

CONNECTING PRESERVATION PLANNING AND PLATO WITH DIGITAL REPOSITORY INTERFACES

Steve Hitchcock, David Tarrant, Les Carr
Electronics and Computer Science
University of Southampton, UK

Hannes Kulovits, Andreas Rauber
Vienna University of Technology
Austria

ABSTRACT

An accepted digital preservation workflow is emerging in which file formats are identified and those believed to be at risk are migrated to what are perceived to be less risky formats. This raises important questions about what to convert and when, if at all. In other words, how to connect file identification and migration. This area has become known as preservation planning, and seeks to take account of a wide variety of factors that might impact preservation decisions. Broadly there are two approaches to preservation planning. One provided in some digital preservation systems is to simplify and reduce both the number of file formats stored and therefore limit the number of preservation tools needed based on accepted recommendations. A more thorough, flexible and possibly complex approach, supported by the Plato preservation planning tool developed by the Planets project, allows decisions on preservation actions to combine analysis of the characteristics of different file formats with specific local requirements, such as costs and resources. This paper shows how Plato can be integrated with digital repository software, in this case EPrints, to enable this powerful approach to be used effectively to manage content in repositories of different sizes and with varying degrees of preservation expertise and support. These tools are accessed via a common repository interface to enable repository managers, and others who do not specialise in preservation, to moderate decisions on preservation planning and to control preservation actions.

1 INTRODUCTION

Progress has been made in the development of a framework and tools for digital preservation, but so far there has been little join-up or integration of these tools to create a workflow that is accessible from within digital repositories. Typically, support for digital preservation has been aimed at national libraries and archives or enterprise-level digital libraries that might have the scope and expertise to adopt complex and costly procedures. This does not apply to all digital repositories seeking to collect and provide access to the digital outputs of research and teaching of a single institution, at universities for example, and which

are now diversifying in terms of content collection and focus. This paper identifies a preservation workflow and tools that can be applied to digital repositories. We show how these tools can be accessed via a common repository interface to enable repository managers, and others who do not specialise in preservation, to moderate decisions on preservation planning and to control preservation actions.

A range of factors is driving the growth of repository content and the promise of long term preservation, and these in turn are driving the demand on the types of content a repository is expected to handle. As a result institutional digital repositories are now collecting not only peer reviewed publications and open access research but also scientific data, teaching and learning materials as well as arts and multimedia content. It is important to realise that as the range and diversification of these types of content increases, so do the problems with managing and preserving these resources. Likewise the number of tools, services and required infrastructure will also increase.

Digital preservation is now supported by a wide variety of tools, each with their own distinctive interfaces, as revealed and visualised by a series of detailed reviews of a selection of these tools by [10][11]. We are already seeing preservation tools that 'bundle' other tools to provide a specified workflow, e.g. File Information Tool Set (FITS)¹, and the emergence of preservation systems such as RODA² and Archivematica³, that seek to manage complexity via a single management interfaces. These aggregated tools and systems have not yet connected preservation support with the places where most new digital content is currently being deposited, stored and accessed, in the institutional repositories.

The importance of the interface in a digital system is clear from mass market consumer adoption. When launching the much publicised Apple iPad earlier this year, Apple CEO Steve Jobs said:

"75 million people already own iPod Touches and iPhones. That's all people who already know how to use the iPad."⁴

Familiar and successful interfaces reduce barriers to entry for systems and devices and enable users to make

¹ FITS - <http://code.google.com/p/fits/>

² RODA - <http://redmine.keep.pt/projects/show/roda-public>

³ Archivematica - http://archivematica.org/wiki/index.php?title=Main_Page

⁴ Steve Jobs, launching the iPad, January 27, 2010 http://news.cnet.com/8301-31021_3-10440943-260.html

faster progress and become more productive. For digital resources, many Web based repositories have been built on widely-used open source software such as DSpace, EPrints and Fedora. These have a common, and often underestimated, resource: their interfaces. In fact, repository software is essentially a series of interfaces for deposit, search, browse and management tasks performed by content authors and contributors, users of the information they provide, repository administrators and third-party service providers. It seems likely that additional repository services, such as preservation, should be provided through the familiar repository interface, rather than through the native interfaces of a disparate set of tools.

Although nascent repository policies don't yet state it explicitly, it is unlikely that repositories which grow on the basis of institutional requirements can escape the consequent expectation of effective content management over timescales consistent with the institutions planning horizons.

The JISC KeepIt repository preservation project is working with four specific repositories, chosen for the variety of content types which these repositories hold, to deploy an exemplar toolkit capable of helping and performing digital preservation on these repositories. By integrating a set of tools and services into existing repository software interfaces, we demonstrate not only the value to repository managers but also how taking this approach lowers the barrier to understanding and applying digital preservation.

In this paper we highlight how preservation workflow, and one particular part of that workflow, preservation planning, has been integrated within a repository interface.

2 PRESERVATION WORKFLOW AND REPOSITORIES

Digital libraries have long acknowledged that preservation is a vital part of the role of a repository. However, preservation is often sidelined due to the practical constraints of running a repository. Dealing with institutional-scale ingests and quality assurance with minimal staff and investment rarely leaves sufficient capacity for engaging with a preservation agenda when the creation of a concrete plan for preserving an institution's collection of digital objects may require the detailed evaluation of possible preservation solutions against clearly defined and measurable criteria.

Digital preservation is the process of storing and managing content for the purpose of continued access through changes in the technology framework over time, both to present the essential content or data (the digital bits) and, ideally, to be able to continue to represent the author's original intent and meaning through other features. Broadly, preservation has been modeled as a set of administrative processes allied to more technical processes for digital content management and storage. Underlying the latter are the computing applications and platforms that are used to create, distribute and access digital content, now includ-

ing repositories, the Web, and so on. This analysis of the purpose and practice of digital preservation has produced a consensus on a practical preservation workflow that, while it may differ in terminology, has a common core that can be represented with respect to digital objects and their formats as follows:

identification - characterisation - risk assessment - planning - action

The first and last elements of this workflow, covering actions such as format conversion, or migration, to safer formats, are the simplest to understand and tools are available to implement these processes. The key requirement now is joining these two end-processes through the more difficult, and subjective, steps of risk assessment and planning, to determine whether, and when, a preservation action such as a migration should apply.

To implement this workflow for EPrints digital repositories, KeepIt and its predecessor JISC Preserv projects have been applying tools for preservation workflow produced by some of the constituent partners in the European-wide Planets project⁵ such as the National Archives (of the UK, TNA), the British Library and the Vienna University of Technology (TU Wien):

- Format identification: DROID (TNA)
- Characterisation: XCL (XCDL, XCEL), a means of recording the significant characteristics of a digital object in an XML-based format
- Risk assessment, planning: PRONOM (TNA), PLATO (TU Wien)

KeepIt is building these tools into EPrints software through a series of plug-ins that provide access to the tools. EPrints has offered this modular application architecture since version 3.0 in 2007, and the latest version 3.2 is required to access these preservation plugins. The processes and services these tools provide can be accessed via a common interface that allows repository managers to moderate format risk assessment, and set parameters the repository software can use to make decisions on taking preservation actions such as migrating formats. Two successful workshops have been conducted to present first-hand experience of these tools and interfaces to repository managers [5][7].

The project has also contributed to the workflow by creating a format risk registry to show how format risk can become more open based on linked data principles [12]. The aim is to integrate the ability to process all preservation-related information within the repository. This includes extending to new means of representing provenance, such as the Open Provenance Model [9], and is all handled by tools that can pass XML-based information.

⁵ Planets, Preservation and Long-term Access through Networked Services <http://www.planets-project.eu/>

3 IDENTIFICATION & CHARECTERISATION

The first stage in the preservation workflow, identification, has a number of established tools. Using PRONOM-DROID[4], for example, the first Preserv project [3] created a central format profile service for repositories (available via roar.eprints.org), which revealed a heterogeneous range of formats were in use.

In Preserv2 we realised this service needs to be bought within the repository to play a useful part in a ‘smarter’ repository preservation workflow [6]. Repositories may not be able to reveal all stored content, and a service such as PRONOM-ROAR cannot provide information on files which are not publicly available. Additionally, some repositories may expose content incorrectly to Web harvesters such as PRONOM-ROAR - a typical example would be a PDF document that requires authorisation to access, would return an HTML page but without the required error code (HTTP 402), thus not making it clear to the harvester that this isn’t the resource it requested. Managing this service from the repository administration screen provides more detailed and trusted results. Figure 1 shows this screen as it stood at the end of the Preserv2 project.

What this approach demonstrated is how repository software can work with different preservation tools, in this case for format identification, and can import, process, display and export preservation-related data in XML-based formats, and this is the basis of the latest work to integrate preservation planning with the repository.

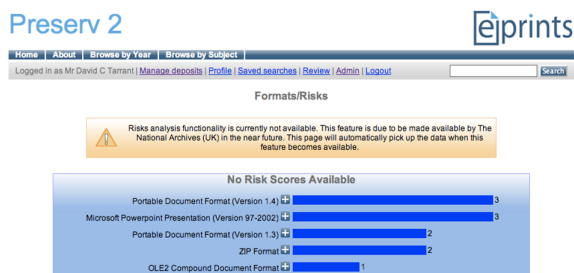


Figure 1. EPrints: Preservation interface showing file classifications

Figure 1 shows early developments on the preservation workflow in EPrints and this screen forms the basis of the rest of the work presented in this paper. At each stage of the preservation workflow this screen has been updated to become the central preservation control interface in the repository.

The identification of a file format can obscure as much as it can reveal of the essence of a digital object if we simply rely on this process to name the format. The format will have been created as the result of using one or more software applications, which will have allowed the creator to embed certain required features in the object, and also allows the user to recreate these features. Given the power of modern applications, it is possible that creator and user, or other interested stakeholders such as archivists, may seek to exploit different features in the object. This char-

acterisation of digital objects, according to the viewpoint of different stakeholders and the significance they may attach to features, is an emerging area of interest in digital preservation and presents additional tasks in the preservation workflow [8].

4 RISK ANALYSIS

While the preservation workflow has become clear, the basis for making decisions on how to implement each stage of the workflow has not because in many cases a detailed risk analysis is not available. What risks are posed by a given file format, by an alternative format and by the tools used for conversion, and how can these risks be quantified? To some extent these questions can be answered by registries such as PRONOM. As part of the Preserv2 project the National Archives (UK) constructed a series of risk categories and also a schema for assigning a risk value to each of these categories. The idea is to enable registries to generate numerical scores for format risk.

Using these categories and a hypothetical scoring system, the preservation interface in EPrints was enabled to obtain these risk scores from PRONOM and then display these in a traffic light scale depending on the score returned. Figure 2 shows the same preservation screen in EPrints, this time displaying information on format risk. Note that risk information depicted here is for demonstration purposes only and should not be considered an indication of actual format risk.



Figure 2. EPrints: Preservation interface with risk score data

Quantitative research in this area is still fairly new, so it remains more difficult to provide dynamic risk information as opposed to the comparatively static information describing file format. For example, the number of available tools and software products which can read a particular format, which are relevant factors in a risk analysis for that format, is likely to change more regularly than the documentation and encoding of the format itself (which typically only changes with a new version of the format).

[12] looked at possibilities to crowd source such information from publishers of linked data [2] and demonstrates the immediate benefit of combining information from PRONOM and dbpedia (wikipedia) and, potentially, other sources of format information. Ongoing work in this area is being performed by the Unified Digital Formats Registry (UDFR) project.

5 PRESERVATION PLANNING

By this stage in the preservation workflow we have identified a digital object by its format, we have analysed the characteristics we consider to be significant and that might impact preservation decisions on the object, and we have begun to quantify the risk associated with different formats and actions. All of this helps us to connect preservation action decisions with specific digital objects. Now we have to consider how this process might scale for the growing content of a digital repository. Digital content has been shown to grow at great rates, e.g. Planets market survey, so a way is needed to plan, record and, when necessary, update the whole analysis, and then to apply the plan when and where required to automate the outcomes. This is the preservation planning stage of the workflow.

Preservation planning involves evaluating available solutions against clearly defined and measurable criteria and arriving at a concrete plan of action based on these evaluation results. It is important to realise that this is both a static process which should be undertaken at the beginning of any project (alongside risk analysis) and also reactively as risks arise during the project itself.

Both Planets and EPrints believe the most effective form of file format preservation is done reactively. This implies that risk analysis is a constant process which identifies specific cases where a preservation plan may need to be constructed. This also avoids the risk of a preservation plan becoming outdated at a later stage, potentially causing a greater problem.

The remainder of this paper focuses on this small part of the preservation workflow, in effect the preservation planning sub-workflow, shown below, and considers enhancements to the repository and preservation planning tool to enable these to work together to provide a fully capable preservation solution.

**collection gathering - planning - management - action
- review**

6 COLLECTION GATHERING

A preservation plan relates to a set of digital files. This is the first role of the repository, enabling the user to identify the set of files at risk, using the interface described earlier, and to select a collection of these files ready to import into the planning tool.

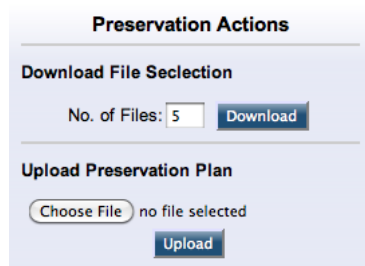


Figure 3. EPrints: Preservation Actions Panel

Figure 3 shows the Preservation Actions panel within EPrints. This panel corresponds to a single format that is identified as being at risk. Inspection shows this interface is also used in later stages of the planning process. Using this panel the user can choose the quantity of files required from the risk category, which can then be download and imported into the Plato planning tool.

To ensure the preservation plan is robust to the significant characteristics of a file format, it is important to provide the planning tool with a selection of files exhibiting as many of these characteristics as possible. While this can be done by performing deep analysis on the files, initially it was decided to settle on the following simple set of criteria, applied by EPrints, to select the files:

```
If more than 1 file requested:
    Provide Newest and Oldest
If more than 3 files requested:
    Also provide Largest and Smallest
Then
    Provide a random selection
```

Although simple, this approach should reveal a variety of factors. It is envisaged that further tools could be built into the repository to examine files in more detail, to provide a greater selection of files containing a wider range of significant characteristics. However, this should be done objectively such that the characteristics can be clearly identified, ensuring users understand why each file has been selected. A loss of understanding about why each file has been chosen could potentially be more wasteful than beneficial. This is certainly an area for further investigation and with the proliferation of tools in this area should certainly be one of the easier targets to achieve.

7 FORMING A PRESERVATION PLAN

With a collection of files gathered and imported to the planning tool, the next stage follows the workflow set by Plato, which is designed to guide the user through a set of experiments to make decisions and to formulate a preservation plan. The basic Plato workflow has four stages:

define requirements - evaluate alternatives (run experiments) - analyse results - build and validate preservation plan

Plato encourages users to think carefully about their decisions and allows improvements to be made iteratively at any stage of the workflow to obtain the best result. Each of these stages can be replayed at any time if it is realised that data is missing, but the eventual target remains the choice of one of the alternatives and making it the basis of the preservation action plan.

While there are many valid preservation actions, including do nothing, hardware/software emulation and migration, in the context of this report we are only considering migration. While “do nothing” is already supported as a no action plan, emulation with a repository context may require a substantial amount of further work to achieve.

With a great number of tools available to migrate one format to another, the choice of which new format to use can be quite complex, thus the need for a well defined set of requirements becomes ever more important.

Plato [1] defines three main stages in the preservation planning workflow:

1. Requirements definition: The important first step that defines the later evaluation criteria. Requirements from a wide range of stakeholders and influence factors have to be considered for a given institutional setting. This involves curators and domain experts as well as IT administrators and consumers. Requirements are specified in a quantifiable manner, starting with high-level objectives and breaking these down into measurable criteria, thus creating an objective evaluation tree. The requirements stage is also used to specify the significant characteristics and sustainability requirements which any plan must fulfill.
2. Evaluation of potential strategies: a series of tools are picked which suit the requirements outlined. Each file in the test set is migrated, and each successfully migrated file (e.g. an output was obtained) is stored for evaluation.
3. Analysis of the results: the results of the test migrations are evaluated against the requirements. As the requirements are weighted, this allows the planner to produce a well-informed recommendation for a preservation solution.

Finally, Plato allows a preservation plan to be exported. Although this could be the final plan, a plan can be exported at any time during the workflow and includes details of all requirements, tools selected as well in-line encoded copies of the files that were uploaded as the test set. The resulting action plan constitutes a small part of the total preservation plan, which is represented in XML and designed to contain a full record of all decisions and criteria that were evaluated in Plato. This plan can be reloaded into Plato at a later date for re-evaluation to ensure the defined action is still the most applicable.

Once a final preservation action is selected, this data is then also exported as part of the XML preservation plan, this small part of the plan conforms to a schema which has been specifically designed by the EPrints and Planets collaboration to be parsed easily by third party tools. This ability to output a clear and concise action plan that can be interpreted by other tools is a critical feature of Plato. A sample action plan is depicted by figure 4.

Figure 4 represents the minimal amount of information a repository such as EPrints needs to understand and to apply the action plan. Outlined below are the main elements of this plan and the reasons for their inclusion:

ActionPlan key:action values:migrate,emulate,none

Broadly outlines the plan and the chosen strategy. Note, it is important not to assume just because the preservation plan does

```
<ActionPlan action="migrate">
<tool>
<toolIdentifier uri="http://dbpedia.org/data/ImageMagick" version="6.5.1-0"/>
<parameters>
-verbose -compress None -quality 100 %INFILE% %OUTFILE%
</parameters>
<targetFormat mimetype="image/png" extension="png"/>
</tool>
</ActionPlan>
```

Figure 4. Plato: Example action plan

not contain an ActionPlan section that the preservation action is to do nothing. An action plan which does nothing should state this implicitly.

ToolIdentifier key:uri values:Semantic URIs

Identifies, using a globally unique identifier, the tool used in the action plan. In our example this tool is ImageMagick.

ToolIdentifier key:version values:Version Number

This is a vital field if you wish to verify the translation achieved by the tool is exactly the same as the one performed by the evaluation in Plato. A different software version here is likely to generate a slightly different file as it will write its own version information into the file.

Parameters

This field defines the parameters used in the execution of the tool. For the purposes of simplicity we have chosen the constants %INFILE% and %OUTFILE% to represent how files are parsed to the tool.

TargetFormat keys:mimetype,extension

The mimetype and extension define the resultant format of the migration. Although only one should be required, some mimetypes can have multiple extensions which each tool may choose to use, thus it is handy to include both.

Plato thus forms one part of the preservation workflow which takes a set of inputs, in this case a set of files of a single format which have been identified as being at risk, to help the user produce a preservation plan, the output. The key to making these components work in a repository environment is to be able to handle both the inputs to the planning process and the subsequent output.

8 PRESERVATION PLAN MANAGEMENT

Once the preservation plan has been formed, we now require the repository to accept this plan to be managed, preserved and acted upon. As shown in figure 3, EPrints allows the preservation plan to be uploaded directly via the same preservation actions user interface used to download the original at-risk repository files to Plato. Each file format can be related to a single preservation plan that can be uploaded via the preservation interface, and these can be managed using the screen section shown in figure 5. This shows a preservation plan has been defined for the GIF image format from 1987, and that this plan, uploaded in March 2010, has performed an action on a single file in the repository, defined by the action plan outlined in the previous section.

Preservation Plans				
ID	Import Date	Related Formats	Quantity	Actions
50	25 March 2010 15:00:35 +01:00	Graphics Interchange Format (Version 1987a)	1	Download Plan Delete Docs

Figure 5. EPrints: Preservation plan management panel

This successful migration means that EPrints has been able to find the tool defined by Plato and act upon the plan in order to migrate not only the files that were given to Plato in the test data set but also all other files of this type in the repository. By handing back the task of mass migration of all files of the identified type to the repository alleviates the scalability issues. Here the repository is already handling this quantity of content and may as a result already be linked to services and service providers who can assist with on-demand processing for larger operations such as migration. Similar concepts for mass identification, that can be mapped to mass migration, are discussed in [6].

9 VALIDATING RESULTS

EPrints displays preservation results for the administrator via a screen that tracks file formats, showing the quantity of files in each format classified according to the associated risks on a traffic light scale, where red represents high risk and green low risk. This screen also presents formats resulting from actions performed by the preservation plan. Figure 6 is a snippet from this interface showing a set of high risk (GIF 1987a) files in the low risk category, reflecting the new risk category for the migrated files, with the red bar reflecting which category these files would be in had the migration not taken place.



Figure 6. EPrints: Preservation interface showing migrate files

A single file of the same format remains in the high risk category because it has no migrated version yet. This could be a newly uploaded file or one which failed the migration. Pressing the '+' button would allow the plan to be manually executed on this file, and also allows the repository administrator to see when the file was created and provide other information about recent related processes. Note that EPrints never deletes any files, i.e. the originals, without explicitly being asked to do so by a depositor, editor or repository manager.

EPrints also updates the record, or abstract page relating to each item, to show that a preservation action has taken place. Figure 7 shows a single file, in PNG format, that has been migrated from the original GIF version.

Both the migrated and original files are shown in this instance with the relation between them clearly displayed. This also demonstrates part of the provenance information stored within EPrints relating any migrated files not only to the originals but also to the preservation plan that caused that migration.



Figure 7. EPrints: Abstract screen for migrated record

Figure 7 shows a public-facing EPrints abstract page containing details of a migrated file which is the result of a preservation plan. The original, pre-migration, file is kept but simply has less prominence.

10 PROVENANCE AND PRESERVATION METADATA

Provenance is an important aspect of digital preservation to establish the authenticity of objects. By migrating an object the repository is creating a new version which, for it to be authentic in the eyes of the user, needs a full set of preservation metadata detailing why and how this new version was created. We have shown how the repository stores the full preservation plan containing all the requirements and decisions made during the planning process. It should also be possible to find which files resulted from that plan or, vice-versa, which plan a file was a result of.

EPrints uses the Open Provenance Model (OPM), and data stored by EPrints relating to preservation and migration can be easily serialised according to this model. OPM [9] defines a minimal set of core elements, including the following which are detailed in terms of their application in EPrints:

- **WasDerivedFrom:** relates the original to the migration file. In EPrints this is the two-way relation `isMigrationVersionOf` and `hasMigratedVersion`.
- **WasInformedBy:** relates the preservation plan to the action which took place. In EPrints this is a one-way relation defined by `isMigrationResultOf`.
- **WasGeneratedBy:** holds many important roles, both to define which tool generated the plan, but also which tool was used to perform the migration. In EPrints this information is left in the preservation plan from where it can be sourced.

With EPrints and many other repository platforms accepting arbitrary linked data (triples) relations between objects, adding this type of data to an existing record is well supported. Both the objects and the preservation plan (which is also an object) obtain a persistent identifier, which can be used to relate the objects.

Figure 8 provides an overview of the relations between the objects in EPrints that are part of this preservation

process, including the original and migrated files as well as the preservation plan and file format which relates the original file to the preservation plan. This shows some of the key actors involved in this plan and the relations between them. For clarity, this omits the preservation actions related to the migrated file via the preservation plan.

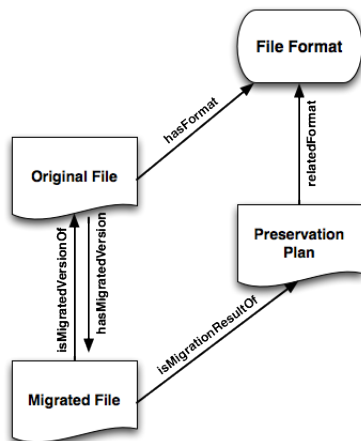


Figure 8. Provenance and preservation metadata structure

11 CONCLUSION

A preservation workflow has become established to manage the file formats of digital objects and take preemptive actions to ensure the objects remain accessible and usable as originally intended. Critical new developments have been described in this paper that enable this workflow to be managed from sources such as digital repositories that are seeing rapidly increasing volumes of content deposited yet are often managed with few resources for preservation. Key to implementing the workflow within these repositories is to use and adapt familiar repository interfaces rather than require administrators to learn new interfaces for many preservation tools required to implement the workflow. We have shown how the results from a powerful preservation planning tool, Plato, can be applied and controlled using this repository preservation interface. Importantly, through a series of workshops, we have also shown that by making these tools look familiar to the repository managers, that the barrier to learning the issues with digital preservation and understanding the responsibilities is lowered. Subsequently through the KeepIt project, these tools have now been rolled out to a number of partner institutions as well as being made available freely online. While it is clear there is still work to be done in some areas, completing the join up of the preservation workflow from characterisation to preservation action within a familiar interface represents a huge leap forward for digital preservation.

12 REFERENCES

- [1] Becker, C., et al., "Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans" *International Journal on Digital Libraries (IJDL)* December 2009 <http://www.ifs.tuwien.ac.at/~becker/pubs/becker-ijdl2009.pdf>
- [2] Berners-Lee, T., "Linked Data" *W3C Design Issues* 2006
- [3] Brody, T., et al., "PRONOM-ROAR: Adding Format Profiles to a Repository Registry to Inform Preservation Services", *International Journal of Digital Curation*, Vol. 2, No. 2, December 2007 <http://www.ijdc.net/ijdc/article/view/53>
- [4] Brown, A., "Automatic format identification using PRONOM and DROID" *The National Archives, Digital Preservation Technical Paper* 2005
- [5] Field, A., et al., "Digital Preservation: Logical and bit-stream preservation using Plato, EPrints and the Cloud." In *13th European Conference on Digital Libraries (ECDL)*, 27 September 2009, Corfu <http://eprints.ecs.soton.ac.uk/17962/>
- [6] Hitchcock, S., et al., "Towards smart storage for repository preservation services" In: *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects* 29-30 September 2008, London, UK <http://eprints.ecs.soton.ac.uk/16785/>
- [7] Hitchcock, S., et al., "Digital Preservation Tools for Repository Managers 4: Putting storage, format management and preservation planning in the repository" In: *KeepIt course module 4*, 18-19 March 2010, Southampton, UK <http://eprints.ecs.soton.ac.uk/21029/>
- [8] Hitchcock, S., et al., "Digital Preservation Tools for Repository Managers 3: Describing content for preservation" In: *KeepIt course module 3*, 2 March 2010, London, UK <http://eprints.ecs.soton.ac.uk/21001/>
- [9] Moreau, L., "The Open Provenance Model Core Specification (v1.1)" 21 Dec 2009 <http://eprints.ecs.soton.ac.uk/18332/>
- [10] Prom, C., "Archive for category Software" *Practical E-Records*, Curious Entries 2009-10, <http://e-records.chrisprom.com/?cat=3>
- [11] Prom, C., "PLATO (Digital Preservation Planning) Software Review" *Practical E-Records*, April 25, 2010 <http://e-records.chrisprom.com/?p=1082>
- [12] Tarrant, D., et al., "Where the Semantic Web and Web 2.0 meet format risk management: P2 registry", In *iPres2009: The Sixth International Conference on Preservation of Digital Objects*, October 5-6, 2009, San Francisco <http://eprints.ecs.soton.ac.uk/17556/>