Project Title:

INNOVATIVE DIGITAL PRESERVATION USING SOCIAL SEARCH IN AGENT ENVIRONMENTS

Project Acronym: DURAFILE

Contract Number: **605356**

D1.1 State of the art in digital preservation and multi-agent systems

Dissemination Level: **PUBLIC**

Lead Beneficiary: Tallinn University (TU)

Submission date: December 2nd, 2013

Document revision history

Version	Date	Comments
1.0	15/11/2013	Josep Lluis de la Rosa (UdG), Rando Rostok (TU), Raivo Ruusalepp (TU)
2.0	02/12/2013	Josep Lluis de la Rosa (UdG), Rando Rostok (TU), Raivo Ruusalepp (TU), Santiago Macho (ATEKNEA)

Table of Contents

1. Executive Summary

This report is an early deliverable from the DURAFILE project that will inform the project through an overview of the state of the art in both digital preservation and intelligent software agents' domain. The report builds on the overview of the state of the art that was provided in the proposal of the DURAFILE project and updates it in several aspects.

The deliverable sets out the intellectual framework for commencing the work in the DURAFILE project. It describes the external trends that the project can benefit from and, indeed, can build upon. It starts by summarising the state of play with software agents and digital preservation at the end of the last project that combined these two technologies – PROTAGE.[1] It then provides an overview of the last trends in the digital preservation domain and a short update on developments in the intelligent multiagent systems' area. Chapter 6 reviews recent projects that have combined the two technologies and looks at their market potential. The report concludes with recommendations for the project.

Digital Preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. It includes all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organization; "born-digital" materials created for a specific purpose (e.g. teaching resources); or the products of digitization projects. Depending on time[2], Digital Preservation can be divided into:

- **Long-term preservation** - Continued access to digital materials, or at least to the information contained in them, indefinitely.
- **Medium-term preservation** - Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely.
- **Short-term preservation** - Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.

This project will be focused to long-term preservation of multimedia objects.

This deliverable is primarily targeting the DURAFILE project partners, but will provide a useful update for anyone who is interested in the use of intelligent software agents in digital preservation. It is distributed as follows. The document begins with highlighting the concept of Digital Preservation including a justification of its necessity. The core of the document consists in the state of the art of the software agents and trends in Digital Preservation. A section on trends in Intelligent Agents software technologies including a study of the potential of these agents for Digital Preservation solutions follows the Deliverable. Section 6 overviews past research and commercial projects using Intelligent Agents for Digital Preservation solutions. An interesting analysis of market potential is described in section 7. The document finishes with the conclusions and all the recommendations gained from the state of the art in Digital Preservation and Intelligent Agents which will be used during the development of the DURAFILE platform.

[1] http://www.ra.ee/protage
[2] http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts

2. Digital Preservation

Digital preservation is defined by the DigitalPreservationEurope project as "a set of activities required to make sure digital objects can be located, rendered, used and understood in the future".[3] A more comprehensive term 'digital curation' is often used in parallel with digital preservation. It has wider coverage and involves "maintaining, preserving and adding value to digital data throughout its life-cycle".[4] The key challenge in preserving usability of digital objects over time is overcoming technology obsolescence but a set of other issues around managing collections of digital objects is also involved.

Economically sustainable preservation — ensuring the ongoing and efficient allocation of resources to digital preservation — is an urgent societal problem. It is urgent because digital information is inherently fragile, prone to information loss and degradation. Preservation insures against multiple risks to information assets over time. Such assets must be actively managed for sustained periods of time, using best practices for data stewardship across the full lifecycle of creation, description and curation, deposit in secure storage, use, and reuse. Some digital materials require relatively intensive levels of preservation to ensure usability, and others much less. But in all cases, access to information tomorrow depends on preservation actions taken today. A fundamental fact of digital sustainability is that without preservation, there is no access. (BRTF 2010, p. 9)

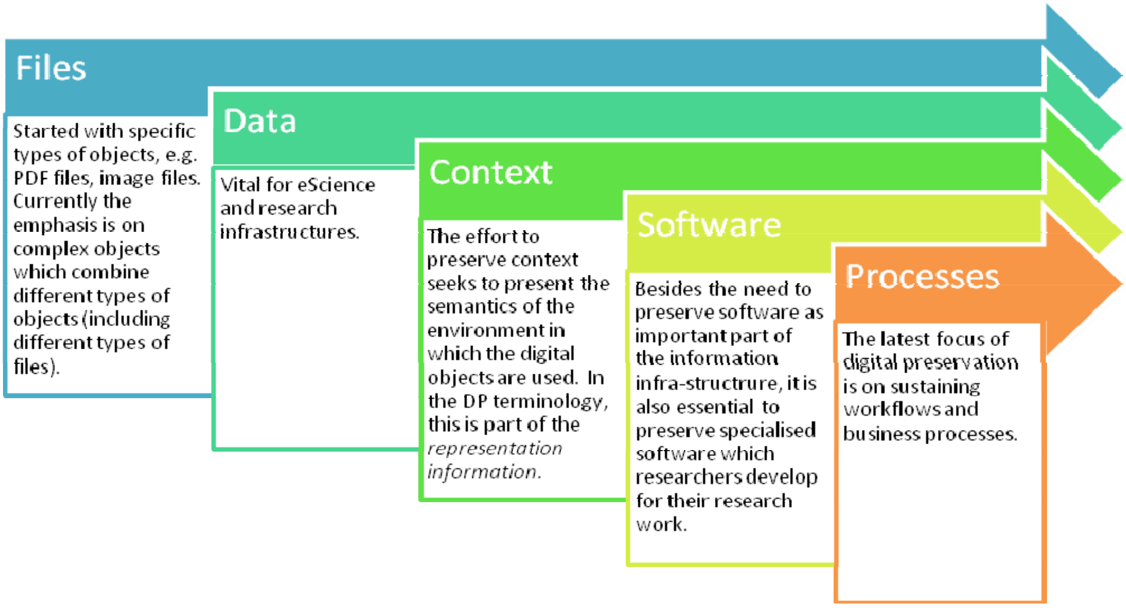

Figure 1.- Evolution of digital objects addressed by digital preservation (from: Dobreva, Ruusalepp 2013).

The importance of long-term preservation was highlighted in report of the Comité des Sages (Reflection group on bringing Europe's cultural heritage online) to the European Commission (The New Renaissance, 2011, p. 6):

- "Preservation is a key aspect in digitisation efforts. Digital preservation is also a core problem for any born digital content. The organisational, legal, technical, and financial dimensions of long term preservation of digitised and born digital material should be given due attention. […]

[3] http://www.digitalpreservationeurope.eu/what-is-digital-preservation/
[4] http://www.dcc.ac.uk/digital-curation/what-digital-curation

- To avoid duplication of effort by companies operating across borders and by the cultural institutions, a system could be envisaged by which any material that now needs to be deposited in several countries would only be deposited once. This system would include a workflow for passing on the copy to any institution that has a right to it under national deposit legislation."

The importance of preserving digital objects is, thus, well understood. Hardware and media obsolescence, lack of support for older computer formats, human error as well as malicious software all can lead to loss of digital information. Preservation, however, is not concerned only with sustaining single digital objects. To be used meaningfully in the future, digital objects should be preserved in context which makes them understandable to the future users. It is often said that digital preservation is interoperability over time. However, preservation is a complex interoperability-providing activity not only because of the increasing complexity of digital objects, but also because the context of use, too, needs to be re-created. This can mean sustaining not only the underlying data, but also any specific software which was used to work with them, and the technological infrastructure. The gradual expansion of preservation towards various types of objects is presented on Figure 1.

3. Trends in Digital Preservation research, tools and services

Without proper maintenance digital assets will fall into disrepair, succumbing to digital "diseases" that impair or limit the ability to use them, such as physical degradation of media so that objects are no longer readable, technological obsolescence, or even outright loss. Consequently, preventive measures must be taken to ensure that physical media is stable and that information encodings can be understood. Digital preservation in practice means provisioning secure storage systems, refreshing aging media, fixity checks, and replication in multiple systems and/or locations, format migration, emulation and other techniques to keep information safe and accessible over time.

Since the last report that analysed software agent technology synergies with digital preservation (Protage 2010) there are a number of trends in the digital preservation domain that can shape the DURAFILE solutions. This chapter looks at latest trends in digital preservation from the following angles:

- Current research topics
- Digital preservation tools
- Digital preservation services

3.1 Current research topics in Digital Preservation

The predominant funding for research into digital preservation issues is coming from the European Commission and to some degree from national funding agencies in a few countries (e.g., UK, Germany, Netherlands, Austria, Spain). In summary, current activities address digital preservation issues at three levels: fundamental research, applied research and development, and networking.

The fundamental research moves beyond the preservation of simple documents and data structures and focuses on interactive objects, embedded objects, ontologies and ephemeral data. For example, projects like LIWA[5] (addressing web archiving), BlogForever[6] (addressing blog preservation), TIMBUS[7] (research on preservation of business processes), Wf4ever[8] (tackling the issues of preserving workflow information) or PERICLES.[9] Fundamental research is also carried out on formal methods for object validation, for example validation of objects according to format specifications and policies as well as results of preservation actions against completeness and correctness (e.g., the PLANETS project[10] work carried forward by the SCAPE project[11]).

Applied research and development in digital preservation focus predominantly on scalable preservation systems. The need stems from the user communities requesting tools, methods and models that perform on realistic heterogeneous large collections of complex digital objects. A second aspect of handling vast amounts of objects effectively is the automation and decision support in a number of stages, ranging from object selection, tool performance, to validation criteria. In the past a number of conceptually well designed modules for digital preservation tasks were developed that required human intervention. Current research is focused on taking these modules to the next level and providing a high degree of automation of preservation processes as well as assist decision making. (Strodl et al., 2011) Examples of such projects include SCAPE (see also Ch. 4.3 below),

[5] http://www.liwa-project.eu/
[6] http://blogforever.eu/
[7] http://timbusproject.net/
[8] http://www.wf4ever-project.org/
[9] http://www.pericles-project.eu/
[10] http://planets-project.eu/
[11] http://www.scape-project.eu/

ARCOMEM[12] (is using the social web for automated information creation and supported appraisal), ENSURE[13] (e.g., scalable pay-as-you-go infrastructure for preservation services for integration into workflows).

Thanks to the networking efforts, the awareness of digital preservation issues now reaches far beyond the traditional cultural heritage and research data sectors to include the industry and enterprise domains. This development is well reflected in recent project consortia with increasing participation of industry players as solution providers as well as problem owners (e.g. SHAMAN,[14] ENSURE, SCAPE, TIMBUS, etc.). Networking is also happening around the establishment of an audit and certification process based on national (DIN 31466) and international standards (ISO 16363). The leading network in this domain is the APARSEN project[15] but other networks like CESSDA[16] are also involved in establishing auditing practices.

The preservation of digital audio-visual content has been a constant topic among EU-funded projects and continues to present new challenges both on technology (storage) and representation (encoding, file formats, access) levels, as well as from the legal (copyright and access models) aspects. The PRESTO-series of projects[17] have tackled most of these issues with the currently on-going project PRESTO4U[18] focusing on bringing the existing tools and expertise together into easily accessible solutions for practitioners.

In summary, the conclusion reached by Strodl et al (2011, p. 4) is still valid: "[There] is a slow shift from addressing questions that help to fix problems in maintaining digital information over time to ensuring that the problem will not appear in its full complexity in the first place, reducing the need for specific *ex-post* fixing. With the progress made in digital preservation research so far, the community has developed a solid understanding of the problems and the approaches needed to fix them, turning digital preservation activities in some areas into a challenging engineering task that requires further attention." On the level of digital preservation practice, there is general awareness of importance and support from a number of solutions and tools on offer, but a significant proportion of memory institutions still have to find their own way for implementing preservation within their specific business model realities. Hence, the situation with digital preservation tools and services also needs to be looked at.

3.2 Digital Preservation tools

Digital preservation is implemented through complex procedures; breaking these life-cycle processes down into smaller manageable tasks is one of the rationales for providing services that address a clear issue and solve it in an efficient way. This approach has become particularly relevant in distributed environments (e.g. cloud services, research infrastructure service providers) and also for smaller institutions or projects that do not have the capacity to develop bespoke solutions covering all preservation functions. Specialised services can also be integrated into in-house preservation systems where they can help to resolve a granular issue without compromising any of the primary functions of the preservation system.

A recent study by the DC-NET project[19] looked at the basic functional entities which a digital archiving and preservation system needs to implement and analysed the current offering of **software tools** for automating these tasks (Dobreva, Ruusalepp, 2010). The report broke the digital archiving workflow

[12] http://www.arcomem.eu/
[13] http://ensure-fp7-plone.fe.up.pt/site/
[14] http://www.shaman-ip.eu/
[15] http://www.alliancepermanentaccess.org/index.php/aparsen/
[16] http://www.cessda.org/about/governance/eric_req.html
[17] http://www.prestospace.org/, http://www.prestoprime.eu/
[18] https://www.prestocentre.org/4u/about
[19] http://www.dc-net.eu/

into entities that feature pre-ingest (including transfer), ingest, storage, digital object analysis, preservation planning, access, and re-use, which represent a life-cycle process-oriented approach in preservation. Based on a desktop study and analysis of some 190 currently available software tools, the report provides a high-level view on the range of instruments available for the preservation life-cycle. The DC-Net study found that digital preservation services are by and large still an experimental area. The distribution of the tools identified by the study shows that most of them deal with smaller practical tasks. Figure 2 below shows four areas of digital preservation with the highest number of tools on offer. The most numerous type of tools is related to metadata; web archiving is another popular digital preservation domain; and digital object analysis is part of both ingest and preservation planning.

The DC-Net report pointed out a significant lack of benchmarks and metrics for comparing preservation tools for both professionals and beginners in the digital preservation business. It also recommended creating business scenarios that would ensure sustainability for the tools or their development into maintained e-services.

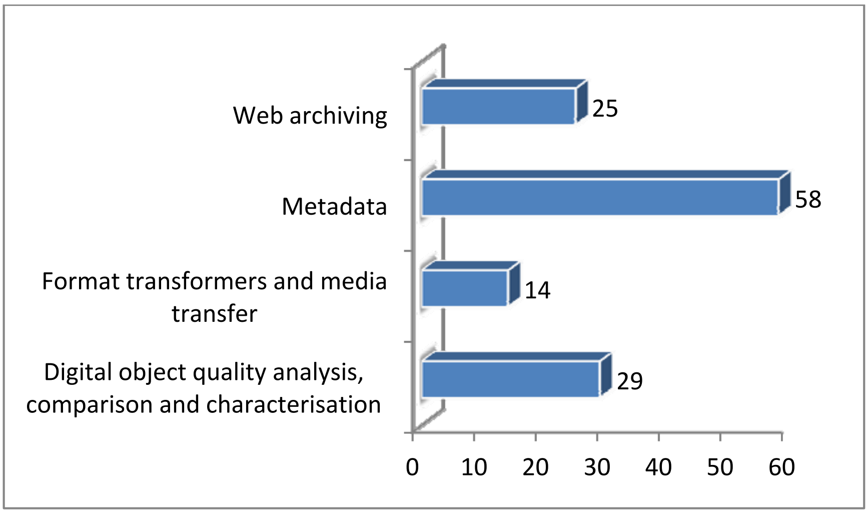

Figure 2: Distribution of digital preservation services according to their purpose.

In the area of digital preservation tools, institutional and project drivers have often resulted in duplicated tools and resources, with value spread thinly around the web. Some examples of tool registries include: the CAIRO project,[20] the National Digital Infrastructure Preservation Program (NDIIPP)[21] in USA, the Library of Congress,[22] the AQuA project,[23] the OpenPlanetsFoundation (OPF),[24] the DigiBIC project,[25] the DCH-RP project,[26] as well as the SourceForge[27] platform for publishing, searching and downloading open source software. A summary of various tool registries was produced by the SPRUCE project.[28] Admittedly, this has led the digital preservation community wondering where they should go to for advice, guidance, best practice and other forms of information necessary for utilizing the tools. In order to induce collaboration rather than competition

[20] http://cairo.paradigm.ac.uk/projectdocs/index.html
[21] NDIIPP Partner Tools and Services Inventory, http://www.digitalpreservation.gov/partners/resources/tools/index.html
[22] http://www.loc.gov/standards/premis/tools.html
[23] http://wiki.opf-labs.org/display/AQuA/AQuA+Mashup+Tool+List
[24] http://www.openplanetsfoundation.org/
[25] http://www.digibic.eu
[26] http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/registry-of-services-and-tools/
[27] http://sourceforge.net/
[28] http://wiki.opf-labs.org/display/SPR/Digital+Preservation+Tools

the Community Owned digital Preservation Tool Registry (COPTR)[29] was recently launched and is expected to become the focal point of information on preservation tools.

3.3 Digital Preservation services

The approach of **microservices** emerged a few years ago as an alternative to integrated digital archive systems. Microservices allow to flexibly combining specialised solutions for preservation depending on the requirements of the institution. They are defined as follows:[30]

> "Micro-services are an approach to digital curation based on devolving curation function into a set of independent, but interoperable, services that embody curation values and strategies. Since each of the services is small and self-contained, they are collectively easier to develop, deploy, maintain, and enhance. Equally as important, they are more easily replaced when they have outlived their usefulness. Although the individual services are narrowly scoped, the complex function needed for effective curation emerges from the strategic combination of individual services."

Microservices for digital preservation are currently under development at the California Digital Library (Merritt repository) (see Abrams et al., 2010), for the Electronic Records Archives at the US National Archives using iRODS (see Rajasekar et al., 2010, p. 60), and are also used in the open archival information system Archivematica.[31]

The debate on the applicability of this approach is on-going, for example (Challis, 2010):

> "I'm not convinced the specs for these are well defined enough for general purpose use yet, but I can see the technique in general being very useful. If the same microservices can be called through multiple interfaces (command line, REST, etc.), then it should in theory make them language agnostic."

Work continues on the thorough decomposition and analysis what constitutes a microservice and how various microservices can be orchestrated so that the major requirements for authenticity and integrity of preserved digital objects are not compromised.

In addition to the services described above, digital preservation is often making use of public registries. The registries collate and unify available information for easy access and to answer specific questions, for example: What format uses a particular filename extension? What platform can render a particular file format? What formats are supported by a particular software tool?

A **format registry** is a collection of records that characterize existing file formats. For example, a file format entry could include name and version number, characterization elements and links presenting dependencies with other formats. There is, as yet, no consensus on how this information should be structured. Three examples of format registries with different approaches are PRONOM,[32] Unified Digital Format Registry[33] and IBM Preservation Manager.[34]

Most of the current advanced technologies in automated file format identification rely on some information from an internal or external format registry. The file identification and validation tools

[29] http://coptr.digipres.org/Main_Page
[30] http://www.cdlib.org/services/uc3/curation/index.html
[31] http://archivematica.org/wiki/index.php?title=Development_roadmap
[32] http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
[33] http://www.udfr.org/
[34] http://www-935.ibm.com/services/nl/dias/is/preservation_manager.html

automatically generate file format data. The simplest mechanism for identification is to analyse the file extension and consult a registry of file extensions (e.g., File Extensions[35]). One problem with the use of file extensions is that they are not unique - e.g. the popular extension DOC has six possible associated formats. In addition, computer users on most platforms can make up their own file extensions or change the existing ones making judgments based on the file extension unreliable.

Another example of a registry is the Trustworthy Online Technical Environment Metadata Database (TOTEM)[36] developed by the University of Portsmouth in the KEEP project.[37] TOTEM allows the checking of dependencies between hardware configurations, operating systems and software tools for the purposes of emulation.

3.4 Conclusions

Since the Protage project concluded its work, the research agenda in digital preservation has clearly taken a turn towards the corporate and business sector which supports the overall concept of the DURAFILE solution. The project will be able to benefit from results of various research and development projects that have explored how digital preservation is applied in businesses. The next chapter outlines some of those examples.

There has also been significant progress in developing new software tools and services that can be combined into full digital archiving workflows based on the user needs. This facilitates the DURAFILE solution architecture that can make use of a greater variety of microservices and tailor them into solutions based on specific user needs.

Several reports have noted the challenges of orchestrating digital preservation tools and services for distributed architectures (e.g. cloud) because of lack of appropriate models. The main digital archive reference model (the OAIS model[38]) remains too abstract for detailed implementations and too focused on single repository solution. The existing preservation services are scattered and mostly atomic and, therefore, it is a challenge to set up a complete digital archiving workflow based on repository software and combine it with a large number of micro-services, and cloud/grid services. This presents an opportunity for the DURAFILE project with its software agents' based architecture that can offer an alternative for the current (emerging) distributed digital archiving models.

[35] http://www.file-extensions.org/
[36] http://keep-totem.co.uk/
[37] http://www.keep-project.eu/ezpub2/index.php
[38] ISO 14721:2003 Space data and information transfer systems - Open archival information system - Reference model

4. Trends in Intelligent Agents Software technologies

Agent-oriented computing has been regarded as a promising computing paradigm for developing and implementing complex, distributed software systems, as this paradigm is based on intelligent agents and multi-agent systems enables software engineers to model applications in a natural way that resembles how humans perceive the problem domains (Brau et al., 2003; Chmiel et al., 2005; Jennings, 2001). Intelligent agent technology has been successfully applied in many industrial and commercial areas, including information retrieval and filtering, electronic commerce, human computer interaction, telecommunication systems, air traffic control, planning and scheduling, process control, manufacturing, workforce management, and military (Luck et al., 2005). It has also gained great success in studying complex physical and social problems. For example, multi-agent systems have been successfully adopted to investigate the impact of climate change on biological populations and the influence of public policy options on social or economic behaviour. Therefore, employing agent-oriented computing as the general methodology to study long-term digital preservation emerges as a promising direction.

The last trends in Intelligent Agents is to employ them into computerize long-term Digital Preservation tasks. The use of Intelligent Agents can make Digital Preservation tasks more automated and easy enough so that users can readily preserve their digital objects, while reducing the preservation cost and increasing the preservation capacity. The fact is there is not a single preservation plan for every problem but several options and different users might have different useful-preferences among the existing solutions making the adoption of a multi-agent architecture instead of a more conventional centralized repository necessary. The dynamics of these preservation plans (that can be updated during its life) also indicates that the adoption of a distributed architecture is the appropriate solution. Similar to the way antiviruses do, the Intelligent Agents can scan continuously a directory for obsolete files. In the case of a digital obsolescence sign is found in the user directory, the agents will warn the user. The file obsolescence found can be solved with collaboration among users through Social Search empowered by the intelligent agents.

The complexity of a multiagent system and the collaboration of several Intelligent Agents produced two new concepts that will be developed during the DURAFILE project: Social Search and the Trust concepts.

4.1 Social Search

The Social Search can be defined as a type of search that takes into account the social graph of the person that initiates the search. This search could be a collaborative discovery of web pages, news, images, videos, podcasts and knowledge sharing. This social search produces a user input such as promoting or demoting results the user feels are more or less relevant to their query.
The concept of social search was applied recently to the Digital Preservation problem. This Social Search using a social graph can be directly applied to the Intelligent Agents field, were agents form a multiagent system with relationships among them. If the agents are specialized in Digital Preservation solutions, the Social Search using Intelligent Agents can provide a tool for connecting agents (that represents people) that search for an appropriate Digital Preservation plan with other agents that have the solution. This concept of Social Search applied to Digital Preservation tasks is a novelty that appeared recently and that was explored in the Protage project.

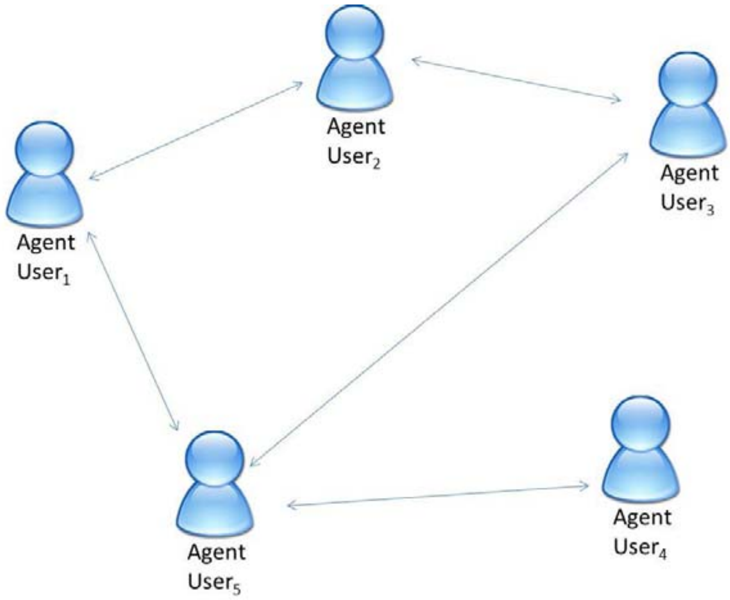

Figure 3: Social Search with Intelligent Agents

Figure 3 shows an example of Social Search applied to Digital Preservation tasks. When the Agent $User_1$ needs a preservation plan, it searches into its own database. In case that no solution is stored in the database, it asks to another agents (Agent $User_2$ and Agent $User_3$) for a suitable preservation plan. The agents that receive the request, search into their own database and if not solution is found they ask for other agents in its contact list. This process can be repeated reiteratively until a certain limit is achieved – this limit is set by a threshold that can be configured.

4.2 Trust

In the previous example, a Searcher Agent searches other user platforms in order to obtain a suitable preservation plan. It is possible that the agent that initiates the search receives several digital preservation plans from other agents. In that case it must select the solution provided by those trust agents and it must select among these solutions the one that it "trusted". The use of trust mechanism is a new trend in intelligent software agent's technology. To obtain this trust index, once an agent returns a suitable Digital Preservation plan, it will be evaluated by the agent that started the Social Search, which will update the trust index depending on the degree of satisfaction of the provided Digital Preservation plan. Following the example of Figure 3, an Intelligent Agent wants a Digital Preservation plan for an audio file. The Intelligent Agent receives 2 different preservation plans from Agent $User_3$ and Agent $User_4$. Since Agent $User_4$ is more "trusted" for multimedia files it is the preferred preservation plan. The system executes the preservation plan and evaluates it, updating the trust index of the agent that provided the solution depending of the satisfaction of the provided solution.

4.3 The potential of Intelligent Agents in Digital Preservation

The FP6-funded research and development project PROTAGE[39] (2007-2010) was the first attempt to combine intelligent software agent concepts with digital preservation requirements. The project considered a number of aspects for demonstrating the utility of agent technology for digital

[39] http://www.ra.ee/protage

preservation and developed three prototypes[40] to test them. The three application scenarios that were supported by the prototypes were the following:

- the use of agents for agencies and companies in pre-ingest activities (first prototype);
- the use of agents in a digital preservation guidance function (second and third prototype);
- the use of agents in a monitoring function (second and third prototype).

A brief summary of each scenario is provided below.

4.3.1 Agents in pre-ingest

Pre-ingest is the stage when digital objects are prepared for transfer and handing over to a digital archive. When looking with the eyes of a public agency or a private company the essential characteristic of pre-ingest activities is that there is a clearly defined long-term repository, whether in-house or external, which will take care of the user's digital assets. The timeframe and requirements for transfers are usually clearly defined by a transfer policy which says when and how the digital assets should be transferred to the repository.

The agents in the first PROTAGE prototype were modelled to consider these characteristics and were to gather information about the user's location and affiliation, then to decide on the legislative requirements in the given environment, decide on the optimal repository to use, download SIP requirements and fulfil all necessary tasks to ensure the necessary quality of the data to be transferred.

The evaluation of the results of the first prototype tests resulted in a decision not to pursue further the use of agents in this scenario. The reasons were related to the complexity of and time required for developing agents that could select and negotiate transfer requirements; the un-dynamic nature of SIP requirements that would mean the transfer process could not benefit significantly from using software agents; and finally the technical complexity of applying agents in a diverse legislative and technical environments. It was also noted that the varying metadata structures and respective metadata requirements of repositories pose a challenge which is not easily solvable by software agents.

4.3.2 Agents for digital preservation guidance

The subsequent Protage prototypes focussed on the application of agents in helping the user to determine the best possible workflow for dealing with some digital preservation related risks and to select the appropriate tools for the workflow steps.

The prototypes implemented agent-based guidance functionality that analyses user's digital objects using standard components (like Droid or JHove). The user can search for suitable preservation tools based on the collection's characteristics through agent communication. The search returns hits for a keyword from trusted users who have similar collections with similar preservation problems. This model supports effective and trusted knowledge transfer from experts in digital preservation (e.g. large memory institutions) to those who do not have the experience or ready access to the know-how required for making decisions on best software to apply to their specific preservation problem. The model also supports re-use of successfully applied preservation plans and tools by other users and providing feedback on them.

[40] http://www.ra.ee/protage/testing-protage/

4.3.3 Linking software agents and digital preservation

Intelligent agent is an autonomous entity which observes through sensors and acts upon an environment using actuators (i.e. it is an agent) and directs its activity towards achieving goals (i.e. it is rational).[41] Intelligent agents may also learn or use knowledge to achieve their goals.

Intelligent agents are often criticized as representing technology that is actively pursued in research labs but rarely appears in deployed applications. In fact, many of the underlying technologies of intelligent agents have migrated into mainstream applications, where they are no longer referred to as agents. Many university departments will revisit the evolution and application of intelligent agents and consider how they are shaping emergent technologies or becoming embedded within applications.

The new generation of systems must be able to behave in an autonomous, flexible manner in unpredictable, dynamic, typically social environments. It is precisely their autonomy that defines agents. Agents are a tool to achieve autonomy to interact in a real world with as early references as (Wooldridge 1999, Jennings 1998) "An agent is a computer system that is capable of flexible autonomous action in dynamic, open, unpredictable and typically multi-agent domains". The most recent definition of agent is: It is a **design metaphor** (Luck 2005).

Agents must decide by themselves whether to execute their methods according to their goals (agents must be pro-active), preferences, and beliefs. Also, agents must be flexible. When designing agent systems, it is impossible to foresee all the potential situations an agent may encounter and specify agent behaviour optimally in advance. Agents therefore have to learn from and adapt to their environments. This task is even more complex when nature is not the only source of uncertainty, and the agent is situated in an environment that contains other agents with potentially different capabilities, goals, and beliefs. An agent must also show a social attitude. In an environment populated by heterogeneous entities, agents must have the ability to recognise their opponents, and form groups when it is profitable to do so. Agents can also assist users teaching or training, making recommendations and helping different users collaborate (Montaner 2003, Adomavicius 2005). In these situations, agents are regarded as the technology in which the IT and telecommunications sectors should converge. They have a role to play at the client-side of such systems, providing customers with personalised, proactive interfaces to new services and products (Aciar 2007). In e-commerce scenarios, they have a role to play as middleware, putting users in contact with the goods and services that best suit their needs. And they have also a role to play as servers, cooperating and negotiating on behalf of organisations and other end users (de la Rosa 2011). Today, agents are being extensively used to implement electronic markets and electronic auctions. Generally speaking, the design and implementation of multi-agent systems is an attractive platform for the convergence of various AI technologies.

Software agents justify themselves most in situations where users are spread and do not have well known and strict requirements to follow. In digital preservation context agent technology justify themselves in situations where new knowledge is needed for taking some actions (for example, maintaining the usability of file collections). As the PROTAGE project discussed (Protage 2010), agents have less to contribute towards innovating established digital preservation routines (for example, transfers of records to archives for preservation).

More recently, the definition of social machines by Hendler and Lee might take the lead of the paradigm that better suit the advent of social networks. Hendler and Lee (2010) stated that real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in

[41] http://en.wikipedia.org/wiki/Intelligent_agent

which the people do the creative work and the machine does the administration of the Digital Preservation stuff. . . The stage is set for an evolutionary growth of new social engines. The ability to create new forms of social process would be given to the world at large, and development would be rapid as it must be proved in DURAFILE.

The application of automated digital preservation tools on corporate information systems is a technically complex task. The “agentification” of singular preservation tasks regardless of their position in a larger workflow is the target approach that will be explored in the DURAFILE project.

4.3.4 Monitoring agents

In order to support users with identifying the preservation risks associated with their digital collections, the Protage project developed a monitoring agent that determined the migration needs by collecting statistics about file formats used in the user’s network. It would be simple to automate this functionality in a way that agents automatically detect file formats at risk (for example when less than 5% of the network uses a given file format) and to automate the search for an appropriate migration tool for such formats.

4.4 Conclusions

The potential of software agents in digital preservation was presented in this section using the Protage F6 project (Protage, 2010). The experience in the development of the project provided the next conclusions:

- the application of software agents on larger workflows like pre-ingest does not seem reasonable, instead the “agentification” of singular preservation tasks regardless of their position in a larger workflow seems more promising;
- the potential of software agents is clearly visible in a guidance function and most in monitoring function as they are able to react on changes in the environment and take into account the different contexts of different users;
- software agents are highly dependent on the number of agents in the network/environment as most of the potential lies in applying quantitative mechanisms and feedback.

The last point was re-iterated by the knowledge-exchange model that the Protage prototypes relied on where trusted knowledge on digital preservation solutions came from a small number of memory institutions to a large number of individual users with very different preservation problems. The downside of such a model is that it requires active take-up by expert users who contribute trusted knowledge to the environment in order to offer a comprehensive set of guidance. The user community should preferably also be large in order for the quantitative calculations based recommendations to function correctly.

It should be noted that the scope and period of the PROTAGE project was limited and did not allow for exploring the potential of combining the software agents with other technologies, e.g. semantic tools. The next chapters look at the developments that have occurred in both the digital preservation and software agents’ domains since the end of the Protage project.

Finding suitable Digital Preservation solutions is a distributed task that involves several participants. Multiagent Systems seems to have a suitable structure (autonomy for executing tasks and connection with other participants for sharing solutions) for use them in these Digital Preservation solutions. Two new concepts arise from the cooperation of Intelligent Agents: Social Search and Trust. These concepts were reviewed in this section including an example in the Digital Preservation domain, which shows the suitability of Intelligent Agents. Both, Social Search and Trust models will be developed and adapted to Digital Preservation solutions during the DURAFILE project

5. Digital preservation using agent-based technologies: update

In a dynamic environment where new kinds of complex file types are invented, organisations and businesses do not have sufficient manpower or resources to ensure their long-term preservation through manually applying standards and technologies. Hence, more intelligent and automated solutions are the direction that the market is striving towards. One of the main challenges with intelligence-based solutions in digital preservation today is that there are not enough automated workflow tasks that are fully supported with reliable knowledge and resources. As the first prototypes demonstrated, agent-based solutions can certainly help to achieve better results in the long-term digital preservation – agents can evolve by communicating with users and getting feedback from other agents, and by acquiring up-to date information from memory institution trusted knowledge bases.

The applications of intelligent agent technology are well documented in literature from a computer science or engineering perspective, but not from a library perspective (Allasia et al, 2012). Some research exists on adapting and developing agent-based technologies and strategies for digital library or hybrid library, to make them more accessible to end users and automating their workflows. The downside from the DURAFILE perspective is that most of these projects do not contain much to help improve preservation tasks. Most of these projects are somewhat dated for today's technological standards and are often based on experimental results or theories. However, a few recent projects exist that have a bearing on how agent-like technologies can be applied to digital preservation.

5.1 Research Projects using intelligent agents

Only a few years ago there was no product on market for Digital Preservation of multimedia objects, but during the last years the trends go into the direction of integrate Digital Preservation tools in software solutions. Recently new research projects aiming at Digital Preservation in different ways have been identified, including European Research Projects FP6 and FP7. Only few of this research projects incorporate the use of intelligent agents for Digital Preservation. This section provides an overview of several research projects which combine intelligent agents and Digital Preservation.

5.1.1 PROTAGE

The mission of the **PROTAGE** project[42] was to investigate and initiate complementary new approaches to digital preservation that make long-term digital preservation easy enough for users to be able to help preserving their own content, while reducing the cost and increasing the capacity of memory institutions to preserve digital information. PROTAGE goal was to experiment hardware and software independent autonomous software agents for different long-term digital preservation tasks. The project was based on latest agent technologies and digital preservation strategies. Three prototypes were developed by this project to test knowledge transfer models in support of preservation planning.

So far, PROTAGE is the only project that has developed a multi agent digital preservation solution and experimented with software agents for achieving best results. Even though it was targeting large institutions it made a breakthrough in the field. Its approach created a benchmark for the future,

[42] http://www.ra.ee/protage/project/

next-generation digital preservation projects. It had ambitious ideas but at the same time suffered from technical drawbacks which can and should be avoided in the future.

5.1.2 PANIC

The aim of the **PANIC** project (Hunter and Choudhury, 2006) was to build a system which dynamically incorporates a range of different preservation services to help either librarian or archivist to select best singe service for their task. All the tools and services are integrated into flexible, extensible web based framework also known as semantic web services architecture. PANIC system combines three main software components: preservation metadata capture, obsolescence detection & notification and preservation service discovery and invocation. Agent based approaches are used to compare stored objects' metadata against real-world registers and if needed notify users when data formats are not supported anymore. Also to compare user made service specification to existing ones and make list of ranked solutions and either automatically execute best ranked service or let the user decide. It supports software independent metadata capture for mixed-multimedia objects. However one of the major drawbacks for this project was that its implementation is only semi-automatic, i.e. technical metadata must be entered manually, real-world registers are not implemented, manual service specification etc.

5.1.3 SCAPE

SCAPE[43] approaches digital preservation through research and development sub-projects: Testbeds, Preservation Components, Platform, Planning and Watch. It addresses scalability of digital preservation workflows for large-scale collections. It aims to develop a framework for automated, quality-assured preservation workflows and integrating these components with a policy-based preservation planning and watch system. **Scout**[44] is a preservation watch system developed within the SCAPE project. One of the main goals of Scout is to provide preservation monitoring and assistance to end users. Scout will notify the user when content is at risk and reduces risks by community engagement and assistance. It discovers information about users who are holding content of a specific type (file format). Scout is combining features from tools like PRONOM, C3PO, REST, SPARQL.

5.1.4 ENSURE

ENSURE[45] addresses digital preservation issues with innovative approaches and tools: cost and value evaluate the benefits of different quality solutions. Preservation Lifecycle Management builds on industry standard lifecycle management approaches to manage the preservation lifecycle, ensuring regulatory compliance, allowing changes in the preservation approach to reflect environmental changes, addressing evolution of ontologies and managing the quality of the digital objects over time.

5.2 Commercial Developments using Intelligent Agents

Although the use of Intelligent Agents for Digital Preservation tasks is still a research field, a few commercial developments appeared the last years. This section overviews some developments that become commercial products.

[43] http://www.scape-project.eu
[44] https://github.com/openplanets/scout
[45] http://ensure-fp7-plone.fe.up.pt/site/

5.2.1 SPAR

SPAR (Scalable Preservation and Archiving Repository) is a digital preservation system designed to help Bibliothèque nationale de France with their digital collection preservation. One of the main challenges, when designing a system such as SPAR, is the risk of obsolescence of the system itself. To prevent that, the whole system was designed as a modular system, in which each part is able to work independently from the others. Also data-first approach was used to ensure best possible long-term protection. The pre-ingest phase is meant to harmonize the different digital documents into a SIP (Submission Information Package) that is SPAR-compliant and can be processed by the rest of the system in a generic way. This approach very much relies on the creation and maintenance of metadata. The system is fully self-describing: descriptions of the processes, agents (including software agents) and formats used in the system are ingested as information packages, themselves to be preserved. A track is a collection of objects that share the same requirements in relation with preservation. In SPAR, each package belongs to a track, which can be viewed as a family of documents with similar intellectual and legal characteristics; each track has one channel per homogeneous technical characteristics. Since SPAR is an OAIS-compliant system, every piece of preservation information has to be submitted and stored in it as an information package. To this end, it uses reference packages, of three different types: context, formats and agents.

5.2.2 PRESERVA

PRESERVA lays the foundations for a new object-centric digital preservation paradigm. It solves preservation issues involving complex digital objects by building new digital preservation environments where objects become active actors with their own budget for attracting know-how and services (Self-Preservation Aware Digital Object (SPADO). This represents a paradigm shift in digital preservation. The research required to develop SPADOs includes the exploration of a number of synergetic areas as a support to the new objects: focusing on computational ecologies, cloud computing, and social networking. The project will deliver a proof of concept of the emergent behaviour of huge communities of SPADOs competing for the best DP services and the appropriate DP know-how available from users.

5.2.3 IT IS FOREVER

This is a gamification approach as an innovative solution with a strong focus on engineering and design, combined with the most modern tools of digital preservation. The gamification of memories socialization is the act of "remembering": Given a picture, remember where and when it was made, who came in the picture, what event happened, who did the photo, story or anecdote about that day, and all that is capable of a conversation to both desktop (especially) internet, is an individual act, close family, extended family, or friends or colleagues. If you get these talks on-line then there is a powerful reason for archiving or store personnel (forever) photos, videos, notes, emails, comments, facebook or Twitter to feed these conversations. Personal Digital preservation is currently a chaotic process, craft, and poorly served in the marketplace. One of the main problems detected in digital preservation is the absence of software tools that allows actions to preserve digital objects stored on their computers such as text, data, graphics, still and moving images, sound recordings, software, etc.. We speak advisedly: there are no simple solutions or automated digital preservation for consumers and this is the main objective of this project, just taking advantage of the power of crowdsourcing gamification caused by a preservation process that requires human participation in high first, and then preservation technology can develop independently. Consequently, the gamification of personal digital preservation solves the appraisal (which is important to preserve), metadata (by managing comments on which is extracted entities, subjects, objects, actions and temporalities of what happened to that digital object), and upgrade and migration of formats for the prevention of digital obsolescence.

http://www.easyinnova.com/index.php/projectes/preservacio-digital/it-is-forever

5.2.4 MIDPoint (Micro trading concept for long term digital preservation)

Following with the current paradigm, large-institution-oriented approach to digital preservation, significant yet insufficient budgets are assigned to the preservation of a rapidly growing number of digital objects of increasing complexity. Following an appraisal decision, the next tasks are to find the appropriate resources that will preserve digital objects and collections. This approach is a sort of **Macro trading** where big institutions assign tasks to big service providers that tends to be less cost effective as there is less competition. This approach is inefficient in the way resources are allocated (what collections and objects are to be preserved) and in the way the tasks are assigned, and clearly unable to cope with the growing costs and preservation and recovery demands when needed. Differently, this project will reverse the resource allocation with **micro-trading**: collections and objects have their own preservation budget and trade for maintenance and recovery services which in their turn bid and compete, negotiate and cooperate to be highly efficient in terms of cost and time (best ratio of service / cost offers are selected). The competitive market situation ensures high quality at best possible cost. **reservation ecosystem with different actor types** (different actor types will allow negotiation with specific ecosystem dynamics). This is a paradigm shift from the current institution or user-centric approach, where the institution or user performs the roles of “caring”, “paying for” and “curating” the digital objects, to an object-centric approach where the object has the role of “caring” for itself, the users or institutions “pay” for its preservation and provide know-how for “curating” it and the DP services compete to “preserve” it. The new role assignment will lead to more balanced preservation decisions on how, when, and what to preserve than the current paradigm.
http://www.easyinnova.com/index.php/projectes/preservacio-digital/midpoint

5.2.5 Pyramid

This is a particular implementation following a new object-centric digital preservation paradigm to address a complex real-world problem of Digital Preservation. This is powered by a social network as an environment that enables their behaviour under the policy of “preservation is to share”. In all the experiments, DOs travel through the P2P network that is build up out of a social network (of users that share their storage space both in PC or cloud) and distribute (migrate) copies of themselves for preservation. These (migrated) copies maintain links to the parent Digital Objects because they correspond to a same object. Each node representing a user’s computer might only be able to read a specific video or image format after suffering a Software Adoption Wave (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.
http://www.easyinnova.com/index.php/projectes/preservacio-digital/pyramid

5.3 Other projects

This section overviews other projects which have explored agent based solutions in digital preservation field include the following. Only a few of them incorporate the intelligent agents to their solution, but we mention them due because they provide interesting solutions for the Digital Preservation problem that could be used in the development of the DURAFILE project.

5.3.1 PrestoPrime

PrestoPRIME[46] developed practical solutions for the long-term preservation of digital media objects, programmes and collections. Its aim was to find ways to increase access by integrating media archives with European on-line digital libraries in a digital preservation framework. One of the framework architectures which was partially funded in the PrestoPRIME project was Autonomic preservation of digital content, the so called **"Access Copies"**. Agents were used here to manage several digital preservation aspects: knowledge, reasoning, planning and scheduling. In knowledge phase, information about metadata and data dependencies is collected. Collected data is evaluated and compared to propose the best digital preservation strategy. However final decision hast to be made by the user, agents will provide assistance in the ranking process. This approach is not as strict as the OASI insists because loss of content is allowed (the original file will be kept and more access files can be made later if needed).

5.3.2 TIMBUS

TIMBUS[47] proposes a semi-automatic architecture that provides a complete framework for preserving business process architecture. It covers all aspects of traditional digital archive system but also addresses enterprise risk analysis and business continuity planning. Agents (DP Agent) are used for capturing resources from different IT systems that are required for digital preservation. After that, resources, dependencies and risk analysis will be compared and in collaboration with expert, best DP strategy will be selected.

5.3.3 AONS & AONS II

AONS & AONSII projects (Pearson, Walker 2007) were software development projects at the National Library of Australia in conjunction with the Australian Partnership for Sustainable Repositories. The projects aimed at developing a software tool that automatically finds and report indicators of obsolescence risks, to help repository managers decide if preservation action is needed. **AONS** (Automated Obsolescence Notification System) is a system to analyse digital repositories and determine whether any digital objects contained therein may be in danger of becoming obsolescent. It uses preservation information from public registries (PRONOM, LCSDF) which contain data about file formats and the software which supports these formats to determine if the formats used by digital objects are in danger. Repository checks will be done periodically and when the repository is found to contain obsolescence formats a notification report will be sent to the repository manager responsible. One of the problems with the first version was that PRONOM and LCSDF registries used occasionally different formats so direct mapping was bit complicated and inaccurate in times.

[46] http://www.prestoprime.eu/
[47] http://timbusproject.net/

5.3.4 APARSEN

The **APARSEN** project[48] has published its vision of a service-based architecture of a preservation system and is in the process of developing the component services:[49] preservation services, persistent identifiers, storage solutions, authenticity and provenance services, annotation, reputation and data quality, as well as brokerage services

5.4 SOA for Migration

SOA (Service-Oriented Architecture) for automatic migration proposed architecture to help cultural heritage institutions to accomplish automatic digital preservation (Ferreira et al. 2006). It is assumed that client institutions already possess a digital repository system capable of storing, managing and providing access to the digital objects they hold. The repository system acts as the client application that benefit from the services provided by the SOA. The goal is to provide a framework for discussion by pointing out the fundamental elements that should be present in such a system. Several interesting and competing research projects are presented as promising candidates to implement some of these elements. The SOA architecture is divided into two major sections: the client and the server-side. On client-side are custom applications (Web client, Linux client) and repository systems (Fedora, Dspace). On the server-side are depicted the main components comprising this framework. Each of these components is an independent application with distinctive roles and responsibilities that co-operate with each other by exchanging messages. This approach makes it possible for each component to be governed by a different organisation and facilitates the distribution of workload. Services in this model are the following: Obsolescence Notifier, Format Detector, Service Registry, Migration Broker, Format Evaluator, Object Evaluator and Migration Advisor which each will help to resolve certain digital preservation task. However the disadvantage for this architecture is that it presumes that client will already have digital repository in place, there for it will be most suitable for cultural heritage and memory institutions.

5.5 Conclusions

The Digital Preservation problem is a new topic that appeared only a few years ago. More recently is still the use of Intelligent Agents for the Digital Preservation tasks. This section provided a study of several research projects and commercial developments. It is clear that these projects (research & commercial) provide a starting point for the DURAFILE project. We mentioned other projects that do not provide a solution using intelligent agents but they offered interesting solutions for the Digital Preservation problem that will be studied during the implementation of the DURAFILE solution.

[48] http://www.alliancepermanentaccess.org/index.php/aparsen/
[49] http://www.alliancepermanentaccess.org/index.php/knowledge-base/existing-practices-services/

6. Analysis of Market Potential for Digital Preservation Services

6.1 The demand side

73% of respondents to the UK Needs Assessment for Digital Preservation stated that "their organisation would fail to comply with legal requirements if it failed to preserve its digital information adequately" (DPC, 2006). The PLANETS project[50] estimated that the value of digital documents produced in the EU which is in danger of digital obsolescence if no action is taken to preserve them, is in excess of €3 billion per year.

Digital preservation is not an entirely separable process from other aspects of digital asset management, and it follows then that the decision to allocate resources to digital preservation is not entirely distinct from the decision to allocate resources to the overall digital asset management process. This suggests that in many circumstances, it may be difficult to treat digital preservation as a separate economic activity with a clearly circumscribed set of costs and benefits, which in turn can be weighed in the context of a distinct economic decision. Instead, digital preservation might be better cast as an "incremental decision" built on top of the larger question of digital asset management.

So for example, the question might not be "should I allocate resources to digital preservation?" but instead, "given what I am doing now to secure the ongoing availability and use of digital materials, what are the additional costs and benefits of either a) extending the time horizon over which I can be reasonably confident the materials will persist in their current condition, or b) reduce the likelihood that the materials will cease to be available or usable within the current time horizon?" In either case, the decision on whether or not to invest in digital preservation is not a completely separate decision, but more akin to a parameter within a broader decision scope.

The economics of sustainable digital preservation is currently being explored by the 4C project (Collaboration to Clarify the Cost of Curation)[51] and the Economic Sustainability Reference Model it is developing for digital curation.[52]

This chapter will provide rough estimates of the potential market of the DURAFILE solution in three synergic areas: the antivirus tools market, and the digital preservation consultancy services market.

Digital preservation is inherently linked with the storage business. Storage focuses on "protecting information from physical problems", and similarly the digital preservation software is about "protecting information from digital obsolescence". Therefore, digital preservation can be thought of as yet another feature of storage systems to be sold. But it is stronger to see the image of antivirus that "fights the obsolescence of digital objects". This is further worked in the following subsections.

6.2 Anti-Virus Tools Market

Another possible comparison for digital preservation is to link it to the business of antivirus tools. The antivirus software is about "protecting information from virus attacks", similarly the digital preservation software is about "protecting information from digital obsolescence". It can be thought of as another feature of the antivirus tools.

A rough estimate of the digital preservation market can be made by comparing it to the anti-virus tool market of GBP 5.1 billion (2013 forecast).[53] If we assume that even 10% archival/availability

[50] http://www.planets-project.eu
[51] http://4cproject.eu
[52] http://4cproject.eu/community-resources/outputs-and-deliverables/esrm-summary
[53] http://www.computerweekly.com/Articles/2007/07/13/225550/anti-virus-software-market-grows.htm

value will be achievable via antivirus tool resellers then the DURAFILE potential market for personal digital preservation systems would be about EUR 600 million in 2013 terms.

Given that the digital preservation software features will be similar to those of an antivirus tool - scanning the computer looking for files affected by computer virus and "curing" it from viruses. The DURAFILE solution will be searching for obsolete files and taking necessary preservation actions to preserve their usability. We believe that in the long term the market of DURAFILE will be comparable to the market of antivirus.

6.3 Consultancy Services Market

Digital Preservation consultancy is one of the potential usages of the DURAFILE platform that addresses previously inherent business scaling problems in offering digital preservation services. Digital preservation issues are often too complex for individual institutions to address independently, which is the main reason why institutions try to cooperate,[54] and why there is market for digital preservation consultancy. Libraries are cooperating on this issue with other memory institutions (for example, other libraries, museums, and archives) and with research institutions, publishers, software developers or IT companies that might come to use DURAFILE as their digital preservation knowledge base provider.

The PROTAGE project pointed that 67% of digital preservation users think that the solutions they have today are not good enough, are insufficient (43%) or scarce (23%). Hence, they continuously look for new digital preservation solutions. Having studied how expert users behave regarding consulting and answering other expert or individual users, how they behave regarding using search engines or consulting and contributing to digital preservation web sites, we can firmly state that digital preservation today is a collaborative activity, since the know-how is widely spread and fragmented. Even expert users look for solutions provided by other institutions and consult trusted colleagues. This constitutes a market for digital preservation consultants with access to appropriate tools.

Usually big agencies provide their own digital preservation and document distribution services as well as tasks for related research and development, such as the U.K.'s National Archive.[55] This which does not mean that they have not worked with consultancy groups, but especially for smaller entities, consultancy companies offer their services for archiving, updating, or restoring data. Most of these consultancy companies are not focused only on digital preservation, but they provide it as an additional service, offering among other things:

- Document digitisation.
- Document migration and updating.
- Business intelligence.
- Storage, data repositories, storage migration, data recovery.
- Data security.
- Training.
- Consultancy in data standards, protocols.
- Consultancy for data creators oriented toward preservation.
- Consultancy in decisions, resources, planning, and risk management oriented toward preservation.
- Infrastructure support.

[54] See for example the digital preservation co-operatives in the US: Metaarchive (http://www.metaarchive.org/), Chronopolis (http://chronopolis.sdsc.edu/), HathiTrust (http://chronopolis.sdsc.edu/)

[55] http://www.nationalarchives.gov.uk/preservation/digital.htm

Most of the consultants in DP have considerable experience with big organisations, public and private. Some organisations offering consultancy in digital preservation are: King's College Digital Consultancy Service,[56] The Digital Archiving Consultancy,[57] Charles Beagrie Ltd,[58] Serco Group,[59] and others.

6.4 Conclusions

The market demand, analysed here from three perspectives, is a critical success factor for the DURAFILE solution. Other critical market success factors are to have a clear target group, an attractive context of use, a compelling value proposition and an approach for sustainable customer retention. These will be fully analysed in later deliverables.

[56] http://www.digitalconsultancy.net/
[57] http://www.d-archiving.com/
[58] http://www.beagrie.com/
[59] http://www.serco.com/

7. CONCLUSIONS AND RECOMMENDATIONS FOR THE FOLLOW-UP WORK

This document has presented the state of the art in Digital Preservation and Intelligent Agents. Several past projects that combine both concepts were presented. The deliverable shows clearly the necessity of Digital Preservation solutions for preserving our digital objects, including a description of the trends in preservation. It is crucial to increase the awareness of the need of digital preservation. For this purpose, the National Archives can act as demand pullers, their employees as expert users will be the prescriptions and digital preservation know-how providers, while massive usage tools (like DURAFILE) are require to enable common users be digital preservation aware. Misalignment of incentives among stakeholders may occur between communities that benefit from preservation (and therefore have an incentive to preserve), and those that are in a position to preserve (because they own or control the resource) but lack incentives to do so. Digital Preservation tasks can be distributed over several actors; therefore the use of intelligent agents is justified and described how new solutions as projects or commercial products that appeared recently combine both Digital Preservation and Intelligent Agents. The DURAFILE platform will use Agent Technology for the automation of social networking on digital preservation. This technology is ready to provide the incentives (Moreno et al., 2009) that are necessary for digital preservation. This approach for DURAFILE is promising regarding the huge impact of Web 2.0, that digital preservation expert users seem eager to use. The mission of the personal agents will be to work on behalf of users to boost the contributions of digital preservation knowledge. Expert users seem willing to pay for such a service.

We observed that the market demand is a critical success factor for the DURAFILE project; therefore we need to identify a clear target group that could benefit of the new product. Without well-articulated demand for preserved information, there will be no future supply. Stakeholders for digital materials are often diffuse across different communities. The interests of future users are poorly represented in selecting materials to preserve. Trusted public institutions - libraries, archives, museums, professional organizations and others - can play important roles as proxy organizations to represent the demand of their stakeholders over generations. A decision to preserve now need not be thought of as a permanent or open ended commitment of resources over time. In cases where future value is uncertain, choosing to preserve assets at low levels of curation can postpone ultimate decisions about long-term retention and quality of curation until such time as value and use become apparent.

The recommendation for the DURAFILE project would be to focus on the 2.0 approach of Digital Preservation, notably the exchange of Digital Preservation knowledge by empowering a growing community of non-expert and expert users that have in common that *preservation is to share.* The agent technology behind will get accurate Digital Preservation solutions by increasing the awareness and the resources. We propose a similar technological and business model of the virus software that will fight obsolescence aiming at business opportunities of Digital Preservation rather than fight attacks of malware.

8. References

Abrams, S., Kunze, J., Loy, D. (2010) An Emergent Micro-Services Approach to digital Curation Infrastructure. International Journal of Digital Curation, 5(1), pp. 172-186

Allasia, W., Falchi, F., Gallo, F., Meghini, C. (2012). Autonomic Preservation of 'Access Copies' of Digital Contents. The Memory of the World in the Digital Age: Digitization and Preservation Conference, Vancouver, pp. 976-987

Bermes, E., Fauduet, L., Peyrard, S. (2010). A data first approach to digital preservation: the SPAR project. World Library and Information Congress: 76th IFLA General Conference and Assembly

Braubach, L., Lamersdorf, W., &Pokahr, A. (2003). Jadex: Implementing a BDI-infrastructure for JADE agents. EXP - in search of innovation, 3(3), 76–85.

BRTF. (2010). Blue Ribbon Task Force on Sustainable Digital Preservation and Access. Final Report

Challis, D. (2010). Notes on SITS – the Scholarly Infrastructure Technical Summit. Available: http://blogs.ecs.soton.ac.uk/webteam/2010/11/15/notes-on-sits-the-scholarly-infrastructure-technical-summit/

Chmiel, K., Gawinecki, M., Kaczmarek, P., Szymczak, M., &Paprzycki, M. (2005).Efficiency of JADE agent platform. Scientific Programming, 13(2), 159–172.

Curtis, Joseph (2006). AONS System Documentation

de la Rosa, J.L., Trias, A., Ruusalepp, R., Aas, K., Moreno, A., Roura, E., Bres, A., Bosch, T. (2010). Agents for Social Search in Long-Term Digital Preservation. In: Semantics Knowledge and Grid (SKG), Sixth International Conference, Beijing, China, 2010

de la Rosa, J.L. del Acebo, E., Trias, A., Aciar, S., Quisbert, H. (2009). Crew Intelligence Systems for Digital Objects Preservation. Swarm Intelligence Algorithms and Applications Symposium Conference, Edinburgh, United Kingdom, pp. 23-30

DPC. (2006). Mind the Gap: Assessing digital preservation needs in the UK http://www.dpconline.org/docs/reports/uknamindthegap.pdf

ESG. (2007). Digital Archiving Market Forecast. http://www.computerworld.com/s/article/9061238/Digital_Archiving_Market_Forecast

Ferreira, M., Baptista, A.A., Ramalho, J.C. (2006). A Foundation for Automatic Digital Preservation. Ariadne, vol. 48. http://www.ariadne.ac.uk/issue48/ferreira-et-al

Galushka, M., Taylor, P., Gilani, W., Thomson, J., Strodl, S., Neumann, M.A. (2012). Digital Preservation Of Business Processes with TIMBUS Architecture. Proceedings of the 9th International Conference on Preservation of Digital Objects (IPRES2012), Toronto, Canada, pp. 117-125

Gantz, J. F. The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011: International Data Corporation (IDC), 2008

Guoying Liu (2011). The application of intelligent agents in libraries: a survey. In: Program: electronic library and information systems, Vol. 45, No. 1, pp. 78-97

Hedstrom, M., Digital Preservation: A Time Bomb for Digital Libraries, Ed. Springer Netherlands, ISSN 0010-4817 (Print) 1572-8412 (Online)

Hunter, J., Choudhury, S. (2006). PANIC – An Integrated Approach to the Preservation of Composite Digital Objects using Semantic Web Services. International Journal on Digital Libraries, Vol. 6, No. 2, pp. 174-183

Hägerfors, A., Quisbert, H., Nilsson, J. (2009). Agent Technology Supporting DigitalPreservation. International Multi-Conference on Engineering and Technological Innovation Conference, Orlando, United States

Jennings, N. R. (2001). An agent-based approach for building complex software systems. Communications of the ACM, 44(4), 35–41

Luck, M., McBurney, P., Shehory, O., &Willmott, S. (2005). Agent technology: Computing as interaction (a roadmap for agent based computing). AgentLink.

Moreno, A, de la Rosa, J.Ll., Szymanski, B.K., Bárcenas, J.M. (2009). Reward System for Completing FAQs. In: Artificial Intelligence Research and Development. Vol. 202, pp. 361-370

The New Renaissance (2011) The New Renaissance: report of the Comité des Sages. Reflection group on bringing Europe's cultural heritage online. Available: http://ec.europa.eu/information_ society/activities/digital_libraries/doc/reflection_group/final-report-cdS3.pdf

Pearson, D., Walker, M. (2007). Report of the Format Notification and Obsolescence Service (AONSII) https://digitalcollections.anu.edu.au/handle/1885/46645

PROTAGE. (2010). Briefing paper: Potential of software agents in digital preservation. PROTAGE Project http://rahvusarhiiv.ra.ee/files/Potential%20of%20agents%20in%20DP.pdf

Rajasekar, A. et al. (2010) iRODS Primer: Integrated Rule-Oriented Data System. Morgan & Claypool

Ruusalepp, R., Dobreva, M. (2012) Digital Preservation Services: State of the Art Analysis. http://www.dc-net.org/getFile.php?id=467

Strodl, S., Petrov, P., Rauber, A. (2011). Research on Digital Preservation within projects co-funded by the European Union in the ICT programme. http://cordis.europa.eu/fp7/ict/telearn-digicult/report-research-digital-preservation_en.pdf

Aciar S., Debbie Zhang, Simeon Simoff, John Debenham and J. Ll. de la Rosa acknowledged (2007), Informed Recommender: A Recommender System That Bases Recommendations on Consumer Product Reviews, IEEE Magazine in Intelligent Systems, Vol. 22, issue 3, pp: 39-47, May-June 2007, ISSN 1541-1672, Vol: 22, pp 39-40,May-June 2007

Adomavivius G., and Alexander Tuzhilin (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on knowledge and data engineering, 17[6]:734–749, June 2005.

de la Rosa J. Ll. Nicolás Hormazábal, Silvana Aciar, Gabriel Lopardo, Albert Trias, and Miquel Montaner (2011), A Negotiation Style Recommender Based on Computational Ecology in Open Negotiation Environments ISSN: 0278-0046, IEEE Transactions on Industrial Electronics, Vol. 58 (6): 2073-2085, June 2011

Hendler J. and T. Berners-Lee (2010). From the Semantic Web to social Machines: A research challenge for AI on the World Wide Web. Artificial Intelligence 174 (2010) 156-161.

Jennings N. R. and M. J. Wooldridge (1998), Eds. 1998 Agent Technology: Foundations, Applications, and Markets. Springer-Verlag New York, Inc.

Montaner M.; López B; de la Rosa J. Ll (2003); A Taxonomy of Recommender Agents on the Internet, ISSN 0269-2821 Artificial Intelligence Review Vol 19, pp: 285-330,June 2003

Wooldridge M. & Jennings N. R. Agents (1995): Theory and Practice. The Knowledge Engineering Review, vol. 12, No 2, pp 115-152, 1995