

A close-up photograph of vibrant green grass blades, some of which are covered with small, clear water droplets. The background is a soft, out-of-focus green, creating a sense of depth. The overall image has a fresh, natural feel.

EPrints: Repositories for Grassroots Preservation

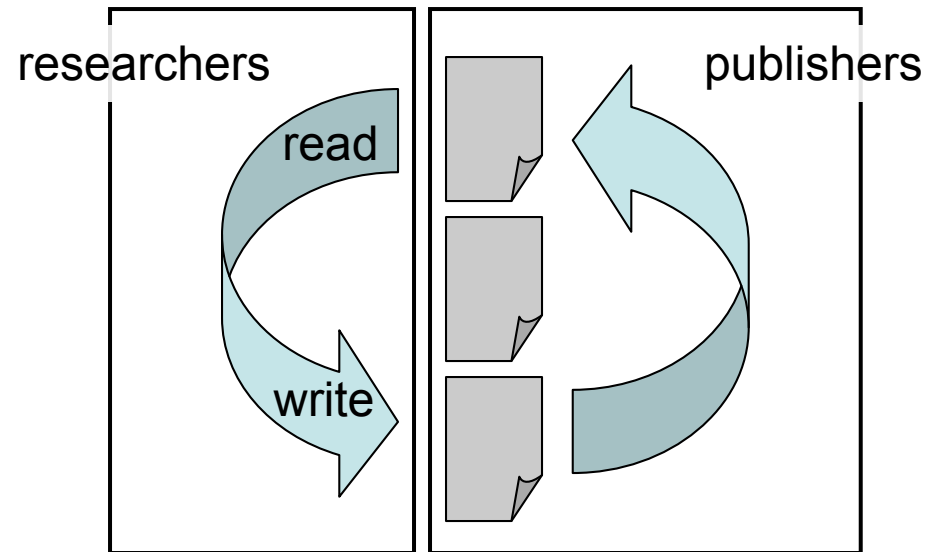
Les Carr, www.eprints.org

Grass roots: preface and précis

- The aim of this presentation is to tell a story. In the context of a meeting which has mainly dealt with the issues of national libraries and enormous digital collections, this is a presentation that addresses a different scale. It is a scale that is both smaller and larger at the same time. It is about collecting individual items from individual researchers - the so-called grass roots - through institutional repositories. Although this seems small and insignificant in comparison to tales of humongous digital collections, the day-by-day aggregated collection of individual items from a community of knowledge producers adds up to the entire scholarly and scientific literature - as well as its supporting data, experimental analyses, discussions and commentaries.
- This story about EPrints focuses on the challenges of acquiring data and documents in order to build up a global collection. The challenges listed are those relating to changing the working practices and use patterns of individuals and their host institutions in order to support long term preservation.
- The story necessarily enlarges on the need to make things easier and more useful for the author/depositor/knowledge producer in order to encourage the first stage of preservation: acquisition.

Problem Space (1)

- Universities and researchers are knowledge producers and knowledge consumers
- Scholarly communications have been outsourced
- Literally nothing to show as evidence of research activities



Problem Space (2)

- Researchers have have hard disks which are just organised enough to support daily activity
 - Disk crashes
 - Stolen laptops
 - Software upgrades that go wrong
 - Backups that never quite get restored
 - Draws and folders full of old stuff that eventually fall off the radar
- “Lost in some research assistant’s computer, the data are often irretrievable or an undecipherable string of digits”



Lost in a Sea of Science Data. S. Carlson,
The Chronicle of Higher Education (23/06/2006)



Arts & Humanities
Research Council



Science & Technology
Facilities Council



Engineering and Physical Sciences
Research Council

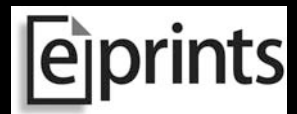
**Congratulations
on your new
research project!**



This is
where your
hardware
will end up

Make sure your data doesn't!

Research outputs go in research repositories

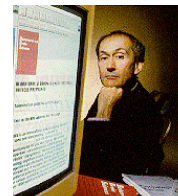


UK Experience

- UK Council of Research Repositories
 - platform agnostic
 - group of repository managers
 - speaks for repository managers
- Most repositories
 - have a part time manager
 - receive little or no technical support

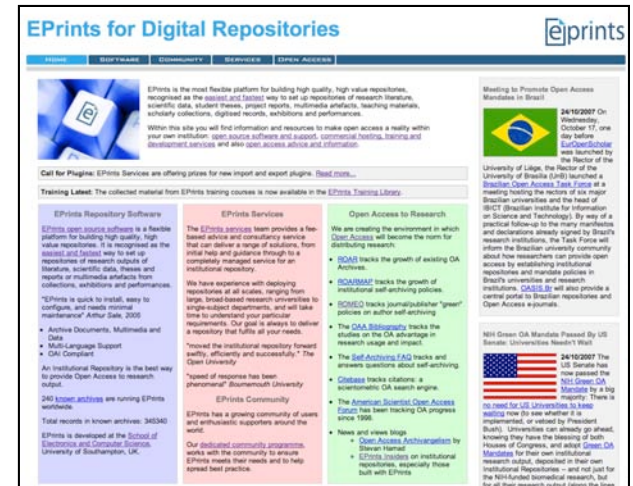
EPrints History

- Open Archiving Initiative - October 1999
 - Originally called UPS
- Among the Participants
 - Paul Ginsparg (Los Alamos, arXiv)
 - Carl Lagoze (Cornell, NCSTRL)
 - Stevan Harnad (Southampton, Cogprints)
- EPrints
 - proposed as a ‘build your own repository’ solution
 - enable institutions and groups to participate in OAI metadata sharing initiative



EPrints History

- First released April 2000
 - to co-incide with OAI-PMH
- Version 3.0 released in Jan 2007
 - at Open Repositories 2007
- Strongly backs *Open Access*
- Used by over 240 registered repositories



The
LINNEAN
SOCIETY
of London

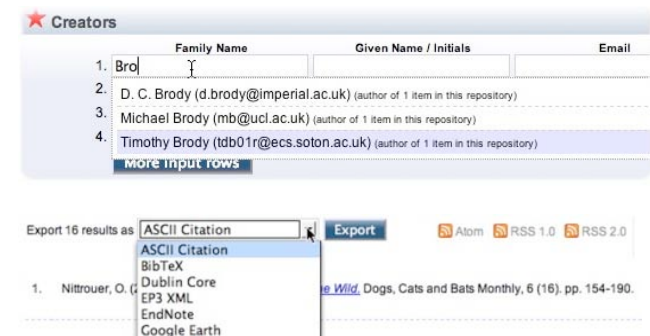


EPrints Management

- Open source (GNU license)
- EPrints development model is more centralised than DSpace / Fedora
 - *c.f.* the original problem statement
 - pros and cons *e.g.* faster turnaround on development cycles, more focused, easier quality management
 - All of these platforms are hybrid open source - they were initially bankrolled!
- EPrints Commercial Services
 - repository hosting, bespoke development & training
 - sustain the development team

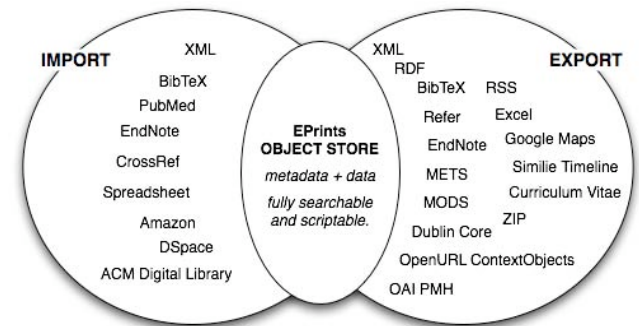
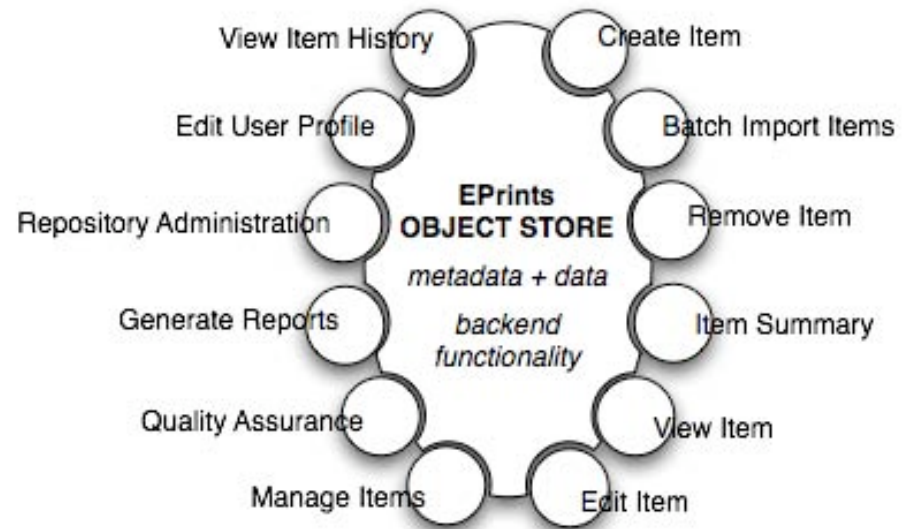
EPrints Core Objectives

- Lower the barrier for depositors while improving metadata quality and ultimate collection value
 - Time saving deposits
 - Import data from other repositories and services
 - Autocomplete-as-you-type for fast data entry
 - Name authorities
- Enter once, reuse often
 - Works with bibliography managers, desktop applications and new Web 2.0 mashups
 - RSS feeds and email alerts keep you up to date
 - Easily integrate reports, bibliographic listings, author CVs and RSS feeds into your corporate web presence
 - Used for corporate reporting and national Research Assessment
- Simple platform for open source contributions
 - Tightly-managed, quality-controlled code framework
 - Flexible plugin architecture for developing extensions

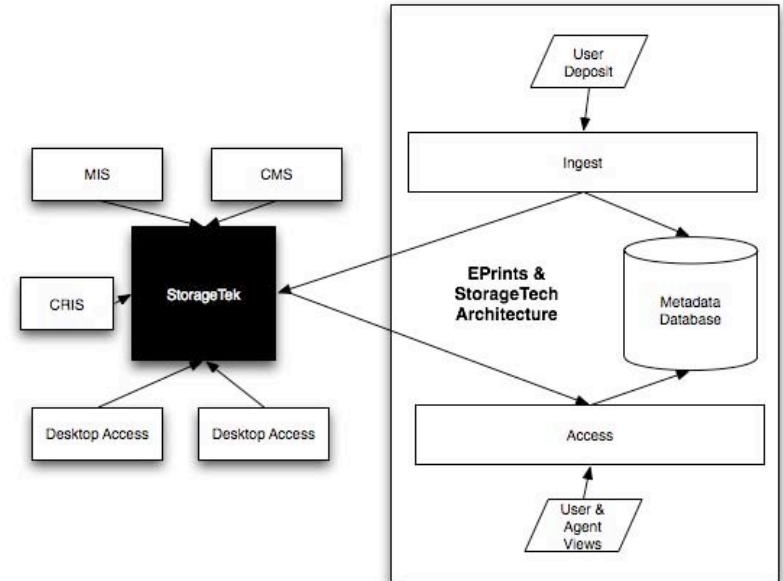
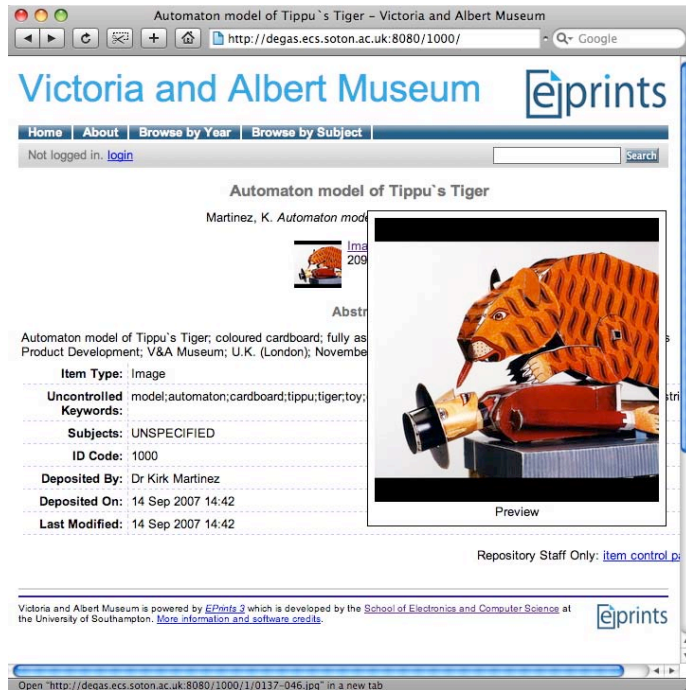


EPrints Flexibility

- EPrints backend
 - object store
 - API
- EPrints frontend
 - Screen plugins
 - User interface + methods + REST interface



EPrints + Honeycomb



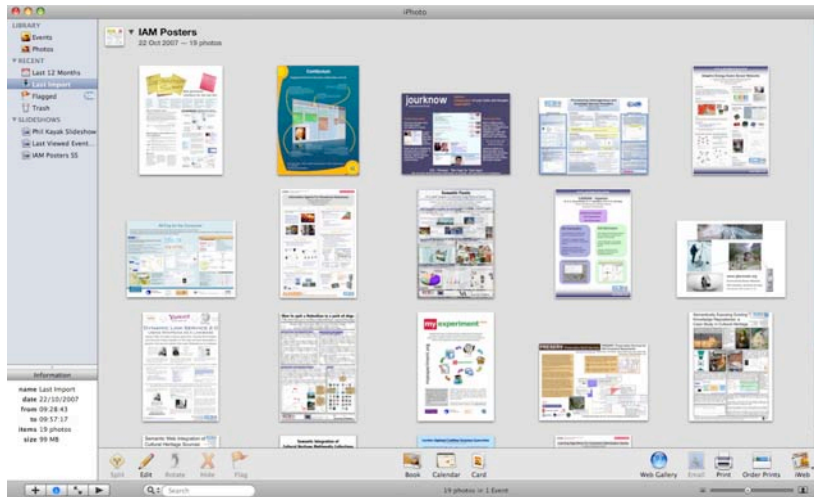
- **Jam today** - large self-managing storage extends repository bang for library buck
 - New chemistry & artistic objects to be collected
- **Jam tomorrow** - potentially take over part of repository responsibility

EPrints Challenges

- Small science > big science
 - Data from Big Science is easier to handle, understand and archive. Small Science is horribly heterogeneous and far more vast. In time Small Science will generate 2-3 times more data than Big Science.
- Lots of inexperienced users
 - Give individuals the tools to become responsible curators of their own intellectual output
 - Give institutions the tools to manage, assist and leverage
 - Give users the tools to access the global literature data - to use and reuse for many, many purposes in many, many contexts by many stakeholders

(Tim O'Reilly)

It's the Data, Stupid



ECS
Electronics & Computer Science
University of Southampton


Site Search Enter keywords here

Home | UG Study | MSc Admissions | PG Opportunities | Research | Business | People | Alumni | Contact | Intranet

University of Southampton > ECS > People > Leslie Carr

Leslie Carr
School of Electronics and Computer Science
University of Southampton
Southampton
SO17 1BJ
United Kingdom

Position: Academic staff in **Intelligence, Agents, Multimedia Group**
Extension: 24479 ([edit](#))
Telephone: +44 (0)23 8059 4479
Fax: 023 80592865
Homepage: <http://www.ecs.soton.ac.uk/~lac/>
Email: lac@ecs.soton.ac.uk
Vcard: <http://vcard.ecs.soton.ac.uk/person/60>




The group secretary for the Intelligence, Agents, Multimedia Group is **Jane E Morgan**.

facebook Profile edit Friends Networks Inbox home account privacy logout

Search

Applications Photos Groups Events Marketplace more

**My Posted Items** Posted Items

Displaying the only posted item.

You posted a link at 9:42am Jul 23

[Investigating the Use of Replica Morpho Butterfly Scales for Colour Displays - Portfolio](http://portfolio.ecs.soton.ac.uk/)
<http://portfolio.ecs.soton.ac.uk/>

The reflective properties and bright colours of butterfly wings are of great interest in many applications. The brilliant blue of the Morpho genus is due to multilayer interference effects originating from tiny lamellae on ridges present on the butterfly scales. ...

I've been experimenting with ways of publishing our students' work at ECS, and we have set up a repository for them to use. here's an example of a great MEng project. >>

Add a comment | No comments

Post a link

<http://>

Post

Share on Facebook

Drag the button above to your Bookmarks Bar to quickly share content with your friends.

For more details, click here.

Subscribe to Posted Items

My Posted Items
Subscription Help >

Posted Items Help

Explore how Share works >
Add Share to your site >



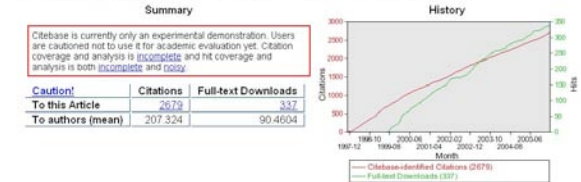
EPrints - beyond the repository

- OAI PMH services
 - Citebase - citation analysis for the Open Access literature.
 - Unfunded PhD work (outshoot of OpCit)
 - 4 million sessions per month
 - Destroy 1 RAID disk every 6 months
 - Celestial - OAI-PMH harvesting proxy
 - Supports Citebase and other services
 - ROAR - registry of Open Access repositories
 - Tracks size and daily deposit profiles over time

This Article's Citation/Hits History (explain?)

[Top](#) [Reference List](#) [Cited By](#) [Co-Cited](#) [All Citations](#) [All Co-Citations](#)

Use the [Correlation Generator](#) to explore the correlation between download impact ("hits") and citation impact.

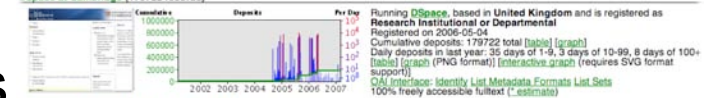


This Record's Reference List (explain?)

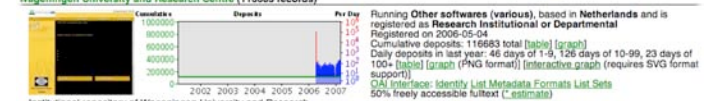
[Top](#) [Reference List](#) [Cited By](#) [Co-Cited](#) [All Citations](#) [All Co-Citations](#)

G/A [1] G. 't Hooft, "A Planar Diagram Theory For Strong Interactions," Nucl. Phys. B72 (1974) 461.
eprint [2] A. M. Polyakov, "Strong Theory And Quark Confinement," hep-th/9711002.
G/A [3] G. Gibbons, Nucl. Phys. B207 (1982) 337.
eprint [4] R. Kallosh and A. Peet, Phys. Rev. D46 (1992) 5223, hep-th/9209116.
eprint [5] S. Ferrara, G. Gibbons, R. Kallosh, Nucl. Phys. B500 (1997) 75, hep-th/9702103.
G/A [6] G. Gibbons and P. Townsend, "Vacuum Interpolation in Superstrings," Phys. Rev. D 51 (1995) 6232.

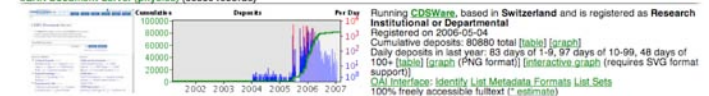
DSpace at Cambridge (179722 records)



Wageningen University and Research Centre (116663 records)



CERN Document Server (physics) (80880 records)



EPrints - Preservation Services

- Format profiling using PRONOM-DROID
 - JISC PRESERV project
 - Initially to be applied to two pilot repositories
 - Ultimately applied to over 200 repositories
 - DSpace & EPrints
 - Applied via OAI
 - Delivered through ROAR
- Add Honeycomb to the mix
 - We can ‘preserve’ repository contents too
 - JISC PRESERV II project

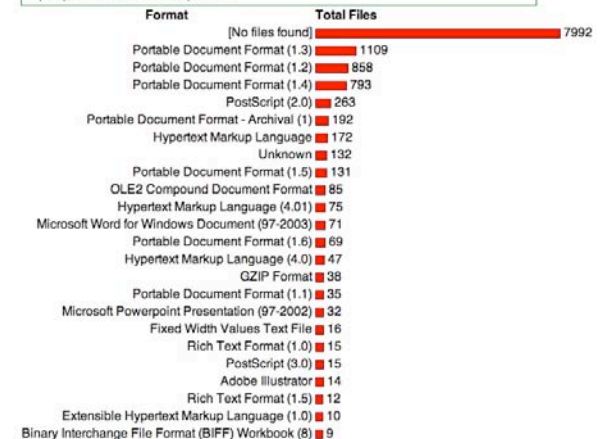
Summary Preserv Profile

For more information about the Preserv project and repository content profiling see <http://preserv.eprints.org/>. Pronom's file format identification is currently under development, which may result in some strange format identifications.

The following chart shows the breakdown of all files contained in the repository(ies), based on an automated crawler. All files (apart from the abstract 'jump-off' page) were downloaded and then identified using the Pronom file format identification tool (Droid). Does not include files that are not hosted on the same server as the OAI interface or files that are over 2 megabytes in size.

You can subscribe to alerts for this profile by going to [Celestial's repository page](#).

OAI Repositories Included (1)
<http://reprints.ecs.soton.ac.uk/pert/oai2>



The background of the slide is a close-up photograph of green grass blades. The blades are long and thin, with many small, clear water droplets resting on their surfaces. The lighting is bright, creating a vibrant green color and highlighting the texture of the grass and the glistening of the water.

**The challenges of human scale
institutional repositories versus
the challenges of industrial-scale
processing of humongous
collections.**

**Lawnmowers vs Combine
Harvesters?**

**How do you manage an entire
nation's grass clippings?**