Time Machine: Projecting the Digital Assets onto the Future Simulation Environment

Jose Antonio Olvera(✉) and Josep Lluis de la Rosa

TECNIO - Centre EASY, Agents Research Lab, VICOROB Institute,
Univ. of Girona, 17071 Girona, Spain
joseantonio.olvera@udg.edu, peplluis@eia.udg.edu

Abstract. In this paper we present the Time Machine (TiM), an environment that simulates the digital assets onto the future and allows studying what self-preservation behaviors need the digital objects. This is based in computation intelligence and related methods of cost management under their own budget, powered by a social network as an environment that enables their behavior under the policy that preservation is to share. This approach contributes to achieve the following digital preservation requirements: adaption to unexpected situations, scalability, and efficient cost management, through an agent-based simulation. The key differentiation feature of TiM is that digital objects become active actors in their long term digital preservation, which has a digital preservation budget devoted to funding the replication of the objects and other operations such as format migration or finding a safe storage within a social network of users; in all, an environment where they will live. Its design considerations and implementation details are presented and, finally an example to illustrate some of the functionalities of the simulator.

Keywords: Multi-agent system · Simulation · Digital preservation · Computational intelligence

1 Introduction

The need for research into new ways of long term digital preservation (LTDP) has been evident over two decades. Memory institutions and academic organizations have taken the lead in defining solutions to LTDP issues; although the digital preservation (DP) problem is not limited to cultural and scientific information. Legislation is placing an increasingly large burden on commercial, industrial, cultural and governmental organizations to ensure that their digital records are retained, made accessible, and effectively preserved for the long term. The PLANETS project estimated that the value of digital documents produced in the EU that are in danger of digital obsolescence if no action is taken to preserve them, is in excess of 3 billion € per year[1].

The challenge in LTDP of complex objects – consisting of text, video, images, music, 3D information, sensor data, etc. generated throughout all areas of our society – is

[1] http://www.planets-project.eu

Y. Demazeau et al. (Eds.): PAAMS 2015, LNAI 9086, pp. 175–186, 2015.
DOI: 10.1007/978-3-319-18944-4_15

real and growing at an exponential pace. An already old study by the International Data Corporation (IDC) found that in 2012 the information created and replicated broke the zettabyte barrier growing by a factor of 9 in just five years [10]. The LTDP of such information will become a pervasive as well as ubiquitous problem that will concern everyone who has digital information to be kept for long time, implying a shift in at least a couple of software and hardware generations.

Currently the level of automation in DP solutions is low. The preservation process currently involves many manual stages but should be approached in a flexible and distributed way, combining intelligent automated methods with human intervention. The scalability of existing preservation solutions has been poorly demonstrated; and solutions have often not been properly tested against diverse digital resources or in heterogeneous environments [20].

We try to meet the following requirements involved in DP:

- Scalability: The exponential growth of digitally born objects requires of scalable solutions from the technological point of view.
- Cost: Associated to the exponential growth because there are limited resources to cope with DP.
- Uncertain future: DP is about heuristics of what results we will get in the future, only.

These problems, together with the rapid obsolescence of software and hardware due to frequent update of private vendors, make DP one of the most challenging application areas for MAS.

As was explained at [18], the prevailing paradigm is centralized, top-down, where institutions are the main players. We propose studying a change of paradigm, mainly bottom-up, where the digital objects *self-preserve*.

Our research is based in studying what self-preservation behaviors need the digital objects (DOs), based in computation intelligence (CI), and related methods of cost management under their own budget, powered by a social network as an environment that enables their behavior under the policy that “preservation is to share”. It means that the more you share your DOs, the most likely you will keep them in the future: one computer or server can fail but does not matter because your DO is spread around the world. In this concept, DOs become active actors in their own LTDP, here named the Self-Preserving Digital Object (SPDO) [5], which has a DP budget devoted to funding the replication of the objects and other operations such as format migration or finding a safe storage within a social network of users; in all, a controlled environment where they will “live”.

The principle behind is to think of a SPDO as an autonomous entity that interacts with the environment as well as with other SPDOs with which it competes or cooperate and thus creates good living conditions to keep alive as long as possible to have descendants who follow its existence. Under this expectation, the research in this topic will pursue self-organization paradigms, including multi-agent systems (MAS) with interesting emergent behaviors. In this paradigm, three approaches for agentification has been proposed [17]: the SPDOs; the collective cognitive networks on DP: the Sites; and the DP Services [23]. We agentified the first two: the SPDO and Sites.

With the purpose of providing the scientific community with a new tool suitable to carry out analysis of novel digital preservation strategies, in this paper we present the product of an iterative work carried out in the last years, the Time Machine (TiM), a simulation environment for the SPDO, written in Java, and implemented with the purpose of:

- Studying what are the social behaviors that supports the task of preservation of the DOs (rules of collaboration),
- Finding out an efficient cost management of the DOs,
- Analyzing what are the topologies of social networks that back better the DOs.

As a consequence, TiM is a more abstract preservation simulator than other similar-purpose tools. Our simulator does not reach the level of detail of other preservation simulators (i.e. bit preservation, ingest or access). The paper is structured as follows: section 2 presents related work and the differentiation of the current work regarding the literature; section 3 are the design considerations and implementation of the agent-based simulator by presenting the most relevant diagrams; section 4 shows the TiM simulating a concrete case, presenting the Graphical User Interface (GUI) of the simulator and the obtained results; section 5 presents where the digital preservation simulation tool is accessible; finally section 6 is devoted to conclusions and future research.

2 Related Work

There are works that use MAS with the aim to simulate the processes involved in the DP. For instance, the EU FP7 PROTAGE project [23] opened up a novel approach to DP by utilizing agent ecosystems for automation of preservation processes. The dissemination results of the project are [13][14] [15]. Another approach named "Shout and Act" [7], is a type of swarm intelligence for communication and coordination of agents inspired by rescue robots: the files, all digital objects (DOs), that need preservation are called the “victims”, and there are teams of preservation agents, whose main goal is to detect files as potential victims that need migration actions and curate them. In addition, a recent work of Pellegrino [12] demonstrates how MAS can either perform an autonomous preservation action or suggest a list of best candidate solutions to the user. It describes an agent-based model aimed to simulate those processes in which a DO faces the risk of obsolescence, a migration process has to be performed and the most appropriate file format has to be adopted.

All the aforementioned works have in common with the current work that use MAS for automation of preservation processes, but none of them agentifies the DO themselves. We claim that, although preservation is currently perceived as being as repository level, in fact it occurs at object level, so mechanisms are required to enable preservation management of objects, and for this to succeed, objects must be self-preserving. Very few works are conducted in this line, among them our previous work [5] and a remarkable recent study with a data-centric perspective through the use of Unsupervised Small World (USW) graph creation algorithm [2], but they focus on

Web Objects (WO), they do only copies but not migrations, as we do, and their networks is of WO whereas our network is composed of sites of users; both of them, have a promising level of success.

As well, there exist works that concretely use simulation principles to analyze digital repositories. In [4], authors presented a simulation tool (ArchSim) able to evaluate a repository system implementation. The focus is on comparing different options such as disk reliability, error detection and preventive maintenance to estimate the Mean Time to a Failure (MTTF) of a whole repository. Similar work was done in [3] where a simulation tool was built to evaluate potential design of the Danish web archive. More recent work was conducted during the PrestoPrime project[2]. In [1] authors describe a tool able to simulate the costs and risks of using IT storage systems for the long-term archiving. The tool is now known as iModel[3]. The main focus of these simulation tools is "bit preservation"; there is a limited support for "logical preservation".

Regarding Rabinovici-Cohen et al. [21], "Bit preservation is the ability to retrieve the bits in the face of physical media degradation or obsolescence, corruption or destruction due to errors or malicious attacks, or even environmental catastrophes such as fire and flooding. Logical preservation involves preserving the understandability and usability of the data, despite unforeseeable changes that will take place in servers, operating systems, data management products, applications and even users. Additionally, logical preservation needs to maintain the provenance of the data, along with its authenticity and integrity so that current and future systems can ensure that only legitimate users access that data".

Within this approach of logical preservation there is the SCAPE simulator [8] that is an environment with a focus on logical preservation, such as migrations, format evolutions, ingest and potential collection change over time, and is also considering aspects such as resources. Our work is more similar to the SCAPE simulator than others explained before. We focus on logical preservation, yet with our goal of not trying to analyze digital repositories as a difference; in our model, the environment is a social network devoted to the DP.

3 Design Considerations and Implementation

There are different digital preservation approaches. Regarding Feeney [9], "*Three approaches to digital preservation have been developed:*

- *Preserve the original software (and possible hardware) that was used to create and access the information. This is known as the technology preservation strategy. It also involves preserving both the original operating system and hardware on which to run it.*
- *Program future powerful computer systems to emulate older, obsolete computer platforms and operating systems as required. This is the technology emulation strategy.*

[2] http://www.prestoprime.org/

[3] http://prestoprime.it-innovation.soton.ac.uk/imodel/download/

- *Ensure that the digital information is re-encoded in new formats before the old format becomes obsolete. This is the digital information migration strategy."*

We focus on the third one to simulate DP activities of refreshing and migration while SPDOs sharing using different computational intelligence methodologies. We conceived the migration of SPDO formats as a replication (copies) of files in different formats and refreshing as a copy in the same format [11][18]. The SPDOs will be distributed over a network of computers or devices as an environment that enables the behavior of the SPDOs in an attempt to preserve them.

Migrated copies of SPDOs are created in various formats following a migration strategy to ensure their survival against Software Adoption Waves (SAW), which occur regularly. SAW are defined as massive format changes that DOs suffer throughout their lifetime after the shifts of software and hardware, resulting in the likely "disappearance" of a percentage of the SPDOs when they become unreadable or inaccessible due to the accumulation of technological changes that provokes their obsolescence.

In our model, a network with nodes represents the users' computers (the sites) and the connections among nodes determine which users are friends with who else, resulting in a social network devoted to the DP[6][13].

In all the experiments, DOs travel through the network that is build up out of a social network and distribute copies of them for preservation. These copies maintain links to the parent SPDO because they correspond to a same object. Each node representing a user's computer might only be able to read a specific format of files after suffering a SAW. Formats range from *oldest* to *newest* to simulate the processing of files in a computer when these files can only be read by special software installed in the computer (the site). The files in older formats become *unreadable* when they are no longer compatible with the new software versions installed on a given computer after several migrations. That formats are taken as a representation of pre-web and non-web formats, because there is no danger of forgetting how to read a file format that was in use after web became available, it exist a migration path forward, as David Rosenthal argues[4]. Further details of the model are provided in the following sections.

3.1 The SPDOs Agents

The SPDOs in our experiments are different file types in several formats. For the sake of simplicity and abstraction their representation ranges from 1, the newest, to 5, the oldest.

Initially, a SPDO belongs to a particular user, and lives in his filesystem, i.e., on his computer, a site that is a node in the network structure. Any SPDO has the following behaviors with subtle variations depending upon the CI implemented: make a copy, move on to other nodes in the network, or remain in the site where it is. The aim of the SPDOs and their descendants is traveling from sites all over the net, mak-

4 http://blog.dshr.org/2013/02/rothenberg-still-wrong.html

ing digital copies of themselves when they can afford to, accordingly the restrictions of the CI algorithm that is applied. The cost for a site to host a SPDO, is directly related with the concrete format of the SPDO: the older, the more expensive (harder to have an adequate software to host it) than if it was newer. Finally, the SPDOs must determine whether the original SPDO or any of its copies are alive in the network.

The following describes the three CI algorithms that have been implemented over the SPDO behaviors, which are not detailed because they are out of the scope of this paper:

- Multi-Population Genetic Algorithm (MPGA),
- Ant Colony Algorithm (ACA),
- and Virus-Based Algorithm (VBA), approach that we implanted it similar to a computer *worm*. SPDOs try to move through the entire network selecting connections randomly but attempting to spread as far away as possible replicating themselves, with some limitations for not collapsing the system.

3.2 The Agents Site

The users have preservation services installed in their own computers for dealing, at some extent, with DP of the SPDOs hosted there. The services are limited in nature and it is represented by a percentage of employment (0%-100%) that describes the capacity of DOs that can be hosted in the user's site and a data structure that indicates the social proximity of several other nodes as users who may be available to help sharing DP efforts, for example, by sharing their SPDOs. The users' computers are connected each other for fulfilling the social and intelligent environment to support the LTDP of the SPDOs. The computers will be named nodes after the creation of a full network of LTDP resources available for SPDOs. Finally, the user's node clearly displays what formats of each file type are supported. All users have a contact list that is represented by nodes (other users' nodes) to which they are connected. These nodes are their friends' computers, resulting in a network through which SPDOs can move onto for new LTDP services, hosting, copies support, or do any other operation. Users can update the formats of their SPDOs when they change format support at the time of their software updates. We will refer to the users' computers as nodes or sites indistinctively.

3.3 Integration of the Agents in the Platform

Fig. 1 shows the framework that has been applied in the presented MAS. In this framework, the communications of the agents are performed through Java methods. The Simulation class can be executed given a certain number of times or a percentage of stability (explained in next section), and contains all the agents within a list. All the agents have a main method, in which they perform their activities in every step of simulation. Finally the platform can be executed via GUI or via console and it has access to the statistics as well agents and Simulation class can save them though the Statistics class.

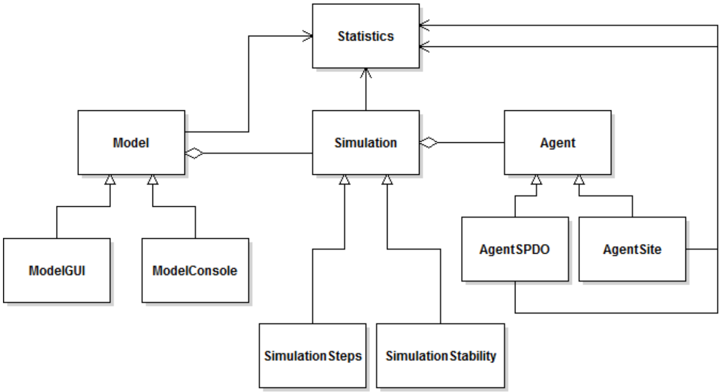

Fig. 1. Excerpt of the simulation framework

4 Simulation Experimentation of Time Machine

4.1 The Measure of Expected Resilience and the Reliability of the Experiments

If we subscribe that *preservation is about sharing*, and knowing that SPDOs when shared might be object of LTDP manipulations, notably migrations for enhancing their *preservability*; here we come with a growing diversity of objects that might contribute with higher resilience in the long-term after suffering several SAWs.

Thus, after acknowledging that these heuristics will be applied to our SPDOs, we need measures of expected resilience that might be explained by the diversity of SPDOs. To do so we use the Shannon entropy (Equation 1) for the evaluation of the results. High entropy indicates high expected *preservation* and resilience after successive SAWs according to our strategy for format migrations. The fact that several copies exist in diversified formats provides better resistance and recovery in the case of a SAW (improved *resilience*).

$$H(x) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i \qquad (1)$$

Equation 2 is the DP application of Equation 1, where n is the total number of original SPDOs, j are the formats SPDOs can migrate from and to (with a value ranging from 1 to 5), and *pi,j* is the percentage of the copies in format j among the total number of copies of an original SPDO i. Equation 3 defines *Pi,j*, where k are the formats from 1 to 5.

$$H(x) = \frac{-\sum_{i=1}^{n}\left(\sum_{j=1}^{5} p_{i,j} \cdot \log_2 p_{i,j}\right)}{n} \qquad (2)$$

$$p_{i,j} = \frac{f_{i,j}}{\sum_{k=1}^{5} f_{i,k}} \tag{3}$$

For the reliability of the experiments, the average resulting entropy is calculated after the number of executions is enough to achieve a stabilized result. We measured this reliability of the experiments using Equation 4, which defines $\overline{x_n}$ as the average of the steady entropy value x_n in execution *n*, such that the sum converges when an increasing number of experiments to yield a reliability of 1 – ε = 99.3%.

$$S_n = \sum_{i=1}^{n} \left(\overline{x_n} - \overline{x_{n-1}} \right) < \varepsilon \tag{4}$$

4.2 Brief Overview of the Platform Functionality

The multi-agent platform can be executed via GUI or via console and it allows to generate the desired topology of social network or choose some defined by default, to set up the different parameters that act in the simulations, and to see (via GUI) a screen that displays everything that is happening in real time (shown in Fig. 2). The executions via console will get all the input parameters from a file.

Fig. 2. General view of what is happening in real time

As a result of the simulation, a number of useful statistics are saved in files for future analysis:

- The entropy of each execution and the average entropy of all executions;

- The percentage of accessible SPDOs in each execution and the average accessible SPDOs in all executions – accessible SPDOs means that at the end of the executions maintain at least one copy of themselves ;
- The percentage of readable SPDOs in each execution and the average readable SPDOs in all executions – readable SPDOs means that have a copy with a format compatible of the site they live (they are immediately readable in that node and thus technically straightforward preserved);
- The percentage of occupation of the sites in each execution and the average percentage of occupation in all executions;
- The number of copies of each SPDO in different formats for each execution and the average in all executions;
- The parameters used in the simulation.

4.3 Examples of Experiments

TiM simulations are deployed over 35-year period in which a SAW occurs every 5 years. After every SAW it is the time at which SPDOs show severe symptoms of obsolescence and an urgent need of DP or curation. We use the month as the temporal unit at every step for our simulations, and thus, there is a SAW every 60 steps (12 months × 5 years), and an execution has a total length of 420 steps (12 months x 35 years). The three CI algorithms as a SPDO behaviors have been simulated and Table 1 shows all the parameters of the simulations.

Table 1. Formulation of the parameters used for the TiM experiment

Parameters	MPGA	ACA	VBA
Total simulation time	35 years		
Equivalence of 1 simulation step	1 month		
Total waves	6 (for 35 years)		
Percentage of DOs involved in a SAW	72%		
# DOs associated to any user	Between 1 and 5		
Preservation service for each user	100%		
Percentage of statistical stability	99.3%		
Random seed	123456789		
Percentage of mutation	2% of the population		
Percentage of exchange	2% of the population		
Percentage of nodes of the network where ants can make copies		20%	
Period in which nodes are changed where ants can make copies		3 years	
Maximum number of copies that a DO can have at each site			1

The function watts.strogatz.game() from R package igraph was used to create the small-world graph of 1000 nodes. The graph was checked to ensure that it was simple and connected. The clustering coefficient (CC) resulting from the graph is 0.48 and the average path length (APL) is 4.41. It was added as default network that can be chosen with the simulator.

The results of the experiments are shown in Fig. 3, that show the evolution of the average entropy over the simulation period, where the X axis are the steps of the simulation (months) and the Y axis is the entropy value, explained above. Looking the results obtained with this experiments, in the end of the simulation the highest entropy value is obtained with the ACA, but the period before is also of the same importance (SPDO must be accessible and reproducible by users at any time) and in that case both VBA and MPGA are better solutions than ACA. In this case is difficult to select the best because the three executions have positive entropy values during all the simulation, but the one that have higher entropy as a mean is VBA (high entropy value indicates high expected preservation).

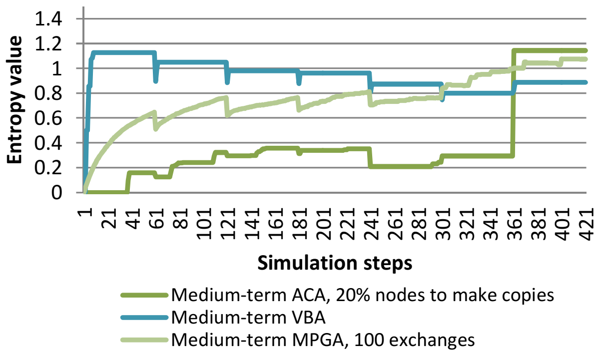

Fig. 3. Comparison of CI algorithms over 35-year period

5 Availability

The software is publicly available at http://timemachine.jaolvera.test.easyinnova.com, along with the documentation, some examples and the source code that can be downloaded. We expect to enrich the site with more source code examples, and illustrative simulation scenarios. Besides, as contributions from the community are welcomed and desired, we plan to encourage the development of new components that can be kept in a public repository.

In this paper we have described the main functionalities of TiM; others have not been mentioned, for example those pertaining to electronic auctions of preservation and curation services for these objects to be preserved, explained at [19], as well as using real social network topologies and the interaction between the users obtained from Facebook, Google+ and Twitter developed in [16]. There are initial classes that implement them, but further development is anticipated in those areas.

6 Conclusions and Future Work

In this paper we have presented the Time Machine (TiM), a simulation environment that projects the digital assets onto the future and allows studying what self-preservation behaviors need the digital objects, based in computation intelligence, and related methods of cost management under their own budget, powered by a social network as an environment that enables their behavior under the policy that preservation is to share. The key differentiation feature of TiM is that digital objects become active actors in their own long term digital preservation. The related work provides several MAS with the aim to simulate the processes involved in the digital preservation as well as well-known digital repositories simulators. Then, TiM design considerations and implementation details are presented. Additionally, an example of experimentation illustrates some of the functionalities of our proposal.

Future work will follow in the direction of studying variations of the CI algorithms that we performed, adding electronic auctions of preservation and curation services for these objects to be preserved and use real social network topologies and make the topology varying with time to have more realistic simulation.

We hope that TiM evolves with the addition of new functionalities both by the authors and the contributions of other developers and that this software becomes a helpful tool for the digital preservation research community.

Acknowledgements. This research is partly funded by the TIN2013-48040-R (QWAVES) Nuevos métodos de automatización de la búsqueda social basados en waves de preguntas, the IPT20120482430000 (MIDPOINT) Nuevos enfoques de preservación digital con mejor gestión de costes que garantizan su sostenibilidad, and VISUAL AD Uso de la Red Social para Monetizar el Contenido Visual, RTC-2014-2566-7 and GEPID Gamificación en la Preservación Digital desplegada sobre las Redes Sociales, RTC-2014-2576-7, the EU DURAFILE num. 605356, FP7-SME-2013, BSG-SME (Research for SMEs) Innovative Digital Preservation using Social Search in Agent Environments, as well as the AGAUR 2012 FI_B00927 awarded to José Antonio Olvera and the grup de recerca consolidat CSI-ref. 2014 SGR 1469.

References

1. Addis, M., Jacyno, M.: D2.1.2 Tools for modeling and simulating migration-based preservation (2010)
2. Cartledge C.L., Nelson M.L.: When should I make preservation copies of myself? In: Digital Libraries 2014, London, UK, September 8-12 (2014)
3. Christensen, N.H.: Preserving the bits of the Danish Internet. In: 5th International Web Archiving Workshop, Vienna (2005)
4. Crespo, A., Garcia-Molina, H.: Modeling archival repositories for digital libraries. In: Borbinha, J.L., Baker, T. (eds.) ECDL 2000. LNCS, vol. 1923, pp. 190–205. Springer, Heidelberg (2000)
5. de la Rosa J.L., Olvera J.A.: First studies on self-preserving digital objects. In: Artificial Intelligence Research & Dev., Procs 15th Intl. Conf. of the Catalan Assoc. for Artificial Intelligence, CCIA 2012, Alacant, Spain, vol. 248, pp. 213–222 (2012)

6. de la Rosa, J.L., Trias, A., Ruusalepp, R., Aas, F., Moreno, A., Roura, E., Bres, A., Bosch, T.: Agents for social search in long-term digital preservation. In: The Sixth International Conference on Semantics, Knowledge and Grid, SKG 2010, Ningbo, China, November 1-3 (2010)
7. de la Rosa, J.L., Trias, A., Aciar, S., del Acebo, E., Quisbert, H.: Shout and act: An algorithm for digital objects preservation inspired from rescue robots. In: Proceedings of the 1st Intl Work. on Innovation in Digital Preservation, Austin, Texas, June 19 (2009)
8. Duretec, K.: Final version of the simulation environment.SCAPE. http://www.scape-project.eu/wp-content/uploads/2014/10/SCAPE_D12.3_TUW_V1.0.pdf
9. Feeney, M.: Towards a national strategy for archiving digital materials. Alexandria **11**(2), 107–121 (1999)
10. Gantz, J., Reisel, D.: The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, December 2012
11. Garrett, J., Waters, D, (eds.) Preserving Digital Information, Report of the Task Force on Archiving of Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group (1996)
12. Jacopo, P.: A Multi-agent Based Digital Preservation Model, Master's Degree Thesis, Department of Physics, University of Turin. (2014)
13. Jin, X., Jiang, J., de la Rosa J.L: PROTAGE: Long-Term Digital Preservation Based on Intelligent Agents and Web Services. ERCIM News **80**, 15–16 (2010)
14. Jin, X., Jiang, J., Min, G., de la Rosa J.L., Quisbert, H., Bengtsson, J.E, Ruusalepp, R.: Towards Computerized Digital Preservation via Intelligent Agent Based Approaches. International Journal of Digital Curation (2011)
15. Jin, X., Jiang, J, Min: A software agent and Web service based system for digital preservation. First Monday, [S.l.], October. 2010. ISSN 13960466
16. McAuley, J., Leskovec, J.: Learning to Discover Social Circles in Ego Networks. In: NIPS (2012)
17. Olvera, J.A., de la Rosa i Esteva, J.L.: An Outline of the Application of Agents to Digital Preservation and an Introduction to Self-Preservation Aware Digital Objects. In: EASSS 2011, Girona, Catalonia, Spain, July 11-15 (2011)
18. Olvera, J.A.: Digital Preservation: A New Approach from Computational Intelligence. In: Joint Conference on Digital Libraries 2013, JCDL Doctoral Consortium 2013, Indianapolis, Indiana, USA, July 22-26 (2014)
19. Olvera, J.A., Carrillo, P.N., de la Rosa, J.L.: Combinatorial and multi-unit auctions applied to digital preservation of self-preserving objects. In: Artificial Intelligence Research and Development - Recent Advances and Applications, vol. 269, pp. 265–268 (2014)
20. Quisbert, H.: On Long-term Digital Information Preservation Systems – a Framework and Characteristics for Development. Ph D Thesis. Luleå University of Technology (2008)
21. Rabinovici-Cohen, S., Baker, M.G., Cummings, R., Fineberg, S., Marberg, J.: Towards SIRF: self-contained information retention format. In: Proceedings of the 4th Annual International Conference on Systems and Storage (SYSTOR 2011), Article 15, 10 pages. ACM, New York (2011). http://doi.acm.org/10.1145/1987816.1987836
22. Rothenberg J.: Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources (1999)
23. Ruusalep R., Dobreva M.: Digital Preservation Services: State of the Art Analysis. www.dc-net.org/getFile.php?id=467
24. The PROTAGE Project: PReservation Organizations using Tools in AGent Environments.