Integrating the repository with academic workflow

Ian Barnes

APSR & The Australian National University
The problem

• Example: Demetrius@ANU

  • Search for “electronic publishing”

    • 5 items in ePrints...

    ...all by Colin Steele

• Search for “computer science”

  • 104 items!

  • 102 added on 19 May 2004 (CS Technical Reports 1992-2004, all in ePrints, mass migration to DSpace)

  • Only 2 since then...
Why?

- People don’t know? ...
- People don’t care? ...
- Copyright issues? ...
- IT’S TOO HARD
  - Repositories don’t support the most popular document formats
  - Metadata has to be added
What to do about this?

• Integrate repository deposit into academic workflow
  • Make it normal
  • Make it easy
  • Make it visible
The digital scholar’s workbench

- Prototype environment
- Bring together many services
  - Document conversion to better format
  - Publishing in multiple formats and media
  - Backup & version control
  - Collaboration
  - Course materials & websites
- Authors write in their favourite word processor
  - ...but they must use the template
  - Minimal change to how they write today
High-level Architecture

File system

DSpace

Storage layer

Workbench (Cocoon)

Middleware layer

Cocoon

Application layer

Word processor

Web browser

User
Sample Ice document

Ian Barnes

Thursday, February 02, 2006

Abstract: This is a short sample document created using the ICE template as a test of the Digital Scholar’s Workbench. This abstract should automatically get captured as metadata in the Dspace ingestion process.

Keywords: word processing, electronic publishing, XML, XSLT

This is an ordinary paragraph before the start of the first section. Make sure it doesn’t get dropped by the structure deepening code, OK?

Here we are at Open Repositories 2006. There’s a very stylish green T-shirt on the desk.
Sample Ice document

Ian Barnes

Thursday, 2 February 2006, 12:15:21 AM

Abstract:

This is a short sample document created using the ICE template as a test of the Digital Scholar’s Workbench. This abstract should automatically get captured as metadata in the Dspace ingestion process.

Table of Contents

1. Introduction
2. Lists
   2.1. Bullet lists
   2.2. Numbered lists
2.3. Definition lists
3. Inline formatting
4. Code samples
   4.1. Subheading
5. Blockquotes
6. Headings
   6.1. Subheading
   6.2. Second subheading
7. Publishing
This XML file does not appear to have any style information associated with it. The document tree is shown below:

```xml
- <office:document>
  + <office:document-meta office:version="1.0"></office:document-meta>
  + <office:document-styles office:version="1.0"></office:document-styles>
  - <office:document-content office:version="1.0">
    <office:scripts/>
    + <office:font-face-decls></office:font-face-decls>
    + <office:automatic-styles></office:automatic-styles>
  - <office:body>
    - <office:text>
      + <office:forms form:automatic-focus="false" form:apply-design-mode="false"></office:forms>
      + <text:sequence-decls></text:sequence-decls>
      - <text:p text:style-name="P1">
        <text:title>Sample Ice document</text:title>
      </text:p>
      - <text:p text:style-name="P2">
        <text:user-defined text:name="Author">Ian Barnes</text:user-defined>
      </text:p>
      - <text:p text:style-name="p-center">
        <text:date style:data-style-name="N38" text:date-value="2006-02-02T00:15:21">Thursday, February 02, 2006</text:date>
      </text:p>
    - <text:p text:style-name="Abstract">
```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

- `<article>
  - `<articleinfo>
    - `<title>`Sample Ice document</title>
    - `<authorgroup role="show">`
      - `<author>`
        - `<firstname>`Ian</firstname>
        - `<surname>`Barnes</surname>
      </author>
    </authorgroup>
  - `<affiliation role="suppress">`
    - `<orgname>`The Australian National University</orgname>
  </affiliation>
  - `<email role="suppress">`Ian.Barnes@anu.edu.au</email>
  - `<pubdate role="show">`2006-02-02T00:15:21+11:00</pubdate>
  - `<abstract>
    - `<title>`Abstract: </title>
    - `<para>`
      This is a short sample document created using the ICE template as a test of the Digital Scholar’s Workbench. This abstract should automatically get captured as metadata in the Dspace ingestion process.
    </para>
  </abstract>`
Sample Ice document

Ian Barnes

Thursday, 2 February 2006, 12:15:21 AM

Abstract:

This is a short sample document created using the ICE template as a test of the Digital Scholar's Workbench. This abstract should automatically get captured as metadata in the DSpace ingestion process.

Table of Contents

1. Introduction ................................................................. 1
2. Lists ............................................................................. 2
   2.1. Bullet lists ............................................................. 2
   2.2. Numbered lists ....................................................... 2
2.3. Definition lists ........................................................... 3
3. Inline formatting ............................................................ 3
4. Code samples ................................................................. 4
   4.1. Subheading ............................................................ 4
   4.2. Second sub-heading ............................................... 4
5. Blockquotes ................................................................. 4
6. Headings ...................................................................... 4
   6.1. Subheading ............................................................ 5
   6.1.1. Sub-sub-heading ............................................... 5
   6.1.1.1. Sub-sub-sub-heading .................................... 5
   6.1.1.2. Second sub-sub-sub-heading ....................... 5
   6.1.2. Second sub-heading ......................................... 5
   6.2. Second sub-heading ............................................... 5
7. Lists again, seriously .................................................... 5
8. White space? ............................................................... 6
9. Mixed nested content .................................................... 6
10. Images ...................................................................... 8
11. Citations and bibliographies ...................................... 0
12. Fancy characters ........................................................ 0
Bibliography ................................................................. 0

This is an ordinary paragraph before the start of the first section. Make sure it doesn’t get dropped by the structure deepening code, OK?

Here we are at Open Repositories 2006. There’s a very stylish green T-shirt on the desk.
Please use this identifier to cite or link to this item: http://hdl.handle.net/123456789/1136

Title: Sample Ice document

Authors: Barnes, Ian

Keywords: word processing  
electronic publishing  
XML  
XSLT

Issue Date: 1-Feb-2006

Abstract: This is a short sample document created using the ICE template as a test of the Digital Scholar’s Workbench.  
This abstract should automatically get captured as metadata in the Dspace ingestion process.

URI: http://hdl.handle.net/123456789/1136

Appears in Collections: Dr Ian Barnes
Sample Ice document

Ian Barnes

Thursday, 2 February 2006, 12:15:21 AM

Abstract:

This is a short sample document created using the ICE template as a test of the Digital Scholar’s Workbench. This abstract should automatically get captured as metadata in the Dspace ingestion process.

Table of Contents

1. Introduction
2. Lists
   2.1. Bullet lists
   2.2. Numbered lists
   2.3. Definition lists
3. Inline formatting
4. Code samples
   4.1. Subheading
5. Blockquotes
6. Headings
   6.1. Subheading
Sample Ice document

Ian Barnes
Thursday, 2 February 2006, 12:15:21 AM

Abstract:

This is a short sample document created using the ICE template as a test of the Digital Scholar's Workbench. This abstract should automatically get captured as metadata in the Dspace ingestion process.

Table of Contents

1. Introduction ................................................................. 1
2. Lists ............................................................................. 2
   2.1. Bullet lists .......................................................... 2
   2.2. Numbered lists .................................................... 2
   2.3. Definition lists .................................................... 3
3. Inline formatting ......................................................... 3
4. Code samples .................................................................. 4
   4.1. Subheading .......................................................... 4
5. Blockquotes ................................................................... 4
6. Headings ....................................................................... 4
   6.1. Subheading .......................................................... 5
      6.1.1. Sub-subheading .............................................. 5
      6.1.1.1. Sub-sub-subheading .................................. 5
      6.1.1.2. Second sub-sub-sub-heading .................. 5
      6.1.2. Second sub-sub-subheading ....................... 5
   6.1.2. Second sub-subheading .................................. 5
7. Lists again, seriously ................................................... 5
8. White space? ............................................................. 6
9. Mixed mixed content .................................................. 6
Still to come

• Blog this (really soon)
• Microsoft Word documents (soon)
• Upload to LMS or course website (soon)
• Document version control (soonish)
• Submit to journals (long-term ongoing)
• Quality print formatting for theses & books (maybe soonish)
• Complex objects & multimedia (not soon)
• Bibliography management (very hard)
• Powerpoint Presentations (maybe)
Summary

Q: How do we get people to put their work in the institutional repository?

A: Make it an easy & normal part of their work:
  • Integrate the repository with academic workflow as one of many useful services
  • Provide incentives for authors:
    IF you do it our way
    THEN you get all this for free
Supported By:

- Thanks to:
  - Adrian Burton (APSR Program Manager)
  - Scott Yeadon (DSpace guru, DRS, ANU)
  - Peter Sefton (RUBRIC Technical Manager, ICE team leader, USQ)
  - Peter Raftos (Program Leader, Digital Resource Services (DRS), ANU)
  - Leo Monus (DRS Software Engineer)