Project no. 269977

APARSEN

Alliance for Permanent Access to the Records of Science Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

D16.1 SOFTWARE REPOSITORY

Document identifier:	**APARSEN**-REP-D16_1-01-1_0
Due Date:	2012-09-01
Submission Date:	2012-12-12
Work package:	WP16
Partners:	APA, CERN, ICT, IKI-RAS, CINES
WP Lead Partner:	APA
Document status	FINAL

Abstract: This document reports on the software repository that has been set up by APARSEN and is being populated with information about preservation-related software. It outlines the features that the repository offers, particularly for categorisation and associating evidence of effectiveness in preservation.

Delivery Type	Report
Author(s)	David Giaretta
Approval	
Summary	Report on repository of preservation software being set up by APARSEN.
Keyword List	Preservation software, Repository
Availability	☒ Public

Document Status Sheet

Issue	Date	Comment	Author
0.1	2012-09-25	First draft to represent the software repository as a deliverable document	Simon Lambert
0.2	2012-10-17	Updates to include details of repository	David Giaretta
1.0	2012-12-12	Updates based on comments from internal reviewer, adding details for software updates.	David Giaretta Gerald Jaschke

Project information

Project acronym:	**APARSEN**
Project full title:	**Alliance for Permanent Access to the Records of Science Network**
Proposal/Contract no.:	**269977**

Project coordinator: Simon Lambert/David Giaretta

Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk

CONTENTS

1 INTRODUCTION

This document represents deliverable D16.1, 'Software repository'. The deliverable is classified as 'Other' rather than 'Report'—it is the repository itself—but this report is submitted to explain the intent behind the repository, the way it has been implemented, and to give a brief overview of it.

The repository is located on the APARSEN public website at:

http://www.alliancepermanentaccess.org/index.php/knowledge-base/existing-tools/tools-for-preservation/

According to the Description of Work, the objective of WP16 is simple: to create and populate a repository of preservation-related software. This is motivated by a perceived need to move away from a situation where preservation-related software (understood to be publicly available, at least in principle) is distributed across a wide range of locations and in diverse ways, making it very difficult to retrieve and reuse.

The Description of Work envisages the possibility of depositing software itself in the APARSEN repository. At the present stage of development this has not been found necessary, as the emphasis has been on building up a set of software that is already accessible and so only requires linking. What the APARSEN repository (strictly speaking it is a registry/repository since it is to contain both actual software as well as pointers to software) adds is a uniform analysis and categorisation of preservation software, with where possible a base of evidence for judging its effectiveness in reality. It thus fulfils the expectation in the DoW that it 'will contain the software or pointer to the software together with descriptions, categorisations and annotations which will allow users to find appropriate tools for their requirements. Of particular importance is information about the strengths and areas of applicability of each tool. The system will allow users to provide their own evaluations and annotations.'

The Dow also says 'We will adopt a repository system within which can be deposited the source code, service modules and executables together with the required documentation about requirements that have been addressed, application and user guides, prototypical implementation examples, and metadata sets that allows one to find and retrieve the stored information as well as the licensing and contact details about their originators', As described in section 2, we adopt SourceForge.

The repository is designed to have a growing amount of content that may be added by registered users of the website.

This deliverable was originally expected to be submitted in February 2012 (Month 14 of the project). However it was deferred because of the dependency on WP14 'Common testing environments' because it became clear that it would be sensible to include the rough classification for digital objects described in WP14 in the description of the software[1]. In that way we could provide a view on the applicability of software and their test environments to the various types of data. Again it should be emphasised that this classification is not meant to be exhaustive or definitive, and indeed it will almost certainly evolve over time; it merely provides a guide to help to check that we are considering examples outside our comfort zone. Nor is a specific digital object necessarily pigeon-holed in one classification. However we believe that this at least gives some guidance to users.

[1] See for example http://www.alliancepermanentaccess.org/index.php/knowledge-base/existing-tools/tools-for-preservation/rough-classification-of-digital-objects/

2 DESCRIPTION OF THE SOFTWARE REPOSITORY

The user view of repository has been implemented as a set of web pages under the APARSEN area of the website of the Alliance for Permanent Access. Much of the software we list are already homed in one or other repository and so while that is the case it seems sensible to leave it there. However for examples of orphan software (i.e. ones from closed projects and not being actively supported) we propose to use the SourceForge repository http://digitalpreserve.sourceforge.net/ where we are not forbidden from doing so. Otherwise we will keep the source code in a dark repository associated with the website.

The following screenshots illustrate some of the important features, including the introductory page (Figure 1), and examples of: listing of software with basic information (Figure 2); detailed information for one piece of software (Figure 3) and example of the connection between software and evidence (Figure 4)

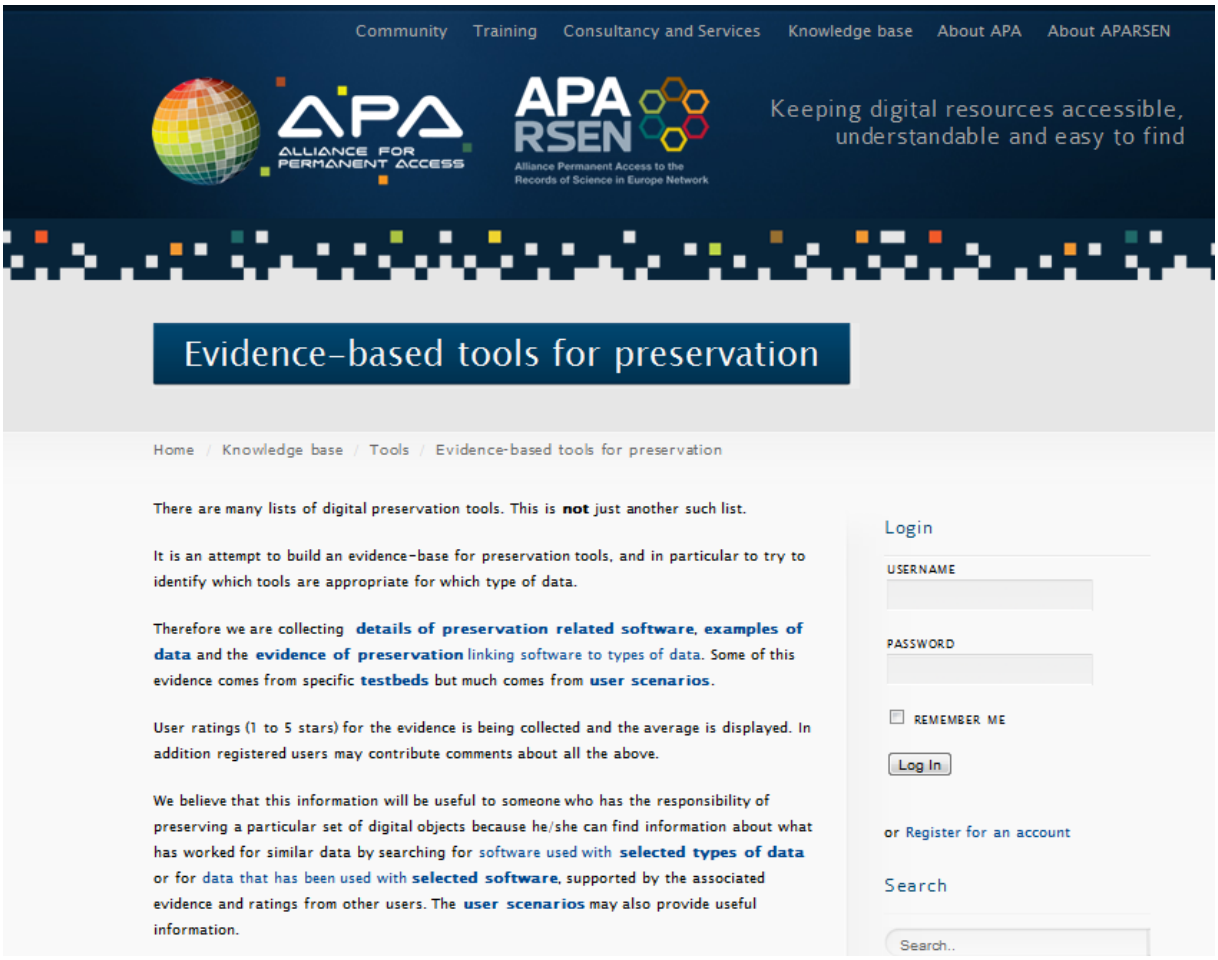


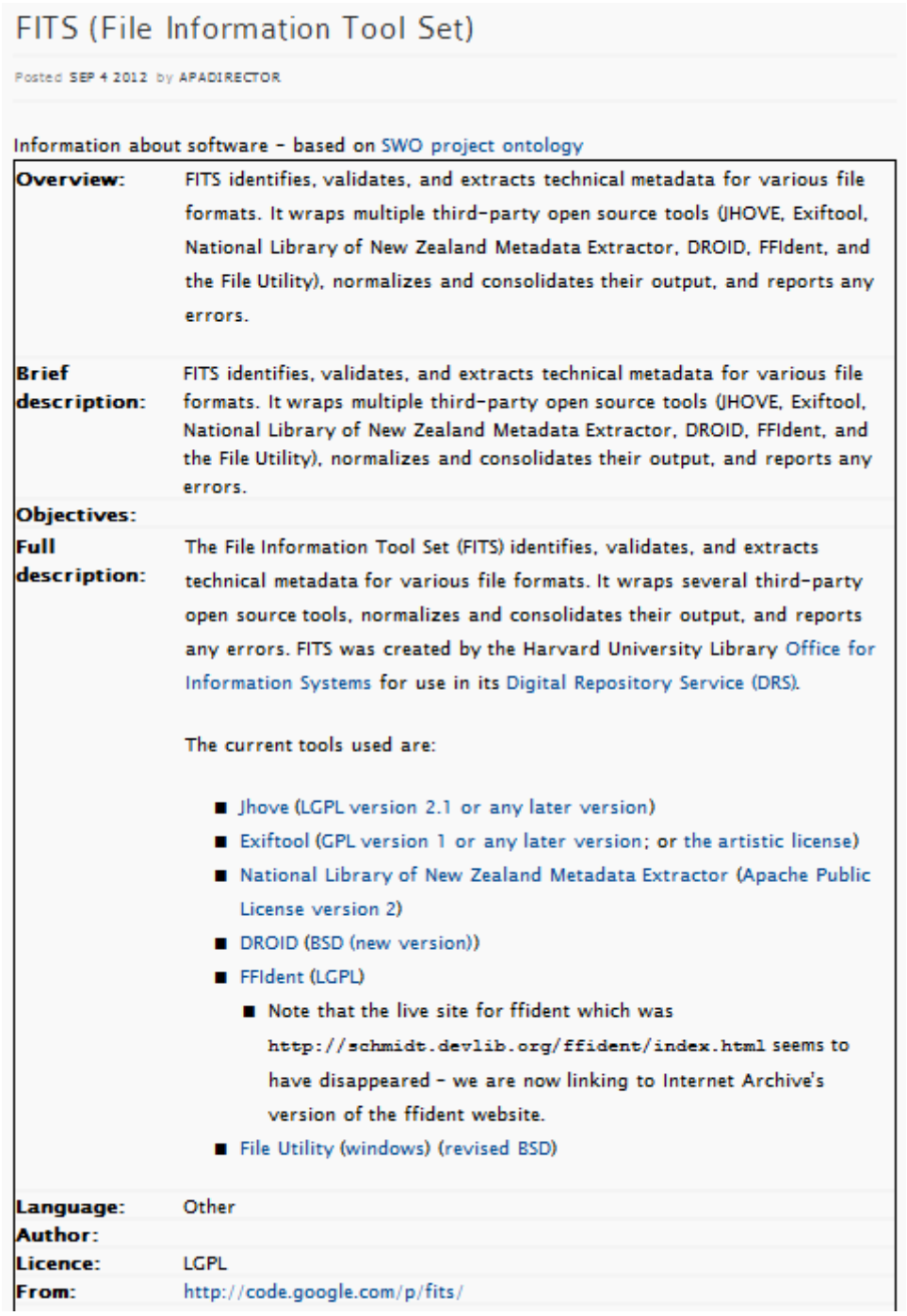
Figure 1 Front page of the repository

The aim is to provide more than just another collection of pointers to software. We aim to provide at least some indication of where that software has been found to be effective in preservation. A number

of different ways of searching are provided: select software to preserve a type of data; select data types which can be preserved with certain software; see the evidence which supports claims of effectiveness. A user rating scheme is also in place which will help users to prioritise the selection preservation strategies. Registered website users can also contribute comments and suggestions.

Link to more details	Short description	OAIS role	Objective
Text print software	Print simple text fie	Use – rendering	Render text
PREMIS in METS (PiM) Toolbox	The Library of Congress and the Florida Center for Library Automation developed the PREMIS in METS (PiM) Toolbox. The project provides PREMIS:METS conversion and validation tools that support the implementation of PREMIS in the METS container format.	Preservation description information	
PLATO – Planets Preservation Planning Tool	Preservation planning tool for migration strategies checked with significant properties (TBC)	Preservation Planning	
MIXED	Migrate content to XML format and then convert to new format.	Migration	Transform digital content to XML
Format Identification for Digital Objects (FIDO)	Format Identification for Digital Objects (FIDO) is a Python command-line tool to identify the file formats of digital objects. It is designed for simple integration into automated work-flows.	Structure RepInfo	Tool to identify file formats of objects which can be used in an automated workflow.
FITS (File Information Tool Set)	FITS identifies, validates, and extracts technical metadata for various file formats. It wraps multiple third-party open source tools (JHOVE, Exiftool, National Library of New Zealand Metadata Extractor, DROID, FFIdent, and the File Utility), normalizes and consolidates their output, and reports any errors.	Data management	
EZID	EZID (easy-eye-dee) makes it easy to create & manage unique, long-term identifiers ■ create identifiers for anything: texts, data, bones, terms, etc. ■ store citation metadata for identifiers in a variety of formats ■ update current URL locations so citation links are never broken ■ use EZID's programming interface for automated operation at scale ■ choose from a variety of persistent identifiers, including ARKs and DataCite DOIs	Reference	
DSPS (Digital Preservation Software Platform)	The Digital Preservation Software Platform (DPSP) is free and open source software developed by the National Archives of Australia. The DPSP is a collection of software applications which support the goal of digital preservation.	Transformation – Nonreversible	
DROID 6	**DROID (Digital Record Object Identification)** is a software tool developed by The National Archivesto perform automated batch identification of file formats.	Structure RepInfo	To identify the format (Structural Representation Information) of digital objects..

Figure 2 Partial list of entries in repository with basic information



FITS (File Information Tool Set)

Posted SEP 4 2012 by APADIRECTOR

Information about software - based on SWO project ontology

Overview:	FITS identifies, validates, and extracts technical metadata for various file formats. It wraps multiple third-party open source tools (JHOVE, Exiftool, National Library of New Zealand Metadata Extractor, DROID, FFIdent, and the File Utility), normalizes and consolidates their output, and reports any errors.
Brief description:	FITS identifies, validates, and extracts technical metadata for various file formats. It wraps multiple third-party open source tools (JHOVE, Exiftool, National Library of New Zealand Metadata Extractor, DROID, FFIdent, and the File Utility), normalizes and consolidates their output, and reports any errors.
Objectives:	
Full description:	The File Information Tool Set (FITS) identifies, validates, and extracts technical metadata for various file formats. It wraps several third-party open source tools, normalizes and consolidates their output, and reports any errors. FITS was created by the Harvard University Library Office for Information Systems for use in its Digital Repository Service (DRS). The current tools used are: ■ Jhove (LGPL version 2.1 or any later version) ■ Exiftool (GPL version 1 or any later version; or the artistic license) ■ National Library of New Zealand Metadata Extractor (Apache Public License version 2) ■ DROID (BSD (new version)) ■ FFIdent (LGPL) ■ Note that the live site for ffident which was `http://schmidt.devlib.org/ffident/index.html` seems to have disappeared - we are now linking to Internet Archive's version of the ffident website. ■ File Utility (windows) (revised BSD)
Language:	Other
Author:	
Licence:	LGPL
From:	http://code.google.com/p/fits/

Figure 3 Detailed entry information using the example of File Information Tool Set (FITS)

Information collected about each piece of software follows the SWO ontology[2].

[2] http://theswo.sourceforge.net/

Link to more details	Description of evidence	TestbedEvidence	Attached evidence	--Rating--	Data link	Software link
ESA use of DRB	ESA have many examples of satellite data in their archives, together with tools which extract the information using the DRB descriptions. DRB is documented and the generic applications can be re-implemented in future, which should give access to all data described using DRB.	See the "measurement.xsd" and "ers-object-types.xsd" in the sample file http://www.alliancepermanentaccess.org/wp-content/uploads/2012/09/GOME_SAFE.zip. The SAFE descriptions can be used to extract the information elements from the data as shown by the applications provided with SAFE.		2 votes Cast your vote now!	ESA test data - various science files	DRB - Data Request Broker

Figure 4 Example of evidence relating to software

3 ENTERING DETAILS OF ADDITIONAL SOFTWARE AND EVIDENCE

After logging in to the website, an authorised user can click on the “Manage Page” button in the right hand panel.

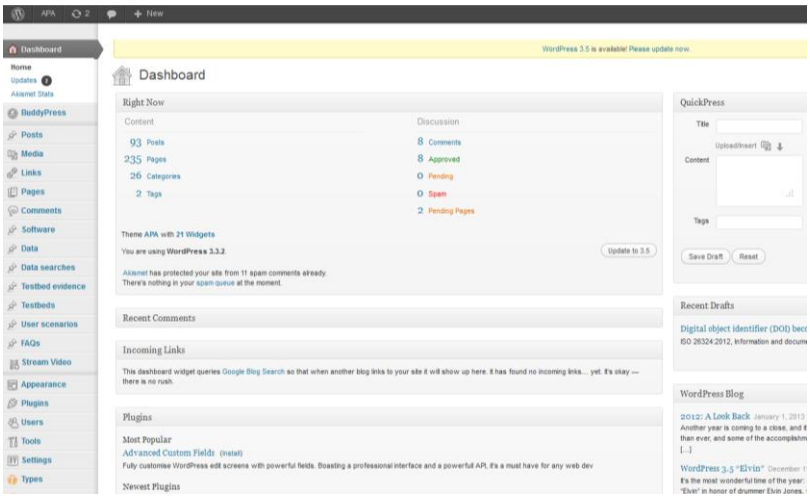


Figure 5 Manage page screen

Then select “Software” then “Add New”

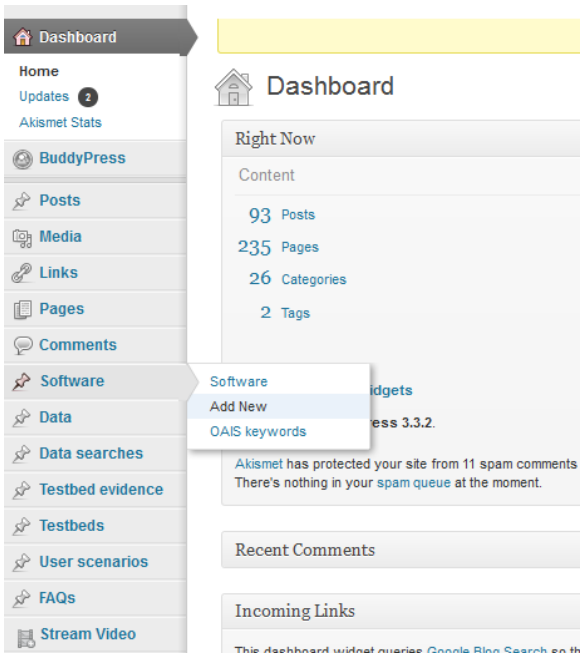
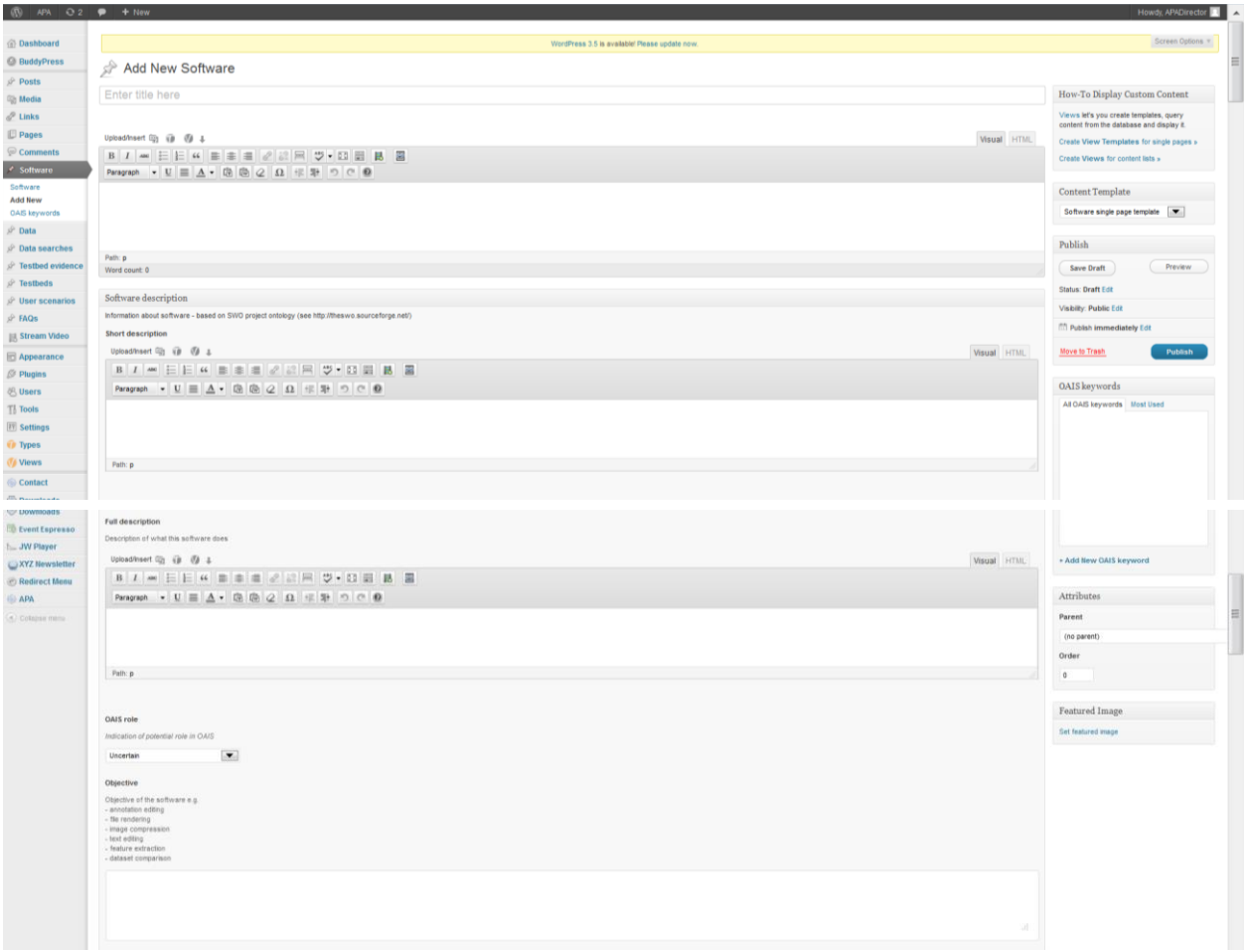
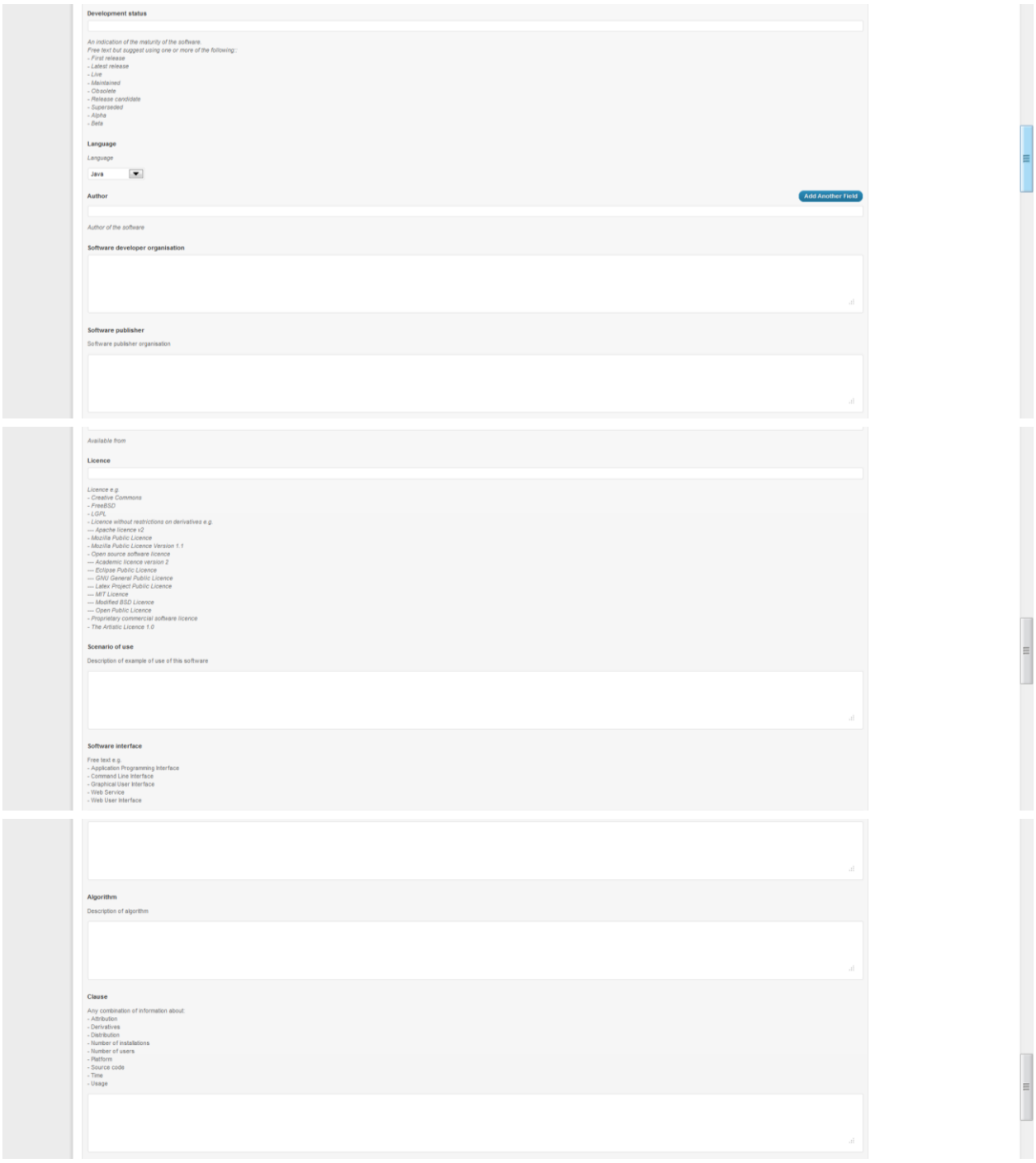


Figure 6 Select "Add New" software

The following screenshots capture the input page in sections.




Development status
An indication of the maturity of the software.
Free text but suggest using one or more of the following:
- First release
- Latest release
- Live
- Maintained
- Obsolete
- Release candidate
- Superseded
- Alpha
- Beta
Language
Language
Java
Author
Add Another Field
Author of the software
Software developer organisation
Software publisher
Software publisher organisation
Available from
Licence
Licence e.g.
- Creative Commons
- FreeBSD
- LGPL
- Licence without restrictions on derivatives e.g.
--- Apache licence v2
- Mozilla Public Licence
- Mozilla Public Licence Version 1.1
- Open source software licence
--- Academic licence version 2
--- Eclipse Public Licence
--- GNU General Public Licence
--- Latex Project Public Licence
--- MIT Licence
--- Modified BSD Licence
--- Open Public Licence
- Proprietary commercial software licence
- The Artistic Licence 1.0
Scenario of use
Description of example of use of this software
Software interface
Free text e.g.
- Application Programming Interface
- Command Line Interface
- Graphical User Interface
- Web Service
- Web User Interface
Algorithm
Description of algorithm
Clause
Any combination of information about:
- Attribution
- Derivatives
- Distribution
- Number of installations
- Number of users
- Platform
- Source code
- Time
- Usage

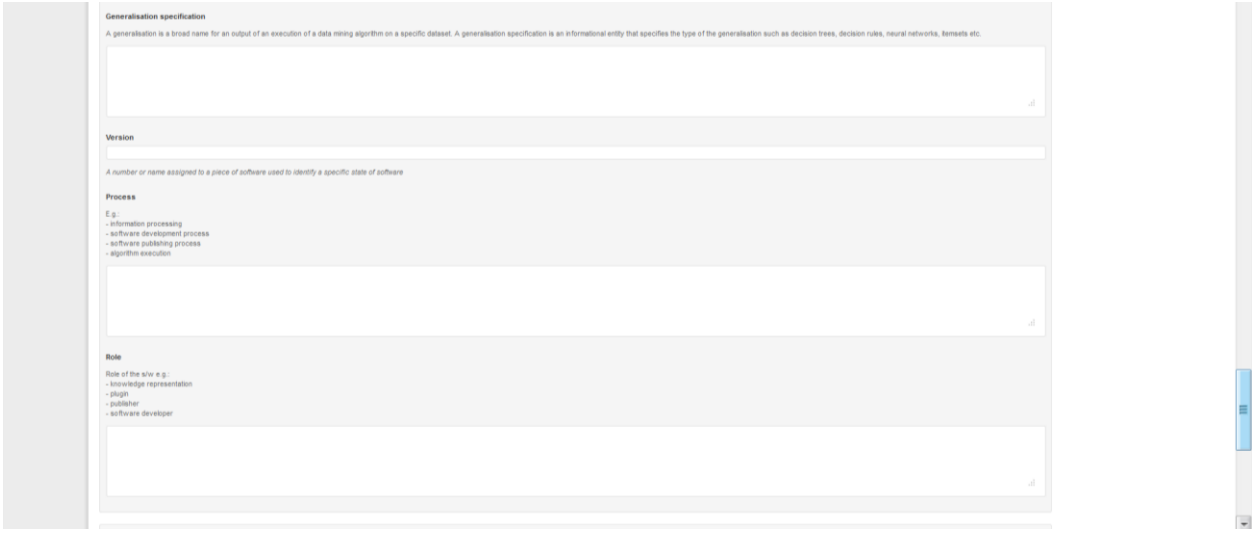
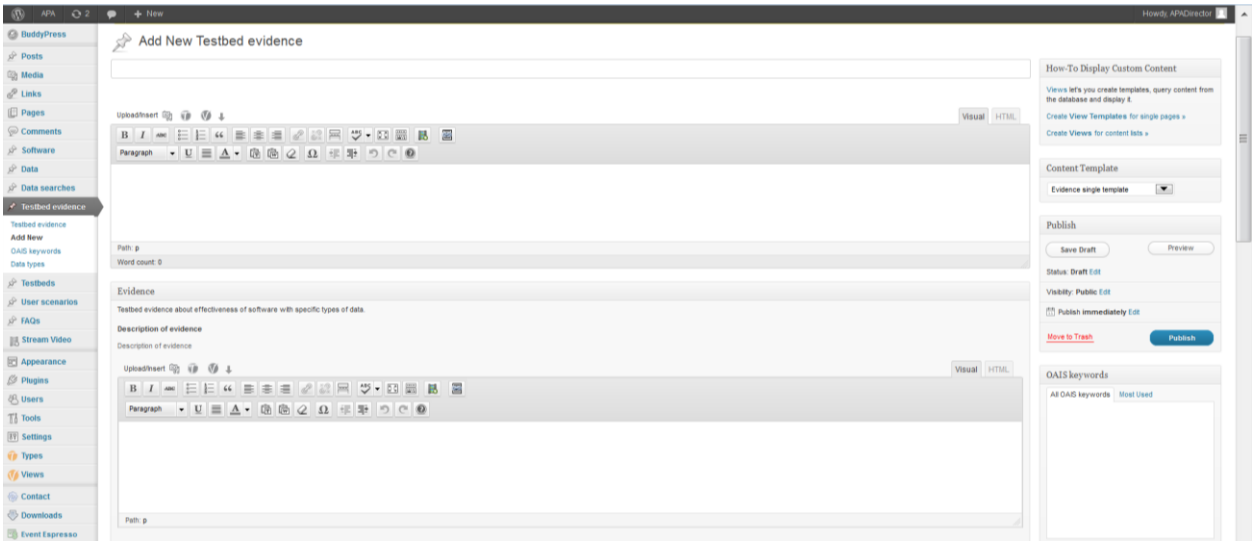


Figure 7 Form for adding new software

To add evidence one can click on “Testbed evidence” then “Add New”

Figure 8 Select "Add New" Testbed evidence

The form is as follows:



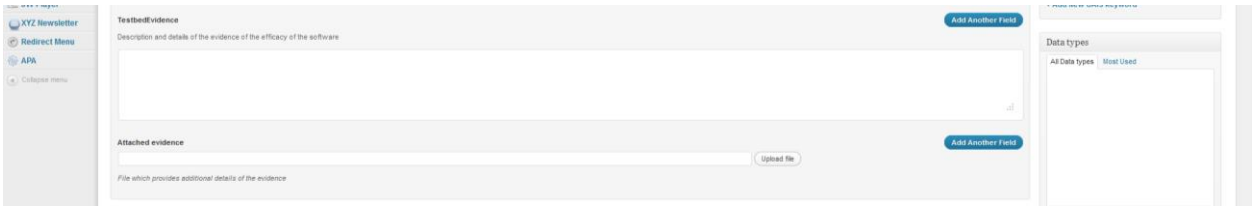


Figure 9 Adding new evidence

After saving this the following appears at the bottom of the form:

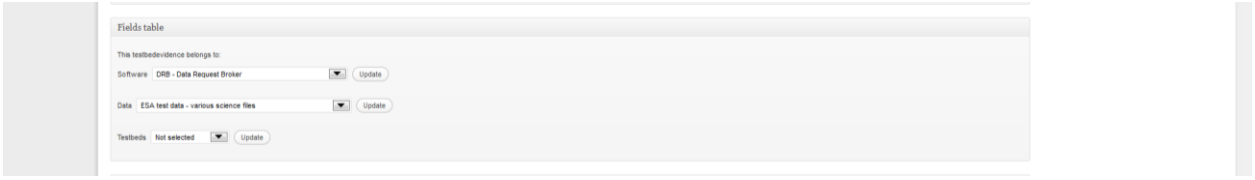


Figure 10 Additional fields to link Data and Software to the Evidence

After selecting from each of the dropdown boxes in turn, the "update" button beside the dropdown should be clicked.

4 UPDATING SOFTWARE DETAILS

Details for any piece of software may be updated/edited by selecting "Software"

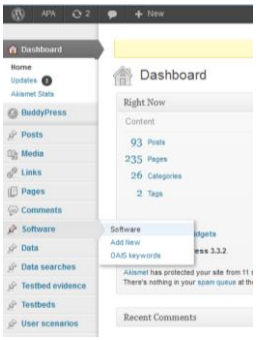


Figure 11 Select "Software" for a list of all software

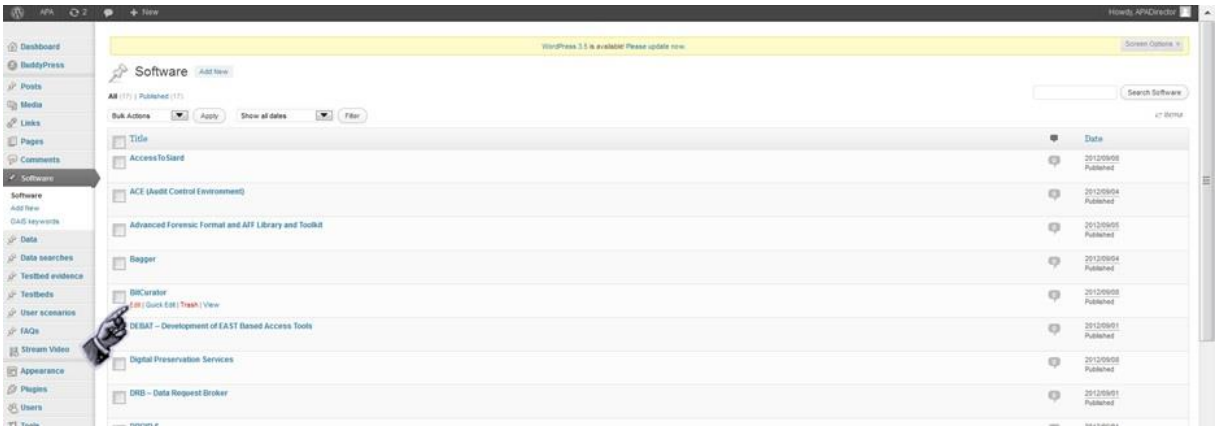


Figure 12 Select the software entry to edit

The software details can be edited and the changes saved by clicking on "Updated"

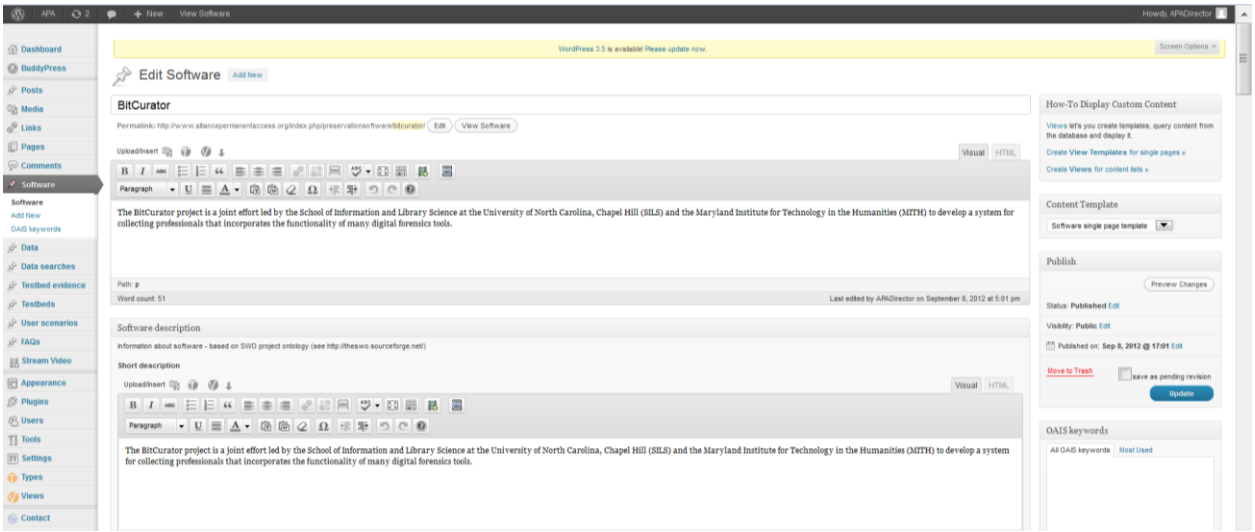


Figure 13 Form to edit software details

5 FURTHER DEVELOPMENTS

WP16 runs until the end of the project in Month 48. In future, the repository will be enhanced by the addition of further content – we aim to harvest information from as many of the extant lists of software and, with the help of comments from the community, to provide useful evaluations. Where we find orphan software which we have reason to believe fill a gap then we will do our best to take over the source code so that it is not lost, but we cannot provide any guarantees of support. We will also try to take source code snapshots even of software that does have a home and store a copy; this copy will be kept in the dark repository associated with the website rather than SourceForge, to avoid confusion.

We will ensure that the software and website can be continued by the Virtual Centre of Excellence.