WP 4
Dissemination Level: Public

Lead Editor: NAE

2010-05-17

SEVENTH FRAMEWORK PROGRAMME THEME
ICT -1-4.1 Digital libraries and technology-enhanced learning
Project full title: PReservation Organizations using Tools in AAgent Environments
Grant agreement no.: 216746
# Table of Contents

1. Introduction.......................................................................................................................3

2. PROTAGE – the project, its' aims and current status ......................................................4

3. Trust and agent-technologies ...........................................................................................5
   3.1. Trust and Reputation ................................................................................................5
   3.2. PROTAGE Trust Model ............................................................................................5

4. PROTAGE Architecture....................................................................................................7

5. PROTAGE Information Model ..........................................................................................9
   5.1. General .....................................................................................................................9
   5.2. User Settings and Preferences.................................................................................9
   5.3. Preservation Activities ............................................................................................10
   5.4. PROTAGE Recommendation and Certification Base .............................................10
   5.5. Archival Knowledge Bases .....................................................................................11
   5.6. Action Plans ............................................................................................................11
   5.7. External Tools .........................................................................................................11
1. Introduction

This document is the introductory document explaining briefly the PROTAGE project and its main concepts.

The document includes:

- a short introduction to the project and its goals
- a section describing the PROTAGE Trust Model
- description of the PROTAGE Architecture
- the Protage Information Model.
2. PROTAGE – the project, its’ aims and current status

PROTAGE is an EC-funded small targeted research project (STREP) in the seventh framework programme. The project consortium consists of seven partners across Europe: National Archives of Sweden (project lead), Lulea University of Technology (Sweden), National Archives of Estonia, University of Bradford (UK), Fraunhofer Gesellschaft (Germany), Easy Innova (Spain) and GiuntiLabs (Italy). The project started in November 2008 and as a three-year project does finish in October 2010.

The name PROTAGE is an acronym for PRreservation Organizations using Tools in AGent Environments. Thus the aim of the project is to explore possibilities for using intelligent agents to help users in their digital preservation related problems. More explicitly, the project expects that software agents are able to gather digital preservation related know-how about technical tools and intellectual solutions; compare those to a user’s particular preservation needs and recommend actions and tools to solve those.

The first PROTAGE prototype was issued early 2009 and tested in spring 2009. The feedback gathered was evaluated and fed into the development process which now (May 2010) has delivered a final PROTAGE prototype.
3. Trust and agent-technologies

The sociologist Niklas Luhmann said "A complete absence of trust would prevent [one] even getting up in the morning" [Luhmann, 1979]. Trust is necessary in our everyday life. It is part of the "glue" that holds our society together. Without trust, governments could not rule and people cannot work cooperatively together. Trust helps to reduce the complexity of decisions that have to be taken in the presence of many risks. Similarly, reputation is a universal concept that has been present in human societies for a long time. From ancient Greeks to modern days, from Vietnamese to Bedouins, the concept of reputation plays a very important role in human social organization. Reputation is one of the most relevant elements that we use to build trust in others.

3.1. Trust and Reputation

Until recently, both concepts were applicable only to human societies and therefore they were a study field for sociologists, philosophers and psychologists. The eruption of Internet and the emergence of virtual (not necessarily human) societies add a new dimension to these old but very important concepts. The study of trust and reputation has many applications in Information and Communication technology. Trust and reputation systems have been recognized as key factors for successful electronic commerce adoption. Reputation is used in electronic markets as a trust-enforcing, deterrent, and incentive mechanism to avoid cheaters and frauds.

The new paradigm of the so called intelligent or adaptive agents and Multi-Agent Systems (MAS) together with the spectacular emergence of the information society technologies (especially reflected by the popularization of electronic commerce) are responsible for the increasing interest on trust and reputation mechanisms applied to electronic societies. These mechanisms are used by intelligent software agents for searching for trustworthy exchange partners and as an incentive in decision-making about whether or not to honour contracts. Another area of application in agent technology is teamwork and cooperation.

3.2. PROTAGE Trust Model

In the PROTAGE prototype trust is mainly used to ask agents for action plans (or "preservation recipes" from both agents and users. More specifically, we consider a scenario where there are several expert users (and their agents which answer on behalf of them) in several areas of expertise and therefore, when asking those agents for recommendation we need to know how useful is the information received from them in regard to our own specific problem.

The trust that an agent has for others depends mainly on three criteria:

- past interactions (have I used this expert earlier and have I found the advice useful?);
- reputation of the agent (how other users and agents rate that expert and its agent?).
Additionally there are a few other considerations taken into account. First, when a user receives an action plan (preservation recipe) it is not straightforward to know its quality if the user is not an preservation expert and hasn’t used it formerly. So – there is the need to try it and later evaluate the quality of the information. Secondly, it has no sense to ask all available agents for action plans. Usually we should ask the people who we think can answer us better about this subject (the ones that we have higher trust in that certain issue).

To give an example on how the trust model works in practice let’s consider the next example.

Example: The user’s agent asks C and P agents (both are experts in that specific field necessary for the user) for a preservation action plan

- Agent C (trust 5/5) gives us an action plan with a “quality” of 3/5 (Agent C thinks that it works fine for him though he is not fully satisfied) trust * quality = 5/5 * 3/5 = 3/5
- Person P (trust 4/5) gives us an action plan with a “quality” of 4/5 (Person P thinks that this action plan is really good, maybe it could run a little bit faster but it has really delivered good output) trust * quality = 4/5 * 4/5 = 16/25
- Our agent’s priority will be to use the recipe received from Person P because 16/25 > 3/5
4. PROTAGE Architecture

The PROTAGE prototype is a distributed system that involves many users and organisations. The key element in PROTAGE is to manage action plans or the single units of it, the actions. The PROTAGE prototype is a multi-agent eco-system and is built on the top of a multi agent community.

Intelligent agents have to be present and distributed in the computer network that makes use of the PROTAGE prototype. In Figure 2, it is possible to see the exchange of communication between the single clients/peers to recommend the best suitable action plan or action to satisfy the digital preservation needs identified by the PROTAGE prototype itself.

Any client (peer) has its local configurations, local plans (already created and possibly used), and actions. The Multi Agent System (MAS) is the agent technology component present in any client and able to manage the local resources (Local config, My plans, My actions), to interrogate the other clients and to make suggestions.

The Access Point or “parent” (ActionPlanRepository) is in charge to manage the certified action plans and actions, and related profiles. The single clients can suggest to the Access Point (ActionPlanRepository) to certify plans or single actions.

![Figure 2: Parent – peers structure](image)

The figure above does not provide a complete overview of the PROTAGE prototype. Any client/peer, drawn on Figure 2 with related local resources, is a client structured as visible on Figure 3):

- Desktop Application with its database and MAS
- External Tools installed in local, such as tools for antivirus check, metadata extraction, file migration etc.
- The MAS is in charge of the communication to the Access Point or ActionPlanRepository (parent) and to the Service Provider or WebServiceRegistry (server);
- The Service Provider is in charge to provide necessary web services to execute tasks in an Action Plan (if necessary tools are not available in the client’s computer)

![Diagram: Servers – client structure]

Figure 3: Servers – client structure
5. PROTAGE Information Model

The information model in the PROTAGE consists of several congruent entities. Each of them has been designed to conglomerate specific informational parts need in the PROTAGE prototype. The overall information model is depicted in (Figure 8 Error! Reference source not found.) and the components of the information model are described in detail, in later sections.

### 5.1. General

The PROTAGE Data Model (Figure 9) consists of three main entities:

- User Settings and Preferences;
- Preservation Activities;
- PROTAGE Recommendation and Certification Base.

In addition PROTAGE benefits of external archival knowledge bases which, being an important information providers for PROTAGE, are also included into the data model.

While in traditional applications the relation between different entities is rather strict in PROTAGE the "glue" between the entities are software agents which by accessing all the four data components are able to provide result sets (and thus preservation solutions) which best fit the individual users needs.

### 5.2. User Settings and Preferences

The User Settings and Preferences part of the PROTAGE data model is the main entity which is used by a user to identify him/her to the general PROTAGE environment and to help agents in making decisions which suite best to the individual users need.

The User Settings and Preferences entity consists of four sub-entities:
• User Profile – general information about the user’s affiliation, location, preservation needs and user type (if the user is an individual or organisation);

• Collection Profile – information about the user’s digital assets which are destined to be preserved with the help of PROTAGE. Every user can describe multiple collections and every collection includes among other information also the descriptions of single computer files (the description is gathered by agents);

• System Profile – information about the user’s computing environment (e.g. network quality, memory and processor). This profile also lists preservation tools which are installed locally at the user’s computer (Local Preservation Tool);

• Trusted Profile – information about the user’s trust towards other PROTAGE users (e.g. which types of users to trust). Particularly a user can set his/her trusted connections: which PROTAGE user groups or individual users are trusted in delivering digital preservation related advice and best practices.

In the PROTAGE architecture the User Settings and Preferences are maintained at a user’s computer and the personal details are not visible to other PROTAGE users.

5.3. Preservation Activities

The Preservation Activities entity assembles information about the preservation tasks a user is supposed to execute in order to preserve his/her collections accessible long term.

The Preservation Activities entity consists of four sub-entities:

• Tool – a singular preservation tool (both executable as a web-service or locally) which is used to execute an action at a given time;

• Action – a singular preservation action like migration, characterisation, etc. An Action is described in a technology independent manner to enable the use of different preservation tools or services;

• Action Plan – composition of one or many Actions and can be related to one or many collections of certain type;

• Execution Detail – detailed information about the execution of an entire Action Plan, singular Actions or Tools;

In the PROTAGE architecture the Preservation Activities information is usually created in the user’s computer. The user can decide to share his/her information about tools, actions and action plans and in that case the information is synchronised into the wider PROTAGE community servers.

Additionally the PROTAGE user group also includes advanced users (digital preservation experts) who contribute with their knowledge and also maintain, check and organise the information delivered by ordinary users.

5.4. PROTAGE Recommendation and Certification Base

The PROTAGE Recommendation and Certification Base gather the user ratings and certificates of all other entities in the PROTAGE environment: action plans, actions, tools, users etc.

The difference between Ratings and Certificates is:
• User Ratings – every PROTAGE user can rate the entities available in the PROTAGE environment (action plans, actions, tools, users), the ratings are visible to other users intending to use the same action plans, actions or tools;

• Certificates – to build additional trust into PROTAGE advanced users (e.g. Archives, Libraries) can issue certificates which assure user that a particular action plan, action or tool has been tested and approved by the organisation.

In the PROTAGE architecture the Recommendation and Certification Base is located in the PROTAGE community servers.

5.5. Archival Knowledge Bases

To access necessary detailed information about preservation tools, web services and archival file formats PROTAGE benefits of external Archival Knowledge Bases which have been set up separately. Of special interest for PROTAGE is the reuse of information available at the PRONOM registry but the data model enables adding new sources of information when those emerge.

Currently PROTAGE has identified three major areas where external knowledge bases should be used:

• information about archival file formats;
• information about preservation tools;
• information about preservation web services.

The use of such external information helps PROTAGE to essentially include more information which can be used by agents to make justified and user-specific decisions.

In future the scope of such knowledge bases can be broadened to include also other types of knowledge (f. ex. information about national legislation).

5.6. Action Plans

An action plan (i.e. a preservation action plan) is a compiled set of instructions (the actions) that instruct what should be done to a collection of digital objects. The user will have the possibility to download action plans from external knowledge bases and in some cases add/remove actions from the plan. The user will also be able to create a new action plan of the actions that is downloaded from the knowledge bases. An action plan will be connected to one or several data profiles to get information about the collections of interest. It will also use the information from the user profile to know for example which kind of tools that are approved for use and in which way the actions will return feedback to the user.

5.7. External Tools

As described in the PROTAGE Information Model (section 4.3) the crucial element in PROTAGE is the definition and execution of action plans.

In the model it is visible that PROTAGE is seeking solutions for two, in some ways contrary, problems at once. On one side the PROTAGE action plans should be technology independent and allow the choice of different tools for the same actions thus ensuring the longevity of these action plans. On the other side it is recommended, that PROTAGE is able to discover and execute the tools automatically, thus making the use of PROTAGE as simple as possible.
To cope with the task PROTAGE is going to use a framework which consists of two parts:

- Preservation tool and service registries;
- Preservation action ontology.

The **preservation tool and service registries** describe in detail the tool and if possible also details for its execution. PROTAGE should be able to automatically contact those registries and find suitable tools for the users actions. Currently there is one promising external registry which could be used for that – the PLANETS Core Registry. The PROTAGE project will investigate possibilities for using the data in the registry to become the primary source for information on preservation tools and services.

To automatically find out which tools or services are suitable for an action PROTAGE is going to develop a **preservation action ontology**. It will be possible for PROTAGE users as well as agents and administrators to map single tools, services and actions to the ontology thus enabling an automated tool discovery for each action.

One possible description on the management of the ontology and adding new tools to the registry is available in the PROTAGE Application Scenarios (Section 3).