



# Open Access: The Web and Scientific Knowledge

COMP6218

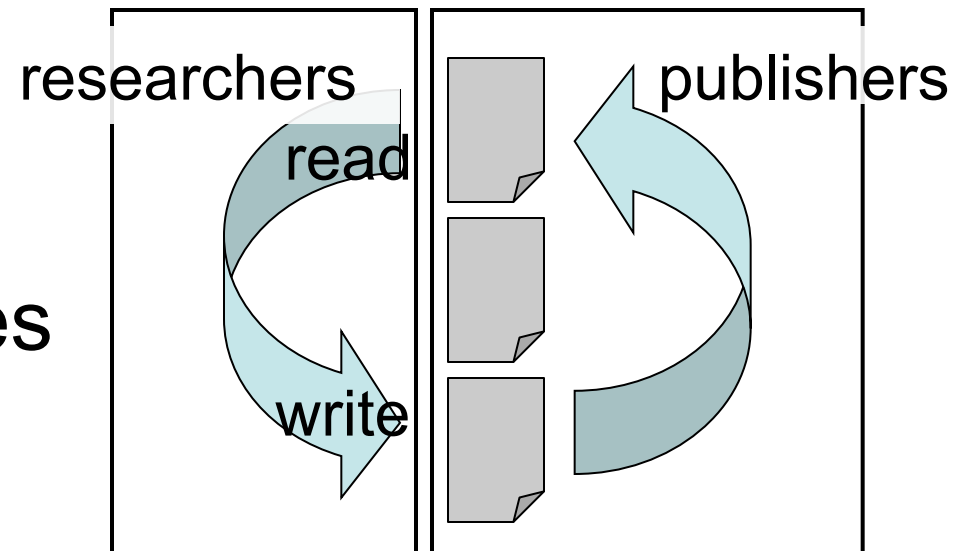
# Open Access (OA)

- Tim Berners-Lee developed the Web as a solution for researchers to share information between themselves
- OA is a current movement for organising and disseminating the world's research knowledge through Web technology



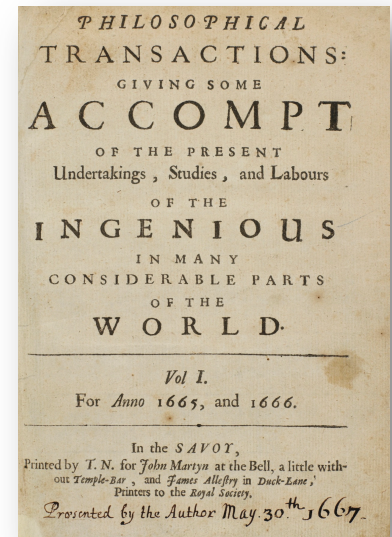
# Open Access: the Problem

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



# Scholarly Publishing 1665-1960

- Scientific and scholarly societies publish their own journals
  - first scientific journal March 1665  
The Royal Society publishes “Philosophical Transactions”
- University Presses
  - Academic and non-profit
- For the benefit of their members, for the benefit of science



# Private Sector 1960-



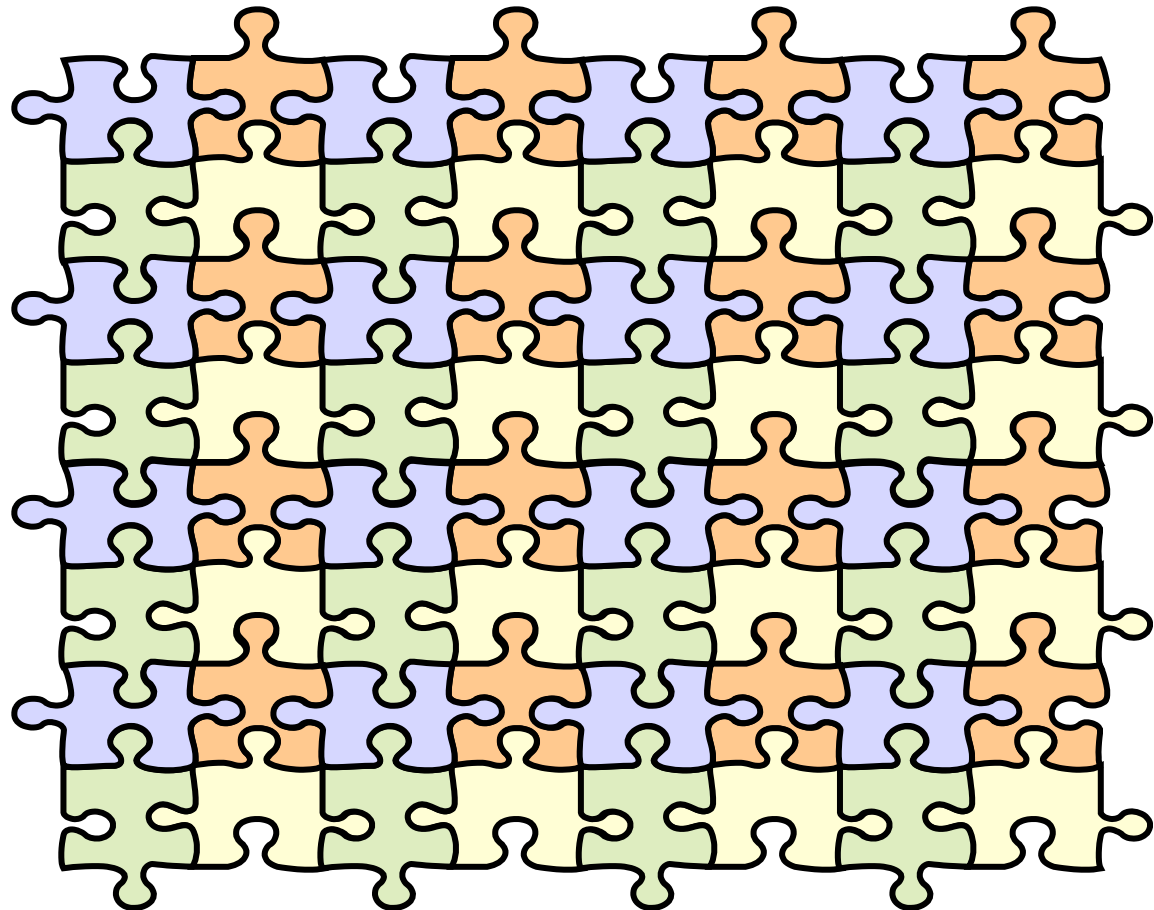
- UK Robbins Report / expansion of higher education & science **budget**
- After the war **Robert Maxwell** decided to publish scientific journals and set up Pergamon Press which was quickly and hugely profitable.
  - bringing “aggressive publishing in science”
- The New Demand made for a very profitable system - with an increasing number of commercial publishers moving into STM.

# Publishing Conditions

- Author writes article
  - Author gives article to publishing company
  - Author signs over copyright to publishing company
  - Publishing company publishes article
- Publishing company receives revenue
- Author receives academic credit
  - Publication glory
  - Citations

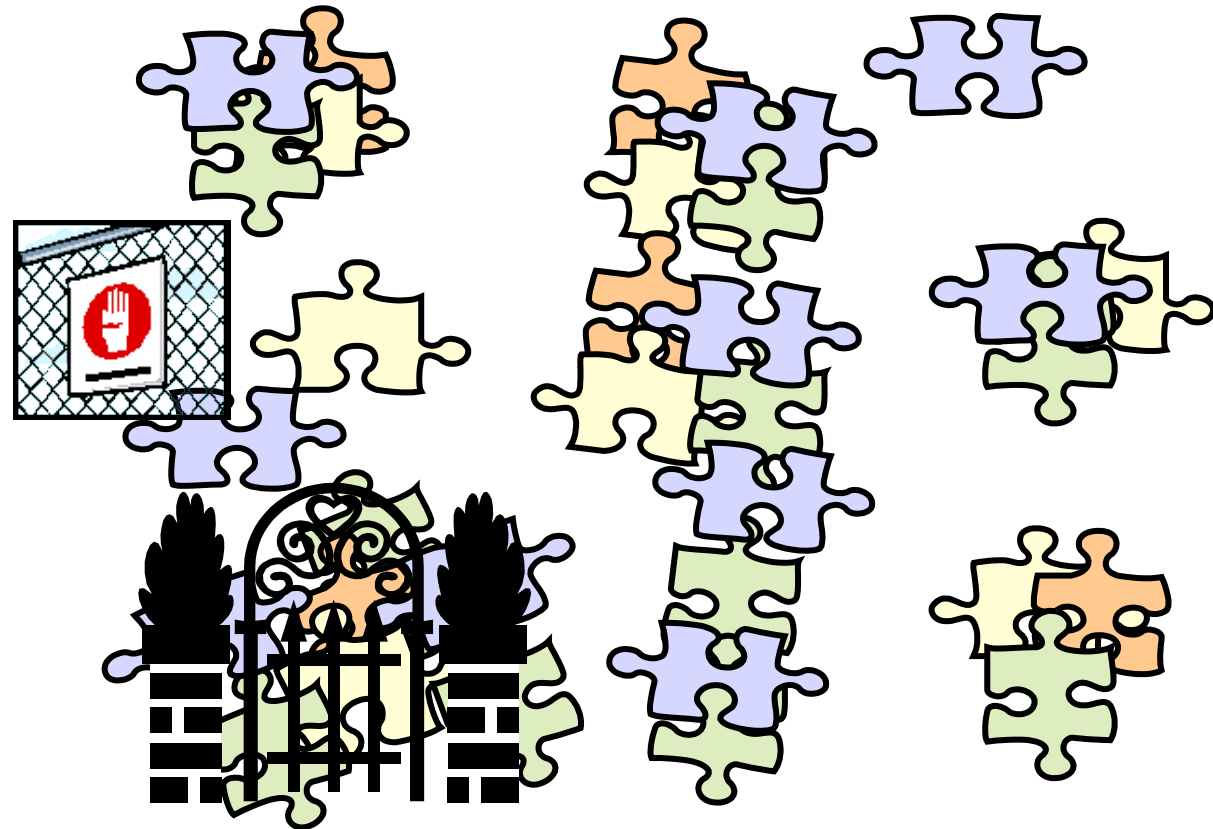
# The Literature: As We Imagine

- Integrated
- Available



# The Literature: As It Is

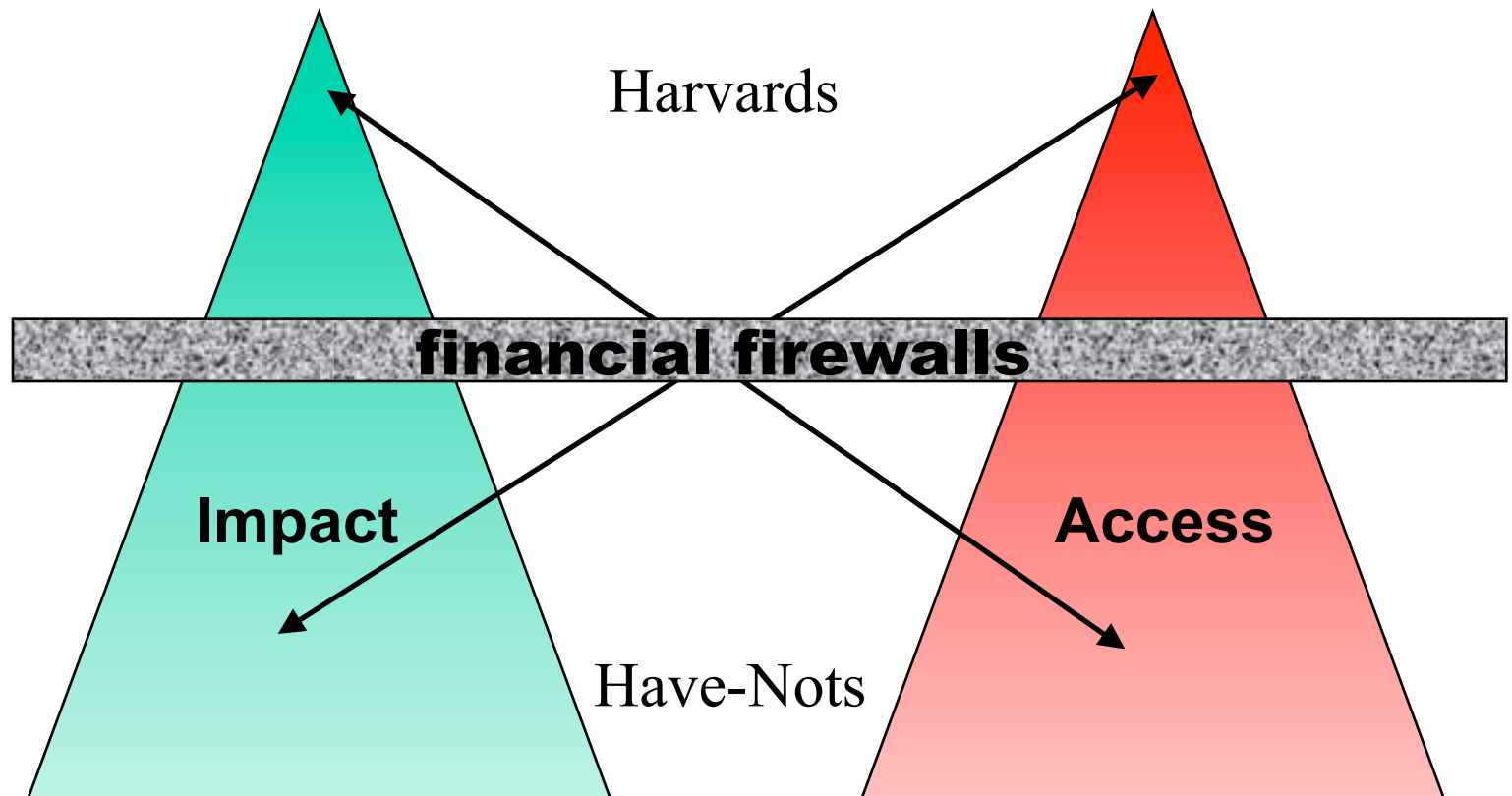
- Inaccessible
- Disjoint





# The Twin Peaks Problem

- 24,000 journals with 2,500,000 articles/yr



# The Budapest Open Access Initiative

- Old tradition of scholarly publishing  
+  
New technology of the Internet  
=  
• Public good: free and unrestricted access to  
peer-reviewed journal literature



Budapest, December 2001

# Open Access Strategies

- **Green: Self-Archiving**

- Journal processes continue as normal
- Authors deposit a copy of their papers into an ‘**open access repository**’
- Public copy is a supplement to the publishers official article for those who can't afford a subscription
- Also an institutional record of its work for sharing, reuse, marketing *etc*

- **Gold: Publishing**

- Journal changes business model
- Readers no longer pay to read
- Instead, authors pay to publish
  - or their funders

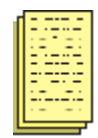
# Limited Access: Limited Research Impact

12-18 Months

**Impact cycle begins:**  
Research is done



Researchers write pre-refereeing "Pre-Print"



Submitted to Journal

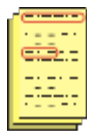


Pre-Print reviewed by Peer Experts – "Peer-Review"



Pre-Print revised by article's Authors

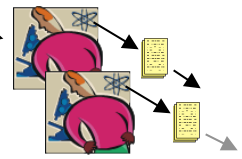
Refereed "Post-Print"  
Accepted, Certified, Published by Journal



Researchers can access the Post-Print if their university has a subscription to the Journal



**New impact cycles:**  
New research builds on existing research



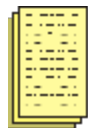
# Maximized Research Access and Impact Through Self-Archiving

12-18 Months

**Impact cycle begins:**



Researchers write pre-refereeing "Pre-Print"



**Pre-Print is self-archived in University's Eprint Archive**



Submitted to Journal



Pre-Print reviewed by Peer Experts – "Peer-Review"



Pre-Print revised by article's Authors

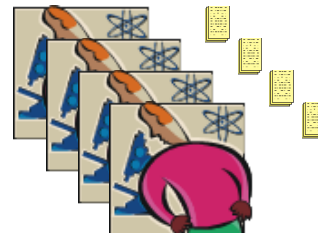
Refereed "Post-Print" Accepted, Certified, Published by Journal



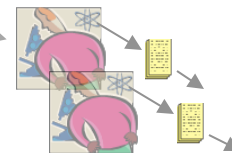
Researchers can access the Post-Print if their university has a subscription to the Journal



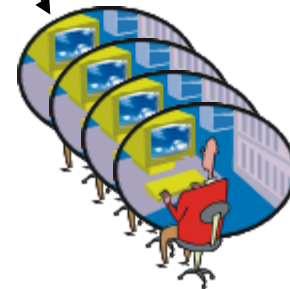
**New impact cycles:**  
Self-archived research impact is greater (and faster) because access is maximized (and accelerated)



**New impact cycles:**  
New research builds on existing research

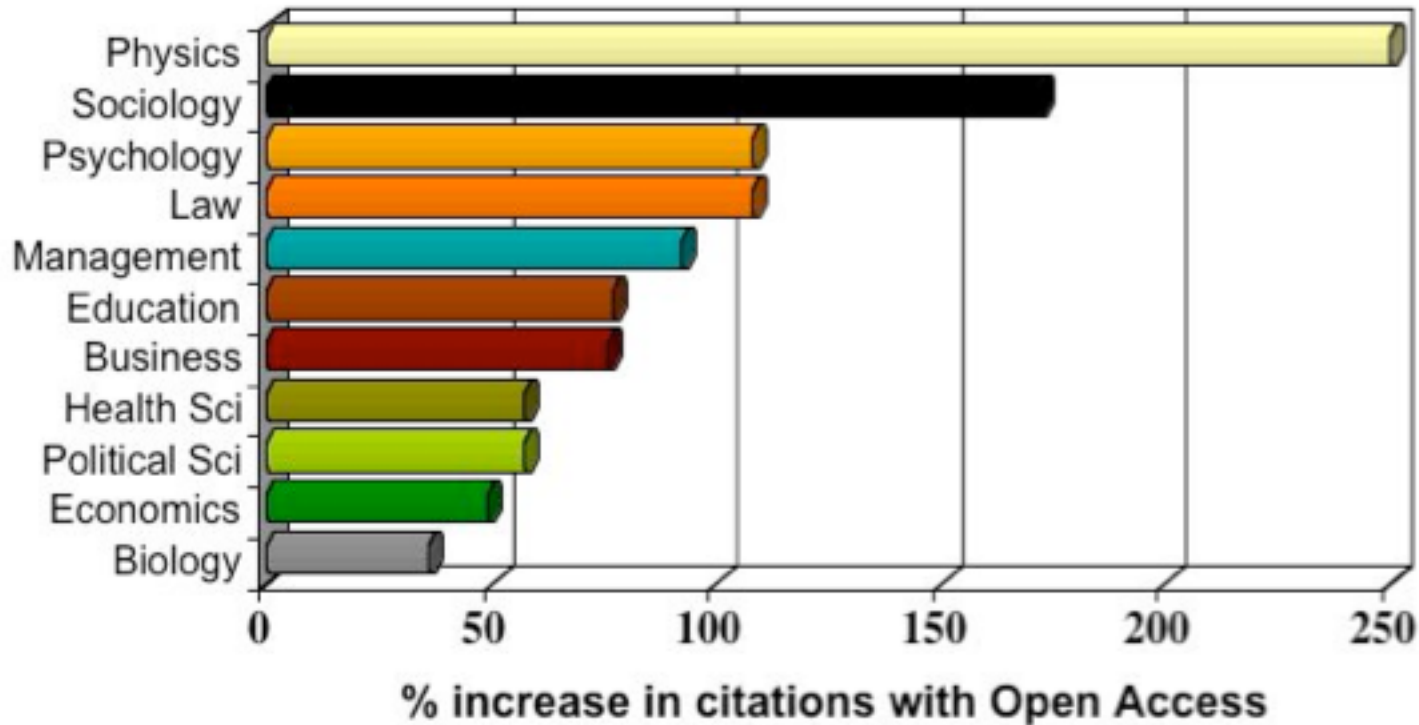


**GREEN  
Open  
Access**



# Open Access Advantage

- OA increases citations



- Full bibliography, see <http://opcit.eprints.org/oacitation-biblio.html>

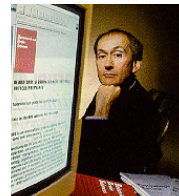
# Contributors to the OA Advantage

$$EA + QA + UA + (CA) + (QB)$$

- **EA: Early Advantage:** Self-archiving preprints before publication hastens and increases usage and citations (higher-quality articles benefit more: top 20% of articles receive 80% of citations)
- **QA: Quality Advantage:** Self-archiving postprints immediately upon publication hastens and increases usage and citations (higher-quality articles benefit more)
- **UA: Usage Advantage:** Self-archiving increases downloads (higher-quality articles benefit more)
- **(CA: Competitive Advantage):** OA/non-OA advantage (CA disappears at 100%OA, *but very important today!*)
- **(QB: Quality Bias):** Higher-quality articles are self-selectively self-archived more (QB disappears at 100%OA)

# Repositories & Green OA

- Open Archiving Initiative - October 1999
  - Agreed OAI-PMH for metadata sharing
  - (2008 OAI-ORE for data exchange)
- Among the Participants
  - Paul Ginsparg (arXiv)
  - Carl Lagoze (NCSTRL)
  - Stevan Harnad (Cogprints)
- EPrints
  - proposed as a ‘build your own repository’ solution
  - enable institutions and groups to participate in OAI metadata sharing initiative





# Example Repository

UNIVERSITY OF Southampton  
School of Electronics and Computer Science

Site Search

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

University of Southampton > ECS > Research > Publications [Login](#)

**ECS Research**

- Research Groups
- Researcher Profiles
- Research Themes
- Graduate School
- Seminars
- Research Facilities
- Centres and Institutes
- Latest Technologies

**Publications**

- Publications Home
- This weeks new additions
- Search
- Browse by research group
- Browse by year

**Intranet Tools**


- Add publications
- Modify RAE Information

[RSS 1.0 Feed](#)  
[RSS 2.0 Feed](#)  
[Atom Feed](#)

## The Semantic Web Revisited

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) *The Semantic Web Revisited*. IEEE Intelligent Systems, 21 (3). pp. 96-101. ISSN 1541-1672

**Download**

 PDF - Requires a PDF viewer such as [GSview](#), [Xpdf](#) or [Adobe Acrobat Reader](#)  
177Kb

**Abstract**

The original Scientific American article on the Semantic Web appeared in 2001. It described the evolution of a Web that consisted largely of documents for humans to read to one that included data and information for computers to manipulate. The Semantic Web is a Web of actionable information--information derived from data through a semantic theory for interpreting the symbols. This simple idea, however, remains largely unrealized. Shopbots and auction bots abound on the Web, but these are essentially handcrafted for particular tasks; they have little ability to interact with heterogeneous data and information types. Because we haven't yet delivered large-scale, agent-based mediation, some commentators argue that the Semantic Web has failed to deliver. We argue that agents can only flourish when standards are well established and that the Web standards for expressing shared meaning have progressed steadily over the past five years. Furthermore, we see the use of ontologies in the e-science community presaging ultimate success for the Semantic Web--just as the use of HTTP within the CERN particle physics community led to the revolutionary success of the original Web. This article is part of a special issue on the Future of AI.

**Creators:** Nigel Shadbolt, Tim Berners-Lee, Wendy Hall

**Item Type:** Article

**Research Group:** Intelligence, Agents, Multimedia

**Deposited On:** 22 May 2006 by Miles-Board, Timothy

**Alternative Locations:** [http://doi.ieeecomputersociety.org/10.1109/9781422204272\\_011](http://doi.ieeecomputersociety.org/10.1109/9781422204272_011)

**ISSN:** 1541-1672

**ID Code:** 12614

**Last Modified:** 16 Nov 2007 04:27

**Performance Indicator:** EZ~03~03~11

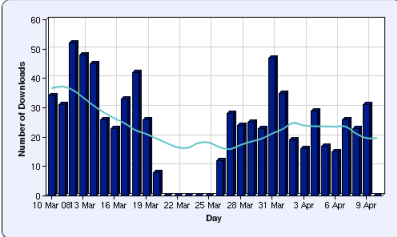
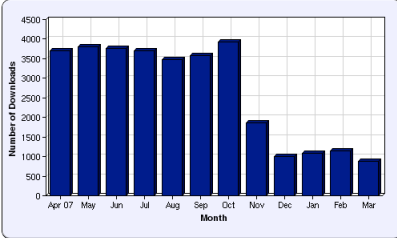
**Tools**

- [Look up in Google Scholar](#)

**Metadata**

HTML Citation

**Download Statistics**

<http://eprints.ecs.soton.ac.uk/>

A repository for a school of Electronics and Computer Science.

It achieves 80-100% full text self-deposit

*(merged with University of Southampton repository in 2011)*

# Fast Forward to Open Access

- The Optimal and Inevitable for Researchers.
  - The entire full-text refereed corpus online
  - On every researcher's desktop, everywhere
  - 24 hours a day
  - All papers citation-interlinked
  - Fully searchable, navigable, retrievable
  - For free, for all, forever

*Stevan Harnad, Les Carr*

*OpCit International DLI Project Proposal (1999)*

# Problems with Green OA

The screenshot shows the ECS EPrints Repository website. The header includes the ECS logo and navigation links. The main content area features a search bar, a welcome message from the Deputy Head of School, and a list of recently added publications. The left sidebar contains navigation menus for Research, Publications, and Intranet Tools.

- ECS repository, 11,000 records, 4,000 full text, 80-100% open access to our research output.

The screenshot shows the Nottingham ePrints website. The header includes the University of Nottingham logo and the text 'Nottingham ePrints'. The main content area features a search bar, a login link, and a description of the repository. The left sidebar contains navigation menus for ePrints Home, Deposit Guide, Browse, Search, and Open Access Support.

- Average repository, 300 items, 200 full text, negligible research output
- Recent NIH request for OA achieved 4% compliance

# Problems with Green OA

- Copyright assignment to publishing company means the author no longer has the right to control the paper
  - Relies on “publisher usage policy”
  - May not allow submission to a repository
  - May not allow copies to be made for teaching, research assessment or career advancement processes
- Romeo service tracks publisher policies
  - <http://www.sherpa.ac.uk/romeo/>

# Problems with Gold OA



- Relies on publishers changing their business model
- Scientific publishing is very lucrative (18% profits)
- Gold publishers making slow advances.

# Role of the Repository

- Who takes responsibility for curating the research knowledge of the world?
  - Privately owned, profit making, publishing industry
  - University, research institution, knowledge creator
- The Institutional repository is a place where the members of an institution can curate their intellectual outputs / knowledge capital
  - Share
  - Use
  - Reuse
- **Web advantage – ubiquitous knowledge**