## Abstract

The thesis describes an agent-based model aimed to simulate those processes in which a digital object faces the risk of obsolescence, a migration process has to be performed and the most appropriate file format has to be adopted. Agents have been designed in order to monitor and control the local system where they reside and its environment. They are able to become aware of obsolescent formats based on global parameters such as their diffusion. They communicate as well with each other to find out the most suitable preservation action to be performed. Agents request suggestions that are evaluated and propagated according to a weighting based on the level of trust assigned to both the agents who identified the problem and proposed the solution. In the current research, the definition of the trust level has been chosen based on the cultural and geographical distances, the expertise of the involved agents and the file format numerosity. The level of trust between two agents is automatically updated after every interaction by the mean of a feedback mechanism profiting of an inter agent communication based on stigmergy. Summing up, the thesis demonstrates how a multi-agent system can either perform an autonomous preservation action or suggest a list of best candidate solutions to the user. It benefits the management of several kinds of digital archive, especially those with limited resources specifically dedicated to digital preservation, such as small personal collections and many public institutions.

## Digital Preservation

Digital preservation is a fundamental issue for those who have to protect and keep accessible any kind of cultural heritage. Both the software environment and the digital object may become obsolete: this process is named “digital obsolescence”. The risk of obsolescence can be estimated from the global environment thus digital preservation cannot be viewed as a local or personal issue anymore, it is instead a collective and distributed concept. There are two main strategies to cope with digital obsolescence: migration and emulation, migration is the process we are going to take into account. This work shows a model carrying out autonomous evaluation of migration processes by means of a multi-agent system.

One of the first efforts to contrast digital obsolescence has been made in 1994 by the Research Libraries Group and the Commission of Preservation and Access that two years later published one of the most important documents about digital preservation: “Preserving digital information. Report of the task force on archiving of digital information”. Based on this report, a reference model named OAIS (Open Archival Information System) was developed. It discusses the

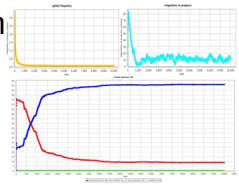
concept of long-term digital preservation and aims to point out the various stages of the life cycle of a digital object and of the related preservation process. Obsolescence identification and metadata extraction are two fundamental processes to elaborate a proper preservation strategy. AONS (Automated Obsolescence Notification System) and AONS II are examples of a system that analyses repositories and detects eventual objects in danger of becoming obsolescent. They take advantage of information about the format which is one of the most relevant metadata. The information is recovered by means of DROID (Digital Record and Object Identification) and JHOVE (JSTOR/Harvard Object Validation Environment), a couple of tools for the format extraction and validation.

Methodology – ABM and MAS

An agent can be defined as a computer system that is situated in some environment, and that is capable of autonomous actions in order to meet its design objectives. The environment and the agents within it are a multi-agent system. In such a system the agents interact with each other and with the environment. In order to satisfy their objectives agents then have to communicate and cooperate to find agreements about which action to perform. An agent-based model could be appropriate to implement the distributed intelligence needed to deal with digital preservation issues. The agents will acquire, evaluate and share a certain set of information in order to understand how serious the risk of obsolescence is and which is the best preservation action to perform.

Model Implementation and  Prototype Design

The model aims to emulate a distributed environment where many archive entities share information about their internal state in order to find solutions to their digital preservation issues. In our im



plementation these archivers are agents named “institutions”, these agent species contain other species inside: four agents named “pastors” that manage the digital objects and a “software manager” as regards the applications. This choice does not represent a constraint since the model can be easily adapted to handle more digital object categories.

The software environment adopted to implement the agent-based model described in the present work is GAMA (Gis & Agent-based Modeling Architecture) version 1.6. In general terms a GAML model is made up by a certain number of actions which consist of a sequence of statements. Each model is made up of a global part that includes all those variables accessible to every agent or that are necessary to the model initialization. The second part, named entities, contains the declarations of all the species of agents that take part in the model. The last part, named experiment is dedicated to the experimental setup.

We decided to verify the stability of the model by monitoring the trend of the number of migrations performed by the agents, the number of migrations in progress and the percentage of both correct and wrong choices. The results in Fig. 1 support the view that the behaviour of the framework is stable and that the goodness of the agents' choices increases exponentially with time.

Conclusions and Future Works

The model discussed in this thesis allows to probe the effects on the stability of a complex distributed consortium of institutional and personal digital libraries due to: different communication protocols and different assumptions to establish reciprocal trust. It is basically a research framework in which different approaches and assumptions can be tested.

Nowadays, the maze of digital formats represents a real issue for our cultural heritage that cannot be addressed locally: a global approach is needed in order to be able to capture immediately every signal notifying, or even suggesting, an obsolescence threat. In the next future different assumptions for the trust scheme, the key parameters and the conditions triggering agents' reaction will be investigated, in order to verify if an automated agent-based world-wide interconnected infrastructure is able to address the digital obsolescence with the needed stability. The statistical analysis of the experimental results lets us affirm that the frequency of migrations faces an exponential decrease and approaches an asymptotic value, the number of migrations in progress becomes constant. Most important the percentage of correct actions increases exponentially with time while the percentage of wrong actions decreases with the same trend. That means that agents benefit from the interaction since they learn how to deal with their preservation issues. In addition, the percentage of lost migrations is equal to 0 so, in the worst case, agents perform a migration that was not completely necessary but they never miss a fundamental one.

For the future it is possible to think of digital objects designed as intelligent agents themselves. In such a scenario each digital object would be able to recognize itself as obsolescent and suggest an appropriate preservation action according to the resources available in the environment. An application will be implemented and tested on a real network in order to verify the effectiveness of the preservation framework presented in this research.

**Jacopo Pellegrino**

**Tesi di Laurea Specialistica**

Autore: Jacopo Pellegrino
Relatore: Marco Maggiora, email: marco.maggiora@unito.it
Presidente della Commissione: Wanda Maria  Alberico
Università: Università degli Studi di Torino
Facoltà: Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso: Laurea Spec. in Fisica delle Tecnologie Avanzate
Data di Discussione: 21/07/2014
Voto: 110
Disciplina: Modellizzazione ad agenti
Tipo di Tesi: Sperimentale
Anno di Iscrizione: 2012-2013
Altri Autori:
Altri Relatori: Walter Allasia, Seamus Ross, Pietro Terna, Daniel Teruggi
Lingua: Inglese
Grande Area: Area Scientifica
Dignità di Stampa: Si
In Collaborazione con: Eurix
Settori Interessati: Archivi digitali, Biblioteche, Broadcasters, Musei, Digital Repositories, Università, Esperimenti Scientifici

**Pubblicata in**: www.pubblitesi.it