Community Involvement Experience from EPrints

Leslie Carr
EPrints Technical Director & EPrints Repository Manager & Information Systems Researcher
Overview

Personal, Partisan & Partial Perspectives on a Process

Stage 0: What is EPrints
- Where did it come from?
- Where is it now?
- Software decisions

Stage 1: What role has Open Source played?

Stage 2: What role has an Open Source community played

Reflections on Open Source in HE

Conclusion

PS EPrints immediate priorities
Stage 0: EPrints Genesis

- In an ideal world of scholarly communication – all research is freely available through research archives

- However, the work of researchers in our institution (even our own research) is often unavailable

- **June 27th 1994** Stevan Harnad’s ‘Subversive Proposal’ leading to the open access vision for scholarly material

- **October 1997** Stevan obtained funding for Cogprints archive (JISC)

- **April 1999** Rob Tansley starts working for SH to generalise Cogprints

- **October 1999**, Beginning of Open Citations JISC/NSF IDL II project

- **October 1999**, Open Archiving Initiative (Santa Fe) between research archive maintainers
  - OAI-PMH metadata interoperability protocol
  - EPrints Archive software announced
Stage 0: EPrints Genesis

- **2000** First release of EPrints, directed by SH, implemented by RT, funded on the sly by OpCit
  - RT leaves for HP & MIT & DSpace
  - Collaboration begun with Cornell team members who seem to be working on some object-based DL called Fedora

- **Oct 2001 - Sept 2002** JISC explicit EPrints funding to make code OS (£30K)

- **Sept 2002 - Sept 2004** JISC follow-on EPrints funding (£72K) for upgrades

- **Oct 2004 - Sept 2006** JISC follow-on EPrints funding (£92K) for open source community developments (and software developments)

- For most of the history of EPrints, it has had a small development team which is highly focused.

- For most of history of EPrints, focus has been on promoting OA, not EPrints!

- **Oct 2005 -** EPrints Services, floating new development and support team.
What is EPrints today?

- EPrints enables you to build a **repository** that allows **individuals** to deposit **important digital items** with **appropriate metadata**
  - for dissemination
  - for curation
  - for marketing
  - for reporting

- It has enjoyed the largest user base of public OAI repositories

- Marketing emphasises low impact and low cost
EPrints Full 10-Step Programme

1. Brand your repository (logos, stylesheets)
2. Define the object types and their metadata
3. Declare the deposit workflows including help text, examples and validation rules
4. Define collections, searches, subscriptions
5. Refine rendering for objects
   6. Expose new metadata fields & deploy embedded applets, media players on abstract page
   7. Create exciting new citation styles
   8. Define object -> OAI mapping
9. Populate repository.
10. Go to step 1 to revisit requirements and ensure job security.
1. Brand your repository (logos, stylesheets)
9. Populate repository.

Hoist by our own petard!
To encourage institutions to adopt OAI repositories in 2000
- Promote EPrints as an out-of-the-box, 2-step, low-maintenance experience

But the message “that’s all you need to do” becomes the message “that’s all you can do”
As the community matures at an uneven pace, marketing message gets mixed!
Step 1 Success

- Looks really different
- But is just a default EPrints install
Step 10 Success - Second Cycle

- Data repository which exposes scientific metadata.
- Schema standards accepted by international subject body.

Available Files

Final Result

Compound Class: Organic
Keywords: Controlled Keywords UNSPECIFIED

Data collection parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical formula</td>
<td>C28H32N2 O11S</td>
<td>17k</td>
</tr>
<tr>
<td>Cell length a</td>
<td>10.9877(7)</td>
<td>11k</td>
</tr>
<tr>
<td>Cell length b</td>
<td>11.9703(8)</td>
<td>1k</td>
</tr>
<tr>
<td>Cell length c</td>
<td>22.4693(18)</td>
<td>15k</td>
</tr>
<tr>
<td>Cell angle alpha</td>
<td>90.00</td>
<td>10k</td>
</tr>
<tr>
<td>Cell angle beta</td>
<td>90.00</td>
<td>49k</td>
</tr>
<tr>
<td>Cell angle gamma</td>
<td>90.00</td>
<td>47k</td>
</tr>
<tr>
<td>Data collection temperature</td>
<td>120(2)</td>
<td>40k</td>
</tr>
</tbody>
</table>

Solution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution figure of merit</td>
<td>0.1386</td>
<td>29k</td>
</tr>
<tr>
<td>R Factor (Obs)</td>
<td>0.0688</td>
<td>8k</td>
</tr>
<tr>
<td>R Factor (All)</td>
<td>0.3088</td>
<td>8k</td>
</tr>
<tr>
<td>Weighted R Factor (Obs)</td>
<td>0.1318</td>
<td></td>
</tr>
<tr>
<td>Weighted R Factor (All)</td>
<td>0.1930</td>
<td></td>
</tr>
</tbody>
</table>

Other Files

Archive Staff Only: edit this record.
What Has EPrints Been Used For?

- **Research**
  - Research outputs: papers, theses, multimedia (music, sculptures, performances, textiles *etc*) [TARDis]
  - Research data [EBank]
  - Research Laboratory workflow [R4L]
  - E-research / E-learning managed environments (learning contracts, research workflow -> journal submission) [CORE]
  - Research administration and assessment [IRRA]

- **Publishing**
  - Journal reviewing management (BBSprints)

- **Preservation**
  - Preservation services [PRESERV]

- **Collections management** (see [www.eprints.org/software/examples](http://www.eprints.org/software/examples))
  - Government and administrative documents
  - Oral Histories
Out of the box (notionally)

Agile
- We still don’t know what a repository should do!
- LAMP - dominant web development environment
- Low commitment programming

Functionality that promotes research (business objective) is paramount
- Open Access (means to a research end)
- Collections management (not so important)
- Preservation (important, but not important enough to hinder immediate dissemination)
EPrints development driven by increasing repository uptake

- The moral high ground of Open Access and citation impact is not enough to motivate researchers on a daily basis

- Need to develop a carrot and stick approach
  - mandates (policies)
  - benefits (software facilities)
Automatically updated CV

On all researchers personal home pages. Also group lists for each research group’s portal.
Benefits: RSS for latest research

Welcome to ECS, the School of Electronics and Computer Science at the University of Southampton. ECS is one of the world’s leading centres for research, teaching, enterprise and innovation in Electronics, Electrical Engineering and Computer Science.

Headline News

ECS improves broadcasting options for Virgin Radio
The School of Electronics and Computer Science (ECS) at the University of Southampton is the first university department to deploy IPv6 in production on its network. Now a team at the School has worked on the technology with Virgin Radio to vastly increase its listening capacity online.

New School Research Brochure highlights advances and successes
The School’s new Research Brochure provides an overview of the research environment within the School, including our two world-leading research groups, our research centres and institutes, and our research opportunities for postgraduate students. You can download a PDF of the Brochure, or order copies by emailing enquiries@ecs.soton.ac.uk.

Multimedia communications come of age at ECS
A group of undergraduate students from...
Benefits: high local visibility
Benefits: high local visibility

Recent ECS publications


Benefits: marketing

Butterflies’ wings dazzle with science

The brilliant dazzle of butterflies’ wings could hold the key to a new type of optical material, called photonic crystals. Over the past 15 years, photonic crystals have attracted the attention of a vast international community, as scientists have begun to realise their potential applications in the field of optoelectronics and telecommunications.

According to Dr Luca Plattner, who undertook research in the School of Electronics and Computer Science at the University of Southampton, our understanding of the way that light is reflected from the wings of butterflies could lead to the fabrication of new photonic crystals.

Dr Plattner investigated the optical properties of a periodic nanostructure found on the wings of a tropical butterfly, Morpho rhetenor. Several decades of scientific investigation had shown that understanding the source of the butterfly’s dazzling blue coloration required the use of the most advanced techniques employed in optical engineering.

Dr Plattner’s study explored the remarkable properties of the nanostructures and the physical mechanisms that produce them, both experimentally through optical measurements which complemented those reported by other scientists, and theoretically via cutting-edge simulation techniques developed for phonics. This enabled him to fabricate optical structures inspired by the butterfly microstructure using silicon-based materials and processes that are common in microelectronics. The work was carried out under the supervision of Professor Greg Parker.

‘The reason for studying the structure on the wings of that particular butterfly was that it has strong similarities to the photonic crystals already fabricated in the ECS-Roelofarends Research Group,’ said Luca Plattner. ‘It was able to explore a biomimetic process, one in which we can learn new lessons from nature which are beneficial to both engineers and biologists.’

Plattner’s work was published in the first print issue of the Royal Society’s Interface magazine, due out on 22 November.

More Information

Benefits: metadata reuse for administrative reports

Carr, Les


Also selected by: Hall, Wendy


Involvements in conferences/workshops (Program Committees, Chairing, etc.)


Responsibilities in professional associations (IEEE)

1. Carr, Les
2. Hall, W.
3. Bailey, C.
4. Miles-Boyadji, T.
5. Lansdale, J.
6. Dewaril, T.
7. El-Beltagy, S.
8. Roue, D. D.
9. Kampa, S.
10. Hall, W.
Stage 1: Role of Open Source

- Jan 2002 Open Source Software (GPL)
  - Adopt a coding standard
  - Adopt a documentation standard
  - Adopt appropriate licensing
  - Upload regular (?) distros

- However, Southampton maintains the initiative, direction and control and effort
Stage 2: Role of Open Source Community?

- These developments were by and large “internal”
  - Come from Open Access imperative
  - Engagement with nascent Open Access community
  - Experiments on our own local users

- Fundamental user interface & functional improvements came from JISC TARDis project
  - Library driven approach for institutional scale
  - Still at Southampton, but genuinely ‘other’
Stage 2a: Community

- Encourage use
- Identify sub-communities
  - Archive managers, maintainers and users
- Support (if possible)
  - E-mail
  - Bug fixes
  - Attend conferences / workshops
  - Run workshops
- Listen
  - Change software
  - Modify priorities
Stage 2a: Frustrations

- “Users” don’t read the documentation
  - Perhaps it’s not very well written
- Users don’t ask questions on the list
  - They assume something is impossible or difficult
- Users would rather recode from scratch than adapt their system
  - Frameworks are very seductive!
- Users share bug fixes and some developments
  - But leave us as the major bottleneck
Stage 2b: Community

- Open Source Community
  - Like stage 2a
  - But we cease to be a bottleneck

- Start to give up control
  - Allow community to define requirements
  - Allow community to establish priorities
  - Maintain executive authority over core coding resource
  - Facilitate community developments

- Still trying!
  - Is this reasonable, given demands on the repository community?
Stage 3: Sustainability

- This is a preservation issue!
- JISC, University of Southampton have been supportive through stages of innovation and uptake
- Recently begun a subscription support service
  - Pays for itself + software development team + institutional overheads
  - Even so, institution has to invest *speculatively*
Reflections on Open Source as an HE funding exit strategy

Many kinds of Open Source model (OSS-Watch)
- Done because the software is a logical necessity, not a business necessity
- Contributed to equally by the community
- Not bankrolled by anyone.

A JISC open source project
- Has a **bigger agenda** (e-research, e-learning…)
- Is bankrolled (in the beginning)
  - And consequently it has a core development team
- Must **innovate** in the face of an agnostic (or antagonistic) community
- Needs to market itself to succeed
- Expected to commercialise to survive
  - Or live off the back of other funded projects?
Repository development is *mission critical* for a University

- MacKenzie Smith can argue from Library POV
- I will argue from a Researchers’ POV

Repository development more often seen as

- Faddish
- A burden to be discharged with least impact

What is required is a high level champion to make Universities change their thinking

JISC is too even-handed to help

- Role of handing out public money
- Afraid of offending industry
- It supports programmes of repository innovations
- **But not repositories**
Open Source reflections (3)

- JISC has responsibility like a good parent
  - Don’t turn your children out on the streets
  - Don’t let them live with you until they are 47
  - Help them to find their place in the world

- JISC should champion repositories as part of university core business
  - Repository managers have strategic case for funding
  - Part of that man-power can be devoted upstream to the repository software

- Result: a genuine open source platform
  - Not just a source of free but unsupported code
Conclusion

- You can’t conclude when you haven’t finished
- Why should open source community building be any quicker than repository population?
- Open source code development = open source investment
- Open source is a cop-out for funders?
  - They de-invest too early in the development cycle.
  - Worse! the funding council is trying to change the landscape, so the software has to continuously innovate and respond

- You wouldn’t do it unless you were convinced that the world has to change!
PS EPrints immediate priorities

- Get v2.4 out the door Summer 2006
  - Improved User Interface
    - Automatic data field completion
  - Metadata quality issues
    - Automatic id entry for authors, journals
  - Web services interface
- More training!
- Distance learning / video conf for Australian universities
  - Capability familiarisation
    - Understand range of capabilities for managers & techies
  - Configuration & customisation
    - Practical, hands on for techies and librarians