Supporting E-Science: The Role of Digital Repositories in Scientific Communication

Radovan Vrana
Department for Information Sciences
Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
rvrana@ffzg.hr

Abstract. *Digital technology has transformed scientific research and scientific communication profoundly. Scientists around the world use digital information resources such as digital repositories for their research and publishing of results of that research. Digital repositories are digital archives of intellectual output of university. They are also building blocks of e-science - data driven, data intensive and networked science. As their number increases, they are being analyzed for their good and weak sides in order to integrate them more firmly into our daily scientific practice.*

Keywords. Scientific communication, Digital Repositories, E-Science

1 Introduction

"What we academics do is we think thoughts, read books, gloss texts, run experiments, make notes, interpret data and put what we know or think we know in the library where others can find it." [6].

While this has been true for many years, decades and even centuries, current situation in the world of scientific communication is quite different. The notion of term scholarly communication itself has changed. According to Thorin [33], "term scholarly communication has evolved to illustrate the breakdown of the process of traditional scholarly publication; that is, as a means to disseminate research results, the present system of scholarly communication can no longer meet the needs of the scholarly community at large." The same author divided scientific communication process into three distinct aspects:

- The process of conducting research, developing ideas, and communicating informally with other scholars and scientists;
- The process of preparing, shaping, and communicating to a group of colleagues what will become formal research results; and
- The ultimate formal product that is distributed to libraries and others in print or electronically.

System of scientific communication has changed profoundly under the influence of digital technology [23]. Earlier in history of our civilization, archivists and scholars traveled to distant places to read publications, while today, they are offered access to digital collections on the Web, reducing the need travel to access scientific knowledge [17]. According to Lawrence [20], improvements to the accessibility of scientific literature helped scientists to locate relevant research and to improve communication and progress in science. Yet, electronic publications in science aren't always welcome, especially in the process of promotion of scientists. A complex study done by Harley et al. [12], confirmed that "rapid dissemination through working papers and preprints in some fields, and the posting of research papers on personal websites and in institutional repositories, will not soon abolish the overriding influence of the formal, peer-reviewed publication system on tenure and promotion requirements in most disciplines". Reasons for resistance to online scholarship are doubts about long-term preservation, fetishism for print, to a lack of leadership, to a dearth of technical expertise [32]. Scientists and scientific institutions that accept the Internet as a credible environment for transfer or scientific information might use blogs, wikis, podcasts. discussion boards, listservs, personal, discipline, and institutional websites, and e-forums to communicate research [32] by publishing it.

Scientific publishing is one of the most important activities in the global system of scientific communication. Outcomes of scientific research become scientific knowledge, and this knowledge has been communicated through the ages primarily via scientific publishing [16]. According to Harley et al. [12], sharing scholarly work at multiple stages of development is universal. For individual scholars, publication provides the basic and often the only channel through which to participate in scholarly activities and make contributions to the knowledge enterprise [30].

As previously noted, until recent history, scientists have been relying on printed information resources, primarily on journals for publication of results of their research. This situation has changed with development of the Internet which now offers abundance of information and information resources in versatile digital formats. The pace of creation of new digital content is becoming faster every day. The volume of scholarly literature has become a strong driving force for development of electronic publishing and digital information resources which can accept diverse digital assets. Gandel, Katz and Susan [10] analyzed the problem of managing the abundance of information. They believe that the current and future era of information abundance will challenge many basic assumptions and practices about safeguarding, protecting, filtering, preserving, evaluating, purging, describing, cataloguing, and vetting information for the purposes of teaching, learning and scholarship. Our future will offer new solutions for the problem of abundance of information by creating new information infrastructures and by changing individual and collective behavior in information environment. Many scientists already have personal digital repositories on their personal computers and / or personal Web sites and store scientific, professional and personal life experiences in them.

While the number of digital information resources increases, printed scientific journals and books are still used. Scientists now use material from two worlds: printed and digital. In this hybrid world, "users desire a hybrid information environment in which online information does not supplant information in print but adds new access opportunities for users to choose." [22]. Pinfield and James [27], called for complete reorientation of scientists towards publishing in digital environment to automate the process of collecting and identification of digital objects in repositories. They concluded that "Once researchers become comfortable with e-prints as a way of disseminating their work, they may start to create files that look more and more like multimedia products.", which will be easier to process than digitized printed articles and books.

Increase in the number of published papers in the 20th century transformed scientists' perception of publishing in electronic environment [23] as some scientists were reluctant to publish in electronic publications because articles published in this type of publications were rarely taken into account for academic promotion.

All these activities are part of the global search of scientific community for new, accessible and stable communication and publishing channels. Digital technology and especially the Internet have been incentives for scientific publishing and system of scientific communication in general to transform into accessible, reliable and usable digital information environment. Two types of digital information resources that appeared in the first part of 1990s, digital libraries and digital repositories introduced new possibilities of publishing, storing and dissemination of scholarly works. Academic libraries, traditional places for dissemination of scientific information, have also undergone changes under the influence of digital technology. According to Breeding [4], they have been increasingly involved in the creation of digital library collections and other content repositories and academic libraries have become increasingly active in building digital repositories of content produced within their institutions and in other aspects of publishing scholarly content. Traditional models of functioning of libraries have transformed in favor of online information resources. Students and researchers are replacing the onsite library for the online environment, because digital repositories offer a chance for libraries to stay relevant and remain a benefit to their community [11].

2 Digital repositories

Digital institutional repository (a digital information repository that is a part of university or other institution) is a digital archive of the intellectual product created by the faculty,

research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access [18]. For Branin [3], digital institutional repository is "a library, an archive, a museum, or even a warehouse that stores for use and safekeeping an organization's records or artifacts falls under the broad definition of an institutional repository." For Hess and Ostrom [13] digital institutional repositories are "a viable solution to the scholarly communication crisis because they provide open access distribution, capture diverse forms of knowledge and communication, and archive and preserve an institution's scholarly record. This new form of scholarly information collection benefits not only information harvesting and preservation, but it is also hoped that a successful institutional repository will reduce reliance on expensive publishers.".

Generally, digital repositories can be divided into at least two categories: repositories that serve single university or several university campuses (they are called institutional repository). Institutional repositories organize themselves along organizational or political jurisdictional lines, and they collect and manage digital assets in a variety of formats and subjects for the constituents within that jurisdiction [3]. Another type of digital repositories is one that serves a scientific discipline or several closely related disciplines (it is called disciplinary repositories) [36]. Disciplinary repositories focus on the collection of digital assets in a subject area [3].

Key part of digital repositories is the management of technological changes and the migration of digital content from one set of technologies to the next [24]. Long term preservation is therefore very important priority for institutions providing technical, financial and administrative support to digital repositories. Wheatley [38] pointed out that a number of requirements must be met to achieve long term accessibility of archived objects. He singled out key functional goals of digital repositories:

1. Data can be maintained in the repository without being damaged, lost or maliciously altered.
2. Data can be found, extracted from the archive and served to a user.
3. Data can be interpreted and understood by the user.
4. Goals 1, 2 and 3 can be achieved in the long term.

To survive, digital repositories need a significant institutional commitment because of the current economic crisis, organized publisher resistance, institutional dysfunction, rapidly changing technology, and confusion about what digital institutional repositories are and what they can and should do [1].

Digital repositories can have many functions and serve many purposes. This part of the paper will put focus on some of them which could have potentially long-term consequences of future development of the global system of scientific communication.

Read [29] singled out three important roles of digital institutional repositories (repositories that are part of universities):

- Publications of staff offering a convenient way for an higher education institution to recognize and store research outputs and make research outputs openly available (with evidence that this increases citation;
- Preserving research data for use by other researchers;
- Holding educational resources such as course material and handouts, especially where these are to be made openly available rather than held within a virtual learning environment (VLE) application for restricted access.

Digital repositories have another important purpose. They present an opportunity for replacing traditional scientific journals. The current system of scientific communication and its important component, scientific publishing based on scientific journals, are now over 300 years old and they need to be changed. Traditional system of scholarly publishing is changing under the influence of the exponential growth of information production, the dramatic increase in subscription fees, the increasing storage cost of printed documents, and the increasing power and availability of digital technology [23]. To find out whether digital repositories can inherit the practice of printed scientific journal, Prosser [28] compared functions of traditional printed scientific journals and digital repositories. First, he singled out traditional functions of journals: registration - the author wishes to ensure that he or she is acknowledged as the person who carried out a

specific piece of research and made a specific discovery; certification - through the process of peer-review it is determined that the author's claims are reasonable; awareness - the research is communicated to the author's peer group and archiving - the research is retained for posterity.

Then he compared journals with digital repositories which showed that digital repositories mirror at least three functions of traditional journals: registration – by depositing in the repository the researcher would make claim to their discovery; awareness – by constructing the repository to OAI standards the institution would ensure that the researcher's work would be found by search engines and available to their peers and archiving – the institution would be responsible for maintaining the long term archive of all the work produced by members of that institution.

This comparison showed that digital repositories are capable of replacing traditional printed scientific journals, but the question is whether they will actually be given an opportunity to replace scientific journals or will they just serve as their long term storage?

Speaking about long term storage, digital repositories can have another important function. As number of digital scientific content grows, it requires adequate and long-term preservation. Moghaddam [25] investigated archiving of scientific electronic journals and concluded that the content of digital files may be lost to future scientists because the physical item deteriorates, the information cannot be extracted and interpreted correctly and because of the short lifecycle of digital media and obsolesce of the technology used. The format of the digital resources can be damaged or lost and may no longer be intact, retrievable, understandable, or displayable. However, the fact that some university has digital repository still doesn't mean that it is oriented towards long term preservation of its content. Levi pointed out that libraries and information centers have digital asset management systems or digital repositories for managing and storing digital objects which are not designed for long term preservation but they focus on access management, or facilitating the day-to-day use of digital content by users [21].

It is rather difficult to collect and present all characteristics of digital repositories in one place. Shearer [31] tried to do so, and enumerated most important characteristics:

- Content of institutional repositories must be digital
- Digital repositories must be institutionally defined – repositories represent the intellectual life and output of an institution
- Digital repositories aim to collect scholarly content exclusively – digital repositories collect many types of content produced at an institution
- The nature of digital repositories are cumulative – they make a commitment to preserve and make accessible digital content on a long term basis
- Digital repositories must provide free and open access to their content
- Digital repositories must be interoperable

Digital repositories are closely related to open access. Open access to scientific and non-scientific literature means its free availability on the public Internet, permitting any user to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited [5]. More specifically its purpose is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge [5]. On April 18th 2010 Registry of Open Access Repositories (ROAR) at http://roar.eprints.org/ listed 1702 active digital repositories around the world. All these digital repositories are open access repositories with occasional restrictions in access or without any restriction. Digital repositories are in many cases realization of the idea of open access to scientific literature. According to Crawford [7],

there are two ways for making deposits in open access digital repositories: green way and gold way. Green way means depositing articles in online digital repositories (after publication) that are harvestable using OAI protocols. Gold way means the journal itself provides immediate full-text online access at no charge and with no restrictions other than attribution (i.e. scientists publish results of their research directly in the repository and not in some printed journal); the online version of the journal is funded by some means other than mandatory subscriptions.

3 Content of digital repositories

Digital repositories are capable of storing versatile content. Regardless of their type (whether they are institutional or they serve a scientific discipline), digital repositories offer mechanism for storage of different file formats. Van Westrienen, and Lynch [34] distinguish several types of material commonplace in repositories they investigated: articles, books and theses, primary data, video, music and the like, course material and other types of material. An institutional digital repository can contain e-prints of scientific papers, research data, but also e-learning materials and other forms of institutional intellectual outputs, which are generally not published or preserved elsewhere [15]. Jones [19] found the following material to be part of digital repositories: archived research paper with the pre-print, the post-print, the PDF and the TeX, journal articles, learning objects, and sub-sections of continuous data. Waters [37] pointed out that these digital assets represent resources for research and teaching with an important role in higher education for training students' research methods. Furthermore, the same author suggested that they should be managed to advance knowledge and improve education. He presented the idea that in time, as scientists in different fields gain experience with and develop discipline-based methodologies for using large-scale digitized content, as well special collection and new media collections, they will need to incorporate the material and train students in the research methods.

4 Digital repositories in Croatia

Scientists who work at the Croatian universities and in scientific institutes are part of the global scientific community and share some common communication habits with their colleagues. To find out more about sharing of scientific knowledge, a research in autumn of 2009 was initiated [35]. The aim of this research was to collect data about existing digital repositories at seven universities in Croatia. Data from six out of seven universities (their central libraries) were collected. The research discovered only nine active digital repositories located at six Croatian universities. Three most represented types of content in digital repositories were educational material, professional papers and scientific papers. Almost half of repositories offer open access without restrictions. However, active digital repositories are still not popular among scientists and students of universities at which repositories are located. Many of these users rarely ask librarians to include material of their interest. On the other side, librarians rarely if ever, evaluate work of their repository. Librarians who participated in the research believe that digital repositories have a great influence on teaching and research. According to librarians' opinion, digital repositories are suitable for long term preservation of professional and scientific works. Results of this research at the Croatian universities showed that digital repositories have become part of the universities' infrastructure. Although their number might be small, one should bare in mind that digital repositories can serve a large number of users, or, in this case, teachers, researchers and students at faculties or universities. Digital repositories have certain future in academic community. The existing digital repositories could be improved if they receive greater financial support and recognition in academic community.

5 E-science

With development of digital repositories scientists around the world got an opportunity of using vast amount of scientific information accessible on the Internet in organized fashion. Digital repositories have been made available to public to help scientists in storing, organizing and using articles, books and research data generated by research and scientific measurement equipment. By making available research data online, a major shift in perception of the final result of scientific endeavor from journal article to data or data sets has been initiated. Research data is now a valuable asset

which could serve as a starting point for a new research. As a result, data from previous research can be accessed and used in subsequent researches. Science which puts research data in focus and enables scientists to do research and communicate on the Internet is frequently called e-science.

E-science is referring to scientific activities supported by high bandwidth computer-mediated telecommunications networks, and particularly to encompass the variety of such digital information-processing applications that are expected to be enabled by the grid i.e. the general purpose network technology which will serve to facilitate new, computationally intensive forms of scientific inquiry [8]. It also means the widespread availability of digital content is creating opportunities for new forms of research and scholarship that are qualitatively different from the traditional way of using academic publications and research data [2]. For Osswald [26] "E-science is based on distributed networks providing the software and computer power necessary to process large sets of data, by interconnecting computers and tools wherever they are available. This is supposed to enhance information exchange and intensify communication between researchers, improving their competitiveness as well as their ability to cooperate, particularly on the international level." For Hey and Hey [14], e-science is shorthand for the set of tools and technologies required to support collaborative, networked science. It will empower scientists to do their research in faster, better and different ways. E-science offers better support collaborative science i.e. the ability to have broader interactions through the sharing of data, experimental approaches and both intermediate and final results in systems that will maintain a history of the data, processes, outcomes and conversations among scientists [39]. Farooq et al. [9] described e-science or e-research as distributed and large-scale scientific collaboration enabled by Internet technologies. The same group of authors singled out the benefits of use of the Internet for scientific collaboration: advanced computational, collaborative, data acquisition services available to scientists and scholars in all disciplines through high-performance networks.

All these definitions demonstrate the need of scientists and universities for quality digital information resources that will help interconnecting scientific knowledge around the world. Digital repositories can help them organize scientific knowledge derived from research done at university where digital repository is located or from research done in some other academic institution in the world. Computer networks are capable of connecting all these digital information resources are enable scientist to forget where they are located at the moment of research and to focus on the research problem at hand using all available scientific knowledge that could lead them to another scientific discovery.

6 Conclusion

The pace of creation of new digital content is getting faster. As a consequence, academic libraries and universities are facing problems of storing and preserving of the increased quantity of digital content. This new content requires proper curation to remain accessible and available in future. Digital libraries and digital repositories are ideal for such task and they are of great interest to scientific community. During the last few years, special attention has been given to digital repositories which have been recognized as crucial storage and preservation facilities for scientific knowledge in academic institutions around the world. Their increasing number meets the need of scientists and professionals around the world for quality information resources which have ability of replacing printed information resources. Growth of digital information resources helps development of e-science, a new type of scientific endeavor supported by computer networks and data curation centers. Next decade will demonstrate practical values of use of digital information resources in science as more scientists will publish and store results of their work in them.

References

[1] Albanese A.R. :Thinking beyond the box, Library Journal, 2009, pp. 26-28.

[2] Arms W., Larsen R.: The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship, available at: http://www.sis.pitt.edu/~repwkshop/SIS-NSFReport2.pdf, Accessed: 15th March 2010.

[3] Branin J.; Institutional repositories, available at:

https://kb.osu.edu/dspace/bitstream/1811/441/1/inst_repos.pdf, Accessed: 1st March 2010.

[4] Breeding M., Maximizing the Impact of Digital Collections, Computers in Libraries, 2009, pp. 32-34.

[5] Budapest Open Access Initiative, available at: http://www.soros.org/openaccess/read.shtml, Accessed: 25th February 2010.

[6] Courant P.N.: Scholarship and Academic Libraries (and their kin) in the World of Google, First Monday, 2006, available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1382/1300, Accessed: 1st March 2010.

[7] Crawford W.: Open Access: It's Never Simple, Online, 2008, available at: http://www.infotoday.com/online/jul08/index.shtml, Accessed: 3rd March 2010.

[8] David P.A., Spence M.: Towards institutional infrastructures for e-Science: the scope of the challenge. The University of Oxford, Oxford, 2003.

[9] Farooq, U et al. Designing for e-cience: Requirements gathering for collaboration in CiteSeer, International Journal of Human-Computer Studies, 2009, pp. 297-312.

[10] Gandel P. B., Katz, R. N., Susan E. M.: The "Weariness of the Flesh": Reflections on the Life of the Mind in an Era of Abundance, EDUCAUSE Review Magazine, 2004, p. 40-51.

[11] Guevara S.: Diving into the blue: a look at Michigan's repositories, Computers in Libraries, 2009, pp. 15-17.

[12] Harley D. et al.: Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines, University of California Press, Berkeley, 2010.

[13] Hess C., Ostrom E: Studying Scholarly Communication: Can Commons Research and the IAD Framework Help Illuminate Complex Dilemmas?, 2004, available at: http://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/2147/Hess_Studying_040512_Paper192.pdf?sequence=1, Accessed 24th February 2010.

[14] Hey T., Hey J.: E-Science and its implications for the library community, Library Hi Tech, 2006, pp. 515-528.

[15] Hockx-Yu H.: Digital preservation in the context of institutional repositories, Program: electronic library and information systems, 2006, pp. 232-243.

[16] Honey S.L.: Preservation of Electronic Scholarly Publishing: An Analysis of Three Approaches, Libraries and the Academy, 2005, pp. 59–75.

[17] Huwe T. K.: Exploiting Synergies, Online, 2009, pp. 14-19.

[18] Johnson R. K.: Institutional repositories: partnering with faculty to enhance scholarly communication, D-Lib Magazine, 2002, available at: http://www.dlib.org/dlib/november02/johnson/, Accessed: 7th March 2010.

[19] Jones R.: Giving birth to next generation repositories, International Journal of Information Management, 2007, pp. 154-158.

[20] Lawrence S. Free online availability substantially increases a paper's impact, Nature, 2001, p. 521.

[21] Levi Y.: Digital preservation: an ever-growing challenge, Information Today, 2008, available at: http://www.infotoday.com/it/sep08/index.shtml, Accessed: 10th March 2010.

[22] Liu Z.: Print vs. electronic resources: A study of user perceptions, preferences, and use, Information Processing and Management, 2006, pp. 583-592.

[23] Liu Z.: Trends in transforming scholarly communication and their implications, Information Processing and Management, 2003, pp. 889 - 898.

[24] Lynch C.: Institutional repositories: essential infrastructure for scholarship in the digital age, ARL Bimonthly Report, 2003, available at: http://www.arl.org/resources/pubs/br/br226/br226ir.shtml, Accessed: 7th March 2010.

[25] Moghaddam G.G.: Archiving Challenges of Scholarly Electronic Journals: How Do Publishers Manage Them?, Serials Review, 2007, pp. 81-90.

[26] Osswald A.: E-science and information services: a missing link in the context of digital libraries, Online Information Review, 2008, pp. 516-523.

[27] Pinfield S., James H.: The Digital Preservation of e-Prints, D-Lib Magazine, 2003, available at:

http://www.dlib.org/dlib/september03/pinfield/09pinfield.html, Accessed: 11th April, 2010.

[28] Prosser D. C. The Next Information Revolution - How Open Access will Transform Scholarly Communications in International Yearbook of Library and Information Management 2004-2005: Scholarly Publishing in an Electronic Era, Facet Publishing, London, 2005, pp. 99-117.

[29] Read M.: Libraries and repositories, New Review of Academic Librarianship, 2008, pp. 71-78.

[30] Shaw V.N.: Scholarly Publishing: Reforms for User Friendliness and System Efficiency, Journal of Scholarly Publishing, 2009, pp. 241-262.

[31] Shearer K.: Institutional Repositories: Towards the Identification of Critical Success Factors, Canadian Journal of Information and Library Science, 2003, available at: https://dspace.ucalgary.ca/bitstream/1880/43357/6/CAIS-IR.pdf, Accessed: 20th March 2010.

[32] Smith S.E., Mercer H.: It's the Content, Stupid, American Libraries, 2010, pp. 48-51.

[33] Thorin S. E.: Global Changes in Scholarly Communication in eLearning and Digital Publishing, Springer Netherlands, 2006, pp. 221-240

[34] Van Westrienen G., Lynch C.A..: Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005, D-Lib Magazine, 2005, available at: http://www.dlib.org/dlib/september05/westrienen/09westrienen.html, Accessed: 20th March 2010.

[35] Vrana R.: Digital repositories at the Croatian universities: providing support for the enhancement of research and teaching, Proceedings of MIPRO2010 33rd International Convention on Information and Communication Technology, Electronics and Microelectronics, 2010. (in print)

[36] Xia J.: Disciplinary repositories in the social sciences", Aslib Proceedings: New Information Perspectives, 2007, pp. 528-538.

[37] Waters D.: Managing digital assets in higher education: an overview of strategic issues, ARL Bimonthly Report, 2006, pp. 1-10.

[38] Wheatley P. Institutional Repositories in the context of Digital Preservation, 2004. available at: http://www.dpconline.org/docs/DPCTWf4word.pdf, Accessed: 1st March 2010.

[39] Wright M., Sumner T., Moore R., Koch T., Connecting digital libraries to eScience: the future of scientific scholarship, International Journal on Digital Libraries, 2007, pp. 1-4.