Modeling of an Environment for Electronic Archiving using a Multi-Agent Approach

Aline Z. Tsague
University of Bamenda
Bambili, 39, Cameroon

Elie T. Fute
University of Dschang
Yaounde, 20269, Cameroon

Laure P. Fotso
University of Dschang
Dschang, 96, Cameroon

ABSTRACT

For several years, the rise of document dematerialization has made that the need for standards or models to promote the collection, communication, storage and retrieval of information within organizations is felt more. The themes of dematerialization and electronic archiving are becoming more relevant to many organizations, whether public or private. In this paper, we analyze sustainable and juridical electronic document storage systems and propose an approach for modelling an electronic archiving system (EAS) directed towards multi-agent systems (MAS).

General Terms

Electronic archiving, multi-agent system

Keywords

Agent, electronic archiving, dematerialization, document, multi-agent.

1. INTRODUCTION

The problem of communication in terms of information or interpersonal relationship is a critical issue in most organizations today. Indeed, these organizations whether public or private, have faced an increase in the volume of paper and/or electronic archives. This exponential growth of information raises at the same time the issues of access to information, its conservation and its internal distribution. Therefore, the need arises for a real awareness of the usefulness of dematerialization and electronic archiving as a natural evolution of human activity. Dematerialization is an operation that aims at managing documents in electronic form as they are in paper form either through a scanning operation, or by reviewing production processes and information management [1]. Thus, archiving must be part of the control policy of business risks [2]. In this regard, [1] defines archiving as the process of organization which aims to identify, put in safety and maintain available all documents that engage a company or organization towards third parties or future activity and whose failure would represent a risk. There is also the issue of integrity of the information stored for commonly used digital media (hard disks, magnetic tapes, optical disks, etc.) do not guarantee conservation beyond a certain number of years, due to their rapid degradation.

An EAS must meet a certain number of constraints (technical, security, legal and juridical, organizational) to ensure the security, authenticity, integrity, perennity and traceability of data from their entering into the system to their destruction. It should also facilitate access to data. Thus, an EAS is a dynamic and complex system [3, 4, 5]. Modelling of such a system requires the use of appropriate tools, hence the interest in the multi-agent approach.

The rest of the paper is organized as follows: section 2 presents a state of the art of electronic archiving as well as that of the multi-agent paradigm. In Section 3, a system model based on multi-agent paradigm is proposed. Section 4 presents details related to implementation and deployment aspects. Finally, Section 5 presents a conclusion of the work.

2. RELATED WORK

2.1 Electronic archiving

The rapid progress of some countries in the digital world has led to the development of several standards often derived from legislation regarding electronic archiving. In addition, with the advent of the electronic signature in many countries, the field of electronic archiving has undergone many transformations. The stakes of electronic archiving are of several types [3, 5, 6, 7]:

1. Juridical: electronic archiving gives juridical value to electronic documents. Thus, it is necessary to ensure that electronic documents constitute valid evidence and having probative value as well as paper documents, and this several years after their entry in an EAS [8]

2. Strategic: it's a matter of data types involved in the electronic archiving. It concerns documents that engage their holder. The retention of these documents allows an organization to keep track of its history

3. Legal or regulatory: at this level the challenge is either legal, that is to say, for compliance with laws, or regulatory in order to comply with the requirements very generic or specific depending on the branch of activity

4. Organizational: it is about the way the data are technically archived to make them easily accessible while respecting the rights of access preset

5. Safety: safety is seen here in two aspects. On the one hand, the storage locality of archives which must be inviolable and protected against disasters (fires, floods, etc.). On the other hand, the access to information: all documents should not be accessed by all users

6. Technology: this stake aims to choose the technology able to reconcile the problems of interoperability between systems and data continuity over the long term, given the rapid evolution of technology and recurring obsolescence of formats, media and restoration tools

7. Financial: the financial stake is double. It is firstly the cost of implementation of the archiving system and its operation and secondly, the cost of risk incurred by the company in case of inability to locate and to return the requested documents.

The stakes of archiving can thus be summarized to the consequences which a company will face if it cannot find the information it has produced a time of activity, meanwhile it needs to communicate or reuse it [3].

The existing standards for electronic archiving are mostly derived from the ISO, the system of reference resulting from the work of a European forum, the French standard, etc. We can cite for example, ISO 15489, ISO 23081, ISO/CEI 27001, ISO/ CEI 27002, ISO/ CEI 27005, Z42-013, Moreq, OAIS (or ISO 14721), the DIRKS method, etc [4, 5, 7, 9, 10]. Each of these standards gives details on technical, organizational and security for the design and operation of an EAS. As pointed out [9], some features such as the integration of fax, workflow, electronic document management (EDM) and collaboration, etc. can be implemented as optional modules insofar as their requirements are not part of the core required features of an EAS.

From now on, there exists within an electronic archiving system the notion of workflow. The MoReq2 standard, a new version of the MoReq, published in 2008, brought little clarification on many issues and introduced specifications for the implementation of a workflow within an electronic archiving system. A workflow in the field of business intelligence enables the management of flow. It is used to study the flow of information in an organization, to know who does what, how, with what information brief, to formalize the working of people and the functioning of an organization.

It should be emphasized that a solution of electronic document management (EDM) is not necessarily an EAS. It is necessary to ensure, in addition to storing the document, the storage of attributes and metadata guaranteeing its origin, authenticity and integrity, as well as all elements of traceability of operations carried out from the reception/creation to the ultimate fate (transfer to final archives or destruction) as required by some standards like NF Z42-013. In addition, for an electronic archive to be valid, it must meet several conditions, including being independent of the system that created it, which is often not the case with EDM solutions integrated into the operational business. Ultimately, whether an EDM or a workflow, the goal is to "make live" a document, while in addition, an Electronic Archiving System uses several features to keep it in a non-editable form [2].

Due to its structure and functioning, an EAS is a complex system. Modelling and implementation of such systems is a major challenge in many areas of science. To provide a satisfactory solution to the problem, interest is charged to multi-agent systems.

2.2 The multi-agent paradigm

In a world where the environment of software is more complex (opening, access, indeterminism, dynamic, ...), the multi-agent approach is to be used as an alternative to classical models. The goal is to solve complex problems in a distributed manner. According to [11, 12, 13, 14], we can define an agent as an autonomous physical or virtual entity operating in an environment which it is able to perceive and act on and can communicate directly with other agents. Figure

1 shows the architecture of an agent. As shown in this figure, an agent uses sensors to perceive its environment and actuators to act on it. The basic unit of perception via a sensor is called a percept. An agent then has a function $f: P^* => A$, where A is a set of actions via actuators and P a set of percepts. This function is a decision module that permits, based on perceptions, to generate actions that will be performed [12]. In Figure 1, we can see that an agent has an internal state, which makes it a dynamic entity. This internal state changes according to its perceptions. We also talk of the behavior of an agent to describe how it reacts to its environment (relation P/A).
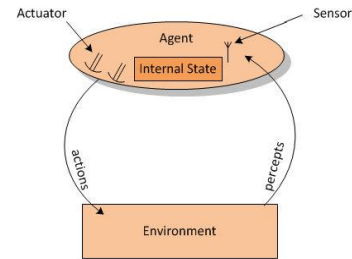

Fig 1: Architecture of an agent

An agent is endowed with autonomy, which differentiates it from an object (in the sense of object-oriented programming) or a process. Indeed, it is guided by a set of personal desires to achieve and is free to respond or not to requests from other agents. In addition to this freedom of action, agents are able to reason and act. Actions they perform change the environment and therefore their future decisions. To solve complex problems, we can put together several agents that interact with each other through their environment. These agents then form what is called a multi-agent system.

According to [11, 12, 13, 15], a multi-agent system (MAS) is an organized set of agents representing the active entities of the system, sharing the same environment, communicating with each other and can be cooperative or protagonists . Agents of a MAS can work together to achieve the same purpose, e.g. the agents of an EAS would cooperate in the capture, storage, access and disclosure of files that it gathers. Moreover, the agents of a MAS can also concur to different purposes, for example, if we model a part of several teams' game by MAS, agents of each team will have different goals. There are two main groups of agents: cognitive agents and reactive agents. These two groups of agents differ mainly in the type of treatment performed on perceptions and rationality of the reasoning performed by agents [12]. Cognitive agents are designed as intelligent entities. They are based on a logical reasoning led from informations of their knowledge base in order to achieve their goals [12, 13]. Agents of this type are often called rational or intentional agents. Generally such agents are used to represent the student model in intelligent tutoring systems (ITS). An ITS is an intelligent system that emulates an actor in front of which is a learner. An ITS should have a good representation of its learners and be an expert in communication i.e. able to act at the right time.

As for reactive agents, they are very simple entities, reacting directly to stimuli from their environment [12, 13]. This type of agents use for solving complex problems, mechanisms of reactions to events not taking into account neither a clarification of the goals nor planning mechanisms.

3. MODELLING

An EAS is the set of materials, software and procedures that organize and control the capture, preservation and destruction, in order to control the risk of non-availability and over-preservation of documents, as well as the provision of engaging documents in a company or organization [1].

It can be an Electronic Safe or an internally developed software solution meeting the requirements of a standard for electronic archiving.

Regarding the implementation of an EAS, the first thing an organization must do is assess its needs. For this, it is recommended to correctly answer the following questions [3]:

- What are the data to archive from the dataset produced?
- What is the criticality of the data?
- What are the requirements of conservation?
- What are the requirements of integrity and security?
- How to deal with volume?
- What access?

The answers to these questions will allow choosing the standard to which the EAS should comply. Depending on the selected archive mode, it's necessary now to determine how to implement an archiving within the organization.

The way the different entities of the system will be assembled is a crucial issue. Regarding an EAS, the main entities or functions are:

- The entries or deposits,
- The storage or conservation,
- The management of descriptive data or metadata,
- The access or consultation.

These entities are usually assembled as a whole, as shown in Figure 2, with no specific accuracy on their nature. On the one hand we have the service side consists of people producing documents to be archived. These documents are then introduced into the EAS where begins the archiving process. During the archiving process, the data and associated metadata will be stored, modified, accessed or disclosed, administered, tracked, etc. On the other hand we have users who access documents of the EAS for consultation and/or modification.
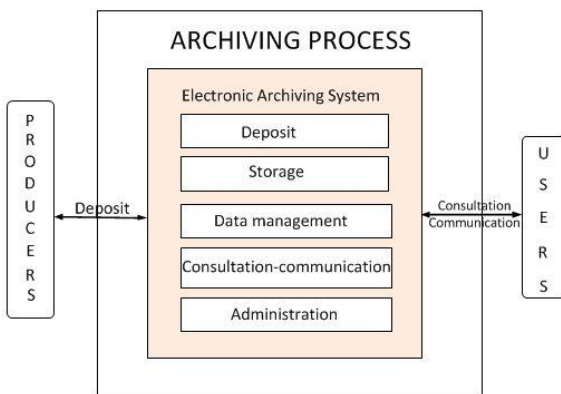

Fig 2: General architecture of an EAS, inspired from [4, 7]

Classical models used to represent dynamic and complex systems such as mathematical models have shown their limits. Indeed, they model the relationships between the various identified entities that make up a system using mathematical equations which do not take into account their individual characteristics. Moreover, the tools and methods used are generally linked to the preexisting information system and are mostly implemented using the MVC technology.

The goal is not only to model entities and their behavior, but also the interactions between these entities [13]. We introduce the multi-agent paradigm in modeling an EAS to take advantage of its modularity, adaptability, reactivity, distribution, simplicity and reusability. The multi-agent approach allows taking into account in the models, qualitative and individual behavior. Several reasons justify the use of multi-agent systems, namely [15]:
- The effectiveness of treatment
- The robustness and dependability
- Flexibility and large scale processing
- A low operating cost
- The interesting cost of developing and reusing

We will also like to build our system according to the rules of software engineering. For this reason, UML formalism is used.

3.1 Identification of agents of the system

From the main entities or functions of an EAS, we have the following agents:

- Agent "**Import**": provides functions and services relating to the acceptance, after an integrity check, of batches of information to be deposit into the archive system, mainly from producers or the administration of the system. This agent is also responsible for preparing the content of batches of information for storage and data management within the archive system,

- Agent "**Storage**": provides the services and functions relating to the storage, maintenance and retrieval of batches of information to be archived,

- Agent "**Data Management**": provides the functions and services related to enrichment, retention and access to descriptive information or metadata (which identifies and documents funds of the archive system) and also "administrative" data used to manage the archive system,

- Agent "**Access**": provides the features and services that help the user to determine whether or not an information exists in the archive system, to find its description, its location if available, and to request and receive information products,

- Agent "**Dematerialization**": provides functions relating to the dematerialization of paper documents. It works with human action. Thus, this agent is not completely independent. For it to fulfill its function, it should as a preliminary that some human actions are carried out. For example, the paper documents should be prepared and a scanner connected to the computer before the agent performs its task.

To this set of agents can be added other cross-functional and complementary agents of the EAS. We can cite essentially agents "**Control**" and "**Perennity**". The agent Control ensures functions and services relating to the overall exploitation of the system while agent Perennity ensures the functions and services relating to environmental monitoring of the archiving system in order to ensure that stored information will remain

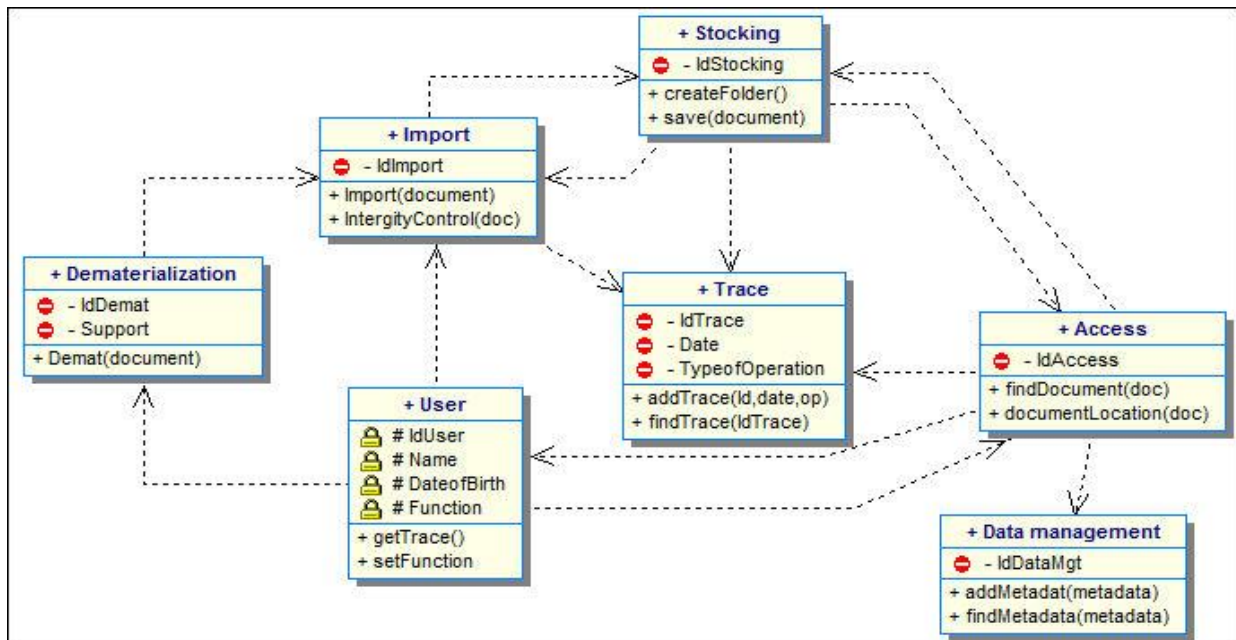accessible over the long term by target users, even if the original computing environment becomes obsolete.

Fig 3: Class diagram associated to the management of an EAS

3.2 Class diagram

Figure 3 shows the UML class diagram of the system being modeled.

In this diagram, the agent Access interacts on the one hand with the Storage agent and on the other hand with the outside particularly with users. These must be taken into account in the internal model of agent Access. Whether for Import agent, Storage or Access, the history related to manipulation of the data is saved. These are operations such as access to documents, editing, and deleting documents, etc. In short, it keeps tracks of all transactions.

3.3 Communicational aspect

Agents interact with each other and with their environment via messages. The messages exchanged can be either an information or a request. Thus, each agent in the system has its own dynamics described by its behavior at each time. In general, an agent can:

- Receive messages

- Send messages

- Update the variables of its scope,

- Make internal deliberation.

Figure 4 shows the communicational architecture of the system where we can see the different interactions between agents and as well as the database.
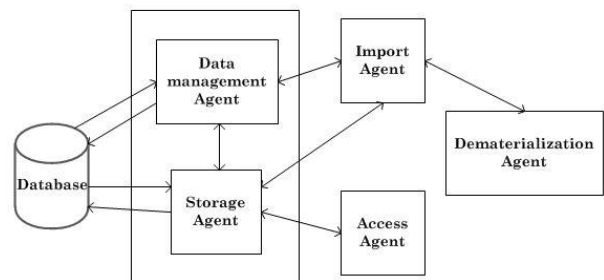

Fig 4: Interactions between agents

3.4 Behavioral aspect of agents

In a MAS, the observed behaviors result from the interactions between the different agents of the system. Indeed, the agents communicate to achieve some functionalities of the system. Thus, to study and describe these features, an emphasis is put on the interactions of all agents. Because most system operations are around the entity document, interactions or communications of agents around this entity should be studied. Figure 5 shows the sequence diagram of the communications of the different agents around a document from its creation to its ultimate fate. The ultimate fate of a document is either its destruction, its transfer to final archives or its transfer of responsibility or ownership.

Due to its place in the modeling of an EAS, the entity document is of great interest. Figure 6 shows the lifecycle of a document in an EAS from its creation to its ultimate fate.
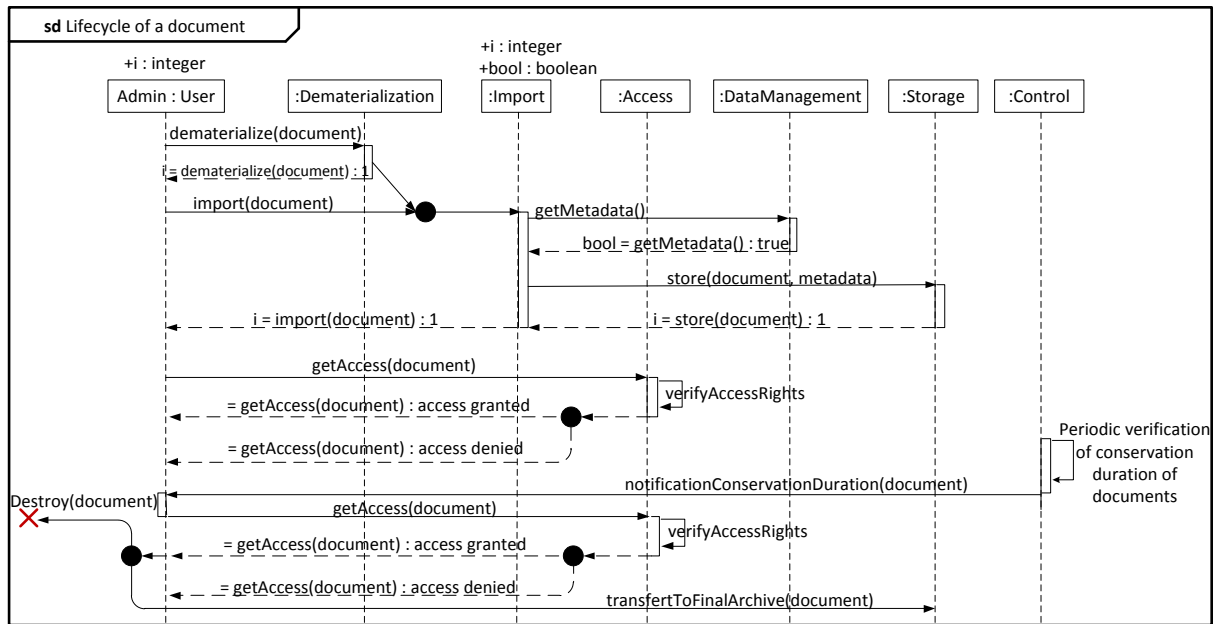
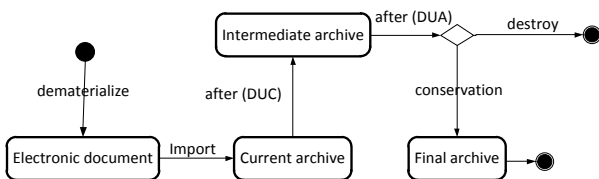**Fig 5: Sequence diagram related to the lifecycle of a document**

Fig 6: Evolution of a document in the system

This figure (Fig 6) shows the state machine diagram of the document entity that passes from its initial state (paper) to an electronic document state after dematerialization. After importing the electronic document, it becomes a current archive which after its useful current duration (DUC) becomes an intermediate archive. In its turn, the intermediate archive becomes, after its useful administrative duration (DUA), either a final archive or a document to be destroyed.

4. IMPLEMENTATION AND DEPLOYMENT

4.1 Implementation of communicational model

The general structure of messages exchanged between the different agents is as follows: **Message (receiver, sender, content)**, where "receiver" indicates the name of the receicer agent or a community of agents, "sender" the name of the issuing agent and "content" of the message content.

To send messages to other agents, an agent must know their identifiers. Each agent has a message box where are deposited all the messages that are intended to it. The algorithm used for the receipt and processing of messages is as follows:

Algorithm ProcessMessage ()

begin

Message m, response;

chain req;

while (1) {

m = waitNextMessage () / * the agent retrieves the next message to be processed in its message box * /

req = readContent (m) / * the agent reads the content of the message and determines the type of treatment to be performed * /

response = process (req) / * the agent processes the message and returns a value that determines whether the agent must return a response or change its internal state * /

if (response.state () == InternalProcess) **then**

Modify internal state of agent

else / * the message processed requires the return of a response * /

sendMessage (m.receiver (), m.sender (), newMessage(response));

}

end

4.2 Deployment

Regarding deployment, it is to describe the physical arrangement of hardware resources that make up the system. Emphasis is also placed on the distribution of components on

these different resources. The main resources are: the application server, the database server and the client(s). The instance of the application server used is JBoss AS 6.0.0 and that of the database server is Postgresql 9.0.1-1. Figure 7 shows the deployment diagram illustrating on the one hand, the communication between several components and on the other hand the physical arrangement of hardware resources that make up the system by showing the distribution of the components of these materials.
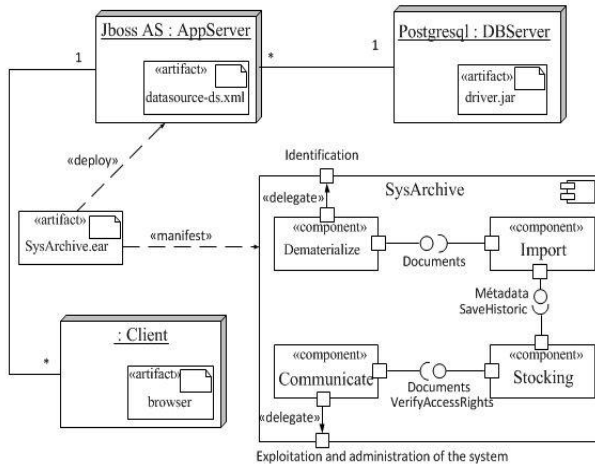
Fig 5: Deployment diagram

5. CONCLUSION

The multi-agent systems have for scope of application, among many others, problem solving in the broadest sense where software agents perform useful tasks for humans. The process of archiving documents requires a great human resource well arranged hierarchically. However, the problem of interpersonal communication is common in most organizations, both public and private. Electronic archiving has overcome this problem. Since an electronic archiving system is a dynamic and complex environment, its modeling requires the use of appropriate tools. Therefore, we propose an architecture of EAS based on MAS to benefit from its properties and that possessed by the agents. We proposed an approach which from the base integrates agents. Within the context of implementation, an implementation of the communicational model based on message exchange has been proposed. Agents interact through messages and have each a message box where are deposited messages that are intended to it. A deployment diagram describing the physical arrangement of hardware resources that make up the system and showing the distribution of these components on different resources was also proposed.

6. REFERENCES

[1] Marie-Anne, C. 2010. New Glossary of Archiving.

[2] Federation National des Tiers de Confiance (FNTC). 2010. Guide de l'archivage électronique et du Coffre-fort électronique. Collection Les Guides de la Confiance de la FNTC.

[3] Marie-Anne, C., Eric, C. and Jean-Marc, R. 2005. Electronic archiving at the use of the leader. CIGREF – FEDISA.

[4] Bureau conseil de la DCSSI. 2006. Archivage électronique sécurisé - Etat de l'art. République Française.

[5] Murielle, S. 2009. State of the art electronic archiving facing its implementation. Training course report for Masters.

[6] Ludivine, M. 2010. How to choose a solution of archives management: the case of the Urban Community of the Porte du Hainaut. Master's dissertation.

[7] Guyon, C. CG10. Conduire un projet d'archivage électronique. L'expérience du Conseil général de l'Aube.

[8] The rights forum on the internet. 2005. The electronic storage of documents.

[9] Eric, M. 2012. Electronic archiving system (EAS) requirements and specifications that reflect an environment [EDM-EAS(EAP)-AMSP]. Archivists – Experts.

[10] Payan, C. et al. 2009. Guide pratique – Comprendre les normes dans le domaine de l'archivage numérique. FedISA.

[11] Jacques, F. 1995. Multi-Agent Systems Towards a collective intelligence. InterEditions.

[12] Arnaud, G. 2011. Study of self-organization in multi-agent patrol algorithms based on digital pheromones. Ph.D. Thesis., University of Nancy 2.

[13] Fabien, M. 2004. Formalism, tools and methodological elements for modeling and multi-agent simulation. Ph.D. Thesis., University of Montpellier II.

[14] Christine, B. 2011. Systèmes Multi-Agents : Modélisation et simulation informatique de comportements collectifs. Université de Nancy.

[15] Elie, T. F., Emmanuel, T., César, V. and Alain, A. 2005. A Model for the Simulation of Complex Phenomenon: The Development of a Telecommunication System, IEEE, SITIS'05, pp. 2-7.