

Open Access Business Model and Financial Issues

Malgorzata Pankowska
 Information Systems Department
 University of Economics
 Katowice, Poland
 malgorzata.pankowska@ue.katowice.pl

Abstract—The paper covers considerations on business model of open access publication for science and research development. The paper is a "state-of-the-art" report about existing open access repositories. The author aims to present that open access movement strongly depends on financial support of research and science development at academic institutions.

Keywords—open access; digital repository; business model; financial support

I. INTRODUCTION

The development of the information society and the widespread diffusion of information technology gives rise to new opportunities for research and learning. Higher education institutions have been using Internet and other digital technologies to develop and distribute education and research results for several years. However, much of that materials were locked up behind passwords within proprietary systems, unreachable for outsiders. The trend towards sharing software programmes (open source software) and research outcomes (open access publishing) seems to be strong and complemented by the trend towards sharing learning resources. The reasons for individuals and institutions to use, produce and share open education and science results can be divided into basic technological, economic, social and legal drivers. The technological and economic drivers include improved, less costly and more user-friendly information technology infrastructure, hardware and software. Legal drivers are new licensing schemes that facilitate free sharing and reuse of content. Government-supported educational institutions allow for free sharing and reuse of resources, assuming that open sharing speeds up the development of new learning resources, stimulates internal improvement, innovation and helps the institution to keep good records of materials and their internal and external use.

There is a need to look for new cost recovery models as institutions experience growing competition. Other arguments cover the altruistic motivation of sharing, personal non-monetary gain, such as publicity, reputation and opportunities to reach the market quickly for the competitive advantage. The increase of volume of research work provided online for free is the natural symptom of science development, because storing the knowledge in closed libraries is fated. Knowledge for its further development must be widely distributed, however the credit should be

given to all who contributed. The main thesis of the paper is that open access does not mean equal opportunities for participation in science and research development. The paper consists of three parts. The first part includes analysis of open access movement premises, the next subchapter comprises the discussion on financial problems at open repository institutions and an analysis of financial procedures to support the selected open access repositories. The third part includes the business model of open access repositories' development.

II. OPEN SYSTEM DEVELOPMENT PREMISES

Openness in the technical domain is characterized by technical interoperability and functionality. Open standards are important since they make it possible for different software applications to operate together. The openness blurs the traditional distinction between the consumer and the producer. The term prosumer is sometimes used to highlight the blurring of roles. To adapt or modify a digital resource it needs to be published in a format that makes it possible to copy and paste pieces of text, graphics or any published media [1]. Development of Enterprise 2.0, Marketing 2.0 and social media marketing are the excellent examples of prosumers' activities.

On the push side, it is announced that if universities do not support the open sharing of research results and educational materials, traditional academic values will be increasingly marginalized by market forces. On the pull side, a number of possible positive effects from open sharing is put forward, such as: broader and faster dissemination, people involvement in the problem solving, rapid quality improvements and faster technical and scientific development. The free sharing of software, scientific results and educational resources is believed to reinforce societal development and to diminish social inequality [2]. According to Dargan [3], open systems offer a building block approach to development that makes effective use of commercial products and open systems are based on standards that define basic system building blocks and provide a foundation for reuse, interoperability and evolution.

The greatest challenge in designing an open software system is selecting which standards to use for an enterprise. Another challenge is finding suitable standards-compliant commercial products. The third challenge is choosing standards that keep pace with technology innovations.

According to Kavanaugh [4], open source offers, in addition to a very rich set of technologies with long histories, a set of new ways to look at certain problems. Issues include:

- A variety of new licensing options and claims.
- Opportunities to deal with the loosely structured community that creates open source software, from selecting distributions.
- The possibility that open source software is built and maintained in different ways.

Open source software is delivered with source code included or easily available. Generally, intellectual property (IP) covers three main branches - copyright (original artistic and literary works of authorship); patent (inventions of processes, machines, manufactures and compositions of matter that are useful, new and non-obvious) and trademark (commercial symbols) [5]. The copyright and patent acts provide protection for intellectual property against unauthorized use, theft and other violations of the rights granted by those statutes to the IP owner. According to Cronin [6], plagiarism is not a legal but an academic offence which may be punishable according to the institution's regulation. It may be a legal offence if there are intentions to benefit from it financially at the expense of the copyright owner.

Since 1990s the open source software licensing is regulated by the activities of Open Source Initiative. The most important licenses are the General Public Licenses, the Lesser General Public License and the Berkeley Software Distribution (BSD) Licenses. Academics worldwide have started to use open licenses to create a space in the Internet - a creative commons - where people can share and reuse copyright material without fear of being sued. The Creative Commons (CC) license gives others permission to copy, distribute, display and perform the copyright work and derivative works based on it, but for non-commercial purposes only. If anyone wants to use the work for a commercial purpose they must do so in agreement with the right's holders. However, there is no clear understanding of what constitutes commercial use. Another problem is the clause called "Share Alike", meaning that any company trying to exploit the author's work will have to make their added value available for free to anyone else [1]. Researchers, as authors, have plenty of opportunities to support open access and get greater reach for their research - through open-access journals, open-access repositories and author rights management. Hine has noticed that in the science fields where book publishing is the dominant mode of communication and reputation building, publishers have a great deal of control over how those fields are represented and when, how and who can access research outcomes [7].

Though technical standards are necessary for interoperability, there has been a resistance to data standards in many humanities fields because they are perceived as necessitating the standardization of research objects and imposing a normative practice. In human sciences, there is a low-degree of functional dependence and the values and

goals incorporated in the technologies of one field are less likely to be shared by another. In physical sciences (e.g., physics, high-energy physics), the knowledge is cumulative, atomistic, concerned with universals, guaranties, simplification, resulting in discovery and explanation. The research works are politically well organized, high publicable and task-oriented. In humanities (e.g., history, linguistics) and pure social sciences (e.g., anthropology, geography), the knowledge is reiterative and concerned with particulars, qualities and resulting in understanding and interpretation. The research works are pluralistic, loosely structured, person-oriented and characterized by low publication rate. In applied sciences (e.g., mechanical engineering), the purposive and pragmatic knowledge is concerned with mastery of physical environment and resulting in products and techniques. The research works are entrepreneurial, dominated by professional values and role-oriented. Patents substitute for publications. In social sciences (e.g., education), the functional and utilitarian knowledge is concerned with enhancement of professional practices and resulting in protocols and procedures. The research works are uncertain in status, dominated by intellectual fashions and power-oriented. Publication rates are reduced by consultancies.

The fields of science have different attitudes towards publishing processes and e-science is differently understood in the particular disciplines. Generally, e-science is defined as the combination of three different developments: the sharing of computational resources, distributed access to massive data sets and the use of digital platforms for collaboration and communication [8]. This is accomplished by transferring the entire research process into the digital environment.

The creation of European collaboration and communication networks in science and scholarly research is one of the key elements of the European Research Area. This means that collaboration and connectivity indicators need to be further developed than they are at present. Scientific collaboration networks i.e. research grids, are constructed to support academic and research community. Some of the notable examples include: TeraGrid, EGEE, LA Grid, and D-Grid [9]. Many different grids have emerged in the last decade. EGEE is used among the European scientific community. BOINC is an open-source software platform for computing using volunteered resources. XtremWeb is an open source software to build lightweight Desktop Grid by gathering the unused resources of desktop computers (CPU, storage, network) [10]. Nowadays, the grid projects provide access to the open repository of research databases and publications, and they create the opportunity to utilize open source software as well as ensure e-publications on projects' deliverables [11]. The open access movement as the worldwide effort was initiated by other organizations to provide free online access to scientific and scholarly research literature, especially peer-reviewed journal articles and their preprints. The open

access movement started out with a series of statements and declarations:

- Budapest Open Access Initiative (BOAI),
- Bethesda Statement on Open Access Publishing,
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [12].

Open access means the permission for any user to read, download, copy, distribute, print, search or link to the full texts of the articles, crawl them for indexing, pass them as data to software for processing and use them for any lawful purpose. The number of open access repositories increases. The Registry of Open Access Repositories (ROAR) covers 2172 active tables of repositories. The other worldwide initiative include EBSCO A-to-Z® Web-based tool, DOAJ, Bentham Science, *OpenDOAR*. EBSCO A-to-Z® is the industry's most complete Web-based tool for organizing and providing links to all of a library's e-resources, including e-journals, titles in full-text databases, e-journal packages, and e-books. The Directory of Open Access Journal (DOAJ) developed and maintained by Lund University Libraries and the e-Depot of the National Library of the Netherlands (KB) have started a cooperation in order to secure long-term preservation of open access journals. The Swedish Library Association is generously acting as sponsor. The aim of the DOAJ is to increase the visibility and ease of use of open access scientific and scholarly journals thereby promoting their increased usage and impact. Currently, the DOAJ collection covers about 4000 journals and is characterized by a very large number of publishers (2.000+), each publishing a very small number of journals on different platforms, in different formats and in more than 50 different languages. Bentham Science is a major journal publisher of 92 online and print journals, over 200 open access journals and related print and online book series. Bentham Science answers the informational needs of the pharmaceutical, biomedical and medical research community. *OpenDOAR* is an authoritative directory of academic open access repositories. Each *OpenDOAR* repository has been visited by project staff to check the information that is recorded there. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories. *OpenDOAR* aims to provide a quality assured list of academic repositories containing full-text materials that are openly accessible. Open access repositories increasingly play a pivotal role in the emerging research information landscape. Repositories are being deployed in a variety of environments (education, research, science, cultural heritage) and contexts (national, regional, institutional, project, lab, personal). They are operating across administrative and disciplinary boundaries and interact with distributed computational services and social communities. Institutions such as universities, research laboratories, publishers, libraries and commercial organizations are creating innovative repository-linked systems for management of digital content to enable use, reuse and interconnection of information. *OpenDOAR* directory

includes 19 Polish repositories. Website design for repositories included in the directory is not the same, there is an acceptance of freedom of Website portal design. Although certain general guidelines were considered and approved, the detailed standardization of portal design is demanded in the interest of end user to enable searching and browsing. Nowadays, site map and navigation do not support searching effectively. The end user can recognize the hierarchical construction of repository content, but the construction of sub-repositories and sub-collections within repositories is not so clear and visible, and different names for sub-collections are applied. Lack of standardization of repository design revealed the general tendency to connect the end user with a particular repository even though in the interests of end user is to increase a number of accessible collections.

Taking into account the only one criterion i.e. Ph.D. works access, it should be noticed the only 8 out of 35 Polish digital libraries enable open access to those works. Some other (i.e., 5 universities) ensure open access at local library to printed copies of Ph.D. works. Only the Rector of Polytechnic Institute in Cracow in 2004 has made decision that Ph.D. works are accessible in open repository for all. According to the survey done in 20 other countries the searching results are similar and the general conclusion is that universities do not strongly support the Ph.D. works to be openly accessible online [13]. For Polish digital libraries included in *OpenDOAR* directory, the unified standard of metadata, known as Dublin Core version 1.1 is applied for all the stored publications' description. Although interface standard for eLibra digital repository Internet portal was widely applied, information retrieval is not easy because of lack of clear classification of repository content and necessity to browse through a mixture of popular daily news, old manuscripts, maps, and scientific publications.

OpenDOAR directory does not cover all of scientific research repositories in Poland. Some universities are overlooked and they develop their digital repositories within other projects. For example, Silesian Polytechnic Institute in Gliwice is involved in Springer Open Choice/Open Access scientific publication programme. Within that programme publication are funded in 100% by the Ministry of High Education in Poland within Springer/ICM agreement. Although ICT allows for high speed transfer and mass data storing, it does not mean a permission for uncontrolled redundancy of information. Unfortunately, digital library content classification are not cohesive. Lack of clear classification of publications results in longer time for searching and low effectiveness of information retrieval. Therefore it can be suspected that some valuable research publication are not quoted.

III. OPEN ACCESS REPOSITORY FINANCIAL PROCEDURE

Repositories can be organized as a place to share and exchange resources, which means that people are either users or producers, or they can promote the collaborative

production of common resources. Some initiatives of open repositories have institutional backing involving professional staff, others build on communities of practitioners or rely on their voluntary work. The survey of 180 repositories out of 2171 included in the ROAR directory allows for identification of some typical procedures for funding of open repositories.

For example, Norikazu Hyodo from Ochanomizu University Library in Japan reports that the management and operation of the ICT resource rely on the grant from the Japanese government. As of last year, Ochanomizu University Library got a grant from National Institute of Informatics (NII) for open repository development. Similarly, the financial support from NII was provided for the digital library at University of Tokyo. Paul Thirion from University of Liege, Belgium reports that their ORBi - Open Repository and Bibliography is completely financially supported by the library of the university. Ruedi Lindegger from Universität St. Gallen, Switzerland responds that for maintaining the open repository they do not need much money. The applied software is purely open source, but the coordinator for the repository is paid by the research department of the university. The input of the data (publications, projects) is done by the researchers themselves. Some additional improvements are done as a part of the administrative budget of the university with cooperation of two partner institutions.

In Finland, Helda - Digital Repository services are a part of the core functions of Helsinki University Library and thus funded from the library's general operating budget - this includes both the repository management work and the application level technical development and maintenance. For the ICT part (servers, disk space, networking etc.) the repository services use the infrastructure provided by the University IT Department. Director of Bibliothèque de l'EPFL presents that the institutional repository is managed and financially supported partly by the library, partly by the IT service. Jorgen Eriksson perceives that the running, maintenance and development costs of the Lund University repository are part of the yearly budget proposal that the university library applies for from the central university management. So economically it is treated like any other task done by the university library.

HathiTrust Digital Library partners pay the infrastructure costs for the content they deposit. The infrastructure is made up of five elements: storage, data centers, tape backup, servers and miscellaneous hardware, and staff to oversee and maintain these elements. To determine the costs for specific amounts of content, each element is converted into a per GB cost. The per GB costs are then added together to calculate a total per GB per year cost. This total cost includes one storage replacement cycle (e.g., storage that was purchased in year 1 is replaced in year 4 to prevent loss of data). Costs for replacement storage are estimated using 10% reductions in storage costs each year from the time the initial storage was purchased. The

total cost also includes the costs of storage and maintenance at two redundant storage locations (one at the University of Michigan and one at Indiana University). Partners are billed on an annual basis and adjustments to the charges are made the following year. HathiTrust is an international community of research libraries consisting of 55 universities.

BioMed Central is a science, technology and medicine research publisher which has pioneered the open access publishing model. Open Repository as a hosted solution from BioMed Central builds and maintains customized digital repositories on behalf of institutions and organizations. Open Repository is a partner of many organizations to support open access and repository development worldwide (e.g., Electronic Information for Libraries, EIFL, which is an international non-profit organization, COAR (Confederation of Open Access Repositories), DuraSpace Registered Service provider that is also a non-profit organization, Symplectic software company, Wjiti as open source software and technology provider).

At the University of Southampton, UK, e-Prints Soton repository is considered to be a core corporate service, alongside HR, student records, finance system and the content management system and is supported as such.

RePEc (Research Papers in Economics) www.repec.org is a collaboration among archive maintainers worldwide who contribute their time to documenting their materials, which are then assembled into a virtual database. RePEc is unfunded. There are various RePEc services, supported by the institutions where they run, but that mainly amounts to provision of hardware. The managers of RePEc services are mainly academics who do not receive compensation nor release time for the work they do to develop and maintain their services. Exceptions may exist for some services, e.g. MPRA at Munich University Library, where those maintaining the service may have that in their job description.

PubMed Central (PMC) as the U.S. National Institutes of Health (NIH) digital archive of biomedical and life sciences journal literature was developed and is operated by the National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM) at the U.S. National Institutes of Health (NIH). PubMed Central charges publishers nothing for including journal content in the PMC archive. A publisher is responsible for any costs it may incur in creating files that meet PMC's technical standards and transmitting them to PMC. UK PubMed Central is a service of the UKPMC Funders Group working in partnership with the British Library, University of Manchester and the European Bioinformatics Institute in cooperation with the National Center for Biotechnology Information at the US National Library of Medicine (NCBI/NLM) It includes content provided to the PubMed Central International archive by participating publishers.

Publishing Network for Geoscientific & Environmental Data (PANGAEA) as an open access library is hosted by Alfred Wegener Institute for Polar and Marine Research in Bremerhaven and Center for Marine Environmental Sciences (MARUM) at University of Bremen in Germany. The PANGAEA is supported with funding by the European Commission, the Federal Ministry of Education and Research, Deutsche Forschungs gemeinschaft, International Ocean Drilling Program.

Hyper Article en Ligne (HAL) is a multi-disciplinary open access archive for the deposit and dissemination of scientific research papers, whether they are published or not, and for Ph.D. dissertations. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

IV. OPEN ACCESS BUSINESS MODEL

Some years ago, the business model was synonymous with the revenue model. The overall business model now demands specifying the following component models:

- Research results model: how the research institutions create or acquire the knowledge products.
- Distribution model: how the research institutions deliver or distribute the knowledge products to the other researchers.
- Marketing model: the persuasive methods researchers and custodians of knowledge use to promote research results in socio-economic environment.
- Revenue model: how the research institutions get revenue.

The research results, distribution, marketing and revenue model can be combined into one business model, specifically tailored to the particulars of knowledge products, audience, distributions and transactions. The issue of who pays for open access article-processing charges is still being discussed. In Figure 1, the business model covering research results generating and distributing is presented. Generally, research works are funded from the national or regional budgets as well as from private sources and they are conducted by research and development (R&D) units, universities and private companies. Today, the complexity of research process demands individuals to be strongly involved in studying the research results of others, therefore researchers cannot work independently on knowledge production and dissemination activities of others, even if they declare as not affiliated to any institutions. However, they can present the research results on their own Websites. Today, only a small percentage of all articles have been self-archived, but universal online access could be achieved through the open repositories. Research results are published by commercial publishers as well, as fully or partly open access organizations. Open access publishers are sponsored by individuals (pay-for-publish approach) or institutions (donation approach), therefore the open repositories can offer publications free of

charge to the readers. Simultaneously, publications' market is supplied by commercial publishers, who sell books and journals to book stores and authorized access libraries. The commercial publishers prefer pay-for-access model as well as pay-for-publish. In the latter, researchers spend money from research grants on publications.

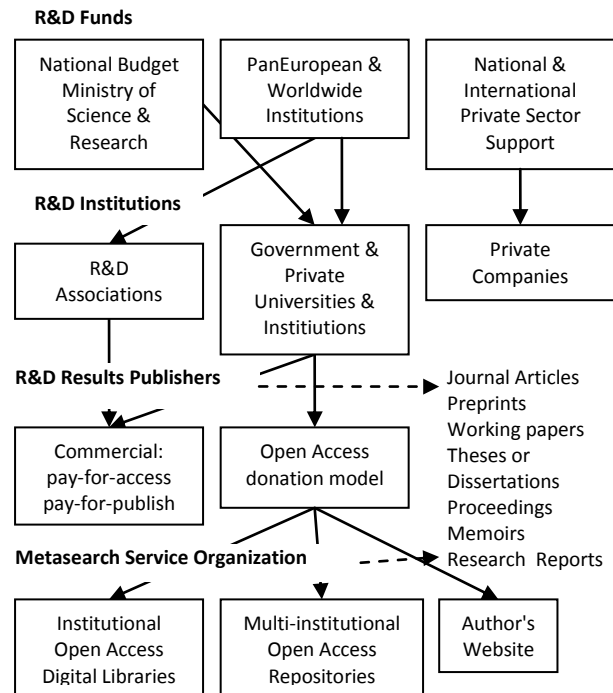


Figure 1. Open access repository financial support.

Although the commercial models are well applied in developed countries, the open access movement encourages the researchers having different research and academic backgrounds from all around the globe to submit their contributions online. The researchers understand that open access is an effective way to reach the appropriate audience. They know that when managing the intellectual property, they are able to maximize the value of the intellectual property, not to maximize the protection of closed resources of knowledge. Commercial publishing house demands authors to be involved in the process of books' and journals' editing. In that way they can reduce the intellectual property dissemination costs. Generally, the open access repositories are financially supported by the university and regional libraries which receive special funds from government institutions (i.e. the Ministry of Science, Research and Education), as well as from international projects as it is visible in the presented above review. However, strong support by the non-profit organization and private companies is still required. The open access movement and open repositories are a way to reduce costs of intellectual property production and dissemination, because here authors are also requested to work on editing and to be involved in

peer reviewing. The commercial publisher implements blended model of distributing books and journals traditionally in printed versions as well as online. However, in the knowledge generating process the costs of production are still increased, therefore in the research results publication process, the budget is spent mostly on research and development process.

V. CONCLUSION

The users need an improvement of access to open resources. The rapidly growing number of repositories makes it important to find the most relevant and highest quality resources. Metadata may improve the function of search engines, therefore approaches such as automatically generated metadata and folksonomies are being tested as applicable. There is an imbalance between the provision of open repositories and their utilization. Taking into account 2172 repositories included in the Registry of Open Access Repositories (ROAR) it should be noticed that vast majority of open repositories is not based on Western European culture, but they are developed in Japan, China and in Latin America countries. A number of projects exists in countries to support open repositories based on their own language and cultures. There is a risk, because this way the researchers share their knowledge among one culture country, taking into account that it can be a small and less developed country. Although there is an idea of repurposing the materials and the interoperability is a key issue for further open access movement development, the worldwide knowledge exchange is not simple, easy or even rational task. The digitalization of knowledge results will be helpful but does not solve the problem of understanding research results. Open standards implementation is necessary to enable research resources searching across repositories and downloading, integrating and adapting across platforms. Therefore, the development of open standards is a specialized task, which requires financial support. Considering the presented above financial procedures, there is a general conclusion that the vast majority of repositories is funded by national and international organizations, therefore there is a risk of consigning less developed countries to playing the role of consumers. The most frequented quoted publications are the results of research done in the well developed countries.

ACKNOWLEDGMENT

Hereby, I would like to thank all managers of open repositories, who supported me in survey on financial procedures.

REFERENCES

- [1] "Giving Knowledge for Free, The emergence of open education resources," Centre for educational research and innovation, Organisation for economic co-operation and development, OECD, 2007.
- [2] A. Swan, "Open Access and the Progress of Science" *American Scientist*, May-June 2007, pp. 197-199.

- [3] P.A. Dargan, "Open systems and standards for software product development," Artech House, Boston, 2006.
- [4] P. Kavanaugh, "Open Source Software, Implementation and Management," Elsevier, Digital Press, Amsterdam, 2004.
- [5] B.T. Yeh, "Intellectual Property Rights violations: Federal Civil remedies and criminal penalties related to Copyrights, Trademarks and Patents CRS Report for Congress," 2008, <http://www.fas.org/sgp/crs/misc/RL34109.pdf>, access date: 4 January 2010.
- [6] C. Cronin, "Plagiarism, copyright, Academia and Commerce," presentation given at Colby College in Proceedings of the Conference on Information Ethics and Academic Honesty, 2003, <http://abacus.bates.edu/cbb/docs/Cronin.pdf>, access date: 4 January 2010.
- [7] Ch.M. Hine, "New Infrastructures for Knowledge Production, Understanding e-Science," Information Science Publishing, Hershey, 2006.
- [8] T. Hey, and A.E. Trefethen, "The UK e-Science Core Programme and the Grid," *Future Generation Computer Systems*, 2002, 18(8), pp. 1017-1031.
- [9] A.B. Mohammed, J. Altmann, and J. Hwang, "Cloud Computing Value Chains: Understanding Businesses and Value Creation in the Cloud," *Economic Models and Algorithms for Distributed Systems*, Birkhauser Verlag Basel, 2009, pp. 187-208.
- [10] S. Naqvi, M. Villari, J. Latanicki, and P. Massonet, "From Grids to Clouds - Shift in Security Services Architectures," in *Cracow '09 Grid Workshop Proceedings*, Bubak M. Turala M., Wiatr K., Eds. Cracow, Poland, October 12-14, 2009, pp. 28-36.
- [11] "Tera Grid. Science Gateway Home," 2011, <http://www.teragrid.org/web/science-gateways/>. access date: 18 July 2011.
- [12] "Budapest Open Access Initiative: Frequently Asked Questions," August 4, 2010 <http://www.earlham.edu/~peters/fos/boaifaq.htm#openaccess>. access date: 10 January 2011.
- [13] M. Pankowska, "Open Access Movement in Science and Research", *International Conference on Information Society, iSociety'2011*, Shoniregun Ch.A. Akmayeva G.A. (eds.) Infonomic Society, London, pp. 347-353.

WEBSITES LIST

- [14] Bibliothèque de l'EPFL, <http://library.epfl.ch>
- [15] BOINC, <http://boinc.berkeley.edu/>
- [16] DOAJ, <http://www.doaj.org/>
- [17] D-Grid, <http://www.d-grid.de/>
- [18] EBSCO A-to-Z[®] <http://atoz.ebsco.com/>
- [19] EGEE, <http://public.eu-egee.org>
- [20] e-Prints Soton, <http://eprints.soton.ac.uk/>
- [21] HAL, <http://hal.archives-ouvertes.fr/>
- [22] HathiTrust Digital Library, <http://www.hathitrust.org/cost>
- [23] Helda - Digital Repository, <https://helda.helsinki.fi/>
- [24] LA Grid, <http://latinamericangrid.org>
- [25] NII, <http://www.nii.ac.jp/en/>
- [26] OpenDOAR, <http://www.openoar.org/>
- [27] Open Source Initiative, <http://www.opensource.org>
- [28] Open Repository, <http://www.openrepository.com/>
- [29] PANGAEA, <http://www.pangaea.de>
- [30] PMC, <http://www.ncbi.nlm.nih.gov/pmc/>
- [31] ROAR, <http://roar.eprints.org/>
- [32] TeraGrid, <http://www.teragrid.org>
- [33] XtremWeb, <http://www.xtremweb.net/>