Open Repositories

Tomasz Neugebauer

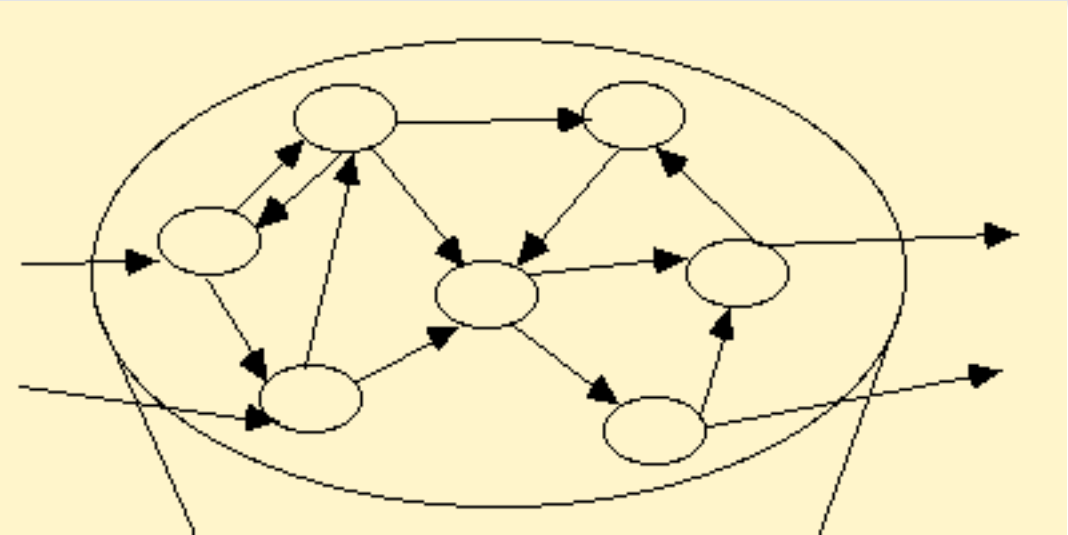
Librarian's Forum Brown Bag– Concordia University Libraries
December 15th, 2008

Open Repositories

- General Systems Theory and Open systems
  - User-centered design
- Open Source
  - History
  - Systems for Digital Libraries
- Open Standards
  - Protocols, Metadata & Interoperability
  - Digital Library Services
- Open Access

System Theory – What is a system?

A system is a complex object every part or component of which is connected with other parts of the same object in such a manner that the whole possesses some features that its components lack – that is emergent properties. (Bunge 1996:20)

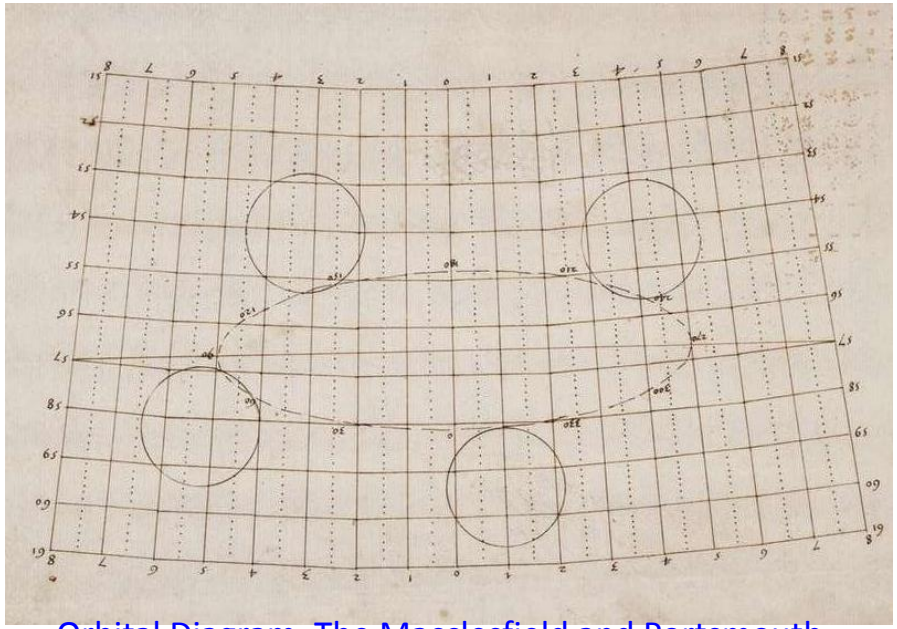


- composition
- environment
- structure

Systems Theory – Closed Systems

- *Closed systems* do not interact with the environment
- Example: assumption that all masses, particles, and forces that affect the system are included in the model.

2nd law of thermodynamics
Any system **which is free of external influences** becomes more disordered with time.



Orbital Diagram. The Macclesfield and Portsmouth collections of Newton manuscripts
http://www.dspace.cam.ac.uk/handle/1810/194207

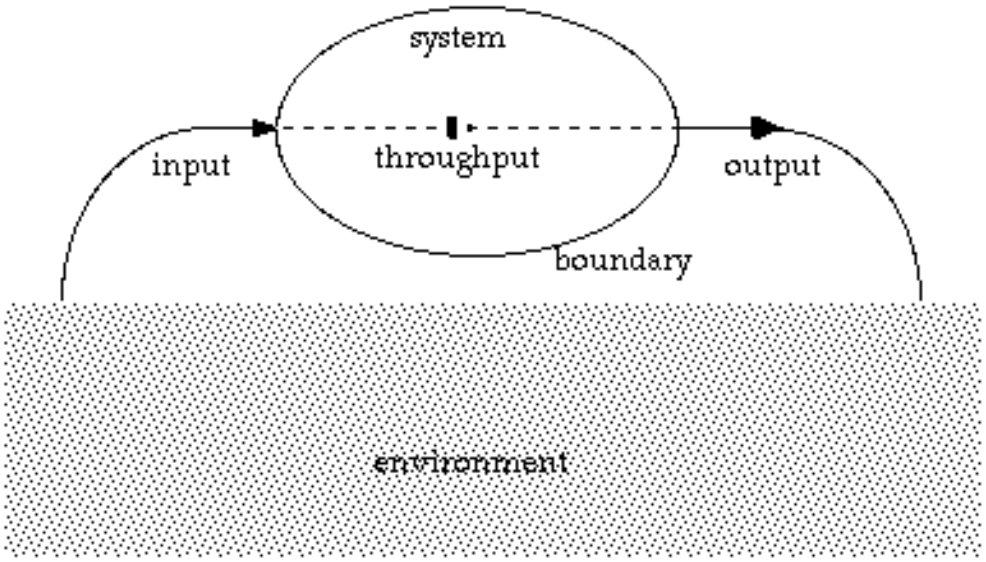
General Systems Theory and Open systems

Open Systems

- **Open system** is defined as a system in exchange of matter with its environment, presenting import and export, building up and breaking-down of its material components" (von Bertalanffy, p.141)
- **Ludwig von Bertalanffy** -
  - formulated the **organismic system theory**
    - modelled the organism as an open system striving towards a steady state
    - two biological principles, namely, the maintenance of the organism in the non-equilibrium, and the hierarchic organization of a systemic structure

Sabine Brauckmann. International Society for Systems Sciences. (1999)
http://www.isss.org/lumLVB.htm

Systems Theory – Open Systems



$S_{t+1}=F(S_t, I_t)$

(Weinberg, p.204)

– Figure. System interacting with environment – F.Heylighen

- **Open Systems** compensate for its deterioration by importing and processing energy/information.

Open Repositories

- Environment (Input/Output): Faculty, researchers, authors, students, online community, other repositories, world wide web.
- Composition: research data / articles / papers / multimedia, etc., metadata (about research data and authors), software, source code, usage statistics
- Structure: metadata standards (OAI-PMH, DublinCore, XML, etc.), editorial policies, harvesting policies, preservation policies, dissemination policies.

User-Centered Design

- ISO 13407 - **Human Centered Design Process for Interactive Systems**
- The standard describes four principles of human-centered design:
  - Active involvement of users
  - Appropriate allocation of function (making sure human skill is used properly).
  - Iteration of design solutions (therefore allow time in projec planning).
  - Multi-disciplinary design

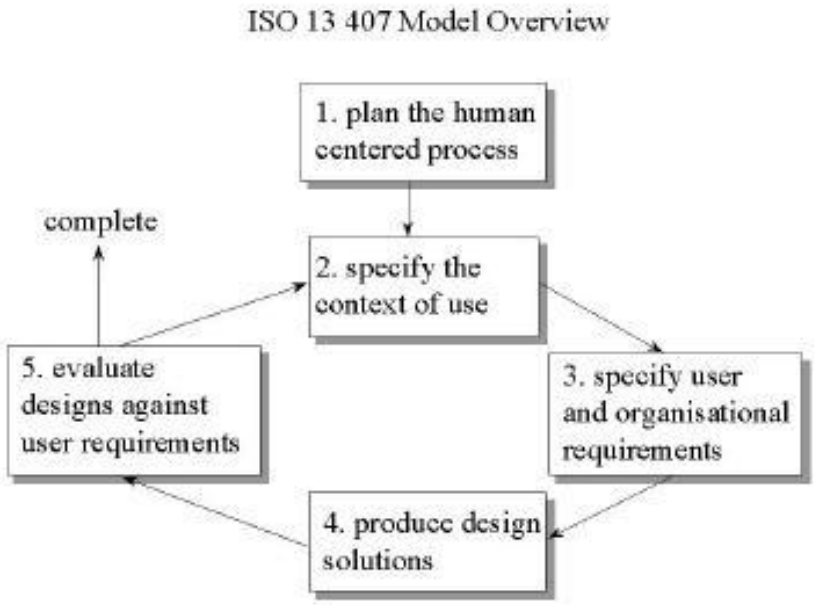


Image: **What is EMMUS?** EMMUS (European MultiMedia Usability Services) (1999) http://www.ucc.ie/hfrg/emmus/methods/iso.html retrieved 15-12-2008

Open Source - History

- Software development has a tradition of sharing and cooperation
- **1960s-70s** –
  - key features of operating systems and Internet developed in academic settings (Berkley, MIT, Bell Labs, Xerox Palo Alto Research Center)
  - sharing of source code by programmers in different organizations was commonplace on an informal basis (little to no effort to delineate property rights or restrict use)

Open Source - History

- **1980s** – AT&T attempts to enforce intellectual property rights to UNIX (many academics and non AT&T researchers made contributions to it)
  - In response, formalized ground rules for cooperative software development process emerged
  - **Free Software Foundation** is formed by Richard Stallman at MIT in 1983
  - **General Public License** (free to use, distribute and modify) introduced for **GNU**; 'copyleft' seeks to keep intellectual property free and available
  - 1989 – Tim Berners-Lee, researcher at CERN, writes a proposal *HyperText and CERN* – includes first definition of **HTML and HTTP.** First web server (1990)

Open Source - History

- **1990s** – acceleration of open source activity
  - 1991 –
    - Linus Torvalds develops Linux
    - First HTTP server released to public (CERN)
    - Rob McCool (NCSA, University of Illinois, Urbana-Champaign) releases more powerful version of HTTP server
  - 1994 – McCool left NCSA
    - many webmasters continue developing HTTP server – small group plan methods for coordinating changes – **Apache Group** is formed.
  - 1995 – first public release of Apache Server – quickly becomes the most popular HTTP server on the Internet
  - 1999 - Apache Group becomes **Apache Software Foundation**

GNU

GNOME

fedora

ubuntu

1997 – 100,000 name servers listed In whois database

debian

Open Source

- History

- **Apache HTTP server (about page)**

  “

  – […] We believe that the tools of online publishing should be in the hands of everyone, and that software companies should make their money by providing value-added services such as specialized modules and support, amongst other things.

  […]

  To the extent that the protocols of the World Wide Web remain "unowned" by a single company, the Web will remain a level playing field for companies large and small. Thus, "ownership" of the protocols must be prevented. To this end, the existence of robust reference implementations of various protocols and application programming interfaces, available free to all companies and individuals, is a tremendously good thing.”

Open Source and Libraries

- Libraries have a long history of dealing with licensed content
- Natural synergy with open source movement – library collections, staff and physical structures usually are available to a wide community of users on a nonprofit, publicly funded basis
- Libraries are frequent users of open source software – though may be unaware of this
  - Internet infrastructure is built on OSS applications but this is ubiquitous and often taken for granted (examples: BIND software that runs DNS server, Apache, Linux, Mozilla)

Open Source and Libraries

- "The key term to connect digital libraries and OSS is *open*. In fact, digital libraries are sometimes referred to as open digital libraries, and "open" models, such as open archives, have emerged at every level of intellectual property sharing. Libraries have derived much support from the premise that an open society is dependent on the free exchange of ideas and opinions, as well as the existence of an informed citizenry**. Digital libraries and open source software are a natural outgrowth of the open models of exchange that help societies grow and prosper."**

  – Art Rhyno, 2004

Open Source

- Open Source Initiative – Definition

1. **Free Redistribution** – can sell or give away software
2. **Source Code** – must be included allowed to be redistributed
3. **Derived Works** – allows modifications to be release with same terms
4. **Integrity of The Author's Source Code** – restrictions on redistribution of modified source code only if license allows distribution of patch files with source code
5. **No Discrimination Against Persons or Groups**
6. **No Discrimination Against Fields of Endeavor** – example: no restrictions against use in genetic research or business
7. **Distribution of License**
8. **License Must Not Be Specific to a Product** – cannot insist on particular software package
9. **License Must Not Restrict Other Software** – example, no insistance that all other programs are open source software
10. **License Must Be Technology-Neutral** – cannot insist on particular technology or style or interface

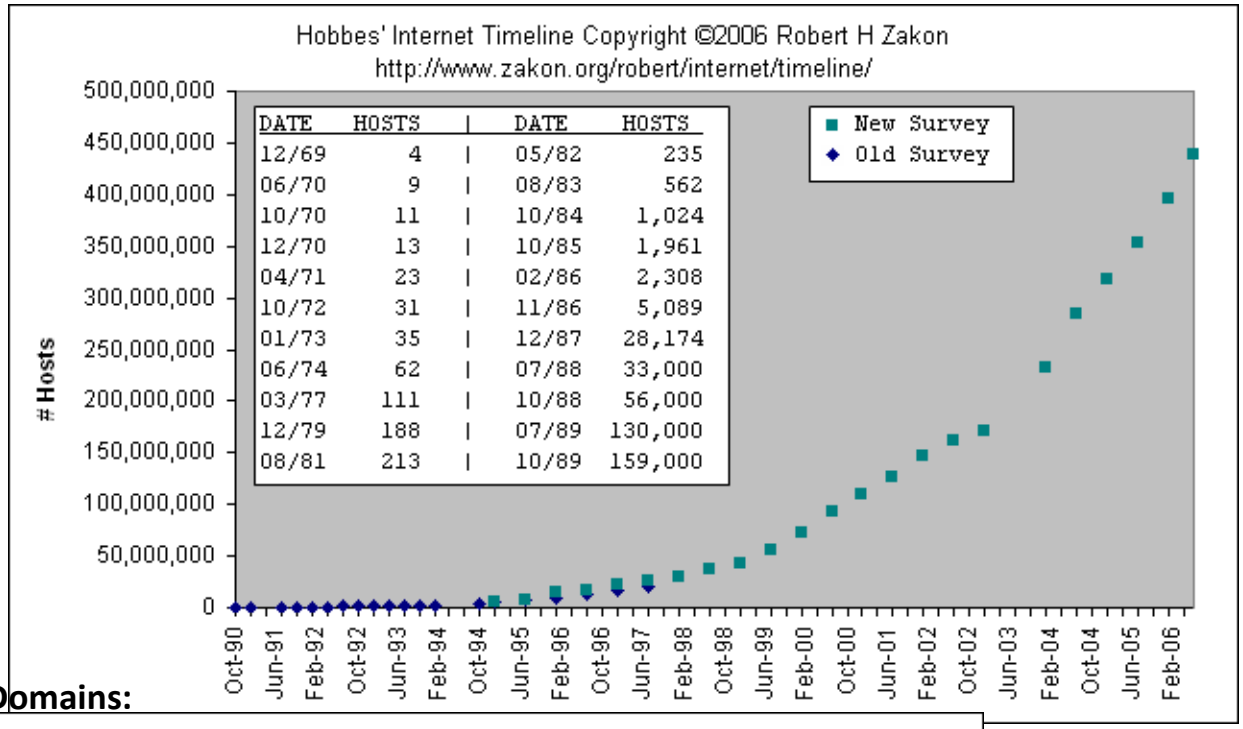
Open Source – “Freedom”

Definition

- **Free software** is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. More precisely, it refers to four kinds of freedom, for the users of the software:
  - The freedom to run the program, for any purpose (freedom 0).
  - The freedom to study how the program works, and adapt it to your needs (freedom 1). Access to the source code is a precondition for this.
  - The freedom to redistribute copies so you can help your neighbor (freedom 2).
  - The freedom to improve the program, and release your improvements to the public, so that the whole community benefits (freedom 3). Access to the source code is a precondition for this.

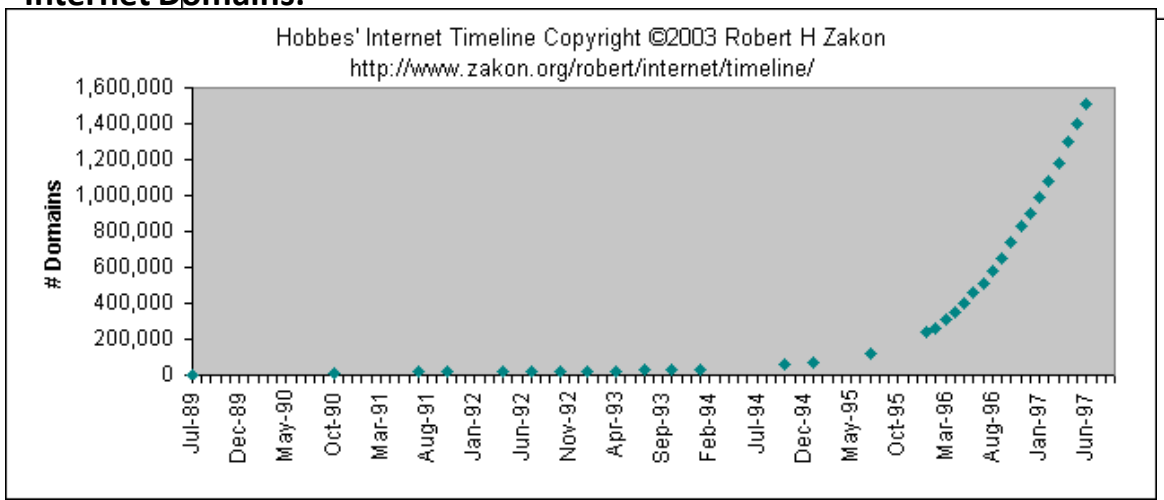
Internet Growth

Internet Hosts:

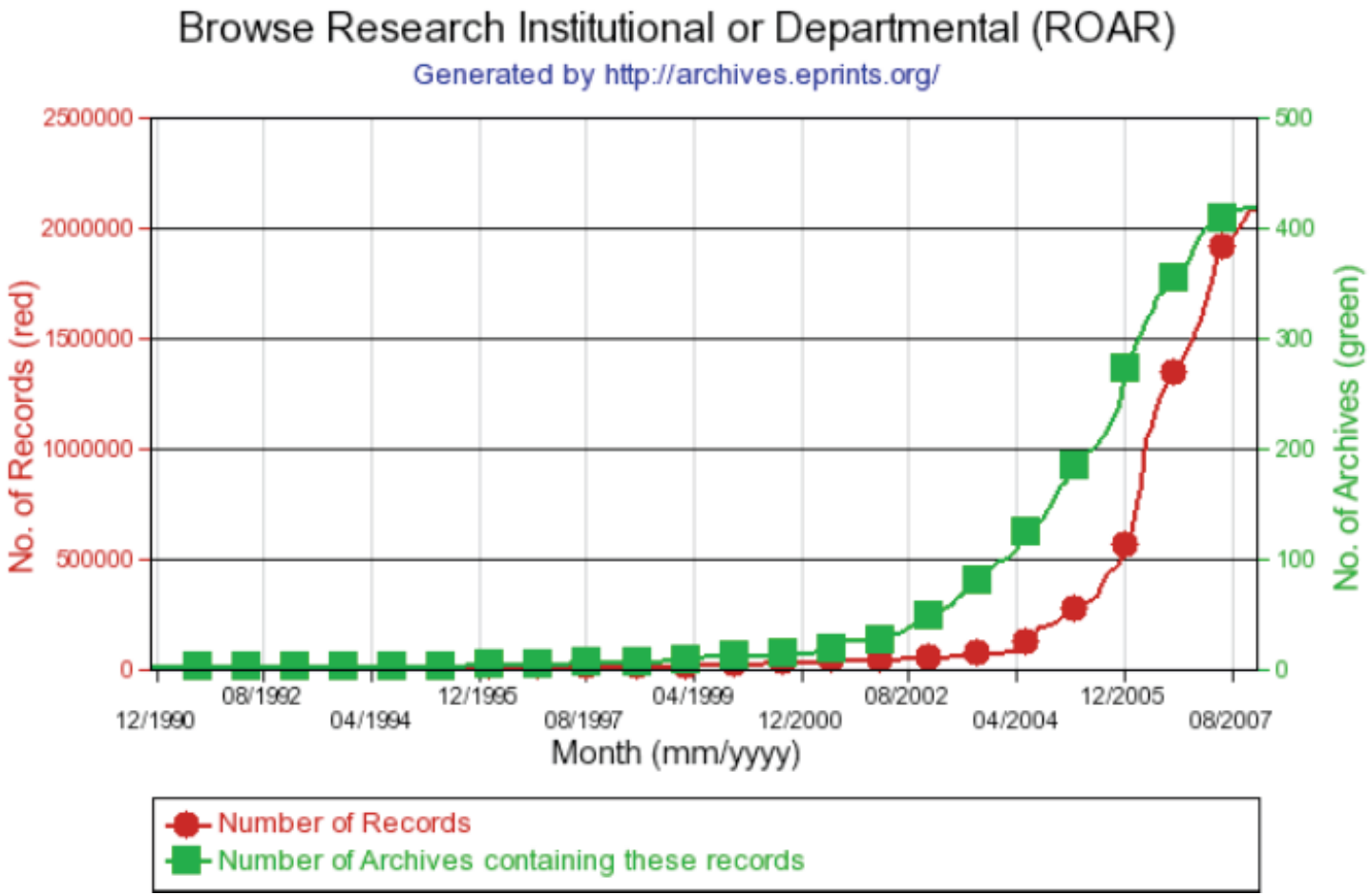


DATE	HOSTS	DATE	HOSTS
12/69	4	05/82	235
06/70	9	08/83	562
10/70	11	10/84	1,024
12/70	13	10/85	1,961
04/71	23	02/86	2,308
10/72	31	11/86	5,089
01/73	35	12/87	28,174
06/74	62	07/88	33,000
03/77	111	10/88	56,000
12/79	188	07/89	130,000
08/81	213	10/89	159,000

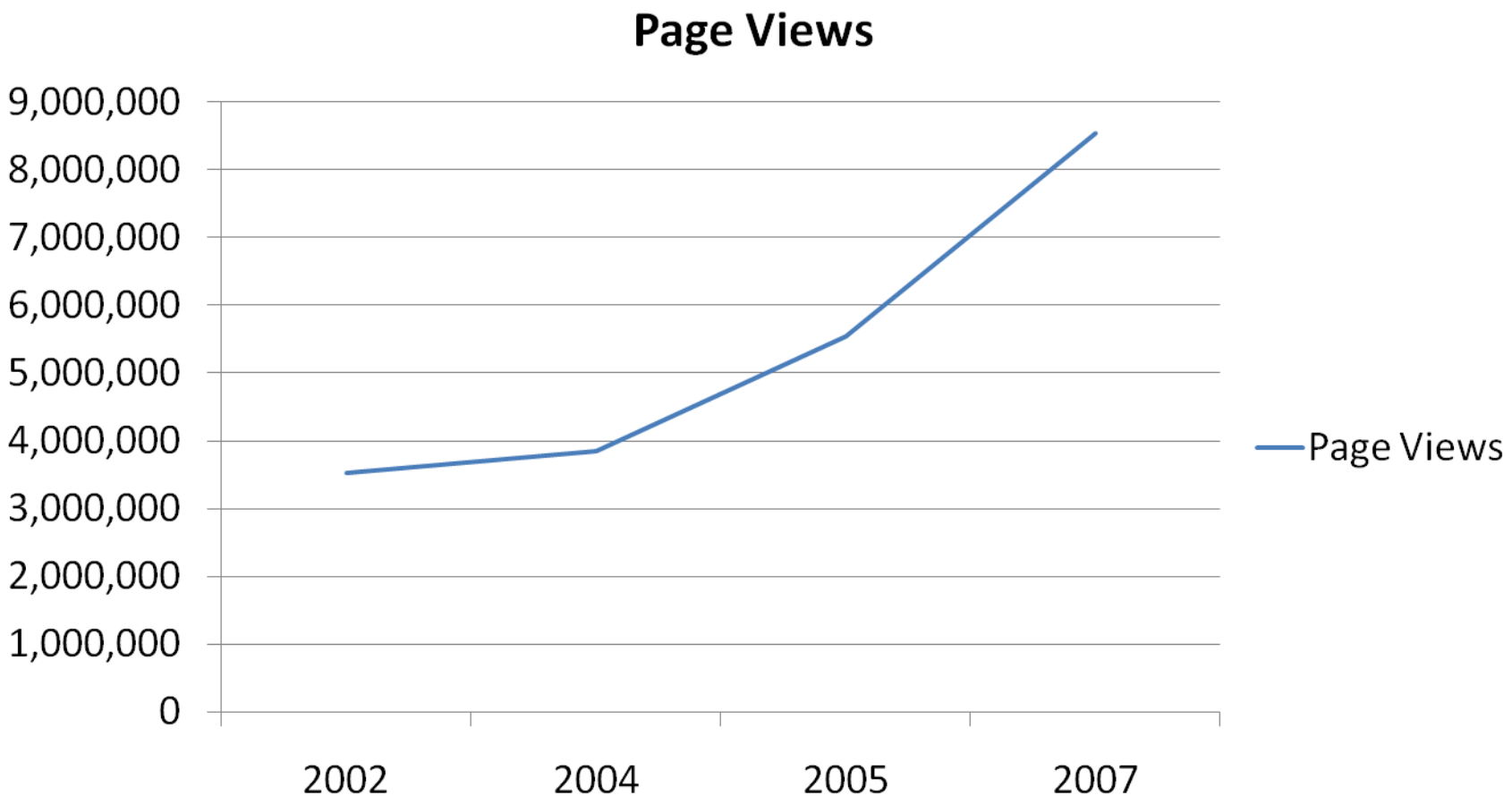
Internet Domains:



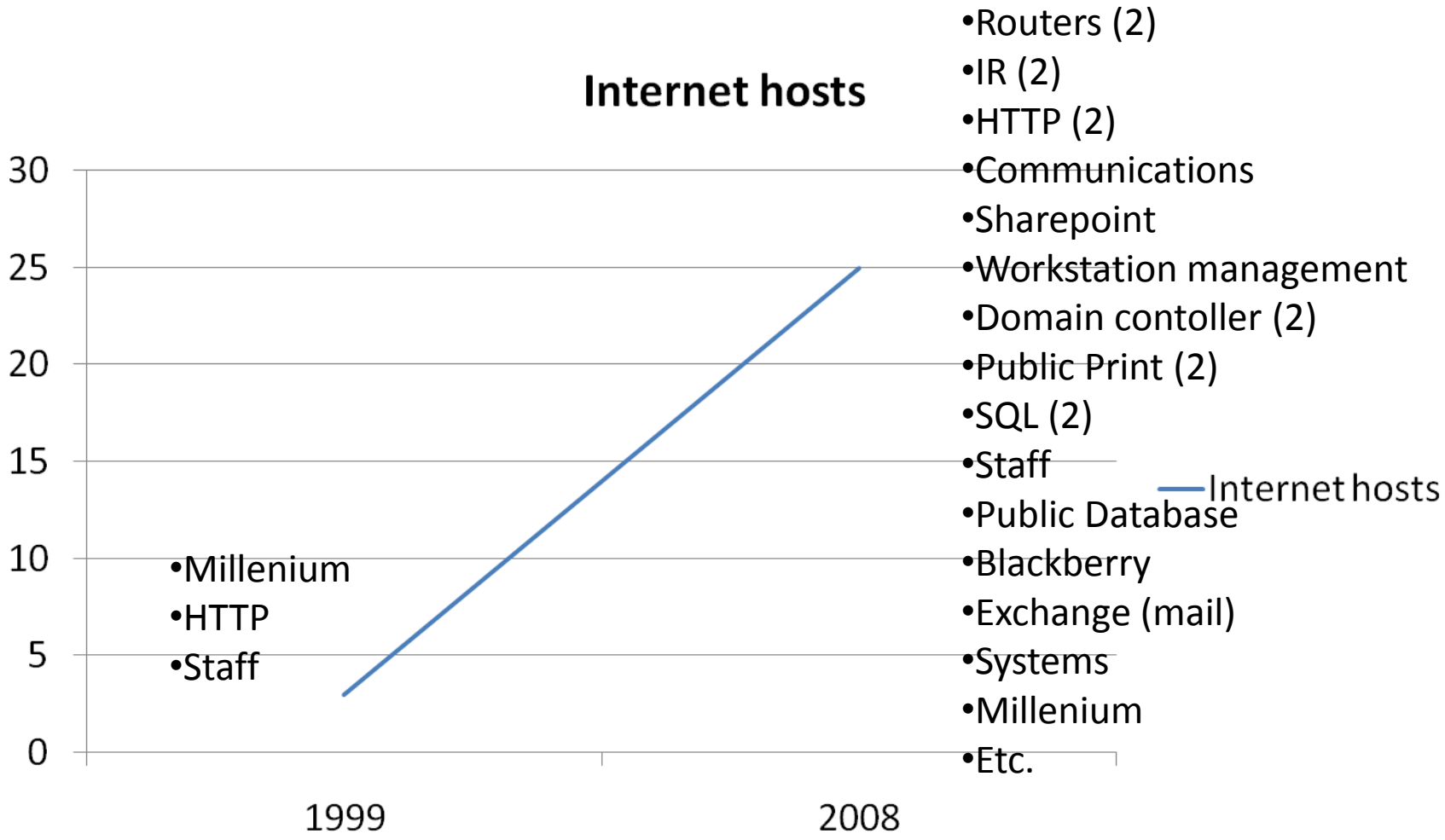
Internet Growth – Open Repositories



Library Web Site



Library Web Servers



Open Source – Systems for Digital Libraries

- http://maps.repository66.org/

Open Standards

- Definition of open standard
  - "a standard that is independent of any single institution or manufacturer, and to which users may propose amendments."
  - 3 characteristics:
    - Anyone can use the standards (utility)
    - Anyone can acquire standard without significant cost (utility)
    - Developed in a way in which anyone can participate (process)
  - Open in utility but not in process: XHTML (W3C)
  - Open in utility and process: Dublin Core

Open Standards

THE Open GROUP
Making standards work®

- Unix specification
- “free flow of information through interoperability”
- Created *Developer Declaration of Independence*

- Open Archives Initiative –
  - promote interoperability standards that aim to facilitate the efficient dissemination of content
  - Created OAI-PMH

Open Standards – Protocols, Metadata & Interoperability

- HTTP/HTML
- XML (and more W3C such as CSS)
- URI (URL/OpenURL)
- OAI-PMH
- Dublin Core

Open Standards
Metadata & Interoperability –

OAI-PMH

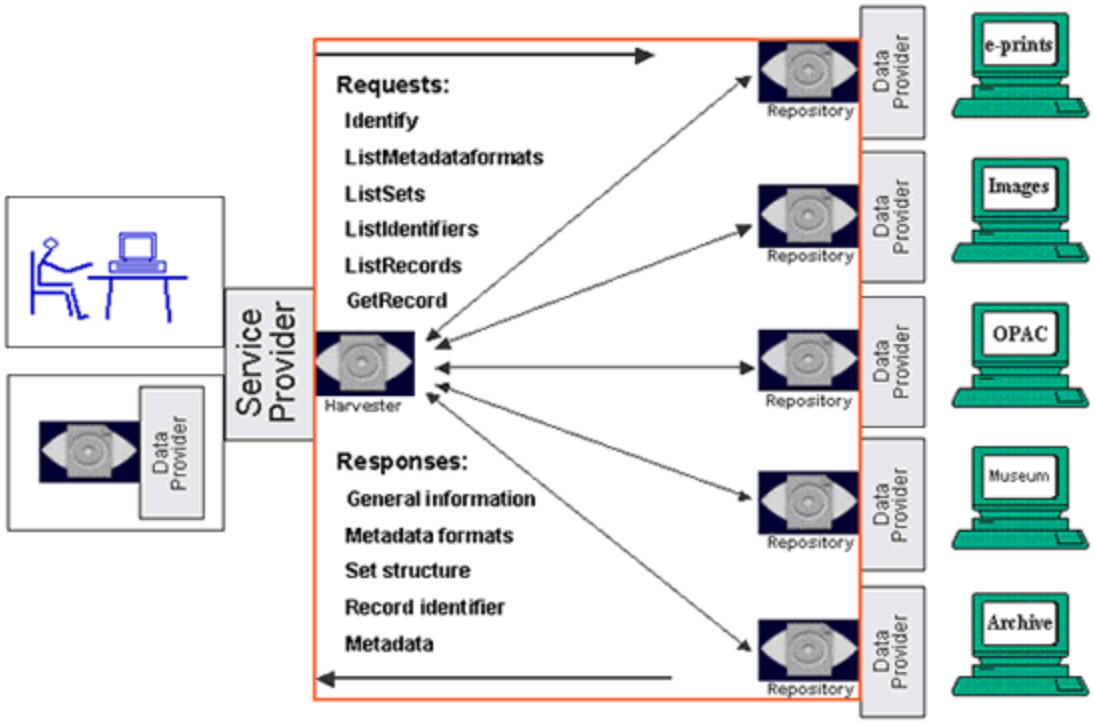
- **OAI Protocol for Metadata Harvesting (OAI-PMH)**
  OAI-PMH is a lightweight harvesting protocol for sharing metadata between services.
- **Protocol**
  A protocol is a set of rules defining communication between systems. FTP (File Transfer Protocol) and HTTP (Hypertext Transport Protocol) are examples of other protocols used for communication between systems across the Internet.
- **Harvesting**
  the gathering together of metadata from a number of distributed repositories into a combined data store.
- **Interoperability**

  allows for useful communications and collaboration between information systems.

Open Standards
Metadata & Interoperability

OAI-PMH

- The **OAI-Protocol for Metadata Harvesting (OAI-PMH)** defines a mechanism for harvesting records containing metadata from repositories.



Open Standards - Digital Library Services

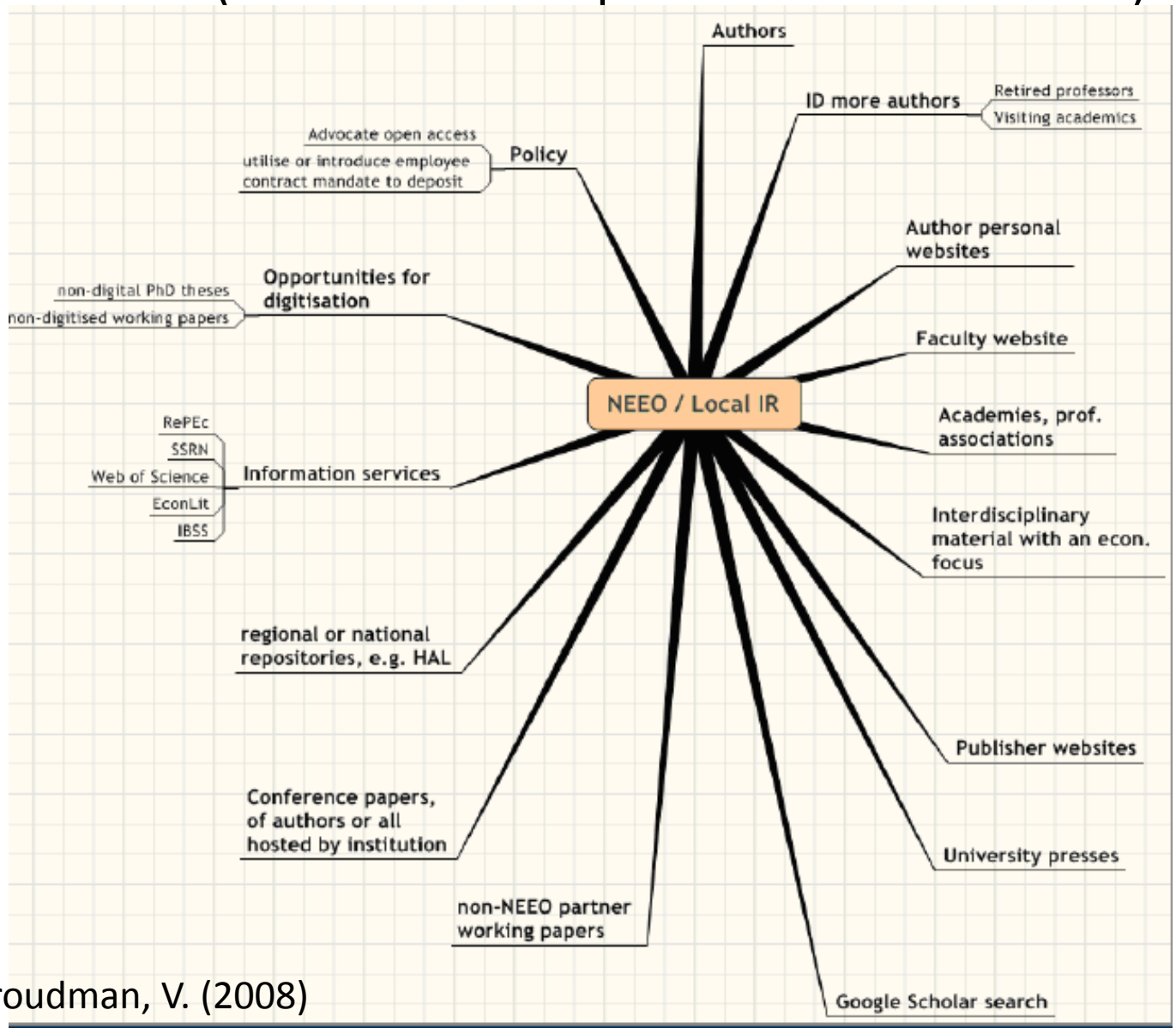
Data Services – Search & Directory

- Worldwide
  - OAIster - OAIster is a union catalog of digital resources, providing access to digital resources by "harvesting" their descriptive metadata (records) using OAI-PMH (the Open Archives Initiative Protocol for Metadata Harvesting.
  - Scirus
  - Registry of Open Access Repositories (ROAR) - Institutional Repositories
  - OpenDOAR - Directory of Open Access Repositories
  - Scientific Commons
- Canada
  - CARL - ABRC - Canadian Association of Research Libraries / Association des bibliothèques de recherche du Canada's institutional repository search
- Europe
  - DRIVER
  - NARCIS, Connecting Africa - Netherlands
  - HAL (Hyper Article on Line) – France
  - NEEO (Network of European Economists Online)
- Australia
  - http://search.arrow.edu.au/

Open Standards – Digital Library Services

- Vanessa Proudman identifies services offered by successful repositories (University of Minho, ePrints Soton, Hyper Article on Line, Cream of Science, CERN Document Server, Connecting Africa):
  - Search and browse
  - Dissemination of repository content to other information services:
    - National ones such as Intute in UK
    - International ones such as GoogleScholar
    - Disciplinary ones such as ArXiv or RePEc
    - Generating automated publication lists with links to digital objects
    - Bibliographic lists for export to personal and departmental web sites
  - Disciplinary customizations of deposit interfaces
  - Usage information/statistics pushed to researchers

Open Standards - Digital Library Services – NEEO (Network of European Economists Online)



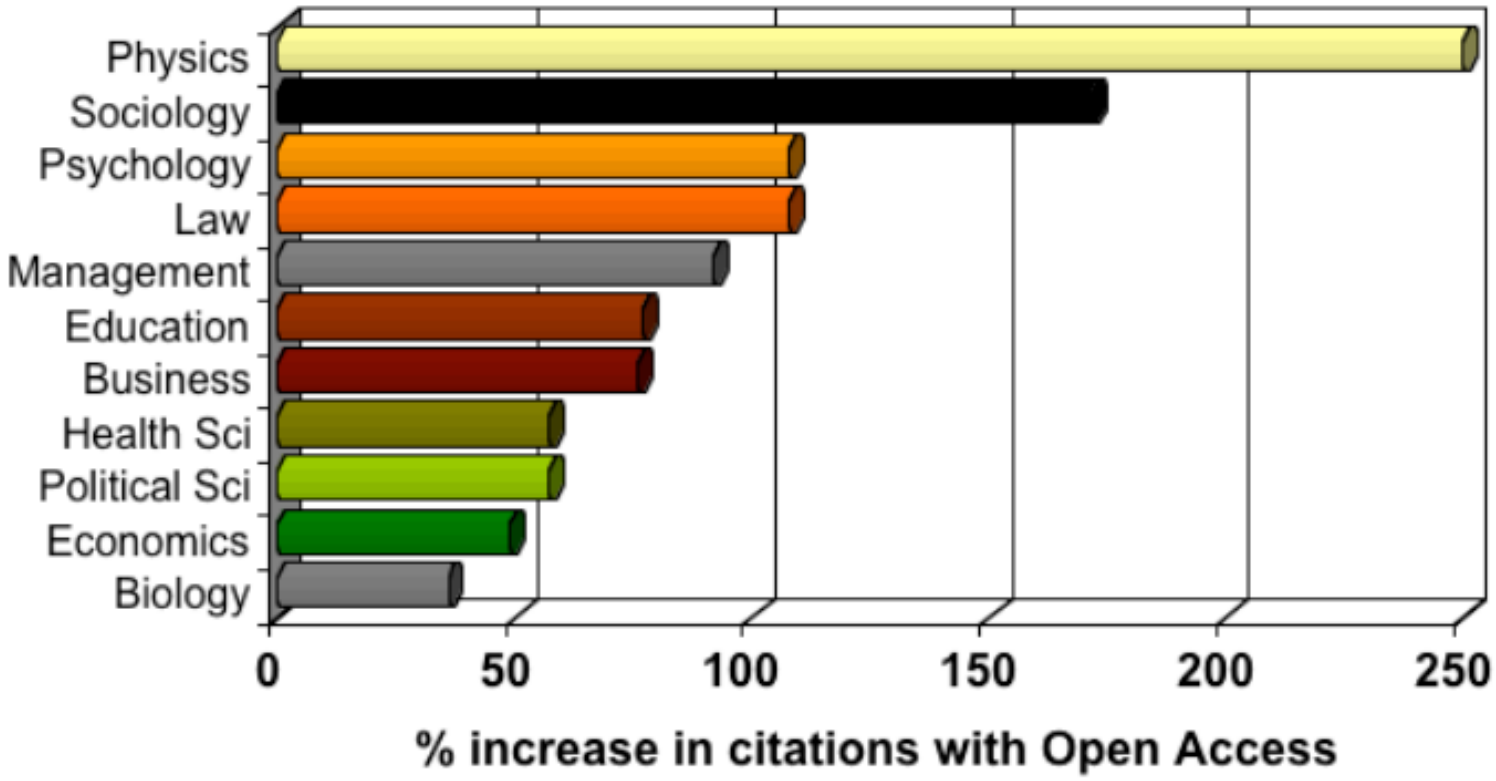
Source: Proudman, V. (2008)

Open Access

- 25,000 peer-reviewed research journals worldwide – 2.5 million articles per year
- **Gold Road** – journals are OA (currently about 10% of all journals)
  - Cost recovery transfers some of the costs to author at publication time
- **Green Road** – authors self-archive in Open Repositories

Source: Harnad, S. et al. (2008)

Open Access



Source: Harnad, S. et al. (2008)

Open Access - Disciplinary repositories

- arXiv.org - Open access to e-prints in Physics, Mathematics, Computer Science, Quantitative Biology and Statistics.

  Cogprints - archive for self-archive papers in any area of Psychology, neuroscience, and Linguistics, and many areas of Computer Science, Philosophy, Biology, Medicine, Anthropology, as well as any other portions of the physical, social and mathematical sciences that are pertinent to the study of cognition.

  Networked Computer Science Technical Reference Library - a collaborative project involving NASA Langley, Old Dominion University, University of Virginia and Virginia Tech.

  RePEc - Research Papers in Economics

  NDLTD - Networked Digital Library of Theses and Dissertations (NDLT)

Why? Support for Open Access

- The following organizations have mandated and/or endorsed open access:
  - The US National Institutes of Health (NIH) Public Access Policy
  - The US Congress: the Federal Research Public Access
  - Canada's Social Sciences and Humanities Research Council
  - The Canadian Institutes of Health Research
  - Wellcome Trust, the UK's largest private biomedical research funder,
  - The Research Councils UK
  - European Commission
  - eutsche Forschungsgemeinschaft (German Research Foundation, DFG)
  - Organization for Economic Cooperation and Development

Open Access - Copyright

Copyright

- **ROMEO** tracks journal/publisher policies (permissions and conditions) on author self-archiving

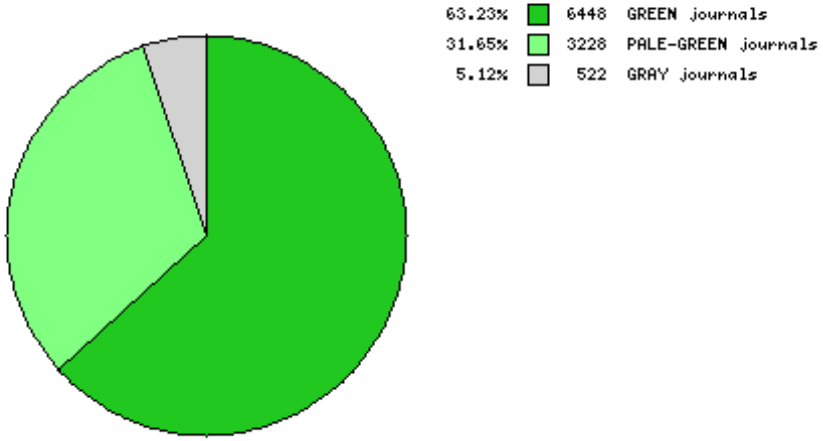
Journal Policies - Summary Statistics So Far

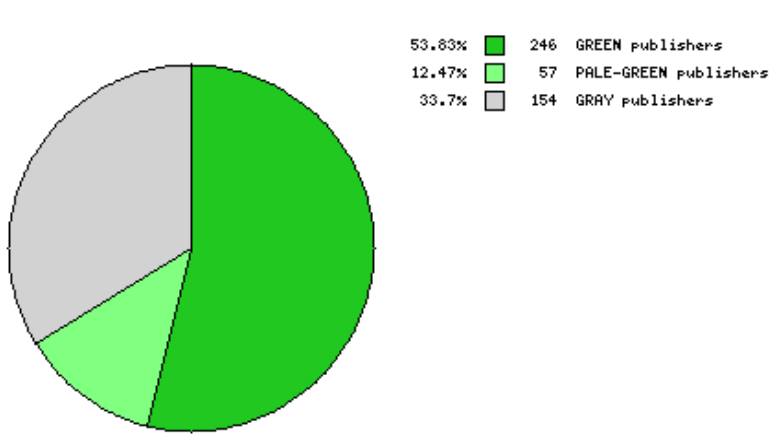
Current Journal Tally: **95% Green!**

FULL-GREEN = Postprint, PALE-GREEN = Preprint, GRAY = neither yet

Total number of publishers registered at ROMEO to date: 457

Source: (retrieved Dec 12, 2008)
http://romeo.eprints.org/stats.php





Open Access

Cyberlaw expert Lawrence Lessig suggests that the creation of a 'commons', wherein the **free exchange of ideas and collaboration** prevail, is fundamental to an open society. This belief stands in stark contrast to that of commercial and other interests, where goal is to control the Internet and its content.

-Wendy Pradt Lougee (2002) (emphasis added)

Open Access

the Open Archives Initiative seeks to address concerns within the scholarly community about certain aspects of traditional journal publishing, especially the notion of a "**gift economy**", wherein intellectual property is ceded to the for-profit sector and then repurchased for community use. These concerns have given birth to new conventions, such as e-print archives, for distributing content.

-Wendy Pradt Lougee (2002) (emphasis added)

Open Access

Doug Greenberg has offered a stark characterization of the contrast between traditional library and Internet techniques of access:

> If the key to the library's power is its **rigid, counterintuitive arrangement of static information in a comprehensible and hierarchical structure**, the key to the Internet's power is its flexible arrangement of dynamic information that permits the human mind literally to jump from one thing to another and back again with no more than stream of consciousness as a guide. It is anybody's guess which of theses systems is better adapted to *human creativity and curiosity.*

-Wendy Pradt Lougee (2002) (emphases added)

Works Cited

- "Basic Concepts of the Systems Approach." Heylighen, F. (1998) retrieved 4-12-2008 <http://pespmc1.vub.ac.be/SYSAPPR.html>
- Bunge, M. A. Finding Philosophy in Social Science. New Haven and London: Yale University Press (1996).
- Corrado, E., M. The Importance of Open Access, Open Source, and Open Standards for Libraries. *Issues in Science and Technology Librarianship*. (2005). 7-12-2008 <http://www.istl.org/05-spring/article2.html>
- Deek, F. P. , McHugh, J. A. M. Open source : technology and policy . Cambridge University Press (2008).
- Giglia E., Vignocci M. Open Access: Trends and Strategies after Berlin5. *Library Hi Tech News*. 24.9/10. (2007) 25-11-2008. <DOI:10.1108/07419050710874160>
- ISO 13407:1999 Human-centred design processes for interactive systems.
- Harnad, S., et al. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update, *Serials Review*, 34.1 (2008), Pages 36-40, ISSN 0098-7913, DOI: 10.1016/j.serrev.2007.12.005. (http://www.sciencedirect.com/science/article/B6W63-4S0HC0P-1/2/273ce267efe063d9e089fef26cd0bb16)

Works Cited

- Lerner, J., Tirole, J. The Economics of Technology Sharing: Open Source and Beyond. *Journal of Economic Perspectives* – Volume 19, Number 2. (2005). < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=629598 >
- Lougee, W. P. Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age. Council on Library and Information Resources. August 2002. <http://www.clir.org/pubs/reports/pub108/contents.html>
- Proudman, V. What can we learn from Europe in our quest for populating our repositories? *Open Repositories 2008*. (2008) retrieved 1-12-2008 <http://pubs.or08.ecs.soton.ac.uk/14/ >
- Rhyno, Art. Using Open Source Systems for Digital Libraries. Libraries Unlimited. (2004)
- "Main Technical Ideas of OAI-PMH". The Open Archives Forum. <http://www.oaforum.org/tutorial/english/page3.htm> retrieved 11/12/2008
- Weinberg, Gerald M. An introduction to general systems thinking. New York ; Toronto : Wiley, (1975)