



Project Number	IST-2006-033789
Project Title	Planets
Title of Deliverable	Report on “description language for PA tools, blueprint for development of new PA tools and gap analysis”
Deliverable Number	D4
Contributing Sub-project and Work-package	PA/2
Deliverable Dissemination Level	External Public
Deliverable Nature	Report
Contractual Delivery Date	May 31 st 2008
Actual Delivery Date	May 22 nd 2008
Author(s)	KB

Abstract

This document contains a report on the status of the PTDL language, the blueprint and the gap analysis. This report will provide a summary of the content and the current status of these deliverables, as well as an overview of the context and dependencies of the deliverables within the project.

Keyword list: Report, description language, blueprint, gap analysis, preservation action tools

Contributors

Person	Role	Partner	Contribution
Madelon Hoedt	Author	KB-NL	
Frank Houtman	Review/author	KB-NL	
Caroline van Wijk	Review/author	KB-NL	

Document Approval

Person	Role	Partner
Frank Houtman	subproject leader PA	KB-NL
Christen Hedegaard	External reviewer	KB-DK

EXECUTIVE SUMMARY

This document contains a report on the status of the PTDL language, the blueprint and the gap analysis, as well as an overview of the position of these three deliverables within the PLANETS project.

The second chapter of this report contains an overview of the relation and the use of these three deliverables within the PA/2 workpackage, the PA subproject and the PLANETS project as a whole, followed by information on each of the deliverables and an outlook to the future in the final chapter.

PTDL (an acronym for Preservation Tool Description Language) is a format for the exchange of messages between the PA tools registry, which will be developed by the PA/3 work package, and the preservation planning tools provided by the PP sub-project.

The blueprint for the development of new preservation action tools consists of a list of functional requirements for tools that are designed within the Planets project, as well as the workflow that needs to be followed when incorporating newly developed tools into the Planets framework.

The gap analysis concerns itself with the investigation of the use of file formats. Based on this study, an analysis can be made of existing preservation action tools that can be incorporated into the Planets framework, as well as an inventory of actions and tools that are missing and thus need to be built by the workpackages PA/4 and PA/5.

The blueprint is now in its second iteration, containing requirements for both migration and emulation tools. The gap analysis is now in its second iteration, as well, but is not yet able to show gaps in the tool provision. The third iteration will contain new analyses of the list of file formats.

Where PTDL is concerned, the function of this language might have been changed during the process of its development. Next to this, it remains to be discussed with IF and Tessella – as developers of the PA tool registry - whether XML is the most efficient means of communication between the planning tool and the PA tool registry. Until more clarification about the technical environment, issues and solutions of IF and Pronom becomes available, further work on PDDL will be suspended.

TABLE OF CONTENTS

1.	<i>Introduction</i>	5
1.1	Objective	5
1.2	Scope	5
1.3	Document overview	5
2.	<i>Context and dependencies</i>	6
2.1	Introduction	6
2.2	Within the PA/2 workpackage	6
2.3	Within the PA subproject	6
2.4	Within Planets	7
3.	<i>Report on PTDL</i>	8
3.1	Introduction	8
3.2	Background	8
3.3	Current status	8
3.4	Next steps	8
4.	<i>Report on Blueprint</i>	10
4.1	Introduction	10
4.2	Background	10
4.3	Current status	10
4.4	Next steps	10
5.	<i>Report on the gap analysis</i>	11
5.1	Introduction	11
5.2	Background	11
5.3	Current status	11
5.4	Next steps	11
6.	<i>Conclusions and outlook</i>	12
7.	<i>References</i>	12

1. Introduction

1.1 Objective

The objective of this document is twofold. First and foremost, it provides information on three of the deliverables in workpackage PA/2. Secondly, this report contains an overview of the context and dependencies of the PTDL, blueprint and gap analysis within the larger framework of the Planets project.

1.2 Scope

Workpackage PA/2 is developing a theoretical model for the development of preservation action strategies and steer the development of preservation action tools and services. The deliverables within this workpackage concern themselves with the description of preservation action tools on several levels: metadata; the functional requirements of the tools; and information on the relation of the tools to the Planets framework.

The aim of this report is to describe the current status and future of each deliverable and to show the function of the deliverables within a larger framework.

1.3 Document overview

This report consists of six chapters. The first chapter contains the introduction to this document. The second chapter holds an overview of the deliverables within the PA/2 workpackage, within PA and within the Planets project as a whole. The third, fourth and fifth chapter provide information on the three deliverables, their current status and the next steps. The final chapter then contains some conclusions and a look at the future.

2. Context and dependencies

2.1 Introduction

Several levels of dependencies can be distinguished where the three deliverables, described in this report, play a part. In this chapter, these dependencies will be examined more closely. Firstly, the role of the three deliverables within the PA/3 workpackage will be described, followed by a discussion on their role within the PA subproject and Planets as a whole.

2.2 Within the PA/2 workpackage

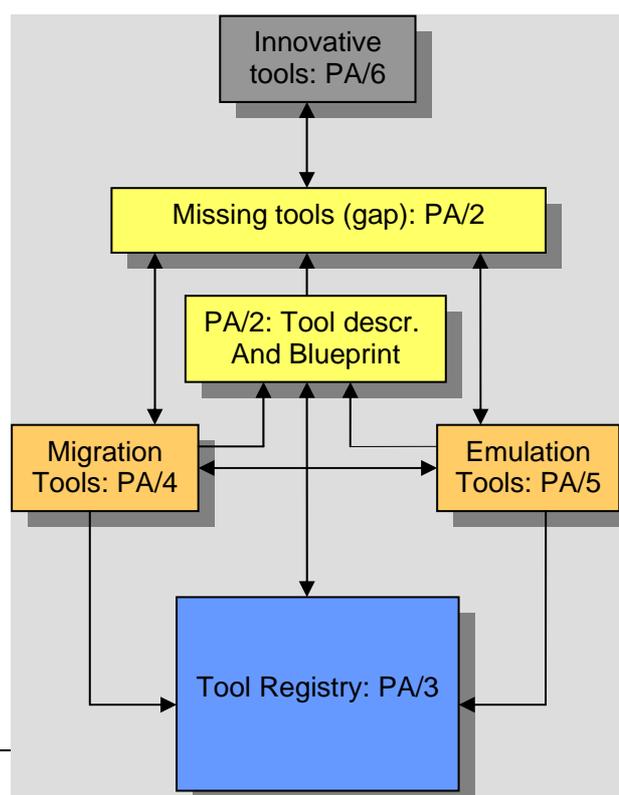
All three deliverables within this workpackage deal with the description of preservation action tools on three different levels. PTDL provides a description of the properties of both migration and emulation tools. The blueprint, in turn, lists the functional requirements to which the different tools should adhere, whereas the gap analysis is an expression of the different tools that still need to be designed, following the guidelines set by the blueprint.

In this way, one could say that the PA/2 workpackage is an important link within the PA workpackage, as well as in itself. On a basic level, all three deliverables are concerned with the same issue, namely description. The description in PTDL is geared towards the properties of the preservation action tools. The blueprint, then, describes the functional requirements, whereas the gap analysis provides a description of the missing pieces and the necessary tools.

The dependencies between the three deliverables can be easily seen: when combined, PTDL, the blueprint and the gap analysis set the rules to which the PA tools developed within the framework of Planets need to adhere, thus forming a basis for all tools designed with the project.

2.3 Within the PA subproject

As stated in the previous paragraph, the three deliverables fulfil an important purpose within the PA/2 workpackage, as well as in the PA subproject as a whole. The following image provides some information on the dependencies within this subproject:



Taking the gap analysis as a starting point, the dependencies here are obvious: by looking at the current situation of preservation action tools, an assessment can be made as to which tools are available (and need to be wrapped as to fit into the Planets framework), and which tools are missing. The issue of missing tools has implications for the PA/4 and PA/5 workpackages, concerned with the development of new migration and emulation tools (as is shown in the graph).

The blueprint, then, functions as a basic rulebook for the output of PA/4 and PA/5, as well as the output of external developers, listing the functional requirements to which the tools need to adhere. It also holds the rules for wrapping these tools so they can fit into the tool registry of PA/3, thus forming a link with this workpackage.

Where PTDL is concerned, this format provides a way to describe the tools, made by PA/4 and PA/5. This description can then be used within the tools registry, providing information on the tools for both users and other systems.

2.4 Within Planets

As shown in the previous paragraph of this chapter, the role of PTDL, the blueprint and the gap analysis is closely linked to the tools registry, which is to be developed by the PA/3 workpackage. Since this registry is one of the more significant future results of the Planets project as a whole, the three deliverables provide additional functionality to the tools registry.

As stated in the first paragraph, concerning the role of the deliverables within the PA/2 workpackage, they can be described as the basic rulebook for the tools, developed within the Planets framework. The results of the gap analysis amount to a list of tools which need to be developed, whereas the blueprint describes the requirements of these tools, concerning their functionality, as well as the way in which the tools need to be wrapped and fitted into the bigger framework of the tools registry. PTDL, then, enables the description of the tools, both on a human-readable and machine-readable level.

3. Report on PTDL

3.1 Introduction

This chapter contains a brief description of the background of the PTDL description language and the achievements so far, as well as some information on the plans concerning the next iteration of this deliverable.[1]

3.2 Background

The preservation action sub-project is responsible for the development and availability of preservation action tools and services in the Planets project. Workpackage PA/2 is responsible for the development of preservation action strategies and the development of preservation action tools and services.

The Preservation Tool Description Language, or PTDL, is an XML-based language, used to describe the properties of preservation action tools. Potentially the language can be used in two ways: PTDL can serve as a means of communication between the preservation planning system and the tools registry, providing machine-readable information on the contents of the PA tools registry, as well as human-readable information on the tool within the registry. The XML schema, in turn, could be used as a generic model for the description of functions, expected performance outcomes, and appropriate uses of preservation action tools.

3.3 Current status

In the first iteration of the development of PTDL, which was delivered in m15 (August 2007), the primary goal was to create an inventory of relevant properties needed for describing preservation action tools. The second iteration, then, provides an extension of this initial model.

The PP/4 working group has already made use of PTDL to describe tools that they have tested. Based on these tests and the model for the PA tools registry (as proposed by the PA/3 workpackage), some modifications have been made in the original model.

The basic model of PTDL is designed in such a way that it can accommodate a description of both tools for objects and tools for environments (i.e. migration and emulation). These more specific descriptions can be derived from a general PA tool description. This model is then validated by testing it, through describing of both types of tools and checking whether the information needed fits into the model. In this way any potential gaps are exposed and it becomes visible whether or not further modifications of the model are necessary.

In the development of the second iteration, the main goal has been to make an inventory of relevant properties needed to describe PA tools. The properties identified so far have been incorporated in a schema, written in W3C XML Schema [2]. In the current version, the language is not based on the RDF or OWL framework [3], [4]. Also, all data types used in the language (except for the basic ones that are part of the XML-Schema definition) are defined in the schema itself. In subsequent versions, the shared use of schemas for data types within Planets might be desired.

3.4 Next steps

The development of the next iteration of the PTDL has been frozen. The reasons for stopping the development of the PTDL for a limited period of time is organisational and with respect to content.

The first reason to temporarily stop the development of the PTDL is its dependency on the progress of other nodes in the Planets system. The PTDL has been developed to perform a communication role between the Planets planning tool PLATO and the PA tool registry following

the work description of the PTDL in the DoW. Unfortunately, the development of the PA tool registry has been delayed and the PTDL can not be adapted yet to the requirements that come from the PA tool registry as this registry has not been developed and tested yet.

Another matter – a matter concerning the content of the PTDL- is that during a PA/3 and IF(ARC) face to face meeting in August 2007, IF mentioned that the XML-based PTDL might not be the most efficient way – performance wise - of communication between the planning tool and PA tool registry. As the PA tool registry is now being developed by Tessella based on PRONOM, the PA/2 working group would like to discuss the usage and efficiency of an XML-based language for communication purposes with Tessella when they have made progress with the development of the PA tool registry. It is possible that another means of communication between the planning tool and PA tool registry will be chosen from an efficiency and performance point of view. Further development of the PTDL would then only be done to provide import and export functionality for the PA tool registry of tool descriptions in XML format.

As soon as the PA tool registry has been developed and is ready for testing, the work on the PTDL can be resumed, provided that the PTDL still has a role in the Planets system. As mentioned before, this could be either as a means of communication between the PA tool registry and planning tool or as the format to which import and export functionality of the PA tool registry will comply.

4. Report on Blueprint

4.1 Introduction

This chapter contains a brief description of the background of the blueprint and the achievements so far, as well as some information on the plans concerning the next iteration of this deliverable.[5]

4.2 Background

The blueprint is used in the development of new preservation action tools, providing a list of functional requirements that the tools, designed within the Planets project, should offer. It also presents the workflow that should be followed when incorporating newly developed tools into the Planets framework.

During the Planets project, different software programs and tools will be developed to perform preservation actions by different companies and institutions. This distribution of tool development could result in a difference in the standards and techniques used. By implementing a framework of required functionalities, a consistent behaviour of the Planets system, as well as a consistent level of quality, can be ensured. The blueprint is used to list these requirements for both migration and emulation tools, developed within the Planets project.

4.3 Current status

Currently, the blueprint mainly consists of a list of the functional requirements for both emulation and migration preservation action tools within the Planets project. An outline exists for the workflow of incorporating a new PA tool within the proposed framework. However, these are both only basic models that still need to be validated, either in cooperation with other Planets partners, or by comparisons with other, similar, models.

4.4 Next steps

The next steps for the blueprint are to provide this validation on several accounts:

- Validation of the requirements lists by Planets partners that are developers and institutions outside of Planets that deal with digital preservation and develop preservation action tools;
- Validation of the requirements lists by comparison of the requirements and preservation action tools (both tools for environments and tools for objects) that were developed by international institutions that have developed these tools for preservation action purposes;
- In this iteration of the Blueprint and for operational reasons, we made a deliberate decision to separate the requirements for tools for objects and the requirements for tools for environments in different tables. However, as many requirements are identical, in a next iteration we will merge the requirements into one table again;
- In a next iteration the (sub) requirements in a measurable manner (SMART).

The validation will then lead to an extension and elaboration of the current model.

5. Report on the gap analysis

5.1 Introduction

This chapter contains a brief description of the background of the gap analysis and the achievements so far, as well as some information on the plans concerning the next iteration of this deliverable.[6]

5.2 Background

Within the Planets project, the PA subproject is responsible for providing the tools that are required to perform preservation actions. In order to do so, existing tools can be wrapped and made available within the Planets framework. If no tool for a certain action exists, new tools can then be built.

To be able to tell what kind of preservation actions should be provided by the system, and thus which tools should be build or wrapped, one needs to know what file formats are used for storing the information that needs to be preserved. The gap analysis provides an inventory of the file formats, used by the various institutions to archive their digital objects. Based on this inventory, several analyses can be made, thus providing information on the existing tools, as well as on the tools that should then be developed within the Planets project.

5.3 Current status

The current version of the gap analysis consists of the so-called “master list”, an inventory of the file formats used by the various types of cultural heritage institutions. This list is based on the results of several surveys that have been carried out in a number of different countries, as well as desktop research. It needs to be said that in a few instances, the response to these surveys has been quite low. This has made it very difficult to obtain the information needed for the gap analysis.

The first iteration of the gap analysis consisted of the results of two surveys, distributed amongst the members of Planets and various cultural heritage institutions in the Netherlands, respectively. These results form the foundation of the aforementioned inventory of file formats. For the second iteration, this model was extended with the results of desktop research and additional surveys, distributed in the UK and Denmark.¹ Based on the information, gathered in this manner, and the list of formats, three analyses have been made. These analyses provide some insight into the amount of files, available in a certain format; the type of content stored within these formats; and the types of institutions that make use of certain formats.

5.4 Next steps

As stated above, the second iteration of the gap analysis consists of three analyses of the gathered data, providing information on the amount of files, available in a certain format; the type of content for which the formats are used; and the type of institutions that make use of a certain format.

However, the inventory of the gap analysis can be approached from many different viewpoints and allows for many different analyses. For the next iteration, the data on the file formats will be expanded and new analyses will be carried out. One of the important issues here is to examine the file formats that may be stored by only a few institutions, but are potentially important for this specific type of organisation. The additional information, obtained from the new data and analyses, can then be combined in order to locate the gaps in the availability of preservation action tools.

¹ We would like to thank Adrian Brown (The National Archives), Helen Hockx-Yu (British Library), Karen Williams and Jette G. Junge (State and University Library Denmark) for their help with the UK and Denmark surveys.

6. Conclusions and outlook

The progress of the PA/2 workpackage is satisfactory in general. New iterations of all deliverables have been produced and finalized. The Blueprint (D2) saw its second iteration. This iteration contains not only requirements for migration tools, but also for emulation tools.

The second Planets year also provided new considerations which in turn lead to different professional perceptions. For example, this was the case with the gap analysis (D3) where a more precise definition of the expected results was responsible for a much greater need for and considerable extension of the data collected so far. This also means that the second iteration of the gap analysis is still not capable of identifying gaps in tool provision.

The workpackage also had to deal with different perceptions and changes on the Planets project level. As a result the function of the PTDL as an exchange language between PA tool registry and planning tool of the PP subproject could have been changed. Next to this, it remains to be discussed with IF and Tessella – as developers of the PA tool registry - whether XML is the most efficient means of communication between the planning tool and the PA tool registry. Most changes have been dealt with. Some deliverables have been slightly adapted as was the case with the gap analysis. The appearances of new iterations of PTDL have been frozen. The new circumstances which lead to this temporarily halt have been analyzed and written down in a paper. Until more clarification about the technical environment, issues and solutions of IF and Pronom becomes available, further work on this deliverable will be suspended.

In the next Planets year, we hope to integrate the results of the different PA/2 deliverables. The next iterations of the GAP analyses should lead to more concrete gaps in tool provision. These gaps, in turn, should provide PA/4 with clear indications about what kind of migration tools should be adapted or developed. It is also intended to test the blueprint for the development of new preservation action tools along these lines.

Finally, depending on available knowledge and understanding about the PA tools registry and developing work within the IF subproject, we will write a new, fully adapted version of the PTDL.

7. References

[1] PTDL, deliverable PA/2 – D1

<http://www.planets-project.eu/private/pages/wiki/index.php/PTDL>

[2] World Wide Web Consortium: XML Schema: W3C recommendation, 28 October 2004.

<http://www.w3.org/TR/xmlschema-0/>

[3] World Wide Web Consortium: Resource Description Framework: W3C Recommendation, 10 February 2004.

<http://www.w3.org/TR/rdf-schema/>

[4] The OWL Services Coalition, OWL-S: Semantic Markup for Web services, DAML, 24 July 2004

<http://www.daml.org/services/owl-s/1.1B/owl-s/owl-s.html>

[5] Blueprint, deliverable PA/2 – D2

<http://www.planets-project.eu/private/pages/wiki/index.php/Blueprint>

[6] Gap analysis, deliverable PA/2 – D3

http://www.planets-project.eu/private/pages/wiki/index.php/Tool_Gap_Analysis