

Issues and Prospects of Cloud Technology Application in Electronic Archiving

Maria Musyichuk
Magnitogorsk State Technical University
Magnitogorsk, Russia
mv-mus@mail.ru

Aleksandra Makarova
Magnitogorsk State Technical University
Magnitogorsk, Russia
alex-makarova@ya.ru

Sergey Musyichuk
Magnitogorsk State Technical University
Magnitogorsk, Russia
sv-mus@mail.ru

Abstract – This article looks at the issue of ever spreading cloud storage and computing technology as a global trend. The article describes efforts to regulate cloud computing both in Russia and abroad. It also describes how cloud was used for electronic archiving by a number of Russian, American, and French organizations. Some difficulties and the main challenges of massive spread of cloud computing for data storage purposes are also discussed in this article.

Keywords - cloud computing, electronic archive, electronic documents, standardization guidelines, electronic archiving, archiving outsourcing.

Application of cloud computing in electronic archiving appears to be an up-to-date issue. Cloud storage and computing may be considered a type of external storage services. The benefits offered by external storage service providers include their capability to provide comprehensive maintenance covering both paper and electronic documents of a customer, including complete or selective paper-to-digital document conversion. To stress how critical the issue in view is, the following practices would be worth mentioning that have been adopted by a number of European countries in their efforts to save the heritage information of the 21st century. The European countries introduced a procedure of copying the images posted on the websites that have a great scientific, cultural, and social value. Each resource adds up to 20 new images daily, since even social networking sites update every hour. The national agencies and archives are still looking for ways to store all this content. The issue of finding storage capacity for archived materials is of great urgency. In these circumstances the proposal to use cloud technology seems to be the best solution.

Cloud computing is an innovative technology that encompasses IT resources of various hardware platforms providing access to them via Internet.

The technology of cloud computing enables users (or devices) to access the common pool of resources through a network. Such resources may be the actual computing facilities, data storage, various services, software, even communication networks. The cloud technology makes such resources immediately available while dramatically reducing operational costs [2].

The cloud technology is designed to provide computing facilities, software, or platforms as a service. That is why the basic cloud computing business models bear the following names: Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Desktops as a Service (DaaS), and others.

There are four basic deployment models:

1. **Private Cloud** is an infrastructure designed to be utilized by one company comprising more than one user, e.g. business unit. It can possibly be utilized by the company's customers and contractors. A Private Cloud can be owned, controlled and operated by the company or by a third party, or a combination of the two, and it can physically be located either inside or outside the owner's jurisdiction.

2. **Public Cloud** is an infrastructure designed for unrestricted public use. A Public Cloud can be owned, controlled and operated by commercial, R&D and governmental organizations, or a combination of the above. A Public Cloud physically exists within the jurisdiction of the owner, who is the service provider.

3. **Hybrid Cloud** stands for a combination of two or more different cloud infrastructures, i.e. private, public, or community infrastructures, which keep their unique identities while being connected through standard or private data and application transfer technologies. An example of this can be a short-time use of public clouds to balance the workload between the clouds.

4. **Community Cloud** is a type of infrastructure designed for a specific community of users that belong to organizations sharing common tasks. They can be specific missions or safety, political, or compliance tasks. A Community Cloud can be owned, controlled and operated by one or more of the community organizations or by a third party, or a combination of both, and it can physically be located either inside or outside the owner's jurisdiction [1].

The above basic cloud computing business models are:

1) **Software-as-a-Service (or SaaS)** is a model that enables a customer to use applications of the provider that operates in the cloud and can be accessed from different clients or via a thin client (e.g. from a browser, which can be a webmail or a program interface). Except for a short list of user configuration settings, the cloud provider controls and manages the basic physical and virtual infrastructure of the cloud, i.e. the network, servers, operating systems, storage facilities, even specific functionality of the application.

2) **Platform-as-a-Service (or PaaS)** is a model that enables a customer to use the cloud to deploy their basic software for further deployment of new or existing applications (they can be proprietary, custom, or commercial off-the-shelf applications). Such platforms encompass tools for development, testing, and implementation of applications, i.e. database management systems, middleware, programming language execution environments. The tools are provided by the cloud provider.

Except for developed or installed applications and, when possible, the environment (or platform) configuration parameters, the basic physical and virtual infrastructure of the cloud, i.e. the network, servers, operating systems, storage facilities, is controlled and managed by the cloud provider.

3) **Infrastructure-as-a-Service (or IaaS)** enables a customer to use the cloud to handle the data processing and storage, networks, and other basic computing resources. For example, a customer may install and launch any type of software, including operating systems and platform or application software. A customer keeps control over operating systems, virtual data storage systems, and installed applications, as well as a number of available services, such as firewall and DNS. The basic physical and virtual infrastructure of the cloud, including the network, servers, the types of operating systems used, as well as storage systems, is controlled and managed by the cloud provider [1].

Among the famous projects involving cloud computing is Google Docs, a service that is well-known to a mass market consumer and that, together with Apple's iCloud, is the most widely used cloud product. But the project launched by Microsoft appears to be the biggest of all. Microsoft means to offer as a service not only software and storage space, but the actual platform and the infrastructure.

The project named as Windows Azure is something similar to a cloud operating system. Only users are not supposed to use its environment. It is a cloud product based on IaaS and PaaS models.

Talking about the efforts to regulate the use of cloud computing in Russia and abroad, the following is demonstrating attempts to identify fundamental rules to be applied in this area. In 2014 the European Commission presented draft guidelines aimed at reducing costs and sustaining trust to cloud service providers. The document titled as "Cloud Service Level Agreement Standardization Guidelines" has 41 pages and describes in detail the requirements set for the service providers in terms of service availability, quality of support, levels of security and data management.

The above guidelines have been elaborated by a Cloud Select Industry Group, or CSIG. Among the Group members there

are such cloud service providers as Atos, Amazon, Microsoft, Google, SAP, IBM, Salesforce, Cisco, HP, Accenture, Intel and EMC, as well as legal experts, like DLA Piper.

The top managers of some of the leading cloud service providers, and namely Microsoft, IBM, Telecom Italia, Cloud Security Alliance, Arthur's Legal, addressed the Vice-President of the European Commission Neelie Kroes. In her official statement the Vice-President said that "This is the first time when the cloud service providers came to an agreement with regard to common guidelines on cloud service level agreement. I believe the guidelines would be to a particular benefit of smaller businesses looking for a cloud service that would fit their purposes" [7].

The subject of electronic document archiving using cloud computing is being also looked at by the industry professionals. In 2014 the French Archive Specialist Association (L'Association des archivistes français, AAF) website announced that a white book on cloud computing and information management had been published by the Association. We are quite pleased with the efforts of the French Association, which they presented as a white book, a term used to describe documents that are supposed to be updated within a certain period of time.

On 16 January 2014, a commission of the French Archive Specialist Association, together with a commission on electronic archives, carried out an R&D seminar with the title "What is information management in the cloud environment?" (Quelle gouvernance de l'Information avec le Cloud Computing?). The purpose of the seminar was to redefine the key features of the cloud computing, as well as to point out what key issues may arise in the area of document and archive management in the context of rapidly changing technology.

As a matter of fact, many organizations are already using cloud computing as a technical solution that requires the Internet access and a web-browser and ensures storage, protection, synchronization and shared use of information and documents.

As Ms N.A. Khramtsovskaja clearly states, cloud technology makes us strictly differentiate between storage, backup, and archiving. Originally cloud storage services do not meet the specific requirements for document handling and electronic archiving, including requirements for data and document safety, disposal, traceability, and criticality [5].

It is remarkable that, according to a poll conducted by the French Institute of Public Opinion (Institut Français d'Opinion Publique, IFOP) during the Documation-MIS 2015 conference, paper remains a preferred carrier of important information. Electronic documents have not gained enough trust yet. So the lack of trust to electronic data is one of the pending issues. One can try to analyze the issue though. What is at the heart of the 'trust – no trust' dilemma: the lack of reliability of electronic storage or it originates in the users' minds that are still living on traditional values of that "White paper takes any impression" and "The pen is mightier than the sword"?

The public opinion polls carried out by the French IFOP during the Documation-MIS 2015 conference, show that the use of electronic technology helps save time, especially on document exchange. That is according to 9 out of 10 French respondents who belong to the superior socio-professional category CSP+ ("Catégorie Socio-Professionnelle Supérieure" stands for business managers, qualified personnel, businessmen, white collars and related professionals. This category accounts for approximately a quarter of the country's population, i.e. about 12 million people). At the same time, it is important to stress the opposite opinion, which is an important part of our discussion. One third of people from the same category expressed doubts about safety and confidentiality of electronic data.

The benefits of the electronic technology are well-known and recognized. Thus, 80% of respondents from the CSP+ category (without any significant gender-, age-, or industry-related variations) consider electronic archiving (which basically implies introduction of EDMS/ECM) as a useful means to handle both documents and emails. About the same percent of respondents agree that Intranet, or local network, can be useful. Two thirds of the respondents from the CSP+ category access their documents and emails from various devices (accès multi-terminal).

According to the forecasts, cloud computing is expected to spread. However, so far the cloud capabilities are not widely used. Apart from safety and confidentiality, there are other aspects of

document conversion that are not very trustworthy. 50% of the respondents did not express any interest in the concept of Collaborative Working Environment, especially in enterprise social networks. People do not yet recognize usefulness of such tools.

Cloud computing raises a lot of questions among potential users of new technologies. Thus, for instance, 20% of respondents failed to say if the companies they work for are using such technologies. And only one third of respondents from the CSP+ category admitted they had access to the cloud tools. At the same time, two thirds of the respondents believe that cloud computing is indispensable for efficient work practices.

The experience of some countries where a number of projects have been implemented on application of cloud computing in electronic archiving systems, is currently of great interest. The experience of government agencies plays a major role in this. As an example, on 28 January 2014, the US company Meritalk published a small but rather important report. The report was about introduction of cloud technology across the US federal executive departments. Three years ago, the Office of Management and Budget (OMB), a part of the Executive Office of the US President, instructed the federal executive departments to move all the data and applications from the department servers to a cloud. As the observers point out, the Cloud First initiative is expected to help the government maximize the efficiency of the available resources, increase IT flexibility and minimize costs.

The main incentives of such transition include regulatory compliance and a desire to cut costs. However, there exist a lot of pending issues. The picture below shows that the federal agencies choose to not move to the cloud the critical applications that support their activities and so far are only using cloud computing for emailing, collaborative work and as storage for auxiliary systems.

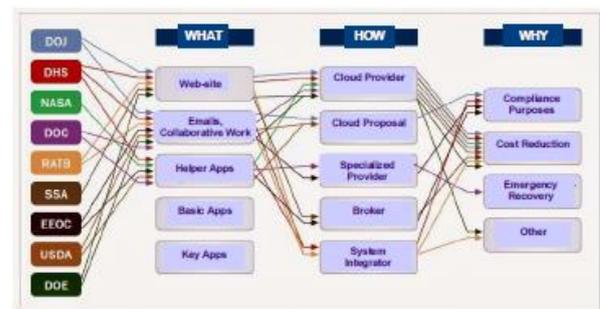


Fig.1. Use of Cloud Computing by US Federal Executive Authorities

This can be explained by the following practical aspects of cloud computing when used for electronic archiving. The researchers highlight that the cloud technology is facing not only technical but also certain legal challenges. For example, the data stored in the cloud should be governed by the norms and regulations of the country of location. On the other hand, such a cloud can be accessed from anywhere in the world, and it can be a distributed cloud [2].

From the legal point of view, it appears to be an extremely challenging situation, especially in the Russian Federation. It causes concern of both business owners and state authorities. 61% of respondents, who took part in a poll by the Russian Cloud Computing Professional Association (RCCPA), named the issue as one of the key challenges on the way to adopting cloud computing.

In 2012, a law was expected to appear that would regulate the use of cloud computing in Russia. The corresponding draft law was submitted for consideration of the Russian Parliament. It may appear that the draft contains more questions than answers. Nevertheless, not only private users and businesses but also the government show interest in adopting cloud computing in Russia. As an example, there have been proposals to use cloud computing during elections.

The following survey may serve as an example of some real steps and ongoing projects on introduction of the cloud archiving technology at the Russian governmental level. If we look at the description of the Russian Federation Public Finance Management Program, we will see that the following measures are foreseen for the implementation of the Electronic Budget system. They include:

1. Creation and implementation of standard forms to be used for collection of electronic information about public organizations involved in the financial management activities;

2. Implementation of a valid EDMS for financial management;

3. Establishment of centralized or cloud data process and storage facilities that would ensure the necessary resilience and disaster tolerance etc. [4]. This measure is also specified among the 5 major activities describing the official targeted programs and the main activities of the minor program.

The Federal Treasury assigned the below amounts to implement this.

TABLE I.
Financing Plan for Cloud Technology
As Part of Support for Public Finance Management Program
To Be Allocated from the Federal Budget ('000 Roubles)

2013	2014	2015	2016	2017	2018	2019	2020
66,812.0	42,000.0	80,000.0	72,470.7	85,163.7	205,409.9	214,653.3	224,312.7
0	0	0	7	6	0	5	5

Today cloud computing is used everywhere – from document storage by users to major businesses and government agencies. Forrester Research is forecasting a profit increase due to the use of cloud computing to 240 billion dollars by 2020 .

In the views of government officials, failure to implement the targeted program would result in the following:

1. Lack of advanced and efficient collaboration between remote participants of the budgeting process; redundant data inputs and processing.

2. Partial automation of an organization’s financial and everyday processes and poor integration of the processes.

3. Lack of information technologies that can help match the budget execution data with the financial results of governmental organizations.

4. Lack of implementation mechanism for the transparency principle set out in the Budgetary Code of the Russian Federation, which implies that the budgetary data should be available to a large pool of users concerned.

Implementation of the targeted program is expected to bring the following results:

1. Comprehensive automation of the public financial control.

2. Integration of the information resources that are essential for the public financial control.

3. Implementation of information services (for centralized financial and accounting control and other services) provided by governmental organizations.

4. Interactive access to the accounting information of governmental organizations available to the public; provision of services by the federal and municipal governments through portals [4].

The annual report on the implementation and efficiency of the Russian Federation federal program “Information Society 2011-2020” indicates that one of the items on the Actual Results of the Federal Program Achieved as of 29 February 2012, is the implementation of a national cloud computing platform that would provide cloud services to the government (SaaS, PaaS, IaaS) [3].

Thus, the minor program #5 “Security in the Information Society” has been realized through the following: 1. A national platform was developed for distributed data processing, which provides cloud services to the government, including SaaS, PaaS, and IaaS. More than 10 cloud solutions were deployed, five of which were launched in a trial mode in the pilot regions of Russia.

2. The National Platform for Distributed Data Processing is deployed on the computing and telecom infrastructure of the Electronic Government and comprises the following components:

- A component that helps manage the deployment of virtual resources and various applications based on the SaaS model;

- An application execution component that allows the developer to create and store the source code depending on the product version;

- An integration component that includes certain elements of the Electronic Government.

3. At the same time, there have been developed basic principles of and fundamental approaches to the safety of the cloud

platform and the services deployed on it; requirements on data centers; a risk assessment technique to be applied when moving the government systems to the cloud; a development concept for the National Distributed Data Processing Platform; and a plan of transition to the cloud for the government.

The other Russian government agencies involved in realization of the program include the Ministry of Telecom and Mass Communications, the Ministry of Economic Development, the Federal Protective Service, the Federal Security Service, the Ministry of Education and Science, the Ministry of Health, the Ministry of Culture, and the Ministry of Regional Development. The number of applications running in the SaaS mode on the cloud platform was 5 apps in 2011; 10 apps in 2012; 20 apps in 2013 [3].

The law requires that all legal documents are safely kept for a certain period of time. This requirement is also true in case of electronic documents. From the experience of some countries, the establishment of a national electronic archive does not only require financial, material and human resources but it also requires a regulatory framework, as well as time to fine-tune the work process. Besides, it would be necessary to reconsider the retention times since both the electronic documents and the related technologies require constant updates.

In this context, the issue of defining the concept of Electronic Archive remains the most urgent challenge, together with legal aspects describing how such an archive should be created and how it should function. Electronic archiving is a relatively new area combining archiving science, IT, document management, and other disciplines. The current re-evaluation of the role of archiving within an organization requires certain standardization. The legal aspect of regulating electronic archiving services, in particular the overall mandatory and industry-related norms and regulations of the federal level, is also of great interest.

Mr. Laptev, an adviser and an instructor of the Department of IT, Development, and Standardization in the area of electronic archiving, data storages and documents of the international department of ICES-UNESCO “Technical and Professional Education & Personnel Training”, Moscow, notes that Electronic Records Management (or ERM) Systems are being actively developed in Europe sponsored by UNESCO. Most European countries have their own standards and guidelines on electronic archiving. In July 2003, Russia saw the birth of an international Committee on Standardization, Education and Certification related to electronic archiving, data storage and documents (www.e-term.ru). The purpose of the Committee is to support the community of professionals in the area.

In Mr. Laptev’s view, an electronic archive is an integrated set of software and hardware means designed to store electronic documents. “Integrated” would be the key word in this definition. Very often an electronic archive is compared to a traditional one, only instead of paper documents it consists of documents carried by media, such as floppy or compact disks. With this approach, the difference between physical and electronic archives would not be great, and the efficiency of such an archive would depend on the human factor, the available storage space etc.

Another definition, which may be the closest to what electronic resources actually mean, describes an electronic archive as a system based on full-text databases (For example, a database of laws and regulations). When searching across full-text databases, a user looks through the abstract and section titles of a document and, if necessary, goes further to the corresponding text. Copies of full texts are also available to the users.

It should be noted here that an electronic archive is not just a separate server, workstation or storage space. Neither is it just a search system capable of finding required electronic data. An archive implies, first and foremost, specific technology and processes that support a complete cycle of document storage – from initial evaluation to utilization through monitoring, definition, safeguarding, and development of references. In this context, an electronic archive may be defined as an integrated set of software and hardware means designed to store, search for, evaluate and make use of electronic documents through monitoring, definition, safeguarding and development of references in compliance with the applicable norms and regulations of the Russian Federation.

It should be highlighted that the use of cloud archiving, as well as cloud storage and processing services, is a leading and beneficial trend in terms of EDMS maintenance costs. Practical

application of the most advanced IT products by the global community demonstrates that the cloud technology has become very popular as an outsourced archiving service.

However, some of the global challengers on the way to expanding the use of cloud storage services include user data vulnerability and new cloud monopolies.

The specific reasons why cloud computing is still not widely used in Russia are quite obvious. The Russian cloud services market is expected to reach 5 billion dollars by 2016. So far Russia is only number 34 in the world by its cloud technology adoption rate. The lack of high-speed Internet coverage in Russia, which is a pre-condition for cloud computing, would be one of the technical reasons for it. Many experts also refer to the mindset that is predominant in the Russian business community when businesses tend not to trust new solutions which have not been extensively proven as practicable [2].

Creating a regulatory environment, i.e. an overall legal framework for interaction between consumers and providers of cloud services, remains a legal challenge that the global community is facing. Terminological and technical aspects also pose a key challenge, according to archive professionals. For example, how to differentiate between the terms 'storage', 'backup', and 'cloud archiving'. Certain aspects of document management, electronic archiving, security, disposal, traceability and data criticality still belong to pending issues.

With cloud computing rapidly developing, the issues of transparency, confidentiality, and control are the key concerns among potential consumers of cloud services. At the current development stage the cloud service consumers are often missing information about how their data is protected and processed on its way to the cloud and what will happen if they decide to use a different provider or if their provider closes down or changes its policies.

The International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC) urged by the European Commission, the National Data Protection Authorities and the information commissioners responsible for elaboration of standards to ensure data protection, have elaborated a new standard, i.e. ISO/IEC 27018:2014 "Information technology - Security techniques - Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors" (See http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61498).

The ISO guidelines are primarily focused on the purpose and the key elements of the standard, and its potential effect. The analysis was carried out from an EU perspective with regard to the EU legislation, both the current and pending legislations. Also, the EU is discussing the need for the new standard and how it should correlate with the other standards on cloud computing, as well as the subject of data protection.

The ISO/IEC 27018 standard sets out guidelines for cloud service providers who are PII processors, and offers a number of control and management measures to be implemented by cloud service providers in order to lower related risks. The standard focuses on lowering risks inherent in public clouds and is aimed at strengthening trust to public cloud service providers while delivering guidelines regarding the providers' contractual and legal responsibilities.

The standard is seeking to help cloud service providers perform their contractual obligations. It also describes the transparency requirements, including the requirement to notify customers in case the law enforcement authorities requested access to the customer's data or the requirement to inform the customers if the provider is using outsourced services.

The standard, which is regarded as the basis for compliance with the national and the EU legislations, contains some elements from the EU Data Protection Directive 95/46/EC, such as the quality processing and accountability principles. Compliance with the new standard can be verified, and the certification bodies, which are independent third parties, can issue compliance certificates to the providers.

However, before the standard can gain momentum at the European market, two main issues need to be resolved. The first one is the difference in terminology between the standard and the EU legislation. And the second issue is a limited area of application as the standard only covers those cloud service providers who act as PII processors.

The new standard is expected to possibly encourage the industry to take measures that would ensure compliance with the data protection directive, and to take a step forward in the efforts to ensure transparency in the relationship between cloud service providers and their customers. Real life will show though if the standard will be accepted and if it will meet the expectations of the authors.

Among the main challenges in the course of expanding the application of cloud computing, as seen by archive experts, there are issues of location and jurisdiction, as well as security and confidentiality issues.

REFERENCES

- [1] PL_Engineering: Software. Cloud computing, URL: <http://pl-e.ru/w/SaaS> (date of access: 30.09.15)
- [2] Bakaldin E. Yu. Cloud: ambiguous weather forecast on the market Modern research of social problems (scientific electronic journal), Modern Research of Social Problems, #6(26), 2013", URL: www.sisp.nkras.ru (date of access: 26.09.15)
- [3] Annual report on the implementation and evaluation of the effectiveness of the state program of the Russian Federation «Information society (2011-2020)» the Ministry of communications and mass communications of the Russian Federation (reporting year 2011. date of report 29.02.2012)
- [4] Passport of the state program of the Russian Federation «Public Finance Management» 2013.
- [5] Khrantsovskaja N. A. The Association of French archivists published a white paper on information management in a cloud computing environment, URL: http://rusrim.blogspot.ru/2014/07/blog-post_3213.html (date of access: 30.08.15)
- [6] Formtek Engineering Data Management Solution certified by Alfresco: Cloud Computing, URL: <http://formtek.com/blog/cloud-computing-eu-drafts-guidelines-for-cloud-slas/> (date of access: 13.04.16)