intelligence services** to build backdoors into encryption technology, its better to use tools that employ open-source or public-domain encryption methods, as they need to work with every vendor's software and, in the case of open-source encryption, can be scrutinised for potential security flaws.

With that in mind, here are some tools worth checking out:

- *Truecrypt* (http://www.truecrypt.org/) for encrypting sensitive files, folders, and entire drives on your PC.

- The *GNU Privacy Guard* (GPG) (https://www.gnupg.org/), an open-source implementation of the OpenPGP protocol used to encrypt email communications.
- *Off-the-record messaging* or *OTR* (https://otr.cypherpunks.ca/), a cryptographic protocol for encrypting and authenticating instant-messaging communications. The protocol uses AES and SHA-1 standards and comes baked into TAILS and is recommended by Schneier even in the wake of the NSA revelations. Here's a list of IM software that supports OTR.

3.5.8.3 Using Secure Hash Function to Check Data Integrity

Data integrity uses a one-way mathematical function, which takes a stream of data and reduces it to a fixed size data. The result is called a *digest* and can be thought of as a fingerprint of the data. The data digest can be reproduced with the same stream of data, but it is virtually impossible to produce a different stream of data that produces the same data digest. A data digest can be used to provide integrity in electronic communications. [35]

A secure hash function is used to generate a Hash-based Message Authentication Code (HMAC), using a secret key that both the data owner and cloud provider share. The resultant hash value is stored together with the corresponding file in the cloud.

When the user requests data from the cloud both the data and its HMAC are sent. He can re-generate the HMAC to protect against changes in the data from any source. Third party can intercept sender's message and replace it with a new message, but he cannot generate an acceptable HMAC without knowing the secret key. [35]

HMAC can be used in combination with any iterated cryptographic hash, such as AES, RSA public key cryptography, and SHA-256 or better.

3.5.8.4 Using a Digital Signature to Check Data Integrity

A digital signature also verifies the integrity of the data. If the data has been changed since the signature was applied, a different digest would be produced. This would result in a different signature. Therefore, if the data does not have integrity, the validation will fail.[34]

3.5.9 Encryption Key Management

The primary difference between encryption key management in an enterprise's data centre versus key management in the cloud is ownership and management of the keys. In a traditional data centre, all key management functions and tools can be configured and maintained by an IT operations team. In cloud environments, the cloud providers and public authorities should make encryption key management a shared responsibility between the cloud provider and the cloud user (citizens and/or public servants). The goal is to reduce costs and improve efficiency as part of a formal key management strategy.

The type of cloud service in use dictates the types of key management available.

It must be noted upfront that in all architectural solutions where cryptographic keys are stored in the cloud, there is a limit to the degree of security assurance that the cloud Consumer can expect to get, due to the fact that the logical and physical organisation of the storage resources are entirely under the control of the cloud Provider. [28]

3.5.9.1 Key Management Interoperability Protocol (KMIP)

The Key Management Interoperability Protocol (KMIP) is a standard that is designed to be a comprehensive protocol for secure exchange of keys between key management systems and encryption devices or applications. By using a standardised protocol, public authorities are able to simplify key management and deploy key management systems that span multiple use cases and equipment vendors.

According to Ponemon, 54 percent of respondents said that KMIP is most important for cloud based applications and storage.

3.5.10 Backups

Backup is the process of making a secondary copy of data that can be restored to use if the primary copy (the production copy which is the official working copy of the data) becomes lost or unusable. Backups usually comprise a point-in-time copy of primary data taken on a repeated cycle – daily, monthly or weekly.

It is the most important means to keep the data from being lost due to intentional or unintentional access. It is also important to encrypt the up-to-date backups. Backup is easiest and the most familiar process for most situations. A backup copy is used to recover data needed to restart an application correctly.

Backup may be required in the following scenarios:

- **Logical corruption**. That can happen due to application software bugs, storage software bugs or hardware failure, such as a server crash.
- **User error**. Where an end user may accidentally or intentionally delete a file or directory, a set of emails or even records from an application.
- **Hardware failure**. In the form of hard disk drive (HDD) or flash drive failure, server failure or storage array failure.
- **Hardware loss.** Possibly the worst case scenario where an event such as a fire results in hardware being inoperable and permanently unrecoverable.

The following backup service levels exist:

1. Recovery Point Objective (RPO).
2. Recovery Time Objective (RTO).

3.5.10.2 Raksha a Data Protection as Service for OpenStack Clouds

Raksha is a scalable data protection service for OpenStack cloud without the burden of handling complex administrative tasks associated with setting up backup products. OpenStack tenants can choose backup policies for their workloads and the Raksha service leverages existing hooks in Nova and Cinder to provide data protection services to tenants. The goal of this service is to provide data protection to OpenStack cloud, while automating data protection tasks including consistent snap of resources, creating space efficient data streams for snapped resources and streaming the backup data to swift end points. Just like any other service in OpenStack, Data Protection as a Service is consumed by tenants; hence, the Horizon dashboard will be enhanced to support data protection service. [11]

3.5.10.3 Automatic Daily Backups

The example OpenStack architecture designates the cloud controller as the MySQL server. This MySQL server hosts the databases for nova, glance, cinder, and keystone. With all of these databases in one place, it's very easy to create a database backup.

```
# mysqldump --opt --all-databases > openstack.sql
```

To automate this process we should create a cron job (a shell script) that runs the backup script once per day (/etc/cron.daily).

```
#!/bin/bash
backup_dir="/var/lib/backups/mysql"
filename="${backup_dir}/mysql-`hostname`-`eval date +%Y%m%d`.sql.gz"
# Dump the entire MySQL database
/usr/bin/mysqldump --opt --all-databases | gzip > $filename
# Delete backups older than 7 days
find $backup_dir -ctime +7 -type f -delete
```

This above script dumps the entire MySQL database and deletes any backups older than seven days.

3.5.10.4 File System Backups

Component	
Compute	The /etc/nova and /var/log/nova directories should be regularly backed up. /var/lib/nova is another important directory to back up apart from the /var/lib/nova/instances subdirectory on compute nodes. This subdirectory contains the KVM images of running instances and this we don't need to backup, since when the compute node was created we took a snapshot of it.
Image Catalog and Delivery	The /etc/glance and /var/log/glance directories should be regularly backed up. /var/lib/glance is another important directory to back up apart from the /var/lib/glance/images subdirectory on compute nodes.
Identity	The /etc/keystone and /var/log/keystone directories should be regularly backed up.
Block Storage	/etc/cinder and /var/log/cinder follow the same rules as other components. /var/lib/cinder should also be backed up.
Object Storage	/etc/swift is **very important to have backed up**, since this directory contains the swift configuration files as well as the ring files and ring builder files, which if lost, render the data on our cluster inaccessible. A best practice is to copy the builder files to all storage nodes along with the ring files. Multiple backup copies are spread throughout our storage cluster.

3.5.10.5 Recovering Backups

To recover backups we must first ensure that the service we are recovering is not running. For example, to do a full recovery of nova on the cloud controller, first stop all nova services:

```
# stop nova-api
# stop nova-cert
# stop nova-consoleauth
# stop nova-novncproxy
# stop nova-objectstore
# stop nova-scheduler
```

To import a previously backed-up database:

```
# mysql nova < nova.sql
```

Once the files are restored, start everything:

```
# start mysql
# for i in nova-api nova-cert nova-consoleauth nova-novncproxy nova-objectstore
nova-scheduler
> do
> start $i
> done
```

3.5.10.6 Supplementing Backups

A *snapshot* is a point-in-time copy of data created from a set of markers pointing to stored data and is effectively a backup. There is a variety of techniques that can supplement backup and provide rapidly accessible copies to which is it possible to roll back.

The following list describes some of these techniques:

a. **Copy-on-write snapshot**. This technique makes an initial snapshot then further updates as data is changed. As long as all iterations of the data have been kept restoration to a specific point in time is possible.

b. **Clone/split-mirror snapshot**. This technique creates reference pointers to the entire contents of a mirrored set of drives, file system or LUN every time a snapshot is made. This technique takes longer than the previous one, because all data is physically copied when the clone is created. However, since the cloning process has to access primary data at the same time as the host application, it may result in performance degradation of the host application.

c. **Continuous data protection (CDP)**. This technique tracks and stores all updates to data as they occur. Like the previous technique there is a price to pay with CDP in terms of the performance impact of storing the data and the cost of storage needed to keep every changed block copy. A solution to this is what it is called *near-CDP*, where snapshots of changed data at set times are taken while changes are consolidated over a longer time period. This means heavily updated data doesn't overwhelm the capacity of the CDP system. Various APIs enable CDP solutions to be implemented by third-party software vendors.

3.5.10.7 Best Practice

A good data protection strategy combines a number of aspects that can include all of the above techniques.

Short-term snapshots are great for dealing with user errors and some data corruption scenarios. They are very fast and very space-efficient.

CDP takes things a step further with more flexible recovery scenarios that trade off backup capacity and performance against restore granularity.

Finally, traditional backup offers a solid rollback solution should a major hardware or site disaster occurs. Although traditional backups don't necessarily provide the flexibility or efficiency of other methods, they offer a better long-term solution for data retention especially where backup policies dictate multiple backup copies in geographically dispersed locations.

An efficient data protection mechanism should make use of a combination of all of these solutions, applying them to different classes of data as necessary.

3.5.11 Automatic Logging and Audit Trails

All uses of personal data by cloud service providers and their subcontractors should be automatically **logged**. The log should be easily accessible to the controller and be designed in a simple, readily understandable form. The cloud service provider and its subcontractors should ensure the integrity of the logs.[17]

Moreover, the cloud service providers and their subcontractors should establish an automatically recorded **copying and deletion audit trail**, showing clearly which copies of personal data the processor or its subcontractors have created and deleted.

To that direction, the cloud service providers and their subcontractors should establish an automatically **location audit trail**, showing the physical locations in which personal data have been stored or processed and when.

These audit trails should be made available to controllers and Data Protection Authorities.

3.5.12 Data Destruction

The deletion of personal data is an important activity in data protection, given Article 6(e) requirements.

It should be ensured that deletion of personal data from disks and other storage media can be executed in an effective way. Measures include immediate overwriting with random data[19], destroying/demagnetising the storage media, physically destroying the media so that it can no longer be used, usage of secure deletion software.

Deletion by dereference of data and later overwriting by reuse of the storage areas is generally not sufficient, as it opens the possibility that data become accessible again by renewed reference before or during the reuse of the storage areas.[17]

[19] special software tools that overwrite data multiple times in accordance with a recognised specification should be used

Is should also be ensured that data deletion also affects backups that are created by the cloud service provider. Since personal data may be kept redundantly on different servers at different locations, it must be ensured that each instance of them is erased irretrievably (i.e., previous versions, temporary files and even file fragments are to be deleted as well).[32]

Data deletion is also of prime importance when a controller terminates a contract with the cloud service provider.[22]

3.5.13 Data Anonymisation

Anonymisation is the process of turning data into a form which does not identify individuals and where identification is not likely to take place.

Anonymising data requires that identifiers are removed, obscured, aggregated and/or altered in some way. The term 'identifiers' is often misunderstood to simply mean formal identifiers such as the data subject's name, address and unique identification numbers e.g. a Social Security or National Health Service number. But, identifiers could in principle include any piece of information, or combination of pieces of information, that makes an individual unique in a dataset and as such vulnerable to re-identification. [39]

Anonymisation is not foolproof. De-anonymisation is variously known as *data intrusion*, *the mosaic effect and jigsaw identification*. The idea is that extra information can be added to the anonymised data, piecing together enough evidence to identify specific respondents, and/or to disclose certain attributes about specific respondents.

However, we should keep in mind that anonymisation may impact on the usefulness of data.

3.5.14 Portability

Cloud provider should make use of standard data formats and service interfaces in order to facilitate interoperability and portability between different cloud providers. This is needed in case a cloud client decides to move from one cloud provider to another. Lack of interoperability may result in the impossibility or at least difficulties to transfer the client's (personal) data to the new cloud provider (so-called vendor lock-in). The same applies for services developed on a PaaS, where the cloud provider should guarantee the portability of data and services.

3.5.15 Contractual Best Practices

The controller should pay attention in the agreement made with a cloud service provider. In particular [17]:

- The controller should secure a complete list of information in advance about all physical locations in which, throughout the duration of the agreement, data may be stored or processed by the cloud service provider and/or its subcontractors, including backup (**principle of location transparency**).
- The controller should ensure that neither the cloud service provider nor its subcontractors transfer data to locations other than the physical locations listed in the contract, regardless of their reason for so doing, and regardless of whether the data are encrypted. This should be supported by technical measures whose existence and dependability the controller has an actual ability to inspect.
- The controller should ensure that the agreement with the cloud service provider does not contain ambiguities or room for interpretations which undermine the principle that the

cloud service provider only processes personal data according to the controller's instructions. Should cloud service providers be able to unilaterally change the agreement the controller should have the right to terminate the contract and to transfer the data to a different cloud service provider.

- The controller should ensure that the agreement explicitly states that the cloud service provider may not use the controller's data for the cloud service provider's own purposes. (**principle of purpose specification and limitation**). This should be supported by technical and organisational measures to mitigate this risk and provide assurances for the logging and auditing of relevant processing operations on personal data that are performed by employees of the cloud provider or the subcontractors. Penalties should be imposed in the contract against the cloud service provider or subcontractor if data protection legislation is breached.
- The controller should have the opportunity to inspect or have inspected all locations that process personal data wholly or partially in the present or have done so in the past, or may do so in the future under the agreement. The agreement should specify that the controller has the right to obtain full insight into all aspects of the cloud service provider and its subcontractors that the controller deems necessary to ensure compliance with the agreement, including ensuring that processing of personal data is done according to instructions, legally and in a suitably secure manner.
- The controller should secure the right to let a trusted third party[20] wholly or partially monitor the processing of personal data by the cloud service provider and its subcontractors, if any.
- The controller should ensure that the cloud service provider and its subcontractors, if any, offers full transparency regarding the data transfer, the locations used for data processing and storage of personal data.
- The controller should request full transparency from the cloud service provider, regarding the subcontractors used and what processing they perform for the cloud service provider.
- The controller should ensure that the agreement contains clear provision for the erasure of personal data.
- EU established controllers whishing to transfer personal data to a country outside the EEA, should follow the Commission model contracts for the transfer of personal data to third countries[21].
- The controller should make it clear, what happens to their information when they close their account. The controller should ensure that the agreement with the cloud service

[20] A recognized auditing firm. However, the prerequisite is that the third party has the necessary qualifications, is independent of the processor, has full access to and insight into the actual conditions and circumstances under which processing by the processor takes place and can reliably report his observations, assessments and conclusions to the controller

[21] http://ec.europa.eu/justice/data-protection/document/international-transfers/transfer/index_en.htm

provider contains details on what they mean by deletion and what actually happens to personal data once they have deleted it.

References

[1] STORM CLOUDS Consortium, "Surfing Towards the Opportunity of Real Migration to CLOUD-based public Services", November 2013

[2] National Institute of Standards and Technology (NIST), Special Publication 800-145, The NIST Definition of Cloud Computing, September 2011.

[3] Data Security vs. Data Protection, http://www.skullbox.net/data-security_vs_data-protection.php

[4] "Data Protection and Data Security Issues Related to Cloud Computing in the EU", Paolo Balboni, Tilburg Law School Research Paper No. 022/2010, August 21, 2010.

[5] Balboni, Paolo, Mccorry, Kieran & Snead, David: Cloud Computing – Key Legal Issues. In: Cloud Computing Risk Assessment. European Networks and Information Security Agency (ENISA), 2009, p. 97 – 111.

[6] Poullet, Yves, Van Gyseghem, Jean-Marc, Gérard, Jacques, Gayrel, Claire & Moiny, Jean-Philippe: Cloud Computing and Its Implications on Data Protection. Council of Europe, 2010.

[7] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data".

[8] 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), Official Journal L 201 , 31/07/2002 P. 0037 – 0047

[9] ARTICLE 29 Data Protection Working Party, Working Party on Police and Justice, The Future of Privacy, Joint contribution to the Consultation of the European Commission on the legal framework for the fundamental right to protection of personal data, Adopted on 01 December 2009.

[10] OpenStack Security Guide (http://docs.openstack.org/sec/)

[11] https://wiki.openstack.org/wiki/Raksha

[12] Trends in Cloud Encryption Study, Ponemon Institute, https://www.thales-esecurity.com/knowledge-base/analyst-reports/encryption-in-the-cloud-english

[13] K. Retzer and S. Kahn, 'Balancing Discovery with EU Data Protection in International Arbitration Proceedings' (2010), New York Dispute Resolution Lawyer, Vol. 3, No. 1.

[14] Privacy-by-Design, 7 Foundational Principles, available at: http://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/

[15] What does the Commission mean by secure Cloud computing services in Europe? http://europa.eu/rapid/press-release_MEMO-13-898_en.htm

[16] Storing your personal data on the Cloud http://epthinktank.eu/2013/02/07/storing-your-personal-data-on-the-cloud/

[17] Cloud Computing http://epthinktank.eu/2012/08/02/cloud-computing/

[18] Working paper on cloud computing – privacy and data protection issues – "Sopot Memorandum" / International Working Group on Data Protection in Telecommunications, 24/4/2012.

[19] Best Cloud IT Certifications for 2014 http://www.tomsitpro.com/articles/cloud-it-certifications,2-537.html

[20] IBM Professional Certification Program http://www-03.ibm.com/certify/certs/cc_index.shtml

[21] Cloud Certified Professional http://www.cloudschool.com/

[22] Top 10 Cloud Computing Certifications http://www.cio.com/slideshow/detail/129043

[23] The Ethics of Cloud Computing A Conceptual Review https://www.dora.dmu.ac.uk/bitstream/handle/2086/5131/2010%20ethics%20of%20cloud%20computing%20IEEE.pdf?sequence=1

[24] Flawed cloud architectures and the rise of decentral alternatives http://policyreview.info/articles/analysis/flawed-cloud-architectures-and-rise-decentral-alternatives

[25] Ostrzenski, Victoria Cloud Computing and Risk: A look at the EU and the application of the Data Protection Directive to cloud computing. Infopreneurship Journal, 2013, vol. 1, n. 1, pp. 29-38.

[26] "Data Protection and Cloud Computing under EU law" – Third European Cyber Security Awareness Day, 13 April 2010, Panel IV: Privacy and Cloud Computing

[27] "Cloud Computing Under The European Commission's Proposed Regulation To Revise The EU Data Protection Framework", Renzo Marchini, Bloomberg BNA: World Data Protection Report, February 2012

[28] "Cryptographic Key Management Issues & Challenges in Cloud Services", Ramaswamy Chandramouli, Michaela Iorga, Santosh Chokhani, NISTIR 7956, September 2013-NIST

[29] "The accountability principle in data protection regulation: Origin, development and future directions", Joseph Alhadeff, Brendan Van Alsenoy and J. Dumortier

[30] "Opinion 3/2010 on the principle of accountability", ARTICLE 29 DATA PROTECTION WORKING PARTY, 00062/10/EN, WP 173, Adopted on 13 July 2010

[31] "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data"

[32] "Opinion 05/2012 on Cloud Computing", ARTICLE 29 DATA PROTECTION WORKING PARTY, 01037/12/EN, WP 196, Adopted July 1st 2012

[33] "Security and high availability in cloud computing environments", IBM Global Technology Services, Technical White Paper.

[34] "Data Integrity In Cloud Computing Security", Dr. Nedhal A. Al-Saiyd, Nada Sail, Journal of Theoretical and Applied Information Technology, 31st December 2013. Vol. 58 No.3

[35] "Introduction to Public Key Technology and the Federal PKI Infrastructure", Kuhn R. D. and others, National Institute of Standards and Technology, 2001. U.S. Government publication.

[36] "The Guide to Data Protection, How much do I need to know about data protection?", UK Information Commissioner's Office, 1 June 2010.

[37] “Deleting personal data”, UK Information Commissioner’s Office, V1.1 February 2014

[38] “Assessing Adequacy International Tata Transfers”, UK Information Commissioner’s Office, V1.0 February 2012

[39] “About Anonymisation: for data about people”, UK Anonymisation Network (UKAN) website

[40] “Protecting Personal Data in Online Services: Learning from the mistakes of others”, UK Information Commissioner’s Office, May 2014

[41] “Working Paper on Cloud Computing – Privacy and data protection issues”, April 2012, International Working Group on Data Protection in Telecommunications

