

PERICLES - Promoting and Enhancing Reuse of Information
throughout the Content Lifecycle taking account of Evolving
Semantics
[Digital Preservation]

DELIVERABLE D10.1
Initial Version of Exploitation Plan



GRANT AGREEMENT: 601138

SCHEME FP7 ICT 2011.4.3

Start date of project: 1 February 2013

Duration: 48 months



Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination level		
PU	PUBLIC	X
PP	Restricted to other PROGRAMME PARTICIPANTS (including the Commission Services)	
RE	RESTRICTED to a group specified by the consortium (including the Commission Services)	
CO	CONFIDENTIAL only for members of the consortium (including the Commission Services)	

Revision History

V #	Date	Description / Reason of change	Author
0.1	4/7/2014	Initial draft	Alex Papadimitriou
0.3	18/7/2014	Structure and content amendments	Simon Waddington Christine Sauter
0.5	25/7/2014	Content review and text additions	Fabio Corubolo Anna-Grit Eggers Jean-Pierre Chanod
0.7	29/7/2014	Content review	Christine Sauter John McNeil
1.0	30/7/2014	Version before final check for submission	Alex Papadimitriou

Authors and Contributors

Authors

Partner	Name
DOTSOFT	Alex Papadimitriou (AP)
KCL	Christine Sauter (CS)

Contributors

Partner	Name
KCL	Mark Hedges (MH)
KCL	Simon Waddington (SW)
SpaceApps	David De Weerd (DDW)
TATE	Pip Laurenson (PL)
TATE	John McNeill (JM)
ULIV	Maureen Watry (MW)
ULIV	Fabio Corubolo (FB)
UGOE	Anna-Grit Eggers (AGE)
XEROX	Jean-Pierre Chanod (JPC)
ULIV	Adil Hasan (AH)

Figures

<i>Figure 1: Digital Data Growth</i>	15
<i>Figure 2: Chart showing the fraction of research data repositories hosted by institutions having the skills for data management</i>	37

Tables

<i>Table 1: Glossary</i>	6
<i>Table 2: A list of exploitation opportunities.</i>	33

Contents

Glossary	6
1 Executive Summary	7
2 Introduction & Rationale	9
2.1 Context of this deliverable.....	9
2.2 What to expect from this document	10
2.3 Document structure	10
3 Preliminary research results	11
3.1 Brief introduction to long-term digital preservation.....	11
3.2 Policies, initiatives and support in EU and the U.S.	13
3.3 The DP market	14
3.4 Market instruments.....	17
3.4.1 <i>Free and Open Source solutions</i>	18
3.4.2 <i>Commercial solutions</i>	18
3.4.3 <i>Training providers</i>	19
3.4.4 <i>Standards, guidelines and certification</i>	21
4 Our exploitation approach	22
4.1 Introduction	22
4.2 Anticipated exploitable foreground	23
4.3 Potential market sectors	25
4.3.1 <i>Space science</i>	26
4.3.2 <i>Archives, Libraries, Museums (ALMs)</i>	27
4.3.3 <i>Higher education</i>	27
4.4 Potential cross-sector opportunities	29
4.4.1 <i>Services and facilities</i>	29
4.4.2 <i>Training</i>	29
4.4.3 <i>Audit and compliance</i>	30
5 Exploitation plan	32
5.1 Introduction	32
5.2 Internal Project Activities	32
5.2.1 <i>Sounding out the exploitation scope</i>	32
5.2.2 <i>IPR planning and auditing</i>	34
5.2.3 <i>Market Scans</i>	35
5.2.4 <i>Benchmarking</i>	38
5.2.5 <i>Internal monitoring and updating the exploitation plan</i>	38
5.3 External Activities	39
5.3.1 <i>Technical solutions</i>	39
5.3.2 <i>Knowledge assets</i>	40
5.3.3 <i>Interaction with facilitators</i>	41
5.4 Implementation	44
6 Conclusion	46

Glossary

Abbreviation / Acronym	Meaning
ALM	Archives, Libraries, Museums
CoP	Communities of Practice: in the PERICLES context this refers to expert groups composed of practitioners in the target communities
DP	Digital Preservation
DPS	Digital Preservation System
Exploitation	Stable sustainability of project results
LTDP	Long Term Data Preservation
OAIS	Open Archival Information System
PET	PERICLES Extraction tool

Table 1: Glossary

1 Executive Summary

PERICLES is an ICT Research and Development project funded by the European Commission. PERICLES is an acronym which stands for Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics.

Investigating the needs of modern preservation and archiving of digital repositories, it quickly becomes obvious that traditional practices and technologies are not cut out to meet the future demands that can easily be anticipated when looking at the fast development of user and technical requirements. Clearly, a new approach is needed to assure that the data generated today is available and useful tomorrow for the next generation of users. As scientific, engineering, and media data and their related metadata are generated across different lifecycle phases, in a continually evolving environment, the concept of a fixed and stable “final” version becomes less relevant. Long-term sustainability mechanisms are needed to deal with technological obsolescence (e.g. of formats, software), “semantic decay” of digital assets (understanding changes over time, e.g. due to changes in use of terminology), and discipline and societal changes as the data evolves across different phases of the information lifecycle. Maintaining the complex relations and dependencies between the components of this ecosystem is key to achieving “preservation by design”, through models that capture intents and interpretation contexts associated with digital content.

The project will develop novel components as a result of their research results and integrate state-of-the-art components such as data management solutions (e.g. iRODS) that virtualise the storage or distributed computation services (e.g. the cloud computation services) to form the basis for exemplars capable of managing data and change through policy and metadata lifecycle management.

This document reflects different ways to exploit the outcomes of the project. It will also place the project activities in the wider R&D and commercial context. It constitutes an initial version of the PERICLES exploitation plan, currently based on the anticipated outcome. It will be adapted to accommodate and implement the actual results from the research as the project progresses. At this stage the exploitable achievements are mainly projected outcomes such as system-level components (e.g. appraisal tools), individual technologies (e.g. semantic analysis, dependency models), methods, and best practice (e.g. guidelines). This document investigates the scope of exploitation and provides an outline of activities to be carried out as soon as the maturity of results comes close to publication. Naturally, exploitation activities rest on dissemination activities which aim at raising the awareness for the project, its goals and results in the respective target communities.

One important consideration for the plan is the fact that, the results of the projects are very different in nature. Besides product entities that could easily be commercialised, we will define approaches and requirements which can stimulate the market to use them for the production of supporting components.

Our approach is two-fold: to create an awareness of the sustainable approach investigated and tested by PERICLES amongst relevant sectors of the market and use domains, and through it instigate a demand which will be the motor for a successful commercialisation of tools, services and systems informed by the PERICLES preservation system prototypes. In parallel we want to identify potential opportunities. This includes identifying potential markets by getting feedback from a wide group of interested parties.

The deliverable establishes the grounds that will guide the PERICLES exploitation activities around developed tools, research results and best practice guidelines:

- Contextualisation and positioning: What has been done so far in the digital preservation community in terms of both research and implementation aspects?

- Target groups and communities¹: What are the anticipated exploitation project results and who may be interested in them in selected markets?
- Strategy: What kinds of activities promise the best exploitation results for the different types of outcome?
- Appraisal: How will the project results be validated against the target community needs?

Within that framework, the consortium has identified four main activity areas to support and enhance post-project exploitation with the aim to maximise the project results. The first area targets the exploration of communities relevant to digital preservation. These communities hold the main target groups to which the results of the project are mainly addressed.

The objective of the second area is to promote, where applicable, the use of tools, technologies and applications developed during the project to identified target groups such as: the academic and research stakeholders of digital objects to be archived and reused over a long-term period, developers and the industry in general. To be able to test the research and development results in practice environments, the project investigates two case studies: Media and art as one use domain, and space science as use case within the science domain. However, partners will also ensure that the results are relevant to other domains that like the two case studies face complex preservation issues.

The objective within the third area is to create two-way communication channels with stakeholders, academic communities and industry for exploiting project outcomes and conclusions within the wider digital preservation context.

Finally, within the fourth area the consortium will ensure that the commercially exploitable products of the project are available to be used in a commercial context, in the research community and in the Open Source community.

Following the scope of exploitation for PERICLES as described in this document, identifying exploitable foreground that results from the project and determine the appropriate activities for exploiting them will be an ongoing process. Finally, it should be mentioned that an important part of the exploitation plan is the prior dissemination activities undertaken in work package 9, which will contribute decisively in communicating the project and making it successful.

¹ Target communities: communities representing user domains, e.g. libraries, museums, science institutions, financial services, health domain.

Target groups: more specific, i.e. stakeholders within or across the communities, e.g. archivists, producers, scientists, or software developers, IT solution and service companies, consultants.

2 Introduction & Rationale

2.1 Context of this deliverable

This document is a Work Package 10 deliverable and contains an analysis of recent research advances in EU projects dealing with digital preservation, a presentation of selected domains and markets where PERICLES outcomes could have an impact on and finally a first version of the consortium's exploitation plan. The document will serve as a guide to project partners in terms of keeping a focus on exploitable foreground, early identification and participation in possible exploitation activities and keeping in touch with actual market needs by utilising structures such as the Communities of Practice.

One of the main advantages of the consortium is based on the very different background of the partners, who as a result can contribute towards addressing diverse market sectors and user domains. Among the members of the consortium are universities, SMEs, public institutions and large organisations, which can all cater for different communities with different exploitation approaches and orientation.

As can be expected SMEs, large industry and other IT partners will firstly exploit the results of PERICLES internally in an attempt to improve their current operation and position within the market, as well as externally by pursuing new markets based on the advantage that these results are offering them, may those be any type of results, including models, methods, white papers and patents.

On the other hand, universities can exploit the results of the project in a way that complements the exploitation plan of the SMEs. The focus of the universities will be to take the research results and incorporate them in their curricula or even as a separate subject that will give them a competitive advantage against other universities. Additionally, universities can use the results of PERICLES for continued research. Another exploitation opportunity for these partners is expert training. For universities there is a potential to offer specialist courses (e.g. Masters) that can be used to generate income as well as widening the uptake of PERICLES results across a wide range of sectors. E.g. KCL runs MA on digital asset management which is attended by students from digital content industries.

Finally, public institutions and organisations will exploit the results in the sense that as partners they will be the first to be able to put into practice the outcomes of the project and customise them early to their own standards and needs. They will enhance their profile as leading institution in their field by setting examples of best practice and act as peer advisor to their colleagues in their domain.

The fact that the nature of the project focuses on both research and development, enables each partner to form the exploitation plan to accommodate the remit of their respective organisation in the best possible way. Based on the twofold nature of the project the strategy which each partner determines will also aim to successfully continuing the project even after its completion both in terms of refining the research areas employed in PERICLES and in developing the system and tools further for application in different fields.

Another key aspect of the project determines further the scope the exploitation activities: in order to validate the results in the practice field and to anchor models and services in the real world of user domain, two case studies form part of the project: media & art represented through TATE and (space) science represented through B.USOC.

Both research and development will, however, go a step further: after instantiating generic results in the two case studies and applying domain-specific parameters in test-beds, the prototypes are then to take into account applicability in other domains as well, in particular fields which share challenges

identified for the two case studies. This route needs to be supported by the dissemination of results into other domains and by more focused feedback from the Communities of Practice (COP²), the Project Advisory Board, the demonstration of the tools and applications in the training sessions (WP8) and lastly by the exploitations activities, in particular in the commercial area.

2.2 What to expect from this document

The main objective of the present exploitation plan is as a first step placing the project in the context of the market, identifying potential opportunities, and scoping out a range of appropriate and promising activities for the anticipated foreground. This document will be updated every 12 months during the project to reflect the developments in project ideas and technologies, the market for potential products and the feedback from potential consumers.

Although the plan is only preliminary, it constitutes a useful tool which will allow the development of exploitation activities during and beyond the project's lifetime leading to sustainable outcomes and exploitation of these.

We should note that the present exploitation plan cannot in under any circumstances be considered as a legally or a binding document for the consortium as to the outcome mentioned.

2.3 Document structure

This section presents the content of the following chapters by providing an overview of what each one entails. Specifically:

In Chapter 3, following the introductory part, the results of the first desktop research on similar EU projects which engage in the digital preservation field will be summarised, outlining what role both research and development have played in forming today's Digital Preservation market.

Chapter 4 focuses on sounding out potential market fields for satisfying current needs, exploring the potential impact of PERICLES outcomes in these markets and comparing them to other existing tools, which will ultimately determine the basic scope of action.

Since PERICLES is both a research and development project, partners need to synchronise their efforts with market needs. It has therefore been decided that the exploitation efforts need to progress in an iterative manner. This iterative process is described in Chapter 5.

² In the PERICLES context this refers to expert groups composed of practitioners in the target communities.

3 Preliminary research results

3.1 Brief introduction to long-term digital preservation

In this section of the deliverable, the results of a first desk research on the evolution of the digital preservation field by examining past and present European and U.S. funded projects, their contributions and results have been summarised

Over a decade ago, and specifically in 2001, the European Commission decided to fund under its FP5 framework, the Electronic Resource Preservation and Access Network (ERPANET) project³. It was the first effort to join memory organisations, as they were then called, such as museums and libraries, a variety of industries, such as ICT and entertainment industries, and commercial sectors with the purpose of raising awareness about the need of digital preservation and at the same time provide the necessary knowledge base to all participating and external individuals and institutions interested in the discipline. The project did not develop any tools of its own but instead promoted co-operation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects. After the project's completion, another project, namely Digital Preservation Europe⁴ (DPE), was funded by the EU with the aim of "securing a shared knowledge base of the processes, synergy of activity, systems and techniques needed for the long-term management of digital material".

The first projects to actually not only research the digital preservation field but to also develop tools in order to facilitate the preservation of digital resources were DELOS⁵, PLANETS⁶, CASPAR⁷, SHAMAN⁸ and PROTAGE⁹. DELOS focused primarily on Digital Libraries (DL) and Digital Library Management Systems (DLMS) and tried to define their exact meaning and functionality as, at the time, those notions were not clearly defined due to their high level of heterogeneity. DELOS tackled this problem by defining a DLMS reference model, identifying key concepts of a DLMS, such as the information space, documents handled, user profile, services, architecture and incorporating this model into DelosDLMS, a prototype for future DLs offering functionalities such as text and audio-visual searching, personalised browsing using new information visualisation and relevance feedback tools and retrieved information annotation and processing.

The PLANETS project's main goal on the other hand was to build practical services and tools to help stakeholders preserve and maintain long-term access to digital content. The final outcome of the project was the Planets Suite, comprising a preservation planning tool, a test-bed and an interoperability framework. The planning tool, named Plato, offered information on digital objects at risk, measured the effectiveness of the developed tools to preserve them and ultimately supported informed decision-making for the preservation procedure. The test-bed provided sample digital content in order to test the preservation tools while the interoperability framework was a platform allowing users to access remote, third-party tools and services across a web-service and execute preservation workflows.

³ Electronic Resource Preservation and Access Network (ERPANET), <http://www.erpanet.org/index.php>.

⁴ Digital Preservation Europe, <http://www.digitalpreservationeurope.eu/>

⁵ Delos Network of Excellence on Digital Libraries, http://www.delos.info/index.php?option=com_frontpage&Itemid=1

⁶ Preservation And Long-Term Access Through Networked Services, <http://www.planets-project.eu/>

⁷ Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval, <http://www.casparpreserves.eu/>

⁸ Sustaining Heritage Access through Multivalent Archiving, <http://shaman-ip.eu/>

⁹ Preservation Organisations using Tools in Agent Environments, <http://www.ra.ee/protage>

Several other projects which dealt with preserving fragile digital information followed and in most instances their outputs were more or less based on the development of some kind of implementation framework and a variety of tools to support that framework. Despite those common objectives, some differences between them could be summarised by the following points:

- CASPAR's willingness to provide training events, in the form of lectures, workshops and training days.
- CASPAR and SHAMAN placed their attention on contributing to standards and policy developments as well as building on the state of the art.
- PROTAGE explored the use of software agents targeting automation of digital preservation processes.
- The focus of some projects on interactive objects, embedded objects, ontologies and ephemeral data. Some examples include the LIWA¹⁰ project, addressing Web Archiving, and TIMBUS¹¹, researching on preservation of business processes.
- Some projects, such as SCAPE¹², carried out research on object validation methods including the validation of objects according to format specifications and policies as well as results of preservation actions against completeness and correctness.
- The ENSURE¹³ project researched scalable pay-as-you-go infrastructure for preservation services for integration into workflows.

Present efforts in digital preservation can be witnessed in projects such as APARSEN¹⁴, PRESTO4U¹⁵, PRELIDA¹⁶, PREFORMA¹⁷ and ForgetIT¹⁸. APARSEN is described as a Network of Excellence, whose purpose, simply put, is "to look across the excellent work in digital preservation which is carried out in Europe and to try to bring it together under a common vision". The project addresses both technical issues as well as economic, legal and organisational issues, which basically derive from merging the outputs of different projects. The project tries to identify the nature and characteristics of each output, classify it accordingly and finally define under what circumstances each one can be used and on what digital object. Similarly but to a smaller extent, PRESTO4U's aim is to identify useful research results into the field of digital preservation, raise awareness and adoption of latest developments of smaller collections, private sector media owners and new stakeholders. PRELIDA project also has a similar mandate but specifically to the Linked Data community, which may not yet be aware of or fully understand the latest outcomes in Digital Preservation. The project's scope is complemented by efforts in identifying challenges in preserving Linked Data and developing a roadmap for tackling them.

PREFORMA distinguishes itself from previously mentioned projects in the sense that its aim is to address the challenge of implementing good quality standardised file formats for preserving data content in the long-term. The main objective is to give memory institutions full control of the process of the conformity tests of files to be ingested into archives by basically establishing a long-term sustainable ecosystem of developed tools.

¹⁰ Living Web Archives, <http://liwa-project.eu/>

¹¹ Timeless Business, <http://timbusproject.net/>

¹² Scalable Preservation Environments, <http://www.scape-project.eu/>

¹³ ENSURE, <http://ensure-fp7-plone.fe.up.pt/site>

¹⁴ Alliance Permanent Access to the Records of Science in Europe Network,
<http://www.alliancepermanentaccess.org/index.php/aparsen/>

¹⁵ PRESTO4U, <https://www.prestocentre.org/4u>

¹⁶ Preserving Linked Data, <http://www.prelida.eu/project>

¹⁷ PREservation FORMAts for culture information/e-archives, <http://www.preforma-project.eu/>

¹⁸ ForgetIT, <http://www.forgetit-project.eu/en/start/>

The ForgetIT project accepts that even if most large memory institutions are aware of the digital preservation problem and may have in fact already taken some measures towards the goal of preserving their content, other organisations as well as home users are still nowhere near accomplishing something similar. The project proposes a combination of three new concepts to overcome this barrier. It first tries to develop “Managed Forgetting” models which assess information and offer “customizable preservation options such as full preservation, removing of redundancy, resource condensation, and also complete digital forgetting”. Secondly, the notion of “Synergetic Preservation” is applied to bridge the gap between active information use and preservation management and provide solutions for smooth bi-directional transitions. Finally, “Contextualised Remembering” combines context extraction and packaging with evolution detection and intelligent re-contextualisation to keep preserved content useful and meaningful. The main expected outcomes of ForgetIT are an intelligent preservation management framework and two application pilots, one focussing on preserving multimedia coverage of personal events and one for smooth preservation in organisational content management.

Research on Digital Preservation has also been conducted on the other side of the Atlantic in various other projects such as InterPARES and DASPOS.

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project¹⁹, funded by a Social Sciences and Humanities Research Council of Canada Partnership Grant, aims at developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity.

The Data and software preservation for open science (DASPOS) project²⁰, funded by the National Science Foundation (NSF), “represents a collective effort to explore the realisation of a viable data, software, and computation preservation architecture for High Energy Physics (HEP)”. The project tries to coordinate the efforts of various fields dealing with preservation in their encounter with Big Data and propose solutions to pressing technical issues. Based on feedback received from other fields, the project’s outcomes will include designed and developed metadata to support preservation of HEP data and a reference architecture for a data preservation system. A test-bed incorporating this architecture will also be constructed.

3.2 Policies, initiatives and support in EU and the U.S.

Ever since the challenges of digital preservation were put on the spotlight at the end of last century, there have been various initiatives, both preservation specific and of a more generic nature, who provided the necessary means and funds to research what needed to be done to ensure long-term preservation and continued access to digital records. Below, a list of such initiatives, which can be used by the PERICLES project to promote and exploit its outcomes, is presented.

The Digital Agenda presented by the European Commission forms one of the seven pillars of the Europe 2020 Strategy. The main aim of the Digital Agenda is to help European citizens and entrepreneurs to benefit from and capitalise on new technologies, long-term, in the best and most effective way, which will ultimately help the Union itself to capitalise on the results, in economic terms. As more clearly stated in the Digital Agenda Initiative the objective of that effort is to “reboot Europe’s economy” and “deliver a smart sustainable and inclusive growth”.

¹⁹ International Research on Permanent Authentic Records in Electronic Systems, <http://www.interpares.org/>

²⁰ Data and Software Preservation for Open Science, <https://daspos.crc.nd.edu/>

The EU leadership perceived the shift that was happening with the financial instability and realised the need to re-position the Union, in an attempt to make it more competitive and to prevent or even eliminate social and economic weaknesses. The decision to reinforce digital technologies is expected to attribute an increase of 5% in the GDP.

The Digital Preservation Coalition (DPC) was established in the U.K. in December 2001 and was officially launched in February 2012. Some of the Coalition's main goals are to produce and disseminate information on all latest research relating to digital preservation, promote and develop services, technology and standards for digital preservation and also to try to get digital preservation on the agenda of key stakeholders. The Coalition tries to attract funding to accomplish its goals and programmes and form strategic alliances both nationally and internationally to address preservation issues. Some of the more recent projects that they have been involved in are 4C, APARSEN, SPRUCE and TIMBUS while some of their more prominent members include the BBC, the National Archives, United Nations and Lloyds Banking Group.

The National Digital Information Infrastructure Preservation Program (NDIIPP²¹) is a U.S. program engaging hundreds of organisational partners across the United States and around the world to preserve digital collections. Founded in 2010, the National Digital Stewardship Alliance (NDSA²²) is one of NDIIPP's initiatives and is a consortium of institutions that is committed to the long-term preservation of digital information. NDSA, with over 150 members representing various sectors, tries to identify, communicate and advocate member needs, convene and sustain a national community of practice and provide professional development opportunities. Their primary means of accomplishing their goals include producing reports and guidance material, hosting meetings and events and arranging monthly or bi-monthly briefings, demos and presentation webinars on tools, techniques and approaches to digital preservation.

The Digital Library Federation (DLF²³) is a U.S. based community of practitioners promoting work on Digital library standards and best practices, research data management, aggregation and preservation services for digital collections and digital library services that expand access to digital resources for research, teaching, and learning. DLF's main outlet is the annual DLF forum when member institutions but also the broader community convene to present their work and share experiences and practices.

The Alliance for Permanent Access (APA²⁴) supports the development of a sustainable European Digital Information infrastructure that guarantees the permanent access to the digital records of science, whether documents or data, across all fields of research, scholarship and technology. It tries to strengthen European and national strategies and policies and be a platform that will assist key stakeholders to cooperate amongst themselves and with other organisations on digital repositories. APA coordinates action to address preservation challenges and promotes the exchange of ideas and preservation developments, mainly through joint projects, such as ENSURE, TIMBUS and SCAPE, which provide the means to bring communities together, understand requirements and develop strategic plans relating to digital preservation.

3.3 The DP market

Towards the end of last century, focus was placed by the so called memory institutions, i.e. organisations which maintained a repository of some kind, such as museums, libraries and archives,

²¹ Digital Preservation Europe, <http://www.digitalpreservation.gov/index.php>

²² National Digital Stewardship Alliance, <http://digitalpreservation.gov/ndsa/nationalagenda/index.html?loclr=blogsig>

²³ Digital Library Federation, <http://www.diglib.org/groups/>

²⁴ Alliance for Permanent Access, <http://www.alliancepermanentaccess.org/>

on investigating challenges in the long-term preservation of digital data. One of the first efforts to tackle the issues raised in digital preservation came in 1994 in the form of a Task Force, which was formed in the U.S. between the Research Libraries Group (RLG) and the Commission on Preservation and Access (CPA). Its purpose was to research what needed to be done to ensure long-term preservation and continued access to the digital records, which became more and more diverse in nature as years went by.

As content is increasingly born-digital (Figure 1), that is content originating in digital form, such as emails, websites and e-books, the need to solve digital preservation challenges is progressing.

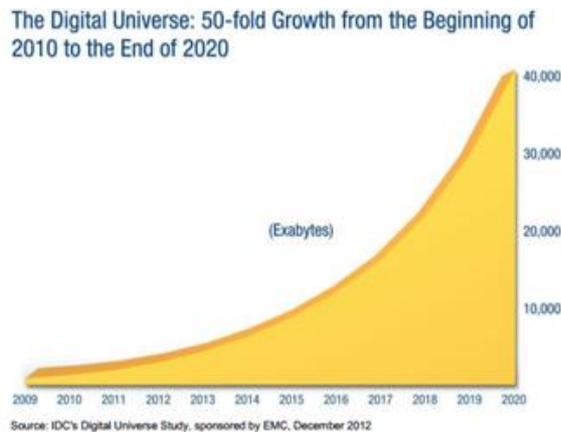


Figure 1: Digital Data Growth.²⁵

It is therefore imperative to preserve valuable digital data and it is precisely this urgency that drives the Digital Preservation market. The advance in technologies, standards and research publications revolving around digital preservation are particularly driven by following motivations:

- Regulatory requirements. Institutions or companies are mandated to retain certain types of data. This is very much true for science data or operational data. This results in opportunities to meet those requirements where there are potential gaps.
- Legislation. Various Data Protection acts which have been enacted across the world.
- Cultural reasons. Maintaining artworks and items of national heritage. This is largely done through public investment and since content is increasingly digital, there is a need to develop new techniques and tools.
- IPR. Protecting Intellectual Property Rights and patents by providing supporting evidence.
- Commercial reasons. Ensuring business continuity by maintaining access to digital data in the long-term.
- Market specific reasons. For example, governments must maintain public records and manufacturing needs to hold onto engineering designs and specifications.

In the last decade the digital preservation market has attracted interest from several projects and initiatives leading to each one conducting a market analysis in order to figure out the state of the field. In 2005 the Digital Preservation Coalition²⁶ surveyed 104 organisations in the UK with their findings showing that even though participating organisations were already aware of the need to

²⁵ IDC Digital Universe Study, <http://gigaom.com/2013/10/02/how-the-industrial-internet-will-help-you-to-stop-worrying-and-love-the-data/screen-shot-2013-09-24-at-4-11-40-pm/>

²⁶ Digital Preservation Coalition, <http://www.dpconline.org/>

preserve digital content and half of them said they were committed to doing so, only 18% had a digital preservation policy in place, and only 20% had funding. Not only that but also half of them had no idea about the volumes that needed to be preserved, about the life span of data while some (50%) decided to print out hard copies as a means to preserve digital information. In another survey in 2006/7, conducted by the Digital Preservation Europe project discovered that out of the 172 organisations surveyed in Europe just one third had implemented a digital repository for the long-term preservation of their digital documents.

Two other more recent projects, Digital Cultural heritage NETwork (DC-NET²⁷) and PLANETS each conducted a market analysis on digital preservation with their findings being presented below.

In 2010, the Planets project produced two white papers, one investigating the digital preservation activities and needs of European organisations, mainly archives and libraries, and one examining digital preservation issues from the perspective of vendors and suppliers.

The first one was based on an online survey, conducted in 2009, when over 200 organisations from around the globe replied to questions regarding their readiness in preserving their data. The main findings of the survey can be summarised in the following bullet points:

- The volume of digital content that organisations expect to archive will increase 25-fold over the next ten years.
- Digital information comes in a range of types, and while over 80% of organisations already need to preserve documents and images, by 2019 over 70% will need to preserve databases, websites, audio and video files as well.
- The digital preservation message has spread far and wide: 93% of respondents indicated that their organisation is aware of the challenges of managing digital information for the long-term.
- Organisations are taking account of digital preservation: 76% include it in their operational planning, 71% in their business continuity planning and 62% in their financial planning.
- By setting out a digital preservation policy, 48% of organisations are actively planning how to tackle digital preservation.
- Organisations are only starting to commit to funding digital preservation, as just 47% have allocated a budget to it.
- There is demand for tools and services to automate preservation planning, to characterise digital collections and to convert digital objects to more accessible formats.
- Organisations which currently need to archive larger volumes, or a wider variety, of content are more likely to have a digital preservation policy in place. However, in ten years' time, all organisations, regardless of whether or not they currently have such a policy, expect to hold similar volumes and range of digital content.

The second white paper that PLANETS produced focused on digital preservation from another perspective, that of IT companies providing such services and are based in the US, Europe, the Middle East and Australasia. These companies were interviewed and their thoughts on the digital preservation market is depicted in the following points:

- The digital preservation market is still in its infancy; there is plenty of potential for growth as it affects all business sectors.
- Currently, engagement is led by the memory institutions at national and international level. Engagement is also high in government and research organisations and emerging in the private sector.

²⁷ Digital Cultural heritage NETwork, <http://www.dc-net.org/>

- Although digital preservation is business critical, many organisations do not have a policy to cover it. Where policies exist, their comprehensiveness is variable.
- There is a lack of information on the costs of digital preservation and its benefits (both tangible and intangible) which makes it hard to put together a convincing business case.
- Migration is strongly preferred to emulation to preserve digital material. However, this may change as emulation has a distinct role to play and there is some interest in emulation tools amongst the briefings' participants.
- Participants thought that the most important factors for a digital preservation solution are that it should: maintain authenticity, reliability and integrity, adhere to metadata standards, and check records have not been damaged.
- Scalability of digital preservation solutions to high volumes of content and high ingest rates are regarded as important but scalability of access is not yet important.
- While attendees thought standards are important in digital preservation, particularly OAIS and ISAD (G), they also thought that there are currently too many standards.
- There is still a need for guidance (particularly training), exchange of best practice, and awareness-raising through conferences and workshops.

In 2011, a study ordered by partners of the DC-NET project, was conducted in order to provide a high-level view on the range of tools and services currently available to support various preservation functions.²⁸ The main outcome of the report was that most tools only play a supportive role in the digital preservation life-cycle while implementation and adaptation of holistic preservation solutions was still low at the time, and although there is no shortage of software tools to aid digital preservation, sustainable services based on well-documented software and offering user support are in short supply.

The study also identified a lack of widely accepted comparison metrics and service maturity models in the domain of digital preservation which would assist the selection of tools and services according to organisation requirements. Furthermore, the report proposes the development of a roadmap that would define digital preservation as an infrastructure service for cultural heritage sector and include benchmarking digital preservation tools as one of its core services. Finally, the report also identified some future trends that would provide a roadmap to digital preservation e-infrastructure being realised.

- Transparent enterprise driven models.
- Launch of self-preserving objects.
- Increased flexibility in digital preservation architectures.
- Clearly defined sets of metrics or benchmarks.
- Real time interoperability.

As was seen in the DC-NET report, back in 2011, software products that offered isolated preservation services and functionality did exist. What was lacking, were complete integrated solutions that would take preservation to the next level of adaptation and implementation. Such products which are currently available and are the most well-known are presented in the next section.

3.4 Market instruments

In the following, we describe examples of existing market instruments used in the preservation field and which may be of interest for integration into PERICLES prototype.

²⁸ Digital Preservation Services: State of the Art Analysis, <http://www.dc-net.org/getFile.php?id=467>

3.4.1 Free and Open Source solutions

The Integrated Rule-Oriented Data System (iRODS²⁹) is an open-source data management software in use at research organisations and government agencies worldwide. iRODS functions independently of storage resources and abstracts data control away from storage devices and device location allowing users to take control of their data. iRODS is released under a BSD license and provides the following:

- Storage virtualisation of different disk and tape storage systems
- A logical namespace across storage locations
- A policy engine to can automate data management according to defined rules
- A method to create and define user specific procedures and functions
- Various client interfaces
- A flexible architecture

Archivematica³⁰ is another free and open-source digital preservation system. It is designed to maintain standards-based, long-term access to collections of digital objects by using a micro-services design pattern to provide an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model. Users are able to monitor and control the micro-services via a web-based dashboard. Archivematica uses best practice metadata standards and implements its default format policies based on an analysis of the significant characteristics of file formats. Archivematica also offers an editable, flexible framework for format identification, package extraction and format normalisation.

Finally, one other open source product trying to provide digital preservation services is the Fedora Repository software³¹, which enables long-term access to digital resources. What distinguishes Fedora from other solutions is its flexibility to support all types of digital content, such as digital collections, e-research, digital libraries, archives, digital preservation, institutional repositories, open access publishing, document management, digital asset management, and others.

3.4.2 Commercial solutions

One of the most well-known commercial solutions that exist in the market today is the Ex Libris DPS³², which is designed to support the acquisition, validation, ingestion, storage, management, preservation, and dissemination of all types of digital objects. It also conforms to standards such as Open Archival Information System (OAIS³³), Metadata Encoding and Transmission (METS³⁴), Preservation Metadata Maintenance (PREMIS³⁵), Machine Readable Cataloguing (MARC³⁶), Dublin Core³⁷, and Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH³⁸).

Another preservation solution used by leading archives, libraries, museums, government organisations and businesses is Preservica³⁹, an active preservation and access technology, which is available both as cloud-hosted services and on-premise editions. It basically includes a full suite of

²⁹ <http://irods.org/>

³⁰ https://www.archivematica.org/wiki/Main_Page

³¹ <http://www.fedora-commons.org/>

³² <http://www.exlibrisgroup.com/category/Home>

³³ <http://www.paradigm.ac.uk/workbook/introduction/oais.html>

³⁴ <http://www.loc.gov/standards/mets/>

³⁵ <http://www.loc.gov/standards/premis/>

³⁶ <http://www.loc.gov/marc/>

³⁷ <http://dublincore.org/>

³⁸ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

³⁹ <http://preservica.com/>

OAIS compliant workflows for ingest, management, storage, access and long-term preservation of digital content, offering:

- easy and customizable public access to the archives
- automated ingest workflows supporting functions such as bulk ingest of digitised content and website crawling
- standard file format registries and migration tools
- data management tools allowing the building of object hierarchies, manageable metadata and a two-way catalogue synchronisation framework
- security and administration functionalities

In contrast to Ex Libris DPS and Preservica, Arkivum⁴⁰ focuses on simply offering data archiving solutions. Arkivum's Archive as a Service provides a fully-managed and secure service for long-term data retention with online access and a guarantee of data integrity. Arkivum has an expertise in data centre operations and storage systems implementation but do not offer the full digital preservation packet that other solutions do. DuraCloud⁴¹ is also one such archiving solution which makes it easy to store and do more in the cloud as it is the only service that makes it easy to move copies of digital content into the cloud and store it with several different providers eliminating the risk of storing content with a single cloud provider.

Other tools, both free and commercial, can be found classified under OAIS categories in Digital POWRR's tool grid which was published in early 2013 after researching tools which could be used in digital preservation and were available at the time.⁴² The National Digital Information Infrastructure Preservation Program in the U.S. also lists a collection of 38 tools and services that were used and/or developed by their partner institutions.⁴³

3.4.3 Training providers

As market surveys have shown, many organisations and institutions are aware of the digital preservation concept and their need to form and comply, sooner rather than later, with a data preservation strategy which will most probably involve the use of one or more of the aforementioned preservation solutions. They do, however, need to receive some training before they can actually proceed with their implementation plan and there are several sources of training available for that purpose such as the following:

- The Digital Curation Centre (DCC⁴⁴) provides expert advice and practical help to anyone in UK higher education and research wanting to store, manage, protect and share digital research data.
- The Digital Preservation Coalition (DPC⁴⁵) facilitates training and support activities and creates practitioner-focused material and events in order to ensure that the coalition's members' staff remain informed of, trained in and confident with the new developments and tools which are released and made available.
- Digital Preservation Outreach and Education (DPOE⁴⁶) is a Library of Congress mission to foster national outreach and education about digital preservation by building a collaborative

⁴⁰ <http://arkivum.com/>

⁴¹ <http://www.duracloud.org/>

⁴² <http://digitalpowrr.niu.edu/tool-grid/>

⁴³ NDIIPP Partner Tools and Services Inventory, <http://www.digitalpreservation.gov/partners/resources/tools/index.html>

⁴⁴ <http://www.dcc.ac.uk/>

⁴⁵ <http://www.dpconline.org/>

⁴⁶ <http://www.digitalpreservation.gov/education/>

network of instructors and partners to provide training to individuals and organisations seeking to preserve their digital content.

- Digital Curator Vocational Education Europe (DigCurV⁴⁷) is a European Commission funded project which brings together a network of partners to address the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections.
- Universities also offer training courses, e.g. the Digital Preservation Training Programme (DPTP⁴⁸) offered by University of London Computer Centre (ULCC⁴⁹) with an aim of providing skills and knowledge necessary for institutions to combine organisational and technological perspectives, and devise an appropriate response to the challenges that digital preservation needs present.
- Training days and conferences are also taking place, e.g. by EuropeanaLocal, British Library, Aparsen⁵⁰.
- Many commercial providers of preservation-related systems, such as the ones mentioned above also provide additional training tailored to their solutions.
- Several EU projects which received funding for exploring the preservation domain also offer training services, for example the Timbus project, which conducted Training Days for individuals interested in digital preservation.

Other European projects that include training services within their scope are the following:

- ARCOMEM⁵¹
- BLOGFOREVER⁵²
- ENSURE⁵³
- SCAPE⁵⁴
- REFORMA⁵⁵
- 4C⁵⁶
- DIACHRON⁵⁷
- FORGETIT⁵⁸
- PRESTO4U⁵⁹
- AXES⁶⁰
- DL.org⁶¹

Also, established institutions such as Independent Media Art Preservation IMAP, the International centre for the study of preservation and restoration of cultural property ICCROM, National Science

⁴⁷ <http://www.digcur-education.org/>

⁴⁸ <http://ulcc.ac.uk/services/research-technology-services/digital-preservation-training>

⁴⁹ <http://ulcc.ac.uk/>

⁵⁰ <http://www.alliancepermanentaccess.org/index.php/aparsen/>

⁵¹ ARchive COmmunities MEMories, <http://www.arcomem.eu/>

⁵² <http://blogforever.eu/>

⁵³ Enabling kNowledge Sustainability Usability and Recovery for Economic value, <http://ensure-fp7-plone.fe.up.pt/site/>

⁵⁴ Scalable Preservation Environments, <http://www.scape-project.eu/>

⁵⁵ PREservation FORMAts for culture information/e-archives, <http://www.preforma-project.eu/>

⁵⁶ Collaboration to clarify the costs of curation, <http://www.4cproject.eu/>

⁵⁷ Managing the Evolution and Preservation of the Data Web, <http://www.diachron-fp7.eu/>

⁵⁸ Concise Preservation by combining Managed Forgetting and Contextualized Remembering, <http://www.forgetit-project.eu/>

⁵⁹ European Technology for Digital Audio-visual Media Preservation, <https://www.prestocentre.org/4u>

⁶⁰ Access to audiovisual archives, <http://www.axes/project.eu>

⁶¹ Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations, www.dlorg.eu

Foundation (with the Digital Library Curriculum Project) offer regular training on techniques and practices related to managing repositories and preservation. The Research Data Alliance has an interest group investigating Education and Training on handling of research data.

3.4.4 Standards, guidelines and certification

Several standards and guidelines exist whose purpose is to set the scene for digital preservation in a way that digital content should be ingested, accessed and managed. Some of the most important ones currently being used the most are the following:

- The Open Archival Information System (OAIS) is a reference model, which has been accepted by the data preservation community as an important standard relating to digital repositories.
- The Curation Lifecycle Model⁶², established by DCC, provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt through the iterative curation cycle.
- European LTDP Common Guidelines⁶³ are also provided by the European Space Agency in terms of preserving historical Earth Observation (EO) data series.
- Trusted Repositories Audit & Certification (TRAC⁶⁴) was created to develop criteria enabling the identification of digital repositories capable of reliably storing, migrating, and providing access to digital collections.

⁶² <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

⁶³ http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue2.0.pdf

⁶⁴ http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

4 Our exploitation approach

4.1 Introduction

Dissemination and exploitation are closely related and therefore their use is often ambiguous. Often “exploitation” is used for both aspects, as they complement each other:

- use or utilisation, especially for profit
- the combined, often varied, use of public-relations and advertising techniques to promote a [...] product etc.

A common misconception as to the meaning of exploitation is the unfair treatment of others to make profit. Historically the term was indeed exclusively used in the pejorative sense. Today the term is still often used purely in conjunction with profit-making purposes. However, as used in marketing and promotion activities, and in particular in funded projects, exploitation aims at a stable and effective sustainability of project results. This need not necessarily, though in many cases the activities in exploitation it will, entail financial revenues as this ensures the longevity of the project results. But in a wider concept of exploitation, the aim is that while dissemination activities during a development aim at creating an awareness, an interest and if successful also a demand, the exploitation activities will directly or indirectly see to meeting this demand. This is best done with a very close collaboration between dissemination and exploitation.

As the PERICLES project will provide results of different natures and cater for a range diverse communities each with their specific remit, there are several approaches that the project could adopt for the exploitation of their results. The exploitation plan will look into the options covering both types of exploitation, the revenue generating one, and the non-profit oriented one. At this point we have a list of anticipated results, which needs to be corroborated by the actual research outcomes. As the research moves on, project results will be clearer. The exploitation plan at this stage of the project investigates possible outcomes and options for activities related to the different types of anticipated results.

The project outcomes that partners have envisioned so far include the following:

- Software products, e.g. component or system level modules.
- Services, e.g. on demand cloud-based preservation services.
- Consulting, e.g. advice on best practices and forming a preservation strategy.
- Training, e.g. commercial training.
- Education courses, e.g. Masters level courses to be taught at academic institutions.
- Technology licensing, e.g. through the use of patents.

The activities supporting the exploitation of the aforementioned outcomes will depend on the assessment of criteria such as the level of technology and its technical feasibility, the progress of the state of the art, the added value for the companies, public and private bodies and the end user’s, possible legal restraints and other legal issues, the timing for a possible launching, the available budget at the moment and many other issues will be more or less resolved or could be easier resolved, thus allowing the partners to adopt some or all of the following profit and non-profit commercialisation approaches:

Generating revenue

- Product development
- Services

- Consulting
- Spin-off company
- Patents
- Training

Non-profit-making

- Publications
- Educational activities

The project partners cover a different experience with respect to exploitation. The SMEs have expertise in business exploitation with deploying products on the markets and in the industry sectors. Technical universities sometimes have their results taken up by spin-offs through which the product enters the markets. Humanity departments and cultural institution are exploiting the results more in the sense of taking up results and investigating them further or influence policies that ultimately will determine the use of products.

4.2 Anticipated exploitable foreground

The RTD work packages will identify the foreground that needs to be developed in PERICLES, which will enable the long-term use of digital objects by addressing the challenges of capturing and managing information that are necessary to make meaningful use of the digital object. Specifically, partners have identified the following technologies from PERICLES, which can potentially be exploited and are planned to be developed:

1. Component technologies
 - PERICLES Extraction Tool (PET) tool.
 - Context and environment analysis tools.
 - Scalable video concept detection.
 - Text analysis for semantic evolution.
 - Encapsulation tools (packaging and embedding).
 - Context analysis tools.
 - Reasoning tools.
2. System level technologies
 - Appraisal tools including semantic analysis.
 - Policy registry and policy editor.
 - Policy-to-process mapping and rule execution.
 - Visualisation and topic maps.
 - Change management and quality assurance tools.
 - Space demonstrator.
 - Media demonstrator.
3. Models, architectures and best practice guidelines
 - Ecosystem models and LRM.
 - Architectural prototypes.
 - Preservation practices for evolving ecosystems and preservation by design.
 - Quality assurance.

Some of the aforementioned outcomes, which have either been developed (e.g. PET tool) or are planned to be developed during the project's lifetime as well as the relationships between them and the context under which they will be implemented, are presented below.

The LRM (Linked Resources Model) focuses on modelling heterogeneous resources and their dependencies. This includes defining a Link Semantics, in order to discriminate, type and classify links based on their impact on the ecosystem. Semantics for Change Management will be further elaborated upon, based on the Linked Resources Model. Dedicated languages and tools will be derived from the defined semantics and models, in order to sustainably manage change and ensure the global consistency of information ecosystem. A first version of the LRM as a static model of digital resources and their dependencies will be available in end July 2014. Change management functionalities will be investigated in a next step. The model will be published as code (in Turtle) with supporting documentation and will serve as a basis for the development of other tools within PERICLES, while consequent tasks will involve the incorporation of elements that enable the model to represent change in the ecosystem. These tasks will result in updates of the LRM that will be published as documents and as implementations of the model that will be held in the project code repository. Dedicated services enabling to create, manage and navigate instances of the LRM will be provided for the test-beds.

The information use of digital objects can be characterised as either digital-object-centric (that is all the information, including relationships, that are necessary to use the digital object) or ecosystem-centric (that is all the information, including relationships, that are necessary to ensure the digital object interacts correctly with other components, such as services or other digital objects, of the environment).

The digital-object-centric view addresses the challenges of understanding what information to collect and manage in the case of long-term use where terminology can change, use can change and users can change. So far the challenge of collecting information as well as the dependencies to the digital object have been investigated. A software tool (the PET tool) with a pluggable architecture has been developed with the objective to collect information in situations where identifying what information to collect for a given use is unknown beforehand or is complex to define, and where a formally defined workflow is not necessarily adopted. In the Open Archival Information System (OAIS) model, the tool's position is before the Ingest process in the context of a sheer curation approach⁶⁵. The underlying model for the PET tool is in-line with the Linked Resource Model (LRM).

At the time of this deliverable, the PET tool has reached a maturity level for being released (as version one). The development of the PET tool has been strongly guided by scenarios coming from the diverse domains of Arts and Science represented by TATE and BUSOC and SpaceApps which anticipate general applicability of the tool. The tool will go through a process of validation by the stakeholders from Arts and Science domains and will be provided to the public under an open source licence. Dissemination of the tool, its purpose and capabilities is already underway with a presentation at the international conference on digital preservation (iPres) conference in Australia in late 2014.

The collected information and dependencies need to be packaged and managed once they have been identified and collected. This work which is currently being planned, will result in a tool capable of incorporating existing packaging structures and information embedding techniques, and augmenting them with functionality to manage the dependencies between the information and the digital object. This tool, belonging to the encapsulation tools that are to be developed, will interface to the PET tool and will form part of the ingest, data management and access OAIS functions. The

⁶⁵ <http://alimanfoo.wordpress.com/2007/06/27/zoological-case-studies-in-digital-curation-dcc-scarp-imagestore>

development of the tool will follow a similar route of dissemination to peers via conferences and also via project training and dissemination events as for the PET tool.

A next step in the development will be dealing with the challenge of change in terminology and community. A framework for semantic analysis that can accommodate changing use and community will be guided in its development by the Science and Media stakeholder scenarios. The framework will manifest as software that will be demonstrated to stakeholders both at training and at dissemination events. A set of approaches aimed at deriving context from content semantics will be developed. The approaches will manifest as software demonstrators that illustrate the feasibility of the approach.

The ecosystem-centric viewpoint which understands a digital object as part of a system of interacting objects and aims at ensuring the long-term use of the digital object, involves capturing and preserving these interactions. The work relates to all functions of the OAIS. The first step is to investigate the challenge of augmenting existing representations to model the ecosystem the digital object exists in. This model of the ecosystem will be communicated to peers through presentations at conferences and project training and dissemination events which will enable further refinement of the model. Any feedback received on the model will lead to its refinement and partners will then amend the LRM to accommodate the possibly new identified functionality. The model will primarily be a document and will be used in the development of the next step which will address the challenge of capturing and managing the ecosystem interactions (represented by policies or processes). This will produce tools for managing the policies and processes that ensure the long-term use of the digital object.

The next step is ensuring that the ecosystem interactions are infrastructure independent. This will result in an approach that is capable of representing the processes in a generic manner and will also produce a tool that will demonstrate the approach which will be part of demonstrations at training and dissemination events. The tool will be held in the project code management system and will be accessible under the project licence. Another concern will be to ensure that changes to elements of the ecosystem do not jeopardise the ecosystem. This will result in methods to ensure that new processes maintain the quality of the processes they replace. In parallel partners will also investigate approaches for conflict resolution between new or updated policies and existing processes or policies. Partners will make use of the project's scenarios to guide the development of tools that demonstrate the methods.

The ensuing investigation into the appraisal process will also make use of the development of a model that will enable automation of the process. The model will be used in the development of tools to aid the appraisal process. These tools will also be held in the project code repository and will form part of the dissemination and training activities. The model will also inform the development of the LRM.

All aforementioned, developed and planned, outcomes will be presented at conferences and will also be part of the project dissemination and training events which are anticipated during the coming years. Also, the software that is to be produced during the project's lifetime will be held and maintained in code repositories associated with the project and will, as much as possible, follow an open source licence.

4.3 Potential market sectors

The reasoning behind the selection of the following markets as potential exploitation targets is mainly that they are represented by the two case studies. Due to partner involvement and therefore knowledge of the sector, we know that these sectors have long-term data retention requirements and are indeed relevant to the PERICLES approach. As will be shown later in the deliverable, it is

within the consortium's plan to investigate other sectors and engage with other potentially interested communities for exploitation purposes.

4.3.1 *Space science*

Europe (especially ESA) has made large investments over the years to develop, launch and operate missions in different fields of science (e.g. Earth Observation, Planetary Science, Astronomy, ISS experiments). This has resulted in long time series and large volumes of data being requested and made available to different scientific user communities. Such data are today archived and made available to the user communities as part of the satellites development and operations programmes.

However, no explicit mechanisms exist today to cover their maintenance long after the completion of the operations phase of the relevant missions and preservation is addressed diversely and only on a mission-per-mission basis. With the continued increase of space science missions producing large volumes of science data, there is an urgent need for ensuring long-term preservation of data.

The current situation in Europe does not guarantee that the large amount of archived space science data and the enormous amount of data coming from future missions, in particular in the earth observation domain, are properly preserved in the long-term and are kept accessible to users together with the associated knowledge (e.g. technical and scientific documentation, algorithms) necessary for their exploitation.

Today data owners are themselves dealing with the preservation and access of their own data without the application of internationally agreed approaches and procedures. As a result interoperability, application of common and standard preservation approaches and policies and the utilisation of harmonised technological solutions and services are key needs to reduce preservation costs and to guarantee that European science data will be available and understandable in future to allow current and next generations of scientists to exploit them for very sensitive applications.

The preservation of science and environmental data derived from Space and other sources has become one of the major challenges today. Knowledge of the past and tracking of the evolution form the basis for managing the future for space data and for information in general.

European space science data is a unique and valuable asset for Europe and for its citizens.

This data is unique and irreplaceable and constitutes a capital for Europe that is fundamental to generate economic and scientific advances today and by future generations. The availability and accessibility of such data will allow European stakeholders to continue supporting initiatives and programmes related to the monitoring of global change, space science and planetary exploration, and to provide European researchers with a valuable asset as basis for their activities.

In the earth science domain, policy makers will be able to exploit the results of the research and environmental applications and activities to take decisions based on solid evidence and to monitor the impacts of these decisions in several application areas like European and global environment (e.g. forest monitoring, soil monitoring, urban development, air quality monitoring), management of energy resources, development and humanitarian aid health, food security and sustainable agriculture, water resources management, civil protection and disasters monitoring (e.g. flood prediction and mitigation, landslides, volcanoes monitoring), global climate change understanding and monitoring, global security and sustainable development. It will also provide new opportunities for researchers and scientists allowing them to re-use and combine the reliable and trustable data from different disciplines and to create new knowledge. As a consequence there is clear urge to preserve this data without time constraints and to keep the data accessible and exploitable.

4.3.2 Archives, Libraries, Museums (ALMs)

The requirements for libraries, museums, and archives (and other heritage sectors) have evolved quite radically during the last fifteen years, as each has moved to an increasing number of digital objects. A key challenge is managing the evolution of web-tools and social media technologies, which have proliferated with the advent of Web 2.0 (including web publishing and shared content storage). The user expectation is that this type of content will be permanently accessible and valuable. Thus, there is a need for archiving dynamic and enduring copies of Web 2.0 records, which are being generated in a transient environment; along with the expectations of the general public that such content can be made available on demand. This type of demand is coupled with the expectation that archived content can be viewed in the “original form”, independently of specific software or hardware technologies, thereby re-creating “authentic” user experiences, even if they are associated with software or hardware that is non-standard or obsolete. Other noteworthy challenges relate to the ownership of content and responsibility of managing records; evaluating whether use may change the value of the objects; operating within a cloud-computing environment; and making the objects available for future reuse. Each of these challenges constitute market-drivers that relate to the use analysis conducted as part of the project; and feed into the emerging demand for an approach that supports relational metadata and its use through a creation, retention, preservation, access, and reuse cycle. Managing these demands through a policy-based approach should contribute to an understanding of more appropriate ways for these sectors to incorporate recordkeeping requirements into their processes; and identify the specific means by which the curators can fulfil their responsibilities more effectively.

Museums are increasingly collecting digital objects. These may be software-based artworks, design objects or digital objects related to the history of science and engineering. In the case of artworks these works will be acquired and preserved during their active life and will in the majority of cases evolve and change over time. In other cases there may be a desire to keep a particular digital object functioning in a way that represents the historical context of that object. In either case, understanding and documenting what those objects are dependent on and how the digital environments and the objects change over time is essential to the mission of the museum. OAIS as a model currently underpins much of the tools and thinking for digital preservation. However as a model it does not reflect the change cycles experienced within a museum.

Within an archive environment traditional collections are not accessioned until the end of their active life. For digital collections it is clear that an approach more closely linked to the continuum theory developed within the records management community will be more appropriate.

The PERICLES project will develop methods and tools which address different types of change within the lives of different types of different digital objects and collecting contexts. In so doing it will support changes in practice within museums and archives that better support the long-term digital preservation of their objects.

4.3.3 Higher education

There is significant potential for higher education institutions to develop new and expand existing teaching programmes, especially at the postgraduate level (e.g. MA/MSc and certificate courses) in areas addressed by PERICLES. These have the potential not only to generate income for the institutions, but to facilitate the uptake of the results of PERICLES across the broad range of sectors in which graduates subsequently go on to work. Indeed, integrating advanced research into such programmes is essential if they are to keep at the forefront of a rapidly changing digital world.

For example, digital curation is a broad disciplinary field that includes digital preservation and other areas addressed by PERICLES, and it has a key role to play not only in domains traditionally associated

with the management of information, such as libraries, archives, and other memory and cultural heritage institutions, but also in a much broader range of market sectors. Digital information is a defining feature of our age, and is a key asset for government and for industries as wide ranging as banking, law and medicine. In particular, there are close links with the parallel field of digital asset management, which arose in the business world, particular the media and content industries, to address the need to manage complex information assets within an enterprise context, and to exploit their social, cultural and commercial value.

Moreover, as individuals we increasingly communicate and record our lives and memories in digital form, creating digital information consciously or generating it as a by-product of broader social, cultural and business activities. This information may be held in informally-managed collections, such as Web 2.0 or social media sites, or it may exist simply as dispersed “digital footprints” on the Web, mobile devices, or hard drives. This has given rise to newer interpretations of digital curation, variously dubbed “content curation” or “social curation”, as well as new business models for exploiting this information for economic activity.

The growing importance of digital curation as a field is reflected in a number of international initiatives addressing university-level curricula. Within Europe, the Digital Curator Vocational Education Europe (DigCurV) project was an initiative funded by the European Commission to investigate the requirements for vocational training in digital curation and to establish a corresponding curriculum framework.⁶⁶ This initiative found that many of the core competences and skills required did not all fall within “digital curation” narrowly conceived, but also included such varied aspects such as management, communication, use of social media, or digital marketing strategies. The framework thus recognises that digital curation is a multi-skilled profession that involves a wide range of competences (see www.digcurv.gla.ac.uk for more details).

Elsewhere, the IMLS-funded DigCCurr projects (DigCCurr I and DigCCurr II) developed an open postgraduate curriculum for students intending to work in digital curation. This initiative was based in the USA (University of North Carolina, Chapel Hill; National Archives and Records Administration), although they also collaborated with partners in Canada and the UK.

This is reflected in the growth of programmes in this area; among the project partners, King’s College London has 2 MAs in this area. The MA in Digital Asset and Media Management⁶⁷ has a particular focus on the media and content industries, and addresses systems, methods and tools for the technical, intellectual and economic exploitation of digital resources; this covers such varied topics as digital publishing, data journalism, social curation, management in the content industries, the use of social media and data analytics in marketing, and the economics, value and impact of digital assets. The MA in Digital Curation is a two-year international programme offered jointly by King’s and Humboldt-Universität zu Berlin in Germany, that combines this industry focus with traditional library and information science aspects, in order to provide the combination of theoretical knowledge and practical training needed to preserve and curate the digital information held by organisations across the public and private sectors. Results from PERICLES will be integrated into these courses, and will thus be introduced to a broad range of sectors by graduates of the courses.

Other examples of relevant PGT programmes include the MA in Digital Curation at Aberystwyth University, the MSc in Information Management & Preservation (Digital) at the University of Glasgow, the Master of Applied Sciences (MAS) in Preservation of Digital Art and Cultural Heritage at the Hochschule der Künste Bern.

In addition to such Master’s programmes, there is great potential for developing more flexible modes of education in the areas of digital curation, digital preservation, and digital asset management.

⁶⁶ <http://www.digcur-education.org/>

⁶⁷ <http://www.kcl.ac.uk/artshums/depts/ddh/study/pgt/madamm/index.aspx>

These include short residential courses (e.g. at summer schools, or as part of Continuing Professional Development programmes), postgraduate certificate programmes, or online/blended learning programmes that make use of virtual learning environments to support remote participation. These have particularly been requested by industrial organisations that are unable for reasons of time to allow employees to attend full MA programmes – they thus represent a good investment, both for increasing income to HE institutions and spreading the results of PERICLES more widely within industry.

Universities can also build upon the results of PERICLES for further research, both in terms of follow-on projects, but also to support and act as a springboard for postgraduate research (PRG) students (e.g. PhDs), especially doctorates that would benefit from collaboration between the academic world and industrial or cultural institutions.

Higher education institutions are to an increasing degree taking on responsibility for managing the research data produced by projects undertaken by their academics and researchers. Partly this is driven by mandates for research data management (RDM) from funding bodies, which view RDM as an essential factor in getting value for money from research grants, by facilitating reuse of data that has been created by means of external funds, and also for enabling published research to be validated against the original data. Such data archives also have the potential to provide a “shop window”, disseminating information about research that has been undertaken and attracting both further funding and students. The results of PERICLES, in particular those related to the scientific data domain, will be of relevance to such initiatives.

4.4 Potential cross-sector opportunities

As has been already mentioned, partners aim at producing outcomes which will not be limited to specific domains. This will allow for a more widespread exploitation strategy involving a variety of stakeholders from multiple domains. As the project progresses the list of outcomes will be augmented with newly developed outcomes and therefore cross-sector opportunities such as the ones mentioned below will also be enhanced.

4.4.1 *Services and facilities*

There are commercial solutions, such as DuraCloud, Arkivum and Preservica, which offer their services related either to preservation as a whole or simply dealing with archiving issues, such as cloud services. These services are generic enough for various kinds of digital data to be ingested and are therefore not domain specific. These types of solutions provide the services needed to ingest, identify, characterise, extract metadata, perform basic safety checks (format validation, virus), store, preserve and analyse a digital collection for risks before the data is actually archived.

Apart from developing online tools and services as cloud services, data infrastructures such as storage technologies can be implemented and exploited as on premise installations as long as they are not designed to serve any specific domain. The advantage of on premise storage is that the organisation retains complete control of the hardware and data, satisfying some security and compliance concerns.

It is envisioned that PERICLES outcomes will be as generic as possible allowing their integration in such cross-domain types of services.

4.4.2 *Training*

The importance of offering training related to long-term preservation of different types of stakeholders has been understood to be an important driver to accepting new technologies, new

policies, new workflows and even new career structures. This is reflected in the many current projects offering training as part of their project work scope, such as the ones mentioned in section 3.4.3.

Training can be provided to the following groups, which should serve as extremely valuable channels for exploitation:

- professionals responsible for archiving and recording,
- personnel in supporting professions such as data librarians, data curators, data managers, data analysts, research administrators, infrastructure providers and developers,
- executives responsible for policies and project management,
- financial managers,
- key players in the field who have an influence on market decisions and best practices in policies and workflow management.

Training is not bound to specific sectors or to particular types of stakeholders. Research projects such as PERICLES react to current demands by extrapolating the related issues further into the future and anticipating a future situation and hence preparing to meet the issues in time. This entails creating an awareness of the issues that the researchers investigate and believe to be an emerging concern to be solved. As decision-making processes are often complex when it comes to changing workflows and policies, the awareness needs to be created on both a wide and a deeper level. Dissemination serves the purpose of generating a widespread interest in the project objectives and results, whereas training offers the deeper level of understanding those objectives and using the results for the decision-making process, of e.g. adapting an institution's archival workflow and preservation policies to managing changing over a long-term. Staff thus instructed will feed back ideas into their companies endorsed by their own work experience. Executive managers will discuss the issues presented in the training sessions or material with their staff and place them in the wider context of the institutional remit and organisation. Financial managers will be able to understand the impact it will have on the long-term and the revenue on investment involved. Key actors in the field act as trustworthy advisors to decision makers and help create a general acceptance of future issues and the need for tackling them now.

Successful exploitation also relies on understanding the needs and potential of the target groups. Training is a kind of user engagement and prepares the ground for the practical application of methods, tools and services. By training future professional users, their current use and work practices can be assessed and used as input to tailor the product for cross-sector market exploitation.

Targeting the training market offers a lot of potential opportunities:

- increased recognition of the project results by data professionals,
- enhanced skills and competencies of trained data professionals which will enable them to employ the results,
- encouraging innovation in organisations' archiving and preservation programmes.

4.4.3 Audit and compliance

One other cross sector activity that could potentially be exploited is offering preservation auditing to meet legal and regulatory requirements. This auditing can be performed in accordance to international standards in digital preservation such as the ones mentioned earlier in the chapter, i.e. OAIS model, DCC's Curation Lifecycle model and ESA's European LTDP Common Guidelines.

The aim of this exploitation opportunity will be to check the standard of an organisation's activities that are involved in preserving digital data rather than providing the tools for carrying out the activities themselves. PERICLES partners could determine whether each organisation is compliant

with the model and aid their own exploration into their compliance with the reference model. The auditing service could also be applied to commercially available products, which could be, for example, either compliant or non-compliant to the OAIS model.

5 Exploitation plan

5.1 Introduction

The PERICLES exploitation plan is a work-in-progress document which will progressively become more detailed in terms of concrete actions and market research results. Every 12 months, we will publish an updated version.

There are many “living” documents informing the exploitation plan, such as our overview of IPR background and foreground, or documents containing information on the State of the Art in preservation and in research fields relevant to PERICLES, and of course most training and dissemination related documents, from conference tracking, to dissemination reports, and collected feedback to publication, blogs, posters, papers, training sessions and documentation, and the CoP meetings.

Based on the reflections described above, this initial exploitation plan will take on the character of a set of instruments and mechanisms, describing the scope of what actions are feasible and make sense at what stage of the project or the maturity of a result following the “SMART” goals principles.

The ultimate goal is naturally the exploitation of our results. As this is subject to having a clear idea of what these results will effectively be, to being able to specify functionalities and services they provide, to understanding what their added-value is in the preservation context, for the users and compared to or complementing existing tools, services and systems, we need to break the overall goal down to more specific goals that are appropriate and feasible for the current phase of development.

Our specific goal for the first phase is to prepare the consortium for an effective exploitation of their results, if not for commercial reasons, then for a sustainable impact of their achievements.

This entails:

- understanding what PERICLES can contribute to the existing concepts, frameworks, technical solutions and practices. This is an ongoing process, initiated with the above described desk research which will be intensified for other market sectors as described below. (4.2.2)
- sounding out a scope of feasible actions (4.2.1)
- determining supporting activities (4.2.3-4)
- defining bespoke forms of engagement with the markets (4.3)
- ensuring communication flow and having appropriate recording and reporting procedures in place (4.2.5)

5.2 Internal Project Activities

5.2.1 *Sounding out the exploitation scope*

Sounding out the scope of potential activities relevant and feasible within the framework of the project is based on understanding how we could influence the market. The following ways have been identified:

- Engagement with existing customers and markets.
- Defining potential requirements and technologies in new markets.
- Engagement with target communities through the CoP.

- Engagement through industry organisations (e.g. DCC, RDA).
- Participation in events (e.g. trade shows, exhibitions).
- Training courses (f2f, online).
- Publications in relevant journals.
- Online presence.

In the table presented below we cite some of the activities and tasks that we have identified so far based on the aforementioned market exploration ways, which partners believe will contribute towards a more complete and solid exploitation plan.

Exploitation activity	Exploitable object	Target group Market segment	Commercial value or not-for-profit
Training	Documentation	All	TBD
Training	Webinars	All	Not-for-profit
Training	Live seminars	All	Not-for-profit
Consultation	Preservation strategy	All	Commercial Not-for-profit
Technical Support	Support on developed components/tools.	All	Commercial
Further development	Enhanced/further developed components.	All	TBD
Marketable product	Developed tools.	All	TBD
Methodology	Design methodologies.	All	TBD
Spin-off project Further funding	Acquired knowledge, expertise on LTDP.	All	Not-for-profit
Patents	Applied patents.	All	Commercial
Educational activities	Courses/PhD degrees on DP.	Academic institutes, professional bodies, coalitions.	Not-for-profit
Publications	Research papers.	Academic institutes	Not-for-profit
Standards	Introduce new standards. Complement existing standards.	Academic institutes	Not-for-profit

Table 2: A list of exploitation opportunities.

Moreover, as an internal project mechanism and in close coordination with WP9 which monitors dissemination activities, WP10 will be keeping an eye on exploitation opportunities that may arise through WP9. Specifically, partners will be contacted via email on a monthly basis with the purpose of providing information of interesting upcoming events, e.g. conferences, workshops, etc., which could serve as means of exploiting PERICLES outcomes.

Furthermore, on the basis of the IPR plan and work plan, a questionnaire detailing the experience and remit in terms of the range of activities identified above, allocating “results” and activity type to individual partners or groups of partners will be iteratively distributed while one on one interviews with consortium partners will also take place as this is probably the best way to look in depth in the vision of each one of the partners with regards to sounding out the project’s exploitation scope. This will lead to an iterative detailed exploitation plan for the upcoming foreground.

As described in the DoW, D10.2 will, among others, report on exploitation activities undertaken by partners throughout the project’s lifetime. Hence, WP10 will be monitoring all these activities to assess progress and effort placed on the exploitation front.

5.2.2 IPR planning and auditing

To be able to plan in a timely fashion exploitation activities, we have introduced an IPR plan that gives us an overview of what IPR-bound objects are being used (Background) or produced in PERICLES (Foreground). Currently most anticipated foreground needs to be substantiated through research results which form the basis for the development of specific IT solutions. However, as the project evolves, i.e. once the anticipated foreground becomes manifest work-in-progress, the IPR object can be more precisely specified and their development can be scheduled (with an anticipated date of release).

Every 6 months following the first completed iteration in M18, the IPR plan will be updated by the partners. This will be followed by technical audits for all technical IPR objects under development stated in the plan to better understand their functionality, their benefit for the user, the target group it is designed for and the level of maturity it has or is aimed at. This will help determine and prepare adequate dissemination and exploitation activities.

In parallel to this deliverable the first version of an IPR object, the PERICLES Extraction Tool (PET) is being prepared for release following the publication of the deliverable D4.1 Initial version of environment extraction tools (due M18). Currently the release as such is being discussed subject to a joint ownership agreement between the two partners ULIV and UGOE. This involves the platform for the publication, the licence attached, and first dissemination activities. Part of this procedure will be the technical audit that will yield the kind of information to contextualise the tool within the technology and preservation market, e.g. how it contributes to existing standards, how it supports current practices etc.

The audit will also help assess the maturity of the results in terms of being ready for a specific type of exploitation activity. And they are in scanning the market for emerging technologies and tools that might affect the PERICLES components anticipated or currently under development.

The audits will be in the form of short interviews with partners and will produce a short report detailing the outcomes of these interviews in order to:

- Understand what the IPR object offers as a basis for scanning the market.
- Assess exploitability of IPR object.
- Identify potential exploitation activities.

One such interview with project partners ULIV led to the following feedback received relating to the developed PET tool:

Description:

In the specific context of the PET tool, it is a piece of software that is really aimed to be used by an end user; IT has been designed in a generalist manner, so it does require some configuration and fitting to specific use cases, but nevertheless it could still be downloaded (pre-configured) by an end user, and run, perhaps with some guidance on the consortium's end.

Usage:

Partners can imagine more than one specific use for the PET tool. These are described in the following cases:

Case 1

It is software that would run on a user machine, but in most cases it would be run in a context such as in an institution where there is interest to capture some implicit knowledge that can be derived by the use context and environment but ultimately:

- Be configured for the specific domain and case
- Be installed on the user machine, with its approval, by an administrator.
- run in the background collecting the environment's information.
- Send the collected info would be sent to the administrator by the user; or analysed first, and then sent.

Case 2:

The PET tool can also run as an extraction tool by an archivist on a machine or virtual machine. This will allow the archivist to extract a set of specific environmental information on pre appraisal or appraisal phases.

Case 3:

There can be other use cases for PET (such as running it as a metadata extraction tool) that are not the main intention in the development but are still possible.

Exploitation event:

It was decided that PERICLES presence in iPRES 2014 will be useful dissemination and feedback occasion for promoting the PET tool.

5.2.3 Market Scans

As has already been mentioned, project partners will be scanning for markets potentially interested in the data preservation domain. Ongoing market reviews will take place based on desktop research and any kind of project team members input emanating from their contact with another market's stakeholders. This process has so far allowed us to identify the following markets as viable candidates for the exploitation of PERICLES outcomes and will be further analysed in the next iterations of the exploitation plan.

Media production environment

This sector employs a large number of freelancers as camera operators, editors, producers and directors. Establishing agreed upon protocols which do not slow down production or raise costs is therefore key. The business case for preservation in this environment is often based on the re-use of material in new productions; avoiding the expensive or at times impossible task of re-capturing

material. Partners will further investigate the exploitation opportunity that lies in the production of tools, which enable the effective discovery of footage for re-use.

Digital libraries

Digital library services provide the infrastructure to underpin teaching and learning; research and scholarly communication; web services; and other discovery services based on resource sharing across university and educational sectors.

For digital libraries, the need to develop strategy and take action in the area of digital preservation and data archiving has grown significantly in recent years, and is now recognised as a priority for managing and preserving digital collections, campus publications (including websites), research data, and other digital assets. These extend, in particular, to cultural heritage institutions that maintain digital collections of unique resources and special collections.

The outputs of the PERICLES project are designed to provide needed solutions for accommodating digital objects of every major format (images, audio, video, text, data sets, etc.), and of individual size. The preservation of these digital resources requires the tools to provide access, support authenticity and integrity, and address the mitigating effects of technology or media obsolescence (change management).

The particular challenges faced by digital library services will need to address the rapid growth and evolution of technology, formats, and dissemination mechanisms; PERICLES provides tools that can be used to manage such change. Within a business context, the need for affordable tools is widely acknowledged, which means that the open source nature of the PERICLES outputs may be attractive, ensuring take-up from digital libraries including institutional repositories.

Science and engineering

Many projects in Science and Engineering are expensive to setup, or the situation for the project are unique (e.g. projects to measure the sea temperature, projects to study solar activity, etc.). Funding agencies are increasingly seeking to ensure the data collected by these projects is kept long enough for any interested groups to make use of the data. For example, the UK Engineering and Physical Science Research Council (EPSRC) have started to require that collected data is kept usable for a minimum of 10 years.⁶⁸ The UK Biological and Biotechnology and Biological Science and Research Council (BBSRC) also require digital data produced by funded projects to be kept usable for 10 years.⁶⁹ Although the funding agencies have started to require long-term data management of collected data many of the funded groups lack the expertise or the tools to meet this requirement (see Figure 2).

⁶⁸ <http://www.epsrc.ac.uk/about/standards/researchdata/expectations/>

⁶⁹ <http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf>

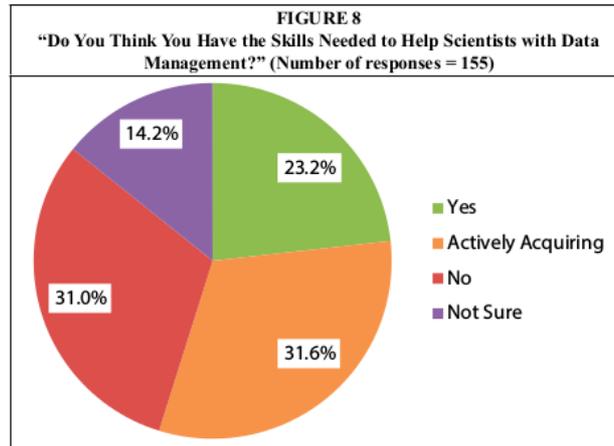


Figure 2: Chart showing the fraction of research data repositories hosted by institutions having the skills for data management.⁷⁰

Many of the respondents cite skills in metadata as well as in the data lifecycle as necessary for the successful storage of research data. Work in the PERICLES project on metadata management and on the lifecycle of the environment and digital object will be highly relevant for those tasked with planning and managing research data.

Health

Increasingly the area of healthcare is adopting ICT to store patient information and to aid in administering treatment.^{71,72} By its very nature patient healthcare information is highly sensitive and subject to stringent policies that often differ across countries, or in some cases regions on how the data must be managed (including who can access the information and how it is disposed) making it very difficult, if not impossible, to distribute records electronically from one domain to another.

However, the average human’s lifespan encompasses multiple generations of ICT which would require the record management system to migrate components multiple times during the patient’s life in order to ensure the records are still accessible. A better understanding of the management and use of electronic patient records will also result in changes to the policies governing their management. In addition, there may be changes in terminology as an increased understanding of the human physiology is achieved. The challenges of accommodating change is one that has been highlighted by a number of groups in the field of electronic health^{73,74} and these are perfectly aligned with the goals of the project.

Financial services

There is increasing interest in digital preservation in the financial services area from a number of perspectives. From a historical and cultural view, banks and other financial institutions are a rich source of documents for scholars. Major UK banks such as Barclays⁷⁵ and Lloyds⁷⁶ have launched public group archives that contain records of permanent business value or historical significance

⁷⁰ Antell, K. et al, July 2014, College & Research Libraries, Vol. 75 no. 4 557-574

⁷¹ http://www.digitalpreservationeurope.eu/publications/briefs/security_aspects.pdf

⁷² http://www.ombudsman.org.uk/__data/assets/pdf_file/0016/24631/Digital-Preservation-Policy.pdf

⁷³ http://www.dpconline.org/component/docman/doc_download/608-datum-cranna

⁷⁴ http://www.dpconline.org/component/docman/doc_download/610-datum-sanders

⁷⁵ <http://www.barclays.com/about-barclays/history/barclays-group-archive.html>

⁷⁶ <http://www.lloydsbankinggroup.com/our-group/our-heritage/our-archives/>

stretching back several hundred years. These describe both the development of the banking institutions themselves as well as containing valuable individual records of potential interest to the general public.

From the perspective of digital content, with the advent of electronic trading, financial institutions manage and process huge volumes of data. Such systems are often partially or fully automated.

At the same time, regulation in the financial sector has become increasingly complex, requiring the curation and long-term retention of records for legal compliance purposes. In the UK, the Financial Conduct Authority⁷⁷ regulates the financial sector. Basel II and III⁷⁸ are international initiatives to strengthen the regulation, supervision and risk management of the banking sector. One of the objectives of Basel is to strengthen banks' transparency and disclosures, which has a direct impact on curation and preservation. Banking data and systems are also critical to the economies of many countries, so ensuring secure and reliable access to data is of great significance. Financial services accounted for 9.6 per cent of the UK's national output in 2011.

PERICLES is relevant to the problems in the finance sector for a number of reasons. The complex and continuously evolving regulatory environment will result in changes to preservation policies as processes to ensure compliance. This poses a challenge in interpreting reusing legacy data. The rapid pace of innovation of financial products means that data will rapidly become not understandable. Financial institutions also maintain highly complex and evolving IT infrastructures, which again pose major challenges for preservation. IT spending for financial services in the U.S. is expected to be approximately \$91 billion in 2013 according to a report by IDC⁷⁹. Hence this also represents a major opportunity for digital preservation methods and infrastructures.

5.2.4 Benchmarking

The PERICLES outcomes need to be benchmarked by comparing them to an industry's best or best practices from other industries, in terms of quality, time and cost. Partners plan to compare results and processes of those studied (the "targets") to their own results and processes. In this way, we will learn how well the targets perform and, more importantly, the business processes that explain why these firms are successful and how we can accomplish the development of better outcomes. In order for this process to be successful PERICLES plans to:

- Continually identify and follow any related projects in search of possibly new outcomes.
- Regularly revisiting the "innovative" aspects.
- Evaluate results against our objective results by the PAB.

It is envisaged that potential outcomes can be assessed based on a common model, e.g. compliance with the OAIS model. For that purpose the table found in the Preserving Digital Objects with Restricted Resources (POWRR⁸⁰) can be used as a baseline.

5.2.5 Internal monitoring and updating the exploitation plan

All above mentioned internal activities are an ongoing effort that will continuously feed into an update of the exploitation plan. In particular the release of results will be monitored, with the technical audits determining the suitable exploitation activities. These will be reported upon in each

⁷⁷ <http://www.fca.org.uk/>

⁷⁸ <http://www.bis.org/bcbs/basel3.htm>

⁷⁹ <http://www.idc.com/getdoc.jsp?containerId=239732#.UTdpn47AW1l>

⁸⁰ <http://digitalpowrr.niu.edu/tool-grid/>

forthcoming exploitation plan. Where possible the reuse will be tracked, e.g. adding code to a released code on Github or research publications citing PERICLES research documentation.

For IT products such as the PET we will set up together with the dissemination leaders and the owner a plan of action for both dissemination and preparing exploitation. This will also identify who is involved in which activity and what the timeline for the exploitation plan is.

The monitoring of the activities and their outcome is particularly important for follow-up activities, as these need build on the success of the initial activities.

5.3 External Activities

5.3.1 Technical solutions

In the consortium, SME partners and IT research institutions are the most likely partners to carry out exploitation of technical results.

With Xerox, the natural channel for the commercial exploitation of the PERICLES outcome is through Xerox Services which focuses on business process and document management, serving a wide range of industries including Financial Services, Government, Healthcare, Higher Education, High Tech and Communications, Transportation. As a provider of data-intensive services, Xerox needs to rely on sustainable data and process management lifecycles across all its business sectors, while primarily coping with the difficulty of accessing and exchanging information and data across customers' and providers' heterogeneous systems, where change is continuous, and often uncontrolled or improperly communicated.

The service sector at large will increasingly require that IT assets (data, services, processes), be managed in a consistent and reliable way over time, with proper change management policies in place that can be enforced and verified. This is all the more challenging as IT assets in a modern service world are distributed over networked infrastructures, owned by diverse stakeholders, internally or externally, within the same organisations or across organisations (service providers, data providers, customers, open data providers, end users) and are subject to continuous change, often without proper notification to the dependent users critically impacted by remote, uncontrolled changes. Properly managing change in a service-oriented organisation is vital to guarantee the quality and continuity of service that customers expect and to ensure that the necessary evolutions of the provided services can be safely introduced without disruption. This is all the more important in critical areas such as transportation or healthcare, where poor performance and service discontinuity may have very damaging effects, to public authorities and broad communities of users.

In that respect, several of the challenges addressed by PERICLES towards a unifying policy-based data lifecycle management approach overlap with challenges relevant to the different lines of business of Xerox Services. In particular, we expect that the project outcome will include methods and tools enabling, by design, the consistent management of dynamic digital resources throughout their life cycle, with an explicit characterisation of the associated change.

In that context, we anticipate that the exploitation of PERICLES results by Xerox will not happen through direct product commercialisation, but rather through the introduction of PERICLES methods, tools and dedicated services so as to optimise and efficiently evolve data-intensive service delivery across multiple lines of business. In addition, means to exploit the results include patent filing (a well-established process in our organisation), as well as the consolidation of research lines explored within PERICLES.

Placing a software tool under an open source licence ensures the longevity of the functionality of the tool as the source code can be examined and duplicated in a supported language (in the case where

the language the tool was originally written in is obsolete). It also ensures longevity as future developers will be able to adapt the tool to future, unforeseen uses. The low barrier to use open source offers, encourages potential users to validate the tool for their needs with no commitments. It allows subsequent researchers in the field to take the tools and extend them instead of having to reinvent the wheel. It also allows commercial enterprises to incorporate the tool in proprietary applications.

DOT plans on exploiting technical results by creating after-market solutions based on the foreground that will be developed during the project's lifecycle. Specifically, DOT anticipates exploiting the Integration framework, any modules that will be developed for the science test-bed, any modules developed for the media test-bed and finally any Application Programming Interfaces (APIs) that are developed for that purpose. The exploitation of the above results may also be achieved by enhancing their initial functionalities or parameterising existing ones in order for them to be applied to several distinct domains. In other cases, full data preservation systems could also be implemented which are based on the above outcomes.

DOT also plans on exploiting the expertise that will be achieved by participating in PERICLES by providing auditing and certification services relating to LTDP systems and installations while DOT would also be able to provide guidance to organisations and institutions which are interested in preserving their digital data.

SpaceApps intends to exploit its long-term (space science) data preservation competencies gained within the PERICLES project by providing services within the future ESA LTDP activities and to set up a prototype for long-term data storage (in collaboration with the Belgian PERICLES end-user partner B.USOC) that should:

- ensure and secure the preservation of archived data for an unlimited time span
- ensure, enhance and facilitate the access by users to archived data and associated knowledge based on solutions developed within PERICLES.

This should initially be addressed at data stemming from the following domains: human spaceflight and operations (ISS and earlier mission), earth observation, science and robotic exploration. This can later be expanded to other domains such as space situational awareness (space weather, etc.).

So far, and based on the IPR plan that has been circulated to partners, they have identified the potential exploitable foreground. Partners will be updating the IPR plan, and then accordingly a technical audit will be performed as was previously mentioned.

5.3.2 Knowledge assets

As a research project, PERICLES will produce new knowledge which will both support the exploitation of technical solutions and promote the implementation of methods and improved practices. The project will use this new knowledge to produce material that would exploit these "knowledge assets".

Research papers are a typical outcome of this kind of new knowledge, which will be exploited by other researchers either as justification, or as state of the art, or as basis for further research. The papers will also be used in higher education or professional training.

In line with the above specified exploitation approach, we will use dissemination to raise the awareness of the results in PERICLES, training and workshops to support the understanding of the outcome, and expert practitioners (in the CoP, see below) will carry the interest in working with the tools, models, applications and services into their respective communities.

For the understanding of the value of the tangible results for long-term preservation, training activities will function as demonstrations to interested stakeholders, with the advantage of having immediate feedback from the users.

Blogs, papers, and presentations are inviting feedback on the ongoing research as much as on the achievements. The pro-active dissemination of getting involved in forums and other blogs, with links to the publication on the PERICLES website, is expected to incite feedback from the interested stakeholders as much as from organisations wishing to take up our results and either adapt, integrate, or develop them further.

5.3.3 Interaction with facilitators

5.3.3.1 COMMUNITIES OF PRACTICE (COP)

Communities of practice are groups of active practitioners who share a concern for something they do and improve their knowledge by interacting with each other on different levels of engagement. They share a domain of knowledge and practice. “They develop a shared repertoire of resources: experiences, stories, tools, ways of addressing recurring problems—in short a shared practice.” Communities of Practice have become associated with finding, sharing, transferring, and archiving knowledge. This knowledge can be considered as a “social capital” of which the domain is benefiting on the long-term. Following Wenger et al (2002), we distinguish the between different levels of engagement which is reflected in our dissemination, training and exploitation approach. With the broad dissemination of more general information on the project (summary on the website, newsletter, leaflets) we target the whole community of practice including the more passive participants of the community. With more targeted dissemination (publications, papers, posters at conferences) and training activities we reach out to the active participants in the community and its leadership group. Finally, the invitation to join our workshops is directed to the leadership level in terms of key players, policy makers and decision influencers of the communities. These latter group is what we refer to as Community of Practice when looking at exchanging expert knowledge embedded in practice, discussing the validity of our approach and outcome in practice, and creating an acceptance and understanding of the importance of following the direction proposed by our research achievements. The members of our CoP may also serve as facilitators in the exploitation of both the technical solutions developed and the knowledge generated during the project.

Working with practitioners in topic-related focus groups (CoP) will provide professional feedback and a validation the project objectives and results. These interests engendered during the focus group workshops where PERICLES members represent the approaches and solutions proposed by PERICLES will be a valuable asset in strengthening an attitude in favour of the solutions and tools supporting and implementing them. “Cultivating communities of practice in strategic areas is a practical way to manage knowledge as an asset”.

“A CoP is defined as successful when its members exchange specific knowledge, practices and/or experiences that contribute to developing a practice (know-how) in a specific field.”

For our purposes, this translates into knowledge created in the framework of the project is being exploited to create know-how in the field of long-term preservation. This in turn will facilitate exploitation of the technical solutions and services after project completion.

5.3.3.2 VENDOR-INDEPENDENT INTEREST GROUPS

The preservation field has gained an increasing importance in public funded projects and has attracted numerous researchers, IT experts, policy-makers and archive, collection care and preservation practitioners to investigate this area.

The consortium has several links to existing or recent projects, namely:

The GLOCAL project - Event-Based Retrieval of Networked Media⁸¹ - a 3 year FP7 IST IP project that ended in Nov 2012 of which CERTH was a partner. GLOCAL outcomes on event-based video analysis techniques presents an interesting approach to adopt within PERICLES (related AV material analysis within WP4).

Another project in which CERTH is currently participating and where a liaising approach might be mutually beneficent is the FP7 ICT IP project i-Treasures - Intangible Treasures: Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures⁸². They intend similarly to PERICLES though without a digital preservation perspective, to deploy ontologies for semantically modelling the domain of interest. Also, the key aspect of change and semantic drift is not part of their scope.

The 4 year FP7 ICT IP project ForgetIT - Concise Preservation by combining Managed Forgetting and Contextualised Remembering⁸³ - of which again CERTH is a partner, addresses current challenges in the area of multimedia preservation, in the personal and organisational context, by introducing three new concepts:

1. Managed Forgetting, which is inspired by the important role of forgetting in human memory i.e. support supporting resource selection for preservation and create immediate benefit from preservation adoption;
2. Synergetic Preservation, which enables a smooth transition between active use and preservation;
3. Contextualised Remembering, which aims at keeping preserved content meaningful and useful along time.

PERICLES would be interested in liaising with them for one of their two pilot applications: "Organisational Preservation focusing on smooth preservation of organisational content management." In turn the semantic drift aspect might offer new perspectives to their work.

CERTH is currently involved in the 2 year FP7 CSA project PRELIDA - Preserving Linked Data⁸⁴ - which bridges the DP and Linked Data communities and addresses challenges of preserving Linked Data as new research interests in DP communities. It is this Linked Data approach that is of interest to PERICLES. In return, the linked resource approach might be of interest to them. However, the project ends in December 2014 and PERICLES is addressing more communities beyond LD communities.

The University of Boras is the Swedish partner in the COST action KNOWeSCAPE⁸⁵ which aims to create interactive knowledge maps. Their end users could be scientists working between disciplines and seeking mutual understanding; science policy makers designing funding frameworks; cultural heritage institutions aiming at better access to their collections. Clearly with the science use case, PERICLES sees one of the markets which would benefit from their outcome in the research data communities. For example: Information visualisation is part of advanced access to DOs. And visualisation of DO dynamics is relevant to PERICLES.

One of the current key FP 7 preservation projects is the 42 months' project SCAPE⁸⁶ - Scalable Preservation Environments (Start in Feb 2011), which aims at developing solutions for the long-term digital preservation of large-scale and heterogeneous collections of digital-objects. Their aim is to develop scalable services for efficient and automated preservation planning and the execution of

⁸¹ <http://www.glocal-project.eu/>

⁸² <http://www.i-treasures.eu/>

⁸³ <http://www.forgetit-project.eu/>

⁸⁴ <http://www.prelida.eu/>

⁸⁵ http://www.cost.eu/domains_actions/mpns/Actions/TD1210

⁸⁶ <http://www.scape-project.eu/>

preservation actions of large (multi-Terabyte) and complex data sets. Several partners are following their progress and have contacts to consortium members. Mutual contact will be strengthened by PERICLES building its research partly on their results regarding models related to preservation policy (see for example D5.1.1 Initial report on preservation ecosystem management).

Another important preservation project currently funded under the Seventh Framework Programme of the EU is TIMBUS -Timeless Business Processes and Services.⁸⁷ The project has started from April 2011. As with SCAPE, PERICLES partners have contacts to their consortium members. Of particular interest are their results on preservation processing and their modelling (again see D5.1.1 Initial report on preservation ecosystem management).

Through the investigation of iRods as potential component in the PERICLES preservation prototype, and the fact that ULIV is closely linked to the iRods user community, there is a link to projects using iRods such as the USA National Science Foundation funded iPLANT project, which offers an infrastructure with data repositories for the research data. There might be an interest in the metadata extraction and packaging approaches and possibly in the policy/workflow aspects (ie managing them). There could also be interest in the context extraction which could help with discovery. As we are looking at tools to enable reuse of data, this is something that will help them in enabling people to better access and use their data.

Another one is the international astronomical community-based initiative EURO-VO - European Virtual Observatory⁸⁸ - developing a portal to research data. Though they are not implementing a preservation system, tools that enable easier long-term reuse of data would be useful to them.

Close contact to the consortium of the SHAMAN project - Sustaining Heritage Access through Multivalent Archiving - exists as UGOE, XEROX and ULIV were all partners in this FP7 project which in many ways fed into the idea for PERICLES. SHAMAN looked at the emulation and data management aspects. The understanding of the importance of investigating solutions related to change seemed for the three partners a natural step forward in preservation research, as SHAMAN didn't look at metadata management and didn't look at the lifecycle management of the digital object.

A ULIV staff is a member of the DEPH - the Study Group on Data Preservation and Long Term Analysis in High Energy Physics⁸⁹ - which is concerned with preservation of particle physics data for a variety of experiments (some Nobel or potential Nobel prize winning experiments). PERICLES is concerned with the lifecycle of metadata and the ecosystem. This can be interesting that may help to guide their approach. They have workshops roughly yearly which is a good opportunity for exploitation openings.

On the case study side, TATE is a partner in PRESTO4U - a PrestoCentre project⁹⁰ involved in the community of practice activities. The project will deliver new tools and services to connect the different constituencies involved in AV media preservation. TATE is also a member of the Digital Preservation Coalition (DPC)⁹¹ which is an advocate and catalyst for digital preservation, aiding members in the delivery of resilient long-term access to content and services.

The FP7 coordination action project CIRCE - Cooperative International Space Station Research data Conservation and Exploitation⁹² has naturally links through our use case partner B.USOC and SpaceApps and provide good opportunities with making contacts to that market.

⁸⁷ <http://timbusproject.net/>

⁸⁸ <http://www.euro-vo.org/>

⁸⁹ <http://www.dphep.org/>

⁹⁰ <https://www.prestocentre.org/>

⁹¹ <http://www.dpconline.org/>

⁹² <http://www.circe-space.eu/>

These are projects and initiatives that are currently on the forefront regarding preservation research and are projects that we closely observe and intend to contact with relevant results from our side in order to liaise with them or use their framework or events for dissemination and exploitation activities:

- Research Data Alliance (RDA⁹³)
- Storage Network Industry Alliance (SNIA⁹⁴)
- APARSEN VCoE⁹⁵
- Open PLANETS Foundation (OPF⁹⁶).
- Digital Curation Centre (DCC⁹⁷)
- ENUMERATE Project⁹⁸

5.4 Implementation

The implementation plan for the coming months includes the next steps that partners will be taking, the schedule and time frame under which these steps will be performed and the necessary milestones that need to exist in order for partners to check that exploitation's internal and external activities and goals that were mentioned in earlier sections are executed appropriately.

As far as forthcoming tasks are concerned, it is envisaged that Year 2 of the project will focus mostly on preliminary community engagement activities (e.g. publications and presentations) and scanning for the identification of:

- Newly developed state of the art software that is used in data preservation. Certain registries such as PRONOM⁹⁹ and COPTR¹⁰⁰ can be scanned for that purpose.
- Other markets and cross markets in need of LTDP services and products.
- Exploitation opportunities, i.e. events, tradeshows, workshops, etc.
- Available exploitation fora where projects, initiatives and other types of platforms collaborate and promote the exchange of ideas in relation to digital preservation. So far we have identified the following dissemination and exploitation channels which could offer PERICLES the chance to propagate its outcomes to all target communities:
 - ENTER Network¹⁰¹ (PERICLES is already a member of the ENTER network and plans on exploiting this membership in the coming months),
 - Innovation Place¹⁰²,
 - MENON Network¹⁰³,
 - DESIRE Project.¹⁰⁴

On the other hand, Year 3 and Year 4 will focus more on participating in events where demonstrations of the developed outcomes can take place and more hands on experiences can be gained. PERICLES partners will attend events where they can showcase existing tools and services

⁹³ Research Data Alliance, <https://rd-alliance.org/>

⁹⁴ <http://www.snia.org/>

⁹⁵ <http://www.alliancepermanentaccess.org/index.php/aparsen/webinars/>

⁹⁶ <http://www.openplanetsfoundation.org/>

⁹⁷ <http://www.dcc.ac.uk/>

⁹⁸ <http://www.enumerate.eu/>

⁹⁹ PRONOM, <http://www.nationalarchives.gov.uk/aboutapps/pronom/tools.htm>

¹⁰⁰ COPTR, http://coptr.digipres.org/Main_Page

¹⁰¹ <http://www.enter-network.eu>

¹⁰² <https://www.innovationplace.eu/>

¹⁰³ <http://www.menon.org/about-menon/>

¹⁰⁴ http://www.ecsite.eu/activities_and_resources/projects/desire

that the consortium can provide to relevant stakeholders who may be interested in LTDP. Having the knowledge from the performed technical audits, events will be selected accordingly.

6 Conclusion

The objective of the PERICLES project is to develop tools and services which address challenges related to the preservation of digital objects, such as dealing with highly dynamic and complex digital objects, automating necessary procedures which at this point are becoming increasingly difficult to be executed manually and also handle the semantic drift of digital assets that is noticed due to changes in terminology.

This initial version of the exploitation plan defines a set of instruments and mechanisms for effective and sustainable promotion of the PERICLES project, its objectives and its results and it relates to the communication and dissemination strategy adopted by the project using established tools and channels.

Specifically, this exploitation plan is addressed to enhance the PERICLES approach towards promoting and exploiting its outcomes by focusing on:

- Developing tools and services that optimally can be easily disseminated and are generic enough for them to be used not only by stakeholders belonging to a specific domain but in a variety of domains as well.
- Exploitation activities that are performed by the Consortium both inside and outside the Consortium itself. Through the project's interaction with other projects, fora and initiatives and together with the help of the dissemination work package, relevant exploitation events will be selected to present PERICLES outcomes.
- Identifying new application areas of the PERICLES outcomes and enlarging the set of users that adopt them. Market scans to identify such opportunities will be performed throughout the project's lifetime for that purpose.