

AWE - Presentation for Open Repositories 2010

Peter Sefton

sefton@usq.edu.au

University of Southern Queensland

Duncan Dickinson

Duncan.Dickinson@usq.edu.au

University of Southern Queensland

The short paper Duncan Dickinson and I put together for this conference is organised around the conference themes, and what our Research and Development group at the Australian Digital Futures Institute is doing about each of them. In this presentation I will pick out some of the the work we're doing and some of the issues we're thinking about, and try to relate this work back to the conference themes.

It's gratifying to be able to make this presentation. Some of the core ideas we're talking about here were the subject of a proposal I submitted for Open Repositories 2007, but it received mixed reviewer feedback and was relegated to a poster; my message that repositories were stuck in “Web 0.5” and needed to be made more webish was not timely.¹

Conference themes

Please tick these off as I go.

1. **The web and the repository & The cloud and the desktop***
2. Knowledge and technology
3. Wild and curated content
4. Linked and isolated data & Ad-hoc and long-term access
5. Disciplinary and institutional systems / Scholars and service providers
6. Ubiquitous and personalized environments

* This is the big one!

So, what is this AWE thing?

AWE: The Academic Working Environment

Duncan: “A set of purposeful technologies brought together by standard interfaces for data exchange?”

Peter: “I needed a single cover-all name for all the different projects we were working on in our [institute](#), so that I could try to report to the powers that be in a more efficient way.”

¹ At dinner on Monday night I was told that the ideas we're pushing here at ADFI are “15 years ahead” of where the users are. Given that this paper was accepted, and doing some quick maths in my head I think that means that we have only 25 months before the things that we're talking about here are considered relevant, given that we're working in Internet time.


The Web and the Repositories

We're **on** the web, but are we **of** the web?

Show and tell

- Repositories need to understand **renditions** and **rendering**
 - Rendering: The ICE service – renders web versions of everything it can and extracts metadata.
 - Renditions: See our site for [USQ policies](#). HTML all the way.
- We're working on small pieces of web-infrastructure that should apply across multiple repositories.
 - Anotar: annotations toolkit designed to be easy to plug in to any web system – even I can do it. See this [WordPress version](#).
(Not as good as [digress.it](#) but much more portable).
 - [Paquete](#): An ePUBesque reader you can embed in any web page.
Also a way for us to support compound objects in a repository.
- We're working on bringing the web to the desktop.
<Insert demo of The Fascinator Desktop>

Screenshot: HTML documents derived from Word documents

 UNIVERSITY OF
SOUTHERN QUEENSLAND

USQ Policy Library (Demonstration)

USQ Home > USQ Policy Library (Demonstration) > Search

NOTICE

This is **NOT** the official USQ Policy Library. This is a **DEMONSTRATION** site which may contain outdated information.

USQ Policy Library

> Policy Library

> **Search**

> Help

Actions

> Clear selection

Policy type

☐ Operational (54)

☒ Human Resources (54)

Title

▲

Advanced Filters ▲

Policy search

Search

☒ Free text search ☐ Title search

Displaying 1 to 10 out of 54 policies

Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [Next](#) [Last](#)

Search Results

[Academic Development and Outside Studies Program \(ADOSP\)](#)
Ensures academic employees are aware of the opportunities and conditions associated with developmental pr

[Acting Appointments](#)
Outlines the processes used to offer temporary higher duties and

[Anti-Discrimination and Freedom from Harassment](#)
Ensure that the University complies with its legal responsibilities in accordance with the relevant Acts.

[Appointment and Reappointment to Senior Positions](#)
Outlines the processes associated with the appointment and reappointment to a senior management or other r

[Appointment by Internal Transfer or Secondment](#)
Outlines the processes involved in the internal transfer or secondment of an employee to a position within the

[Attraction and Retention Loadings](#)

Screenshot: ICE conversion service

ICE conversion service

Document

URL

or Upload No file chosen

Services

Use these links to specify document options, otherwise defaults will be used.

- OpenOffice.org/Power point conversion service [[odp](#)] [[ppt](#)] [[pptx](#)]
- Scalable Vector Graphics conversion service [[svg](#)]
- OpenOffice.org/Excel conversion service [[ods](#)] [[xls](#)] [[xlsx](#)]
- LaTeX conversion service [[tex](#)]
- Converting pdf to html [[pdf](#)]
- Chemical Markup Language conversion service [[cml](#)]
- OpenOffice.org/Word conversion service [[odt](#)] [[doc](#)] [[docx](#)] [[rtf](#)]
- audio/video conversion service [[mp3](#)] [[mp4](#)] [[m4a](#)] [[wma](#)] [[wmv](#)] [[mpg](#)] [[mpeg](#)] [[mov](#)]
- Text to Speech conversion service [[txt](#)]
- Resizing/Cropping images service [[png](#)] [[jpg](#)] [[jpeg](#)] [[gif](#)] [[bmp](#)]
- Wave to MP3 conversion service [[wav](#)] [[aiff](#)]

Screenshot: ICE conversion options for word processing

ICE conversion service

Document

URL

or Upload

Browse...

Reset

Convert

OOo/Word options

Template

```
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html;
  charset=UTF-8"/>
  <title>Default Template</title>
  <style type="text/css">
```

or Upload

Browse...

Rendition

- ☐ Include link to PDF rendition
- ☐ Include link to slide rendition
- ☒ Generate table of contents
- ☐ Do not include title in the HTML body
- ☐ Include source document (including derived ODT from DOC)
- ☐ Generate MODS
- ☐ Generate METS (return XML unless ZIP selected)
- ☐ Generate DC (return XML unless ZIP selected)
- ☐ Generate RDF (return XML unless ZIP selected)
- ☐ Return as ZIP package

Screenshot: Annotations on a (this) document

The short paper Duncan Dickinson and I put together for this conference is organised around the conference themes, and what our Research and Development group at the Australian Digital Futures Institute is doing about each of them. In this presentation I will turn things around and talk about a some of the work we're doing, and some of the issues we're thinking about, and relate this back to the themes.

It's gratifying to be able to make this presentation. Some of the core ideas we're talking about here were the subject of a proposal I submitted for Open Repositories 2007. [TODO – reviewer feedback]

Comment by: dickinso Tue Jul 06 07:45:17 2010 Close Reply

What were your main points in 2007 and where has everything moved since then?

Conference themes

1. The web and the repository & The cloud and the desktop*

2. Knowledge and technology

3. Wild and curated content

4. Linked and isolated data & Ad-hoc and long-term access

Comment by: dickinso Tue Jul 06 07:51:12 2010 Close Reply

I think this is interesting in terms of what we keep and what we throw away

5. Disciplinary and institutional systems / Scholars and service providers

Comment by: dickinso Tue Jul 06 07:52:02 2010 Close Reply

Build it and they will come. Link into real researcher workflows, not institutional report gathering exercises

6. Ubiquitous and personalized environments

Comment by: dickinso Tue Jul 06 07:53:24 2010 Close Reply

Links to your vertical market comment well. Computer as appliance (toaster) has challenges in terms of where to store and how to present. I like the way dropbox handle this.

* This is the big one!

Screenshot: Paquete stand-alone demo

Paquete

- What is Paquete?
- JSON Manifest**
- Install/setup Paquete
- Configure Paquete
- More information?

JSON Manifest

- 2.1 The manifest.json file
 - 2.1.1 JSON file structure

2.1 The manifest.json file

2.1.1 JSON file structure

Paquete requires a json file to store the package structure information. The file called **manifest.json** should have the following structure.

```
{
  "toc": [
    {
      "visible": true,
      "relPath": "c09bf5db57268971eb69ab42e127cc60/Test_document.htm",
      "title": "Test Document 1",
      "children": []
    }
  ],
}
```

Issue: What will happen when vertically controlled computing platforms are the norm?

Does the desktop/lab/home PC become a server? Or will it live in the cloud?

I brought an iPad with me on this trip, but I have found that I can't put music on it or look at the books I have bought across **four** – yes, four – different reader applications using standard file-operations. Imagine the problems if we have to deal with valuable research data which lives on these controlled devices. A whole new era of format lock-in that makes Microsoft look like a free-software hippy.

This is one reason why the web, and delivering stuff in web formats is important.

Idea: How about ePUB as a repository packaging format?

(In my reading of the specs) ePUB is engineered for overloading:

A zip file containing:

- (At least) XHTML, with optional extra elements & a flat table of contents.
 - Can include video, chemistry, whatever if you provide fallback image/text.
- Allows for alternate renditions such as PDF, docx or odt originals.
- We could include an HTML 5 version (using something like Paquete) for modern browsers/devices.
 - Javascript is allowed – but must be ignored by ePUB readers but could be used by web apps.

Idea: ePUB use-cases

Repository: (SIP/AIP/DIP?)

- Serve the content using an in-page eReader (Paquete).
- Let the package handle package semantics (pre-print, published version, presentation), repository can continue to handle streams.
- Support viewers for data types for defined periods, such as JMOL for chemistry using something like ePUB's fall-back mechanism and oEmbed.

Users:

- If all else fails: unzip the package and click 'index.html'
- Use with eBook reader software/hardware.

Developer / repository manager:

- Add more packaging info – ORE, METS etc.

One of the other things we look at at ADFI is ways of bootstrapping the Linked-Data/Semantic Web. I (Peter) have proposed a way of embedding RDF statements in documents using simple interoperable URIs. Duncan has taken this work further with a system that can serve 'proper' RDF. The demo here shows this technique for metadata, but it could also be applied for other kinds of semantics, when you are talking about someone, for example rather than asserting that they are an author or an editor.

Idea: Making linked data, well, links

- Step 1: Approach your repository or ID provider, search for self

<http://nla.gov.au/nla.party-541658>

Note: wrapping this link around some text is essentially meaningless. It's not the semantic web it's the old-web.

- Step 2: Copy the link labelled "Assert Authorship"

[http://ontologize.me/meta/?](http://ontologize.me/meta/?r=http://purl.org/dc/terms/creator&o=http://nla.gov.au/nla.party-541658)

[r=http://purl.org/dc/terms/creator&o=http://nla.gov.au/nla.party-541658](http://purl.org/dc/terms/creator&o=http://nla.gov.au/nla.party-541658)

Word processor-proof linked data - part 2

- Step 3: Paste onto document text

Author: [Peter Sefton](#)

- Step 4: Deposit somewhere that understands

Step 3; Paste onto document text

Author: [Peter Sefton](#)

Step 4: Deposit somewhere that understands

This is the author, ID: [<http://nla.gov.au/nla.party-541658>]

```
<oai_dc:dc>
```

```
<dc:title>AWE - Presentation for Open Repositories 2010</dc:title>
```

```
...
```

```
<dc:creator>http://nla.gov.au/nla.party-541658</dc:creator>
```

```
</oai_dc:dc>
```

Conclusion

We (like many others) are building a toolkit for web/repository construction. The key reasons we rolled our own are:

- We care about having web-resources not just PDF.
- We wanted to be able to deploy the application to the desktop (hence a Java app that can be deployed with Apache Solr).

Copyright Peter Sefton & Duncan Dickinson, 2010. Licensed under Creative Commons Attribution-Share Alike 2.5 Australia. <http://creativecommons.org/licenses/by-sa/2.5/au/>



This post was written in OpenOffice.org, using templates and tools provided by the [Integrated Content Environment](#) project and published to WordPress using [The Fascinator](#).