PERICLES – Digital Preservation through Management of Change in Evolving Ecosystems

Simon Waddington[1], Mark Hedges[1], Marina Riga[2], Panagiotis Mitzias[2], Efstratios Kontopoulos[2], Ioannis Kompatsiaris[2], Jean-Yves Vion-Dury[3], Nikolaos Lagos[3], Sándor Darányi[4], Fabio Corubolo[5], Christian Muller[6] and John McNeill[7]

[1]King's College London, U.K.
[2]Information Technologies Institute, CERTH, GR-57001 Thessaloniki, Greece
[3]Xerox Research Centre Europe (XRCE), 38240 Meylan, France
[4]Swedish School of Library and Information Science, University of Borås, Borås, Sweden
[5]IPHS, University of Liverpool, L69 3GL, U.K.
[6]B.USOC, Brussels, Belgium
[7]Tate, London, U.K.
{simon.waddington, mark.hedges}@kcl.ac.uk,
{mriga, pmitzias, skontopo, ikom}@iti.gr,
{Jean-Yves.Vion-Dury, Nikolaos.Lagos}@xrce.xerox.com,
Sándor.Darányi@hb.se, corubolo@gmail.com,
christian.muller@busoc.be, john.mcneill@tate.org.uk

Abstract. Management of change is essential to ensure the long-term reusability of digital assets. Change can be brought about in many ways, including through technological, user community and policy factors. Motivated by case studies in space science and time-based media, we consider the impact of change on complex digital objects comprising multiple interdependent entities, such as files, software and documentation. Our approach is based on modelling of digital ecosystems, in which abstract representations are used to assess risks to sustainability and support tasks such as appraisal. The paper is based on work of the EU FP7 PERICLES project on digital preservation, and presents some general concepts as well as a description of selected research areas under investigation by the project.

1 Introduction

1.1 Motivation

Existing approaches to digital preservation are heavily influenced by practices that have evolved over many years in the non-digital world. The reusability of digital objects is dependent on their surrounding environment. This can include not only relevant software, but also platforms and documentation, and often the digital objects and their environment have complex interdependencies. Due to the rapid pace of technological change, the environment in which a digital object exists will evolve and some entities may become delinked or even obsolete. This may result in the loss of capabil-

Muller C., Vion-Dury J., Lagos N., Kontopoulos E., Riga M., Mitzias P., Corubolo F., Hedges M., DarÃąnyi S., Waddington S., Kompatsiaris Y. and McNeill J.
PERICLES âĂŞ Digital Preservation through Management of Change in Evolving Ecosystems.
DOI: 10.5220/0006163600510074
In *The Success of European Projects using New Information and Communication Technologies (EPS Colmar 2015)*, pages 51-74
ISBN: 978-989-758-176-2

ity to run software, to interpret information or to render data files. A similar argument can be applied to other types of change. For example, organisational changes may result in digital objects held by the organisation no longer being compliant with current policies and procedures. Evolution of user communities may result in the digital objects being interpreted by individuals and used for purposes that were not envisaged when they were initially created or acquired. This can result in the digital objects not being fit for purpose or even understandable by current users.

Maintaining digital objects in a static form in a repository, as might be done with non-digital assets such as books and paintings, which can remain in a reusable form for centuries, is unlikely to be successful even over time periods of a few years. Thus new approaches are required to managing such digital objects that can deal with both complex dependencies as well as continual change.

1.2 PERICLES Objectives and Approach

The main challenge for PERICLES is to ensure the ongoing interpretation and reusability of digital objects that are heterogeneous, volatile (i.e. subject to continual change) and are complex (i.e. have many interdependencies). By analogy with biological systems, we use the term *digital ecosystem* to reflect an evolving set of interdependent entities, which is subject to influences bringing about change. Digital ecosystems can include any entities that can have a direct or indirect impact on the reuse of digital objects, including data objects, software, user communities, processes, technical services and policies. An important feature of our approach is that a definition of a digital ecosystem includes descriptions of the dependencies between the constituent entities.

Following a widely adopted methodology in science, we introduce computational models to enable the impact of change on a digital ecosystem to be assessed, and in some cases for mitigating actions to be determined, without the need to manipulate the entities in the ecosystem directly. Based on a linked data paradigm, the models make use where possible of existing domain ontologies. The evolution of the models is governed by policies.

In order to populate such models, tools are provided to support the extraction of metadata, such as content, environmental, usage and provenance information. Analysis and visualisation of the models is used to support risk analysis and provide decision support. Finally preservation actions can be determined and translated into executable business processes.

To support the development, testing and real-world deployment of PERICLES components, an integration framework is under development. This includes an Entity Registry and Model Repository to support the storage and retrieval of the models as well as an execution layer to enable preservation components to be wrapped in handlers and run against the stored entities. The integration framework also provides a reference implementation for deploying PERICLES components in real-world applications.

1.3 Digital Preservation Activities in the EU

In this section, we briefly review prior EU-funded activities relating to digital preservation, to place PERICLES in a wider context. In 2001, the EU funded the Electronic Resource Preservation and Access Network (ERPANET) project[1] in the FP5 programme. This was the first attempt to engage with memory organisations, such as museums and libraries, and commercial sector for the purpose of raising awareness about the need for digital preservation and at the same time providing the necessary knowledge base to all participants.

There then followed approximately €100M of EU funding for digital preservation, covering a wide range of topics and including research and development prototypes. The PLANETS[2] project developed the Planets Suite, comprising a preservation planning tool, a test-bed and an interoperability framework. The planning tool, Plato, offered information on digital objects at risk, and supported informed decision-making on preservation actions. The project primarily dealt with simple digital objects, rather than complex dependencies, and used a sampling approach on individual objects to evaluate preservation actions.

CASPAR[3] worked primarily on preservation approaches to validate the OAIS reference model [13] in the cultural, artistic and scientific domains. It investigated the implementation and use of key OAIS concepts such as representation information, knowledge management and preservation description information. SHAMAN[4] studied the incorporation of Product Lifecycle Management within a digital preservation system, and produced an extended information lifecycle model. PROTAGE[5] explored the use of software agents targeting automation of digital preservation processes. The LiWA[6] project dealt with archiving of web content.

TIMBUS[7], addressed preservation of business processes where software and platform are developed and delivered as a service. SCAPE[8] focused on scalable preservation algorithms, extending the results of PLANETS to high volume content. The ENSURE[9] project considered scalable pay-as-you-go infrastructure for preservation services based on cloud computing technology, as well as exploring non-traditional domains for digital preservation such as finance and medicine. Finally the APARSEN[10] network tried to join together work on pervious preservation projects into a common vision, again underpinned by the OAIS model.

PERICLES differs from most preceding projects in that it considers continuously changing environments such as for time-based media, where OAIS is less appropriate. Our approach is based on a continuum viewpoint. Although static dependency models

---

[1] http://www.erpanet.org/index.php.
[2] http://www.planets-project.eu/
[3] http://www.planets-project.eu/
[4] http://shaman-ip.eu/
[5] http://www.ra.ee/protage
[6] http://liwa-project.eu/
[7] http://timbusproject.net/
[8] http://www.scape-project.eu/
[9] http://ensure-fp7-plone.fe.up.pt/site
[10] http://www.alliancepermanentaccess.org/index.php/aparsen/

were considered in earlier projects, such as CASPAR, the use of dynamic models is new. There has also been relatively little work done to date on semantic change, which is an important focus of PERICLES.

1.4 Acknowledgements

This work was supported by the European Commission Seventh Framework Programme under Grant Agreement Number FP7-601138 PERICLES. The authors wish to acknowledge the contributions of our many PERICLES colleagues to the content of this paper.

2 Digital Preservation and Change

2.1 Lifecycle versus Continuum Approaches to Digital Preservation

Lifecycle models are a point of reference for most existing approaches and practices in digital preservation. They provide a framework for describing a sequence of actions or phases, such as creation, productive use, modification and disposal, for the management of digital objects throughout their existence. Such models suggest a linear sequence of distinct phases and activities, which in practice may be non-linear or even relatively disordered. Lifecycle models provide an idealised abstraction of reality, and may typically be used in higher-level organisational planning and for detecting gaps in procedures.

The DCC lifecycle model [10] is one of the most well-known lifecycle models. It 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. The UK Data Archive describes a research data lifecycle[11], which comprises six sequential activities and, unlike the DCC model, is more focused on the data user's perspective. Overviews of lifecycle models for research data are provided by Ball [11] and the CEOS Working Group on Data Life Cycle Models and Concepts [12].

So-called lifecycle approaches typically envisage a clear distinction between active life and post-active life. The Open Archival Information System (OAIS) [3] is a commonly adopted reference model for an archive, consisting of an organisation of people and systems that has accepted the responsibility to preserve information and make it available for a designated community. Although lifecycle models and OAIS provide a useful frame of reference for preservation, they are less suited to dealing with examples where there is a less clear distinction between the active life and archival phases, examples of which will be discussed in section 3.

In [2], we introduced a continuum approach to digital preservation that combines two main aspects. Firstly, there is no distinction made between active life and post-active life; that is, preservation is fully integrated into the active life of the digital

[11] http://ijdc.net/index.php/ijdc/article/view/69

objects. A second aspect is that preservation is non-custodial, that is we do not aim necessarily to remove entities from their environment, both physical and organisational, and place them in the custody of a third party.

Continuum approaches have been proposed in the closely related field of record keeping. The Records Continuum (RC) was originally proposed by Upward in 1996 [4]. An essential aspect is that the content and structure of a record are fixed, but the surrounding context can change over time, so a record is “always in a state of becoming” [15].

2.2 Change Types and Their Impact

PERICLES considers a number of different types of change that can potentially have an impact on the reuse of digital objects. A more extensive review is presented in [16]. The main high-level change types that we have considered are summarised in Table 1.

Table 1. Types of change occurring on digital ecosystems and their impact.

Type of change	Description	Impact
Knowledge and terminology	Changes in semantics that originate from a designated user community.	Different user communities using the same underlying datasets with different understanding and goals.
Technology	This includes hardware availability, software obsolescence, and changes in formats, protocols and interfaces.	Requires replacement of hardware and software components, transcoding of files, redesign of interfaces etc.
Policy	Changes in permissions, legal requirements, quality assurance and strategy.	This can impact how and where digital objects are stored, quality processes they are subjected to, retention periods etc.
Organisation	Change to the organisation due for example to political, financial or strategic reasons. Often organisational changes can be manifested as policy changes.	This can result in different priorities for retaining or maintaining the reusability of digital objects.
Practice	This change originates from new or changed habits of the designated user community (not necessary related to knowledge and terminology changes). It is an indicator that user requirements may change.	This can result in changes to the form in which digital objects are retained, reflecting the changing ways in which they are to be reused.
Requirements	This can include business requirements, functional requirements that a system should fulfil, quality of service and user requirements.	This again reflects the way that digital objects are reused and hence how they should be stored and maintained.
Dependency	Either characteristic attributes of a dependency are changed (e.g. quicker, faster, more flexible, cheaper) or the dependency itself changes.	Evolution in dependencies can reflect different views on the types of change that are being considered.

2.3 Change and Dependency

Change and dependency can in many respects be viewed as dual notions. Thus the types of dependency we may wish to model are related to the types of change that are being addressed. In PERICLES, we say that entity A is dependent on entity B if changes to B have a significant impact on the state of A. A key aspect of PERICLES is that dependencies can have associated semantics and do not merely represent a link between the two objects. The semantics of a dependency are related to the change context under consideration.

A number of notions of dependency exist in the literature. The PREMIS Data Dictionary[12] defines three types of relationships between objects: structural, derivation and dependency. In particular, a derivation relationship results from the replication or transformation of an object. A dependency relationship exists when one object requires another to support its function, delivery, or coherence.

The *Open Provenance Model* (*OPM*)[13] introduces the concept of a provenance graph that aims to capture the causal dependencies between entities. The most relevant concept from our perspective is *process* that represents actions performed on or caused by artefacts, and resulting in new artefacts.

In a preservation context, [17] defines notions of module, dependency and profile to model use by a community of users. A *module* is defined to be a software/hardware component or knowledge base that is to be preserved, and a profile is the set of modules that are assumed to be known to the users. A *dependency relation* is then defined by the statement that module A depends on module B if A cannot function without B. For example, a README.txt file depends on the availability of a text editor (e.g. Notepad). The authors of [8] also define the more specific notion of task-based dependency, expressed as Datalog rules and facts. In [19], the notion of task is extended to *intelligibility*, which allows for typing dependencies. The PERICLES modelling approach goes one step further toward genericity, by allowing any kind of dependency specialisation, and provides a much richer topology for dependency graphs through managing dependencies as objects instead of properties.

2.4 Semantic Change

An important aspect of PERICLES is the study of *evolving semantics* and *semantic change* in particular. The risk of semantic change for digital preservation is that, as a fallout from inevitable language evolution that has been accelerating due to an interplay of factors, future users may lose access to content, either (a) because the concepts and/or the words as their labels will have changed, or (b) because the same concept may have different labels over separate user communities.

Additionally, by better understanding semantic change processes, one can identify ‘at risk’ terminology, which is likely to be hard to understand by future users of the resources. Similarly, one can identify specialist terminology likely to be different across domains and, therefore, difficult for those other domain users to understand.

[12] http://www.loc.gov/standards/premis/
[13] http://eprints.soton.ac.uk/271449/ 1/opm.pdf

Another issue for semantic change might relate to the change over time in the way in which a resource is used.

In the investigation of semantic change, PERICLES is conducting experiments in two major directions: one looking at the interpretability of digital objects over time and over designated user communities and the other focusing on drift[14] detection, measurement and quantification methods. The aim is to eventually relate the two in a common frame of thought.

When considering *drift interpretation* (understandability), one of the key questions is how to represent the knowledge model of a domain, community, or individual. This is a critical question as any subsequent analysis or experimentation will depend upon the quality and 'soundness' of any methods or assumptions made at this initial stage. Regarding *drift quantification* (measurability), vector- versus graph-based measures are computed and ranked. In this way, all kinds of shifts could be analysed, be they community-dependent or temporal.

A key element of our explorations is to address drifts in *word meaning* used for document indexing, and consequent changes in *document meaning* together with *topic shifts typical of evolving document collections*. To this end, we also call in probabilistic methods such as additive regularisation based topic models [20]. Secondly, as in classical mechanics, physical systems in change are typically analysed by *calculus* and represented as *vector fields*. Using physics as a metaphor we introduced a vector field based tool to study evolving semantics [21] and its scalability aspects [22]. This work is in the phase of adding qualitative evaluation of shifts in word meaning to the model, which is novel because typically, only quantitative drifts have been addressed by measurements [24]. Finally, our vector field model of semantic change points in the direction of *social mechanics* [25, 26], thereby paving the way for an integrative metatheory of changes in sign systems as a function of social use depending on evolving sign contexts.

3 Case Studies

The examples selected for study in PERICLES are chosen from the application domains that the project is addressing, namely digital media, and space science.

3.1 Examples from the Digital Media Domain

Within the PERICLES project, the digital media domain covers three different subdomains, namely Digital Video Art (DVA), Software-Based Art (SBA) and Born-Digital Archives (BDA). Several key challenges have been defined within each of these subdomains and corresponding ontologies have been developed; these do not attempt to model the respective subdomains exhaustively, but are primarily aimed at modelling preservation-related risks. Specifically, in DVA, the focus is on the consistent playback of digital video files, with respect to the technical or conceptual char-

[14] In literature, semantic change is covered by expressions like *semantic drift*, *semantic shift*, *semantic decay*, and sometimes as *concept drift*.

acteristics of the corresponding digital components. In SBA, the focus is on the assessment of risks for newly acquired artworks, regarding their technical dependencies but also the functional, conceptual and aesthetic intentions related to the significant properties of an artwork. Finally, within the BDA context, the focus is on the need to be able to access and maintain digital documents as was originally intended, together with all their technical, aesthetic and permission characteristics. A detailed analysis of the media case study is contained in [27].

All key challenges are explored in relation to collections held by Tate[15]. The DVA and SBA collections belong to the main art collection, while the BDA material exists in the Tate Library and Archive collection; these two types of collections are managed by different teams within Tate. The institution has approximately 300 video artworks, including digital video artworks, in the collection. It also has a small but growing number of software-based artworks. The born-digital material in Tate Library and Archive includes material from institutions within the UK, such as records from commercial galleries that come into the archive, as well as artists' personal records. Much of this material comprises standard formats such as emails, spreadsheets, text documents, images, and so on.

3.2 Examples from the Science Domain

B.USOC[16] supports experiments on the International Space Station (ISS) and is the curator of both collected data and operation history. B.USOC chose to analyse the SOLAR payload, in operation since 2008 on the ESA COLUMBUS module of the ISS. These observation data are prime candidates for long-term data preservation, as variabilities of the solar spectral irradiance have an influence on Earth's climate, and the measurements cannot be repeated. The current SOLAR module is built from three complementary space science instruments (see Fig. 1) that measure the solar spectral irradiance with an unprecedented accuracy.

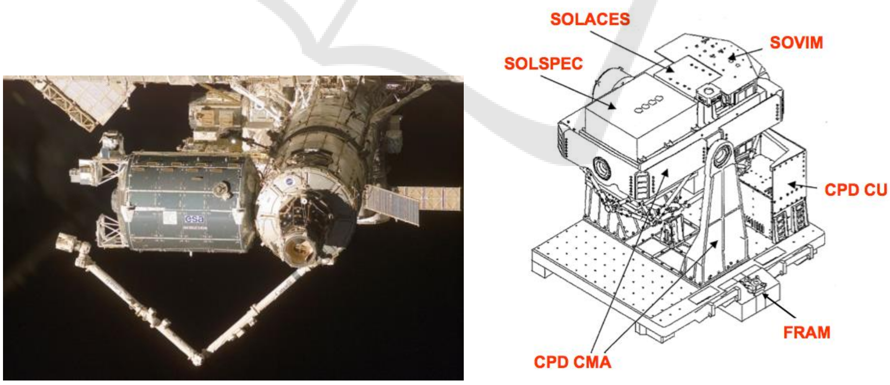

Fig. 1. The International Space Station, and the SOLAR module as part of the COLUMBUS (left) and the SOLAR instrument (right).

[15] http://www.tate.org.uk/
[16] http://www.busoc.be

SOLAR and the phenomena it studies are an example of change in observation data with the evolution of knowledge on the sun. Before the space age (forty years ago), the sun's input to the earth system was called the "solar constant" and great pains were taken to determine it by removing the atmospheric effects. Then in the beginning of the 1980's, several instruments, including first versions of SOLAR, were flown in space and determined that the "constant" was in fact a variable parameter synchronous with the 11 year sunspot cycle, hence its name was changed to "total solar irradiance". Moreover, spectral variations were found not be uniform and that the ultraviolet region, while weak in energy, had important variations relating to solar activity known now as space weather (solar flares and other phenomena).

The SOLAR instruments, which had been designed originally to provide snapshots of the solar spectral irradiance at well-defined parts of the solar cycle, now deliver valuable scientific data relevant to both shorter and longer time scales.

From a digital preservation perspective, the experiments consist of highly complex interlinked digital entities, including raw data and associated telemetry, software, documentation (over a hundred document categories), and operational logs.

4 Model-driven Approach

4.1 Functional Architecture

Following a common paradigm in science, we introduce models to enable the impact of change on a digital ecosystem to be assessed, and in some cases for mitigating actions to be determined. This principle is illustrated in Fig.2, which describes the PERICLES functional architecture, based on [5].

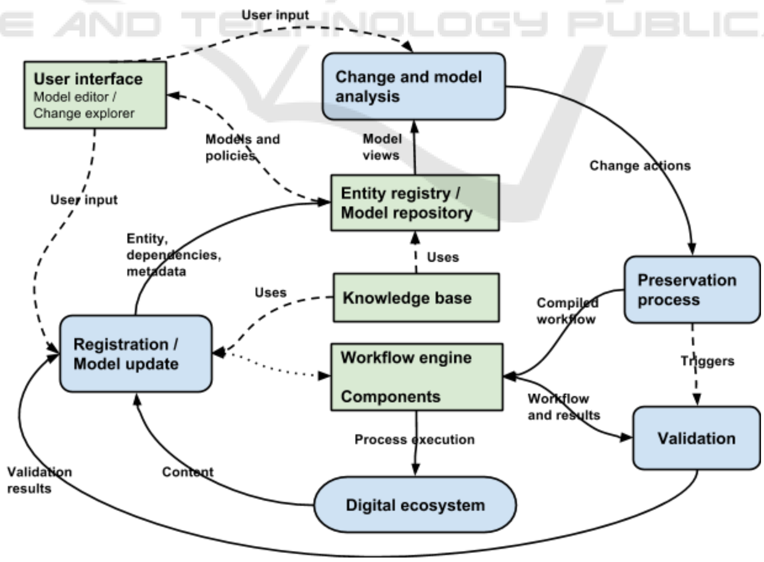

Fig. 2. The PERICLES functional architecture.

The components in blue, joined by solid lines, represent the main workflow for building the models, performing change impact analysis, determining preservation processes or actions, validating the results of the preservation actions, and updating the models. The remaining components support the model-driven workflow. A user interface component is required both to create and in some cases populate the models, as well as to perform change impact analysis and determination of preservation actions. The Entity Registry/Model Repository (ERMR) assigns unique identifiers to the entities in an ecosystem and provides tools for storing and retrieving the models themselves. The Knowledge Base provides the underlying ontologies for constructing the ecosystem models, to be described in the following section, together with reasoning tools. Finally the preservation actions are executed via a workflow engine, using a set of transformation components.

4.2 Digital Ecosystem Models

The PERICLES *Linked Resource Model* (*LRM*) is an *upper level ontology* designed to provide a principled way to modelling evolving ecosystems, focusing on aspects related to the changes taking place. This means that, in addition to existing preservation models that aim to capture provenance and preservation actions, the LRM also aims at modelling how potential changes to the ecosystem, and their impact, can be captured. It is important to note here that we assume that a policy governs at all times the dynamic aspects related to changes (e.g. conditions required for a change to happen and/or impact of changes). As a consequence, the properties of the LRM are dependent on the policy being applied. At its core the LRM defines the ecosystem by means of constituent entities and dependencies. The main concepts of the static LRM are illustrated in Fig. 3. (The prefix `pk` refers to the LRM namespace).

Resource. Represents any physical, digital, conceptual, or other kind of entity; entities may be real or imaginary and in general comprises all things in the universe of discourse of the LRM Model. A resource can be *Abstract* (c.f. `AbstractResource` in Fig. 3), representing the abstract part of a resource, for instance the idea or concept of an artwork, or *Concrete* (c.f. `ConcreteResource`), representing the part of an entity that has a physical extension and is therefore characterized by a location attribute (specifying some spatial information). These two concepts can be used together to describe a resource; for example, both the very idea of an artwork, as referred by papers talking about the artist's intention behind the created object, and the corresponding video stream that one can load and play in order to manifest and perceive the artwork. To achieve that, the abstract and concrete resources can be related through a specific `realizedAs` predicate, which in the above example could be used to express that the video file is a concrete realization of the abstract art piece.

Dependency. An LRM `Dependency` describes the context under which change in one or more entities has an impact on other entities of the ecosystem. The description of a dependency minimally includes the intent or purpose related to the corresponding usage of the involved entities. From a functional perspective, dedicated policies/rules further refine the context (e.g. conditions, time constraints, impact) under which change is to be interpreted for a given type of dependency.

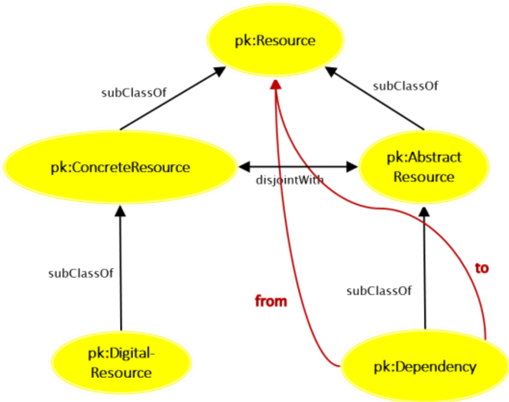

Fig. 3. Main concepts of the static LRM.

For example, consider a document containing a set of diagrams that has been created using MS Visio 2000, and that a corresponding policy defines that MS Visio drawings should be periodically backed up as JPEG objects by the work group who created the set of diagrams in the first place[17]. According to the policy, the work group who created the set of JPEG objects should be able to access but not edit the corresponding objects. The classes and properties related to `Dependency` can be used to describe each such conversion in terms of its temporal information and the entities it involves along with their roles in the relationship (i.e. person making the conversion and object being converted), as other existing models. In addition, the LRM `Dependency` is strictly connected to the intention underlying a specific change. In the case described here the intent may be described as "*The work group who created the set of diagrams wants to be able to access* (*but not edit*) *the diagrams created using MS Visio 2000. Therefore, the work group has decided to convert these diagrams to JPEG format*" and it implies the following.

- There is an explicit dependency between the MS Visio and JPEG objects. More specifically, the JPEG objects are depending on the MS Visio ones. This means that if an MS Visio object 'MS1' is converted to a JPEG object, 'JPEG1', and 'MS1' is edited, then 'JPEG1' should either be updated accordingly or another JPEG object 'JPEG2' should be generated and 'JPEG1' optionally deleted (the description is not explicit enough here to decide which of the two actions should be performed). This dependency would be especially useful in a scenario where MS Visio keeps on being used for some time in parallel to the JPEG entities being used as back up.
- The dependency between 'MS1' and 'JPEG1' is unidirectional. Actually, JPEG objects are not allowed to be edited and, if they are, no change to the corresponding MS Visio objects should apply.
- The dependency applies to the specific work group, which means that if a person from another work group modifies one of the MS Visio objects, no specific conver-

[17] This example is adapted from a use case described in [19], pp. 52-53.

sion action has to be taken (the action should be defined by the corresponding policy).

To enable recording the intent of a dependency, we can relate in the LRM the `Dependency` entity with an entity that describes the intent via a property that we name *intention*, as illustrated in Fig. 4.

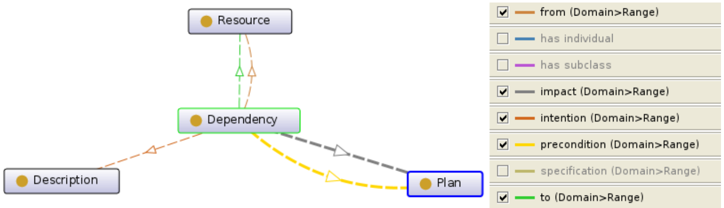

Fig. 4. A view of the `Dependency` concept in LRM.

Let us take once more the example above: we need to be able to express the fact that a transformation to the JPEG is possible only if the corresponding MS Visio object exists and if the human that triggers the conversion has the required permissions to do that (i.e. belongs to the specific workgroup). The impact of the conversion (generating a new JPEG object) could also be conditioned on the existence of a corresponding JPEG object containing an older version of the MS Visio object. The actions to be undertaken in that case, would be decided based on the policy governing the specific operation. Assuming that only the most recent JPEG object must be archived, the old one must be deleted and replaced by the new one (conversely deciding to keep the old JPEG object may imply having to archive the old version of the corresponding old MS Visio object as well).

Plan. The condition(s) and impact(s) of a change operation are connected to the `Dependency` concept in LRM via `precondition` and `impact` properties as illustrated in Fig. 4. These connect a `Dependency` to a `Plan`, *which is defined as a specialized description representing a set of actions or steps to be executed by someone/something* (*either human or software*); this is, thus, a means to give operational semantics to dependencies. Plans can describe how preconditions and impacts are checked and implemented (this could be for example defined via a formal rule-based language, such as SWRL). The temporally coordinated execution of plans can be modelled via *activities*. A corresponding `Activity` class is defined in LRM, which has a temporal extension (i.e. has a start and/or end time, or a duration). Finally, a resource that performs an activity, i.e. is the "bearer" of change in the ecosystem, either human or man-made (e.g. software), is represented by a class called `Agent`.

4.3 Domain Ontologies

A *domain ontology* (or *domain-specific ontology*) is a formal description of modelling concepts in a specific domain in a structured manner. Three media domain ontologies have been developed within PERICLES, aimed at modelling digital preservation risks for the three respective subdomains (DVA, SBA and BDA), via LRM-based constructs (c.f. section 3.1). The key notions adopted and extended by the LRM are:

- **Activity** - represents activities that may be executed during a digital item's lifespan. The media domain ontologies extend the Activity class, in order to model activities that are considered to be important for digital preservation processes (creation, acquisition, storage, access, display, copy, maintenance, loan, destruction, etc.).
- **Agent** - with subclasses for human and software agents. Human agents are in addition specialised for the media domain into artists, creators, programmers, museum staff, etc. and software agents into programs, software libraries, operating systems, etc.
- **Dependency** (c.f. Section 4.2) - indicates the association or interaction of two or more resources in the domain ontology. For the media domain ontologies, we extend the basic notion of LRM dependency with three sub-categories:
  - **Hardware Dependencies** - specify hardware requirements for a Resource in order for it to function properly.
  - **Software Dependencies** - indicate the dependency of a Resource or Activity on a specific software (Software Agent) - name, version, etc.
  - **Data Dependencies** - imply the requirement of some knowledge, or data or information, in order for a Resource to achieve its purpose of existence or function. This kind of data may originate from human input (e.g. passwords), computer files (e.g. configuration files), network connection, live video, etc.

  The context of dependencies may be additionally enriched with the notions of intention and specification (see Section 4.2). For the media domain ontologies, a set of predefined intention types were defined:
  - **Dependencies with a Conceptual Intention** are aimed at modelling the intended "meaning" of a resource (i.e. artwork) by its creator; according to the way he/she meant it to be interpreted/understood. For example, a poem (digital item) belonging to an archival record may not conserve its formatting during the normalization process, something that may be contrary to the intention of the poet regarding the way that the poem is conceptualised/conceived by a reader.
  - **Dependencies with a Functional Intention** represent relations relevant to the proper, consistent and complete functioning of the resource. For example, a specific codec is required to display a digital video artwork.
  - **Dependencies with a Compatibility Intention** model compatible software or hardware components which may operate together or as replacement components for availability, obsolescence or other reasons. For example, the software used for playing back a digital video artwork consistently is compatible with certain operating systems.

A domain-specific instantiation, presented in Fig. 5, describes the following scenario taken from the BDA subdomain: a normalisation activity is applied in a text file (item, digital resource), according to the archival policy defined by the used normalisation software (OpenOffice). In terms of the media ontology, there is (a) a data dependency of the normalisation activity to the item itself, (b) a software dependency of the normalization activity to the used software, and (c) a hardware dependency of the normalization software to the hardware required in order for the software to run efficiently. The intention of all three types of dependencies is functional, meaning that all the required resources modelled in this example impact the functionality of the resources for which the dependencies were implemented.

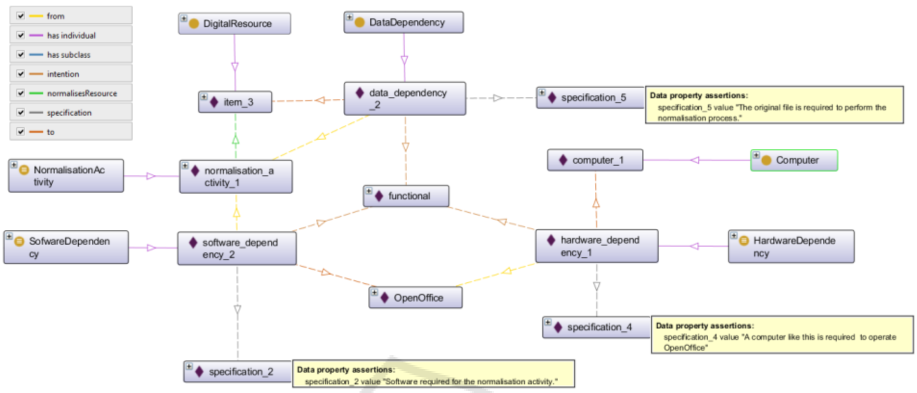

Fig. 5. Dependencies existing within the context of a normalisation activity applied in a digital item.

Within the context of PERICLES and the DVA ontology, an Ontology Design Pattern (ODP) for representing digital video resources was introduced [6]. This work was motivated by the problem of consistent presentation of digital video files in the context of digital preservation. The aim of this pattern is to model digital video files, their components and other associated entities, such as codecs and containers (Fig. 6). The proposed design pattern facilitates the creation of relevant domain ontologies that will be deployed in the fields of media archiving and digital preservation of videos and video artworks.

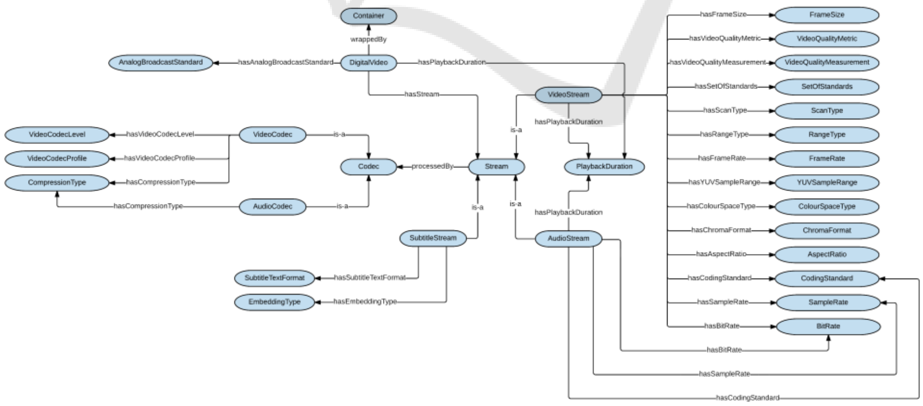

Fig. 6. Digital Video ODP schematic view.

The design pattern illustrates a more general principle, namely that ecosystem models can be constructed from a set of common templates. Such an approach would greatly reduce the effort required to create models by enabling a more modular approach where templates are reused across many different models.

4.4 Environment Capture and Model Population

An important consideration for a model-driven approach to preservation is to minimise the effort required to construct the ecosystem models. The *PERICLES Extraction Tool (PET)* provides one approach. PET is an open source framework for the extraction of the Significant Environment Information of a digital object. Here significance is a positive number expressing the importance of a piece of environment information for a given purpose. The tool can be used in a sheer curation [23] scenario, where it runs in the system background and reacts to events related to the creation and alteration of digital objects and the information accessed by processes, to extract environment information with regard to these events. All changes and successive extractions are stored locally on the curated machine for further analysis. A further extraction mode is the capturing of an environment information snapshot, which is intended for the extraction of information that does not change frequently.

The tool aims to be generic as it is not created with a single user community or use case in mind, but can be specialised with domain specific modules and configuration. PET provides several methods for the extraction of SEI, implemented as extraction modules as displayed in Fig. 7. The configuration has to be done once, after that PET can run automatically in a way that does not interfere with system activities and follows the sheer curation principles. The PET tool can be used to generate models based on the LRM ontology, which can then be used for ecosystem analysis. PET has been released as open source software under the Apache license. PET source code together with documentation and tutorials are available for download on Github[18].

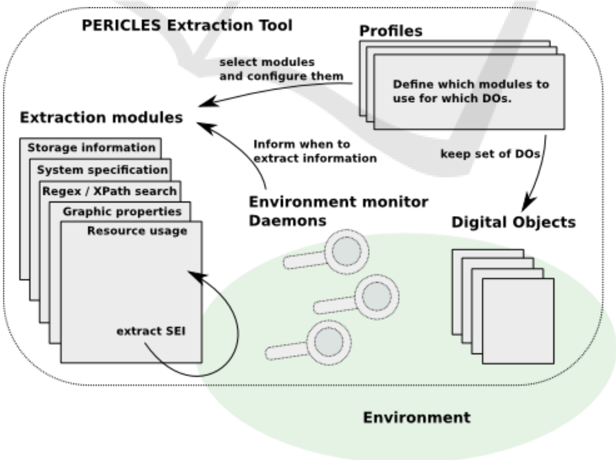

Fig.7. SEI extraction with the PERICLES Extraction Tool.

[18] https://github.com/pericles-project/pet

5 Applications in Digital Preservation

5.1 Monitoring of User Communities

One complex issue facing PERICLES is tracking the use of media (images or video) by different user communities. Although it is possible to apply usage tracking to a web portal making archival content available, once content is downloaded its use can no longer be monitored. Even this presupposes that portal users are registered and provide details of their intended usage, which may also not be feasible. What would be helpful in this context is the possibility to embed metadata and identifiers that would allow mapping and monitoring the diffusion of the media across user communities, in order to help identify their evolution.

Information encapsulation (IE) methods can be distinguished into the categories of packaging and information embedding. Packaging refers to the aggregation of files or other information formats as equal entities stored in an information container. In contrast to this, information embedding needs a carrier information entity (file/stream) in which payload information will be embedded.

The *PERICLES Content Aggregation Tool (PeriCAT)* [28] is a framework for Information Encapsulation techniques. It integrates a set of information encapsulation techniques from various domains, which can be used from within the framework. Furthermore PeriCAT provides a mechanism to capture the scenario of the user, and to suggest the best fitting information encapsulation technique for a given scenario. The tool is available to download on Github[19] under an Apache Version 2.0 licence.

5.2 Quality Assurance

Quality Assurance is defined by Webster[20] as “*a program for the systematic monitoring and evaluation of the various aspects of a project, service, or facility to ensure that standards of quality are being met*”.

In PERICLES we define a series of Quality Assurance (QA) criteria for the entities of evolving ecosystems, in particular for policies, processes, complex digital media objects, semantics and user communities. This will allow us to manage change in the ecosystem by validating its entities, detecting conflicts and keeping track of its evolution through time. Our methods focus on validating the correct application of policies to the ecosystem. When change occurs, the approach will ensure that policies are still correctly implemented by tracing the correct application of the higher-level policies (guidelines, principles, constraints) in the concrete ecosystem implementation. Policies will be expressed at different levels, using a policy model integrated in the Ecosystem Model itself. We support the QA of policies by defining criteria and methods that can validate or measure the correct application of policies through processes, services and other ecosystem entities, so ensuring that the implementation is respecting the principles defined in the high-level policies. The QA methods in turn support the management of change in the ecosystem entities, such as change in policy, policy

[19] https://github.com/pericles-project/PeriCAT

[20] http://www.merriam-webster.com/dictionary/quality%20assurance

lifecycle, change in the processes implementing those policies, or change in other policy dependencies. In this model, we do **not make any strong assumption about the format in which the policy has to be expressed**, be it natural or formal language, nor are we imposing any specific structure on the processes used to implement them (although we are providing exemplar implementations). We are aware that policies and processes in real systems will be implemented using a variety of techniques and we aim to develop a policy layer that can be applied on top of existing ecosystems. This assumption will allow the deployment of such QA methods in systems that are not built using only specific technologies or rule languages, making their adoption simpler.

Other projects have looked at the issue of QA in preservation, in particular the SCAPE project[21], with a focus on the QA of a specific type of digital object (image, audio, e-publications), and also on the implementation of digital preservation policies by collecting metrics on collections; this is a valuable approach that is specific to issues related to digital object QA. Our approach works at the model level and addresses the implementation of digital preservation policies in existing ecosystems, although it takes into account the valuable work done by SCAPE.

More concretely, in [8] we define a policy model, a policy derivation process with guidelines, and a series of QA criteria and change management approaches for policies, taking into account the possibility of conflicting policies. We are currently working on exemplar implementation and refinement of the methods, implemented using the LRM and digital ecosystem models and other PERICLES technologies.

5.3 Appraisal

Appraisal is a process that in broad terms aims to determine which data should be held by an organisation. This can include both decisions about accepting data for a collection or archive (e.g. acquisition) as well as determining whether existing data in an archive or a collection should be retained.

In traditional paper-based archival practice, appraisal is a largely manual process, which is performed by a skilled archivist or curator. Although archivists are guided by organisational appraisal policies, such policies are mostly high-level and do not in themselves provide sufficiently detailed and rigorous criteria that can directly be translated into a machine executable form. Thus, much of the detailed decision-making rests with the knowledge and experience of the archivist or curator.

With the increasing volumes of digital content in comparison to analogue, manual appraisal is becoming increasingly impractical. Thus there is a need for automation based on clearly defined appraisal criteria. At the same time, decisions about acquisition and retention are dependent on many complex factors. Hence our aim here is to identify opportunities for automation or semi-automation of specific criteria that can assist human appraisal.

In [8], we set out our overall approach to appraisal. In Appendix 1, we extend the categorisation of appraisal criteria in the DELOS project. The latter focused on appraisal as the determination of the worth of preserving information, that is, as a means

[21] http://wiki.opf-labs.org/display/SP/SCAPE+Policy+Framework

of answering the question, 'what is worth keeping'? Here appraisal is considered as a process to be revisited throughout the life of the digital object. Consequently, our results differ principally in the breadth of material considered and in the number and breadth of appraisal factors identified.

Within the context of the PERICLES case studies, appraisal can naturally be partitioned into two distinct categories.

- **Technical Appraisal** – decisions based on the (on-going) feasibility of preserving the digital objects. This involves determining whether digital objects can be maintained in a reusable form and in particular takes into account obsolescence of software, formats and policies.
- **Content-based (or intellectual) Appraisal** – acquisition and retention decisions or assignment of value based on the content of the digital objects themselves.

Our focus in this paper is on the technical appraisal aspect. We are primarily interested here in predictive rather than reactive approaches to modelling the impact of change. Projects such as PLANETS [29] used a *technology watch* to detect changes in the external environment, which could then result in changes to archived content. We aim to model risks through understanding longer-term trends to predict the impact of changes in the future. This work follows a number of steps:

- Quantify primary risks to the ecosystem. This is done by analysis and modelling of external data sources to predict the likely obsolescence of software and formats, or hardware failure of entities in the ecosystem.
- Determine the impact of primary risks on entities in the ecosystem. This step aims to identify entities at the greatest primary risk.
- Determine the impact resulting from higher-order risks propagating through ecosystem. In this step we propagate risks through the models.
- Determine potential mitigating actions and their associated costs. In some cases, it may be possible to execute and validate mitigating actions automatically.

The overall goal is to provide a tool for use e.g. by archivists, to analyse a digital ecosystem, determine at what point in the future there is a significant risk to reuse, and the potential cost impact and potential mitigating actions. Such a tool could be applied for example to assess the value of a software-based artwork, by determining how long it can be displayed in exhibitions before elements become obsolete or require refactoring, or the cost of maintaining a set of scientific experiments for a given time period.

In order to enhance the user experience of the model-driven approach, PERICLES is developing a visualisation tool MICE (Model Impact and Change Explorer), which aims to present risk and impact information to users.

6 PERICLES Integration Framework

The PERICLES integration framework, described in detail in [9], is designed for the flexible execution of varied and varying processing and control components in typical preservation workflows, while itself being controllable by abstract models of the overall preservation system. It is the project's focal point for connecting tools, models and application use cases together to demonstrate the potential of model-driven digital preservation.

The integration framework can be deployed in slightly different ways to suit testing and development, or real-world deployment. Fig. 8 shows the configuration for real-world deployment.

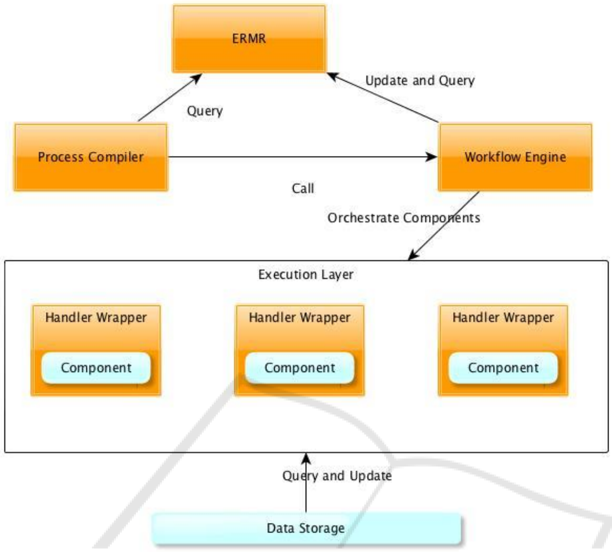

Fig. 8. Real world deployment of PERICLES components, based on the integration framework.

The *Entity Registry Model Repository (ERMR)* is a component for the management of digital entities and relationships between them. Access methods are presented as RESTful services. The ERMR registers and stores entity metadata and models. Agreed metadata conventions provide the necessary registry functionality; the registry is agnostic to the entities and metadata stored with the interpretation of data being the responsibility of the client applications. The ERMR provides a CDMI implementation for HTTP access to entities in the registry.

The ERMR uses a triple store as a database for the storage and retrieval of triples through semantic queries. The ERMR also provides a mediator service to integrate semantic services to extend its reasoning capabilities. It also provides a simple RESTful API for access to the registry. Queries can be expressed in the SPARQL query language to retrieve and manipulate data stored in the triple store. To link entities described as triples to actual digital objects, the ERMR is able to provide unique identifiers used to create a unique CDMI URL for an associated object. This URL is used to link entities stored in data storage components.

The *Process Compiler* (PC) supports the translation and reconfiguration of preservation process models described in the ERMR into executable workflows to be employed by the Workflow Engine. As part of this the PC will transfer information to the ERMR, which is used to update the process descriptions. The current component implementation is targeted towards a BPMN compilation system, though in theory the design can be adapted to any workflow engine language.

The *Workflow Engine* takes processes, compiled by the Process Compiler with descriptions and implementations stored in the ERMR, Data Storage and Workflow Engine cache, and executes them through orchestration of executable components (PERICLES tools, archive subsystems or supporting software) wrapped in Web Services Handlers with REST targets. This is one of the main fixed points of the architecture, which is active at all times.

Processing Elements (PEs) are any pieces of software or hardware that can be used by a PERICLES process to accomplish a given task. They are characterised by a fixed set of parameters including input and output types, platform requirements and versioning information.

Handlers are at the core of the integration framework. They function as the communication points for each major entity within the system. The Handlers deal with the validation of incoming requests, exercise the functionality of the PEs they wrap, store and transfer the results of PE functions and initiate necessary communications with other PEs. The Handlers do not perform any operations that change or alter the data contained in the objects they handle; only PEs can alter and change data. PEs, on the other hand, should not have knowledge of anything in their environment. This means they perform their function and only their function, and the Handlers deal with the rest of the PERICLES system and the outside world.

Data Storage is a specialised long-term Processing Element, a permanent service available to a PERICLES-based system. This component is responsible for storing digital objects, which can be data files, metadata, models or ontologies. Data Storage must be represented in the framework as a long-term service, since Processing Elements are typically transient in nature with limited functionality scope. The Data Storage service must manage data, as required, as bit-level preservation, object replication and distributed storage mechanisms.

7 Application of PERICLES Technologies to the Wider Community

As PERICLES is an inter-disciplinary project cutting across a number of technologies and aimed at diverse communities each with their specific remit, there are several approaches that the project is adopting for the exploitation of the results [1]. These include:

- 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 taught at academic institutions.
- Technology licensing, e.g. through the use of patents.

The two application domains, space science and digital media are the primary targets for technology transfer due to the partner involvement and knowledge.

In 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 users. 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. These data are unique and irreplaceable and constitutes a capital for Europe that is fundamental to generate economic and scientific advances.

The requirements for libraries, museums, galleries, and archives (and other heritage sectors) have evolved quite radically during the last fifteen years. Museums and galleries 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 will be acquired and preserved during their active life and will in the majority of cases evolve and change over time, for example for display in different operational settings. 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. The user expectation is that this type of content will be permanently accessible and valuable. 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 an "authentic" user experience, even if they are associated with software or hardware that is non-standard or obsolete.

Beyond the space science and cultural heritage sectors, media production is a growing sector. 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.

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. Solutions are required address the rapid growth and evolution of technology, formats, and dissemination mechanisms. Preservation requires the tools to provide access, support authenticity and integrity, and address the mitigating effects of technology or media obsolescence.

Projects in Science and Engineering are expensive to setup, or the situation for the project are unique. 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[22], with many other funding agencies taking similar approaches. Although such policies exist, many of the funded research groups lack the expertise or the tools to meet this requirement.

Increasingly the area of healthcare is adopting ICT to store patient information and to aid in administering treatment[23,24]. By its very nature patient healthcare information

[22] http://www.epsrc.ac.uk/about/standards/researchdata/expectations/

[23] http://www.digitalpreservationeurope.eu/publications/briefs/security_aspects.pdf

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.

8 Conclusions

The current paper has provided an overview of some of the work being performed by the PERICLES project, at the end of the third year of the four-year project. The results we have obtained so far illustrate the potential value of models in digital preservation. The final year of the project will be focused on producing integrated prototypes and further developing the user-facing components such as appraisal.

The outcomes of the project align with the EU Digital Agenda for Europe[25] in supporting the digital preservation of digital cultural assets, and are potentially applicable across a wide range of sectors beyond the space science and cultural heritage.

References

1. PERICLES Consortium: Deliverable D10.1 – Initial version of exploitation Plan (2014). http://pericles-project.eu/uploads/files/PERICLES_WP10-D10_1-Exploitation_Plan-V1.pdf
2. Lagos, N., Waddington, S., Vion-Dury, J.-Y.: On The Preservation Of Evolving Digital Content - The Continuum Approach And Relevant Metadata Models, 9th Metadata And Semantics Research Conference (MTSR 2015), Manchester, UK. http://pericles-project.eu/uploads/files/PreservEvolvingDigitalContent-LagosWaddingtonVion-Dury-32-MTSR2015.pdf
3. Vion-Dury, J-Y, Lagos, N., Kontopoulos, E., Riga, M. Mitzias, P., Meditskos, G., Waddington, S., Laurenson, P., Kompatsiaris, I.: Designing for Inconsistency – The Dependency-based PERICLES Approach, First International Workshop on Semantic Web for Cultural Heritage (SW4CH 2015), Futuroscope, France. http://www.xrce.xerox.com/Research-Development/Publications/2015-052
4. PERICLES Consortium: Deliverable D2.3.2 - Data Surveys and Domain Ontologies, (2015). http://pericles-project.eu/uploads/files/PERICLES_WP2_D2_3_2_Data_Survey_Domain_Ontologies_V1_0.pdf.
5. Waddington, S., Tonkin, E., Palansuriyam, C., Muller, C., Pandey, P.: Integrating Digital Preservation into Experimental Workflows for Space Science, PV2015, Darmstadt, Germany. http://pericles-project.eu/uploads/files/PERICLES_PV2015_KCL_Presentation.pdf.
6. Mitzias, P., Riga, M., Waddington, S., Kontopoulos, E., Meditskos, G., Laurenson, P. and Kompatsiaris, I.: An Ontology Design Pattern for Digital Video, Proceedings of the 6th Workshop on Ontology and Semantic Web Patterns (WOP 2015) co-located with the 14th International Semantic Web Conference (ISWC 2015), Vol. 1461, Bethlehem, Pensylvania, USA.

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

[25] https://ec.europa.eu/digital-agenda/en/digitisation-digital-preservation

7. Corubolo, F., Eggers, A., Hasan, A., Hedges, M., Waddington, S, Ludwig, J.: A pragmatic approach to significant environment information collection to support object reuse, iPres 2014, Melbourne, Australia. http://pericles-project.eu/uploads/files/ipres2014_PET.pdf
8. PERICLES Consortium: Deliverable D5.2 – Basic tools for Digital Ecosystem management (2015). http://pericles-project.eu/uploads/files/PERICLES_WP5_D5_2_Basic_Tools_for_Ecosystem_Management_V1_0.pdf.
9. PERICLES Consortium: Deliverable D6.4 – Final version of integration framework and API implementation (2015). http://pericles-project.eu/uploads/files/PERICLES_WP6_D64_Final_Version_of_framework_V1_0.pdf.
10. Higgins, S.: The DCC Curation Lifecycle Model. International Journal of Digital Curation. 3, (1), 134–40 (2008). Available from: http://ijdc.net/index.php/ijdc/article/view/69.
11. Ball, A.: Review of Data Management Lifecycle Models (version 1.0). REDm- MED Project Document redm1rep120110ab10. Bath, UK: University of Bath. (2012). http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf .
12. Committee on Earth Observation Satellites Working Group on Information systems and Services (WGISS): Data Life Cycle Models and Concepts CEOS 1.2 (2012). http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_DSIG_Data-Lifecycle-Models-And-Concepts-v13-1_Apr2012.docx.
13. CCSDS - Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2 (2012).
14. Upward, F.: Structuring the records continuum (Series of two parts) Part 1: post custodial principles and properties. Archives and Manuscripts. 24 (2) 268-285 (1996). http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp1.html.
15. McKemmish, S.: Placing records continuum theory and practice. Archival Science Volume 1, Issue 4, 333-359 (2001).
16. PERICLES Consortium: Deliverable D5.1.1 – Initial Report on Digital Ecosystem Management (2014). http://pericles-project.eu/uploads/files/PERICLES_D5_1_1-Preservation_Ecosystem_Management_V1_0.pdf
17. Tzitzikas, Y.: Dependency Management for the Preservation of Digital Information. Database and Expert Systems Applications, pp. 582-92. Springer Berlin Heidelberg (2007).
18. Tzitzikas, Y., Marketakis, Y., Antoniou, G.: Task-Based Dependency Management for the Preservation of Digital Objects Using Rules. Artificial Intelligence: Theories, Models and Applications, pp. 265–74. Springer Berlin Heidelberg (2010).
19. Marketakis, Y., Tzitzikas, Y.: Dependency Management for Digital Preservation Using Semantic Web Technologies. Int. Journal on Digital Libraries 10(4): 159-77, (2009).
20. Vorontsov K. V., Potapenko A. A.: Additive Regularization of Topic Models. Machine Learning. Special Issue “Data Analysis and Intelligent Optimization with Applications”, 101(1), 303-323, (2015).
21. Wittek, P., Darányi, S., Liu, Y.H.: A vector field approach to lexical semantics. In Proceedings of Quantum Interaction-14, Filzbach (2014).
22. Wittek, P., Darányi, S., Kontopoulis, S., Moysiadis, T., Kompatsiaris, I.: Monitoring Term Drift Based on Semantic Consistency in an Evolving Vector Field. In Proceedings of IJCNN-15 (2015). http://arxiv.org/abs/1502.01753.
23. Hedges, M. and Blanke, T.: Digital Libraries for Experimental Data: Capturing Process through Sheer Curation. In Research and Advanced Technology for Digital Libraries (pp. 108-119). Springer Berlin Heidelberg (2013).
24. Geoffrey I. Webb, G.I., Hyde, R., Cao, H., Nguyen, H-L., Petitjean, F.: Characterizing Concept Drift. Accepted for publication in Data Mining and Knowledge Discovery on December 10, 2015. At http://arxiv.org/pdf/1511.03816v5.pdf

25. Darányi, S., Wittek, P., Konstantinidis, K., Papadopoulos, S.: A potential surface underlying meaning? Yandex School of Data Analysis Conference, Machine Learning: Prospects and Applications, Berlin (2015). At https://www.youtube.com/watch?v=PEnRqv9hyzg.
26. Lerman, K., Galstyan, A., Ver Steeg, G., Hogg, T.: Social Mechanics: An Empirically Grounded Science of Social Media. In Proceedings of International AAAI Conference on Web and Social Media Fifth International AAAI Conference on Weblogs and Social Media (2011). http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/3836/4393.
27. PERICLES Consortium: Deliverable D2.3 – Media and Science Case Study Functional Requirements and User Descriptions (2014). http://www.pericles-project.eu/uploads/files/PERICLES_D231_Case_studies-Functional_Requirements_gathering.pdf.
28. PERICLES Consortium: Deliverable D4.2 Encapsulation of Environmental Information (2015). http://www.pericles-project.eu/uploads/PERICLES_WP4_D4_2_Encapsulation_of_Environmental_Information_V1.pdf.
29. Aitken, B., Helwig, P., Jackson, A., Lindley, A., Nicchiarelli, E., Ross, S.: The Planets Testbed: Science for Digital Preservation, The Code4Lib Journal, Issue 3, (2008). http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/planets-testbed#sthash.mRsLQkcg.dpu.f