2009

# Institutional repositories as sustainable infrastructure supporting e-scholarship

Swanepoel, Marinus

University of Lethbridge

Swanepoel, M. (2009, June). Institutional repositories as sustainable infrastructure supporting e-scholarship. Paper presented at the Educause Southeast Regional Conference 2009: A changing world: rethinking IT services for a sustainable future, Atlanta, GA. doi:10.13140/RG.2.1.2040.1601

http://hdl.handle.net/10133/3698

# Institutional repositories as sustainable infrastructure supporting e-scholarship

Marinus Swanepoel
Librarian
University of Lethbridge
4404 University Drive
Lethbridge
AB
T1K 4M5
Canada
(403) 329-2261
marinus.swanepoel@uleth.ca

The development of digital repositories has been a relative recent one, gathering momentum only in late 2000 with the release a software package called E-Prints. Factors such as the falling costs for online storage and the increase of broadband networking technologies have contributed to their current popularity. The status of institutional repositories is investigated in the context of the Gartner Hype Curve. The available numbers on institutional repositories are reviewed with a view of establishing its current position on the Gartner Hype Curve. Its role in changing the scholarly communication process and the contribution it makes to the open access movement is referred to.

In tracking the development and growth of institutional repositories data was harvested from one of the international registries: the Registry of Open Access Repositories (ROAR), which indicates that there are more than11 million records in more than 900 institutional repositories worldwide. Similar information was gathered from the Directory of Open Access Repositories (DOAR). Data covering the period up to 2005 from the *Institutional Archives Registry* also helped to complete the picture. The data was analyzed to determine the growth over the last decade or so with a view of determining its position on the Product Life Cycle and Diffusion of Innovation Curve as well.

A review of the data shows that institutional repositories are healthy and growing exponentially. It also shows that the majority of records contained within institutional repositories represent traditional research reports and scientific articles. It also indicates that e-theses and e-journals have been part of the repositories almost from its inception and is still part of the backbone as far as content is concerned. According to the numbers, demonstration material, teaching and learning objects, databases, and abstracting and indexing files are relative new types of content types. It is clear that the open access movement is contributing to the use and growth of institutional repositories but it is unclear what the future holds in this regard.

Institutional repositories are no longer just hype. It is a mainstream technology supported by the increasing popularity of the open access movement. Even so, institutional repositories need to be nurtured if its potential to enhance e-scholarship is to be realized.

1. Introduction

According to the statistics being kept by the Registry of Open Access Repositories (2009), institutional repositories (IR's), as a phenomenon, has been on the higher education scene since the early 1990's but has only enjoyed popularity since 2002 when the numbers started to increase significantly. Around that time it started to appear on the radar screen of many universities and many began looking into it as a possible alternative to the traditional scholarly publication model.

This coincided roughly with the rise of the open access movement when, in 1998, the Association of Research Libraries developed the Scholarly Publishing and Academic Resources Coalition (SPARC)(2009). This alliance of academic and research libraries and other organizations came into being to address to develop and promote alternatives to the traditional publishing model, such as open access because of the exorbitant price increases in academic journals over several decades.

In 2004 Cervone stated that: "Institutional repositories significantly extend the role of a library. Such projects are a serious and long-lasting commitment, with extensive benefits' as well.… If libraries step up to the plate, they will fundamentally transform their role from passive transfer agents of information into active partners in the dissemination."

2. Research questions

Specific question that will be answered in this paper are:

a) to what extent are IR's, as a method for communicating scientific and scholarly information, accepted or are they just hype?
b) how mature are IR's as a technology?
c) to what extent are IR's used by faculty and researchers?
d) how does open access fit into the picture?
e) does the evidence support growth or decline?
f) what kind of information is deposited in IR's
g) how are IR's distributed geographically?

3. Methodology

Almost 5 years ago Swanepoel (2005) published an article taking stock of the use and growth of institutional repositories. Four to five years may not sound like much but in the information technology arena it is enough to note significant changes since then. At the time the *Institutional Archives Registry* (2005) was used for data on current repositories. Figure 1 provides a graph of the growth of repositories from 1990 – 2005 as based on the data from this registry. This site is no longer in existence and a new similar source was used for the

latest data: *Registry of Open Access Repositories* (2009). The data from the new source

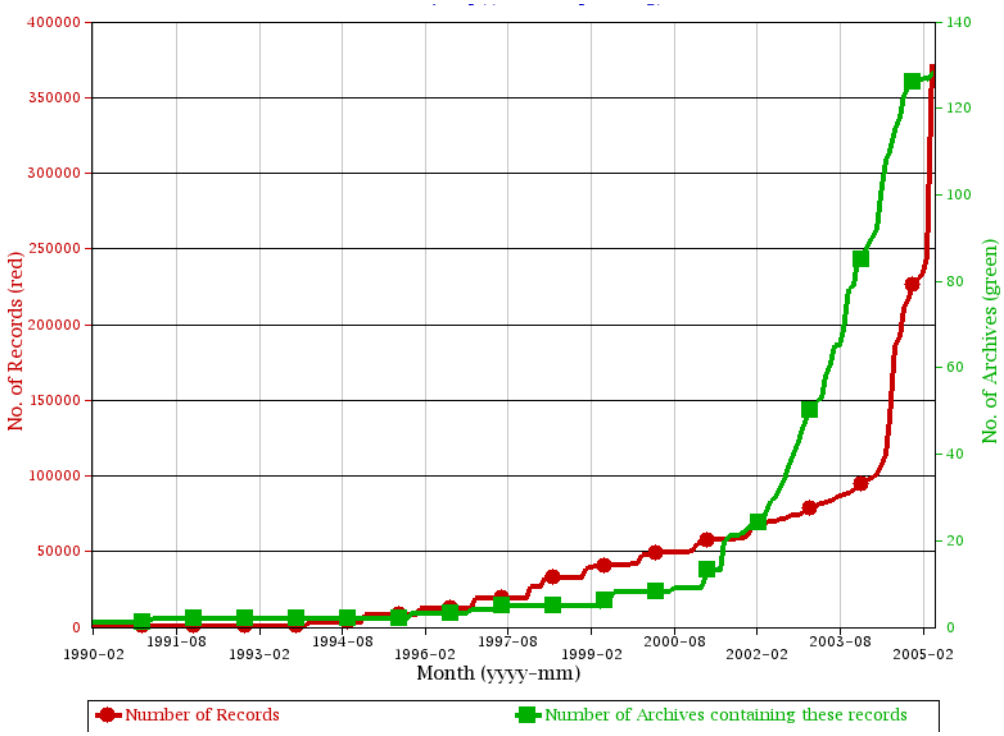



**Figure 1: Growth in Repositories 1990 – 2005 (*Institutional Archives Registry*, 2005).**

includes the period covered by the previous report and correlates with that of the previous site (see Figure 2).

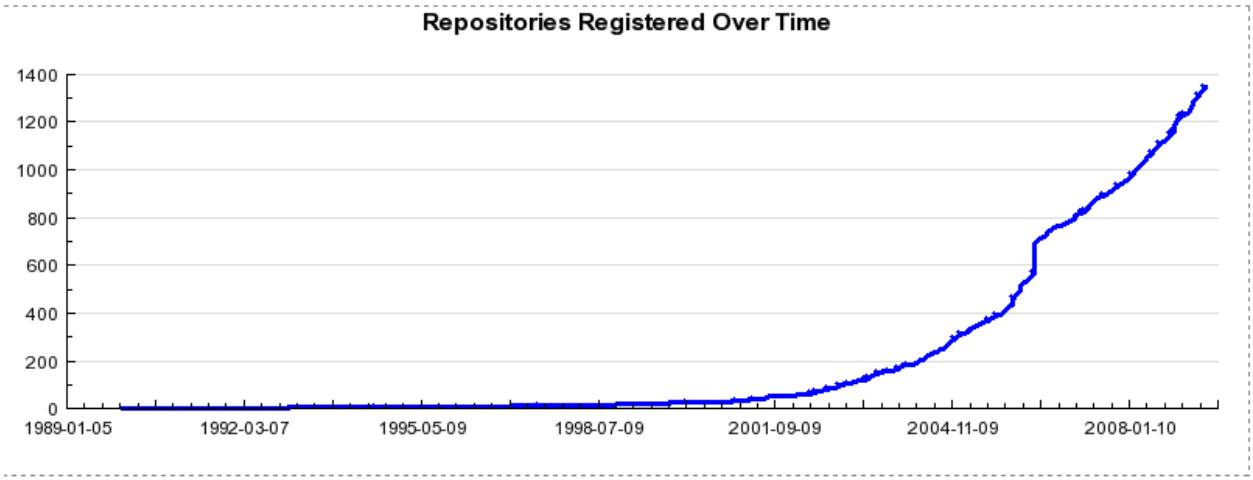



**Figure 2: Growth in Repositories 1990 – 2009 (*Registry of Open Access Repositories*, 2009)**

There is a third set of data that is very similar: the Directory of Open Access Repositories. The goal of this directory is described as: "The OpenDOAR service provides a quality-assured listing of open access repositories around the world. OpenDOAR staff harvest and

assign metadata to allow categorization and analysis to assist the wider use and exploitation of repositories. Each of the repositories has been visited by OpenDOAR staff to ensure a high degree of quality and consistency in the information provided." (*Directory of Open Access Repositories*, 2009).

The graph representing the growth of repositories from this is markedly different, less so because it includes both institutional and subject repositories but rather because the directory database was only started in 2006 and there was quite a backlog of repositories to be registered in the directory. From about 2007 the directory reflects real growth in the number of repositories registered in the database (see Figure 3.)

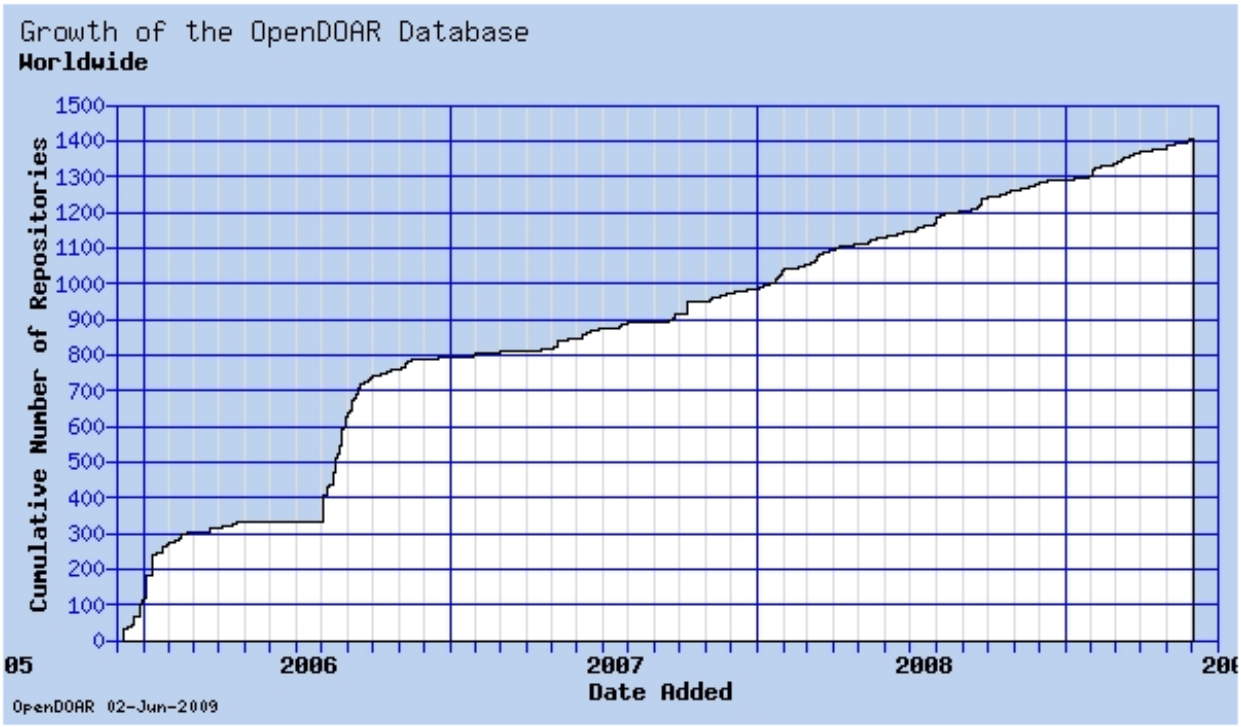



**Figure 3: Growth in the number of repositories registered in the OpenDOAR database *(Directory of Open Access Repositories*. 2009).**

In interpreting these three sets of data use is made of the Product Life Cycle (Sigmoid Curve), the Diffusion of Innovation Theory and the Gartner Hype Cycle.

4. Delimitations

IR's are part and parcel of the Open Access movement. However, for the purposes of this study, this aspect will not be discussed at length. Sufficient to say that Open Access is not entirely reliant on institutional repositories. On the contrary, there are a significant number of open access journals that supports the Open Access movement equally well. Waller (2009) recently reported that: "The Directory of Open Access Journals went over the 4,000 journal mark and now there is a note on the E-LIS home page that E-LIS now contains over

9,000 documents." E-LIS is a well-known subject repository in the area of Library and Information Science.

Nor is the scholarly communication process part of the investigation, even though IR's are part of the new emerging model of scholarly communication. Disciplinary or subject repositories (also referred to by Waller in the above paragraph) are outside the scope of this study as well. It will be limited to institutional repositories. The focus of the paper is on determining the extent to which IR's have become mainstream technology.

## 5. Technology models and their application

### 5.1. Product Life Cycle (2004) (Sigmoid Curve)

The Product Life Cycle or PLC provides a way to gauge the status of a product or service, from introduction to decline.

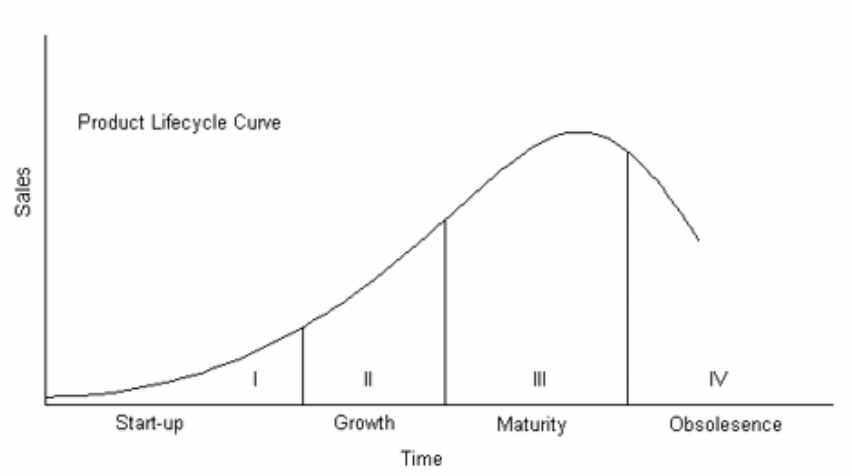



Figure 4: Stages of the Product Life Cycle (Product Life Cycle, 2004)

The four stages of a PLC can be described as follows:
a) Introduction or start-up is the period from a new product's commercialization until takeoff.
b) Growth is the period from a new product's takeoff until a slowdown in the growth of sales is experienced.
c) Maturity is the period from a product's slowdown until sales begin a steady decline.
d) Decline is the period of steadily decreasing sales until a product's demise or obsolescence.

Taking into account the growth in the number of IR's it would be reasonable to deduct that IR's are no longer a start-up technology. It is equally clear that it was not a mature technology in 2005. Therefore IR's must be in the growth phase of the Product Life Cycle, showing signs of reaching the early maturity in the developed countries in 2009. The estimate is that they have made significant progress in this phase but that the degree

of progress and adoption differs from country to country and even from institution to institution.

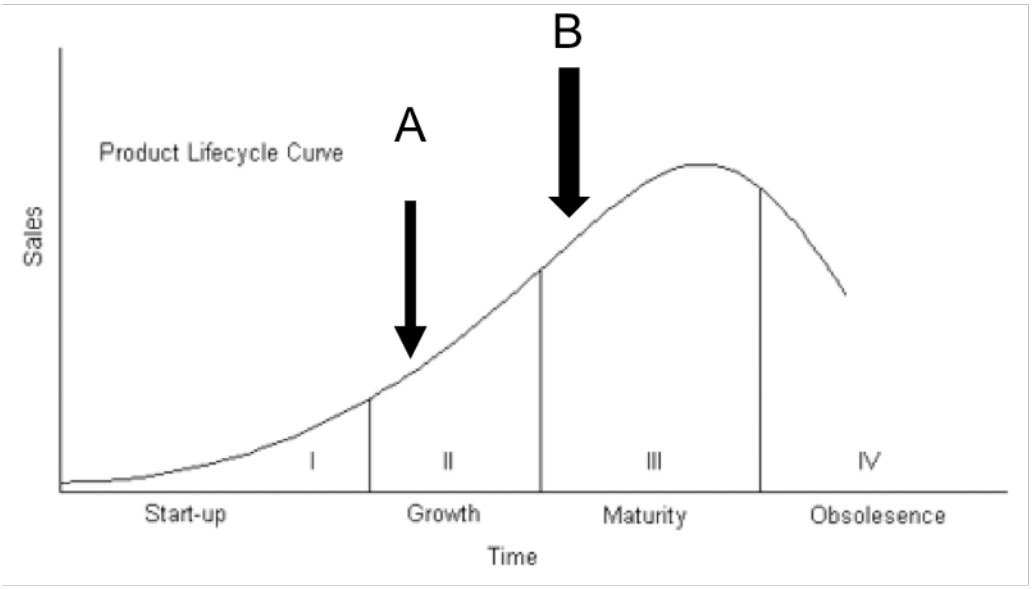



Figure 5: Estimated position of IR's on the Product Life Cycle in 2005 (A) and 2009 (B).

This point of view is supported by Google earth view, showing the distribution and size of repositories in the USA as shown in Figure 6.

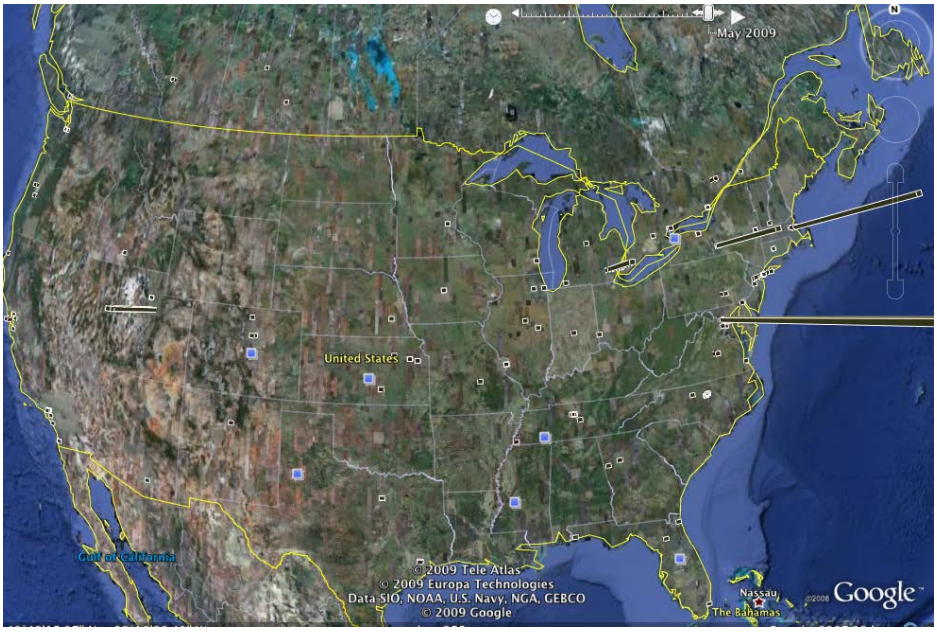



Figure 6: Google Earth view of IR's in the USA

5.2. Diffusion of Innovation Theory (Davis et al, 1989).

Figure 7 shows the bell-shaped distribution of individual innovativeness and the percentage of potential adopters theorized to fall into each category. At one end of the distribution curve are the innovators. The risk takers and pioneers are innovators who adopt an innovation very early in the diffusion process. On the other extreme of the continuum are the laggards who oppose adopting an innovation until relatively late in the diffusion process, if they do so at all.

Earlier research on behavior in general, and technology adoption in particular, has indicated that there could be a common set of determinants of behavior among different segments of the population, e.g., Davis, et al (1989). The Innovation Diffusion Theory shares this basic assumption. It also assumes that everyone will ultimately adopt (Rogers, 1995). These two assumptions are challenged. For example Moore (1991) suggests that people in each of the different adopter categories are different from those in the category to their immediate left. These differences across categories are referred to as "cracks in the bell curve".

This view suggests that innovations that succeed among innovators and/or early adopters may fail among the early majority or late majority since the innovation does not possess the characteristics that appeal to those in these later categories. Further, this implies that factors influencing different categories of adopters are fundamentally different. It also implies that since not everyone will adopt an innovation, it is quite likely that factors influencing non-adoption will be different from factors influencing adoption (Rogers, 1995).

The motivating forces or drivers that result in adoption of the technology by the different adopter categories are known and presented in Figure 6.

Technology adoption decisions have been typically characterized by a strong productivity or utilitarian orientation. Across the different categories, drivers related to the use-productivity possibility (e.g., perceived usefulness, relative benefit, job fit, etcetera.) have emerged as the strongest predictors of adoption (Rogers, 1995).

The role of utilitarian and hedonic outcomes is also supported by motivation theory. Motivation research suggests that there are two main classes of motivation: extrinsic and intrinsic. Extrinsic motivation pertains to achievement of a particular goal whereas intrinsic motivation is the enjoyment and pleasure resulting from a particular behavior (Twist, 2004).

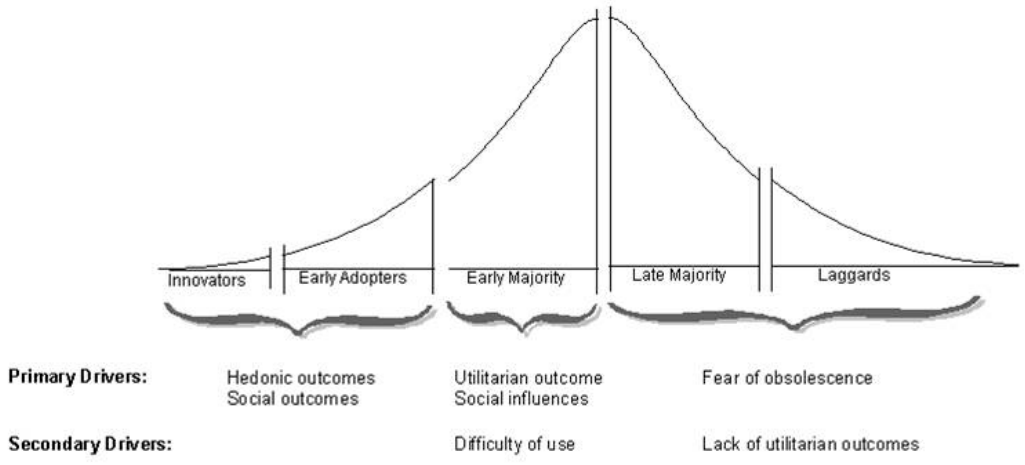



Figure 7: Diffusion of Innovation, mapping adopter determinants to adopter categories of Rogers (1995)

Social outcomes can be thought of as the social rewards (e.g., "public" acknowledgment) that would be achieved as an outcome of adopting an innovation. Adoption may lead to an elevation in power, knowledge, and/or status if the decision is viewed by others to be a good one (Vallerand, 1997). Prior research has emphasized the importance of social outcomes as a determinant of behavior. Similarly, innovation literature suggests that the desire to gain status is an important reason for the adoption of an innovation (Rogers, 1997).

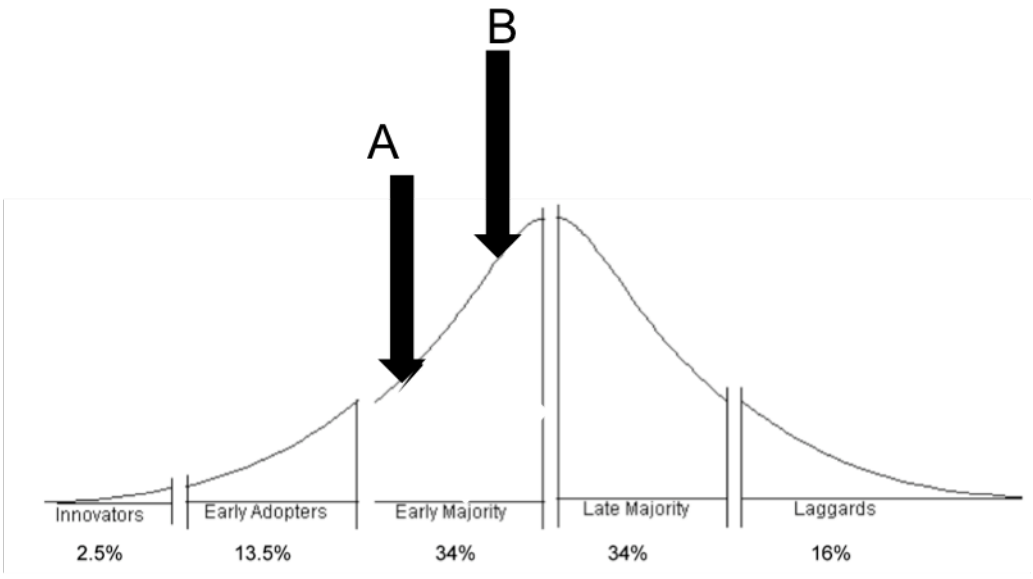



Figure 8: Position IR's on the Diffusion of Innovation Curve 2005 (A) and 2009 (B)

5.3. Gartner Hype Cycle

The Gartner Hype Curve is a tool introduced by the Gartner Group in the 1990's to explain general phenomena of interest and excitement (also called hype) in new technologies. This framework plots the typical evolution of a technology from its early beginning through its maturation to general market acceptance.

The different stages are (Gartner, 2004) (Twist, 2004):

- Technology Trigger: The first phase of a Hype Cycle is the "technology trigger" or breakthrough, product launch or other event that generates significant press and interest.
- Peak of Inflated Expectations: In the next phase, a whirl of publicity normally generates over-enthusiasm and idealistic expectations. There may be some successful applications of a technology, but there are typically more failures.
- Trough of Disillusionment: Technologies enter the "trough of disillusionment" because they fail to meet expectations and rapidly become unfashionable. Consequently, the press usually abandons the topic and the technology.
- Slope of Enlightenment: Although the press may have stopped covering the technology, some businesses persist past the "through of Disillusionment" and climb the "slope of enlightenment". Gradually, through continued experimentation they recognize the benefits and practical application of the technology.
- Plateau of Productivity: During this phase of a Hype Cycle, the real benefits of the technology are established and accepted. Tools and methodologies are more and more stable as they enter their subsequent generations. The final height of the plateau varies according to whether the technology is generally applicable or benefits only a niche market. Approximately 30 percent of the technology's target audience have or are adopting the technology as it enters the Plateau.

Gartner indicated in 2004 that IR's (called E-Learning Repositories by them at the time) were moving down into the Trough of Disillusionment. However the number of successful repositories to be seen in 2009 (Figure 1, 2 and 3) indicates that this technology has moved further along the curve than it seems at first glance. If the number of faculty participating at some individual institutions is taken into account, it suggests that this technology of IR's has moved further down the slope closer to the plateau. Some institutions are reporting a 25% participation rate (9) while 30% penetration (23) is seen as a technology already entering the plateau.

On the other hand, if the flood of literature on IR's that is still being published in professional literature is taken into account, it argues that the rightful position of IR's is still not far from the Peak of Inflated Expectation.

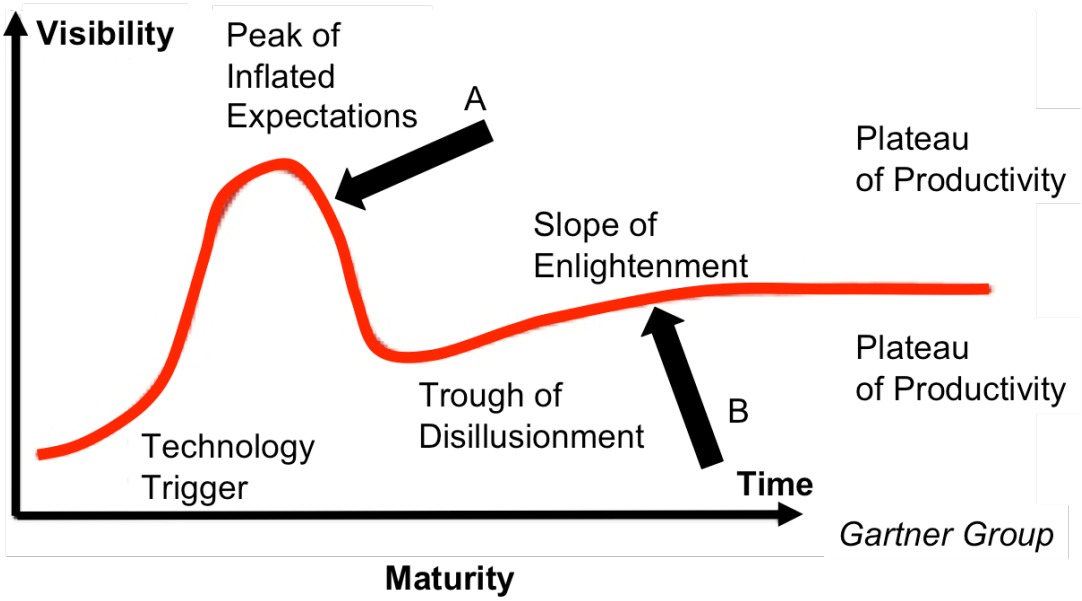



Figure 9: Position of IR's on the Hype Cycle in 2005 (A) and 2009 (B)

6. The questions answered

6.1. The acceptance of institutional repositories

It is quite surprising the number of publishers that allow self-archiving of articles by authors in some repository. A database of publisher policies is maintained by the SHERPA/RoMEO project (2009); of the 602 publishers included:

- 29% allow archiving of both pre- and post-print
- 21% allow archiving of post-print
- 11% allow archiving of the pre-print
- 39% do not formally support archiving.

However taking cognizance that (Mark Ware Consulting, 2006):

- There are about 23,000 scholarly journals in the world,
- They publish collectively around 1.4 million articles a year.
- The number of articles published each year and the number of journals have both grown steadily for over two centuries, by about 3% and 3.5% per year respectively.
- The number of researchers, which has also grown at about 3% per year now stands at around 5.5 million

6.2. The maturity of institutional repositories

In 2006 the University of Michigan conducted a nation wide census of IR's (Markey et al (2007). Questionaires were sent to library directors at 2,147 institutions in the United

States. A total of 446 (21%) participated in the census. More than half of the responding institutions (53%) had done no IR planning. Some 20% had begun to plan, 16% were actively planning and pilot testing IRs, and 11% had implemented an operational IR. The study confirmed some of what other surveys had shown about operational IRs. It found that most IRs have been created at research institutions and that few are found at master's or baccalaureate institutions; that the library usually takes the lead in planning, staffing, and paying for IRs; that faculty and graduate students are the major contributors to operational IRs, but that such contributions are still low; and that DSpace is the preferred IR software for both pilot testing and implementation.

At the same time, the census yielded new findings. For example, the authors concluded that there is a "sleeping beast of demand" for IRs from master's and baccalaureate institutions, based on the finding that, among those respondents who had not begun planning, half intend to do so within 24 months.

6.3. The use made of institutional repositories

The short answer is that there is some proof that it is being put to good use but there is a lot of room for improvement. The Michigan census (Markey, K et al. 2007) on IR's in the US found that 50% of the respondents who have implemented IR's have less than 1000 items in their repositories and only 19.4% of respondents more than 5,000 items.

Authors whose articles are available in open access repositories are more often being cited. This finding was reported by a number of researchers namely: Andrew Odlyzko (2000), "The Rapid Evolution of Scholarly Communication," (http://www.si.umich.edu/PEAK-2000/odlyzko.pdf) (accessed Jan. 3, 2008); Steve Lawrence, (2001) "Free Online Availability Substantially Increases a Paper's Impact," Nature 411 (May 31): 521; Alma Swan and Sheridan Brown (2004), "Authors and Open Access Publishing," Learned Publishing 17(3): 219–224; Kristin Antelman (2004), "Do Open-Access Articles Have a Greater Research Impact?" College & Research Libraries (September): 372–382; and Gunther Eysenbach (2006), "Citation Advantage of Open Access Articles," PLoS Biology 4(5) e157:0692–0698.

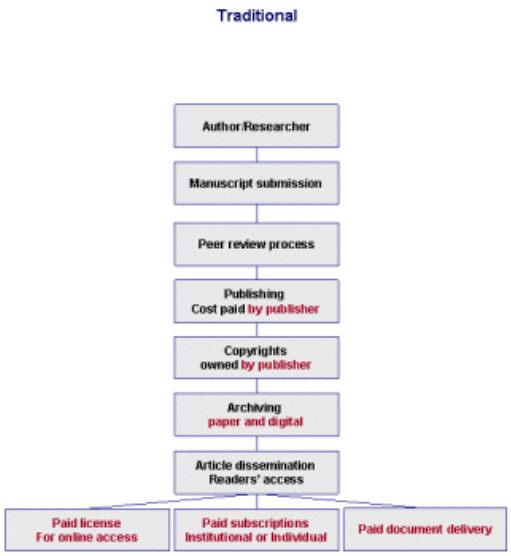
Even so some faculty do have concerns about self-archiving (Markey et al, 2007):

- Absence of campus-wide mandates regarding mandatory contribution of certain material types, e.g., doctoral dissertations, master's theses, faculty preprints
- Contributors' lack of knowledge about how they can benefit from IRs
- Convincing faculty that the IR will not adversely affect the current publishing model
- Contributors' concerns about intellectual property rights for digital materials

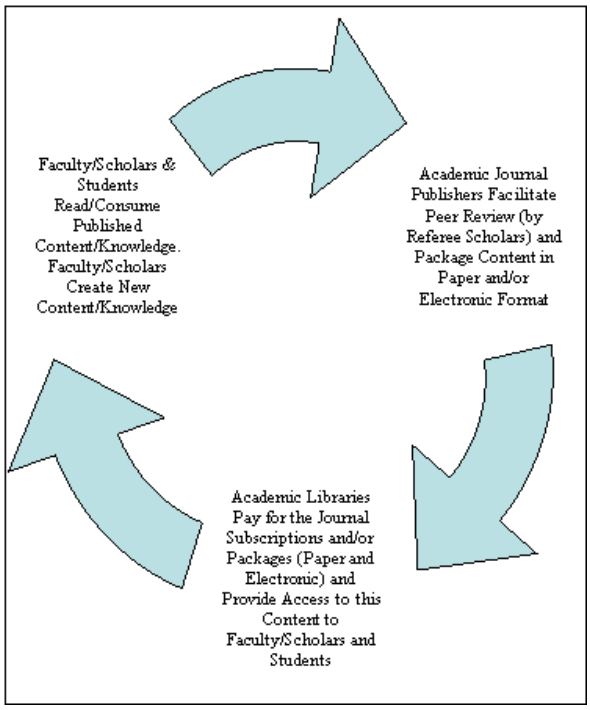
6.4. How institutional repositories and open access fit together.

There area about 2,000 publishers in existence today of which about 600 were approached to allow self-archiving, of these 60% complied in one way or another (Sherpa/Romeo. 2009). This means that 1,740 do not allow self-archiving.

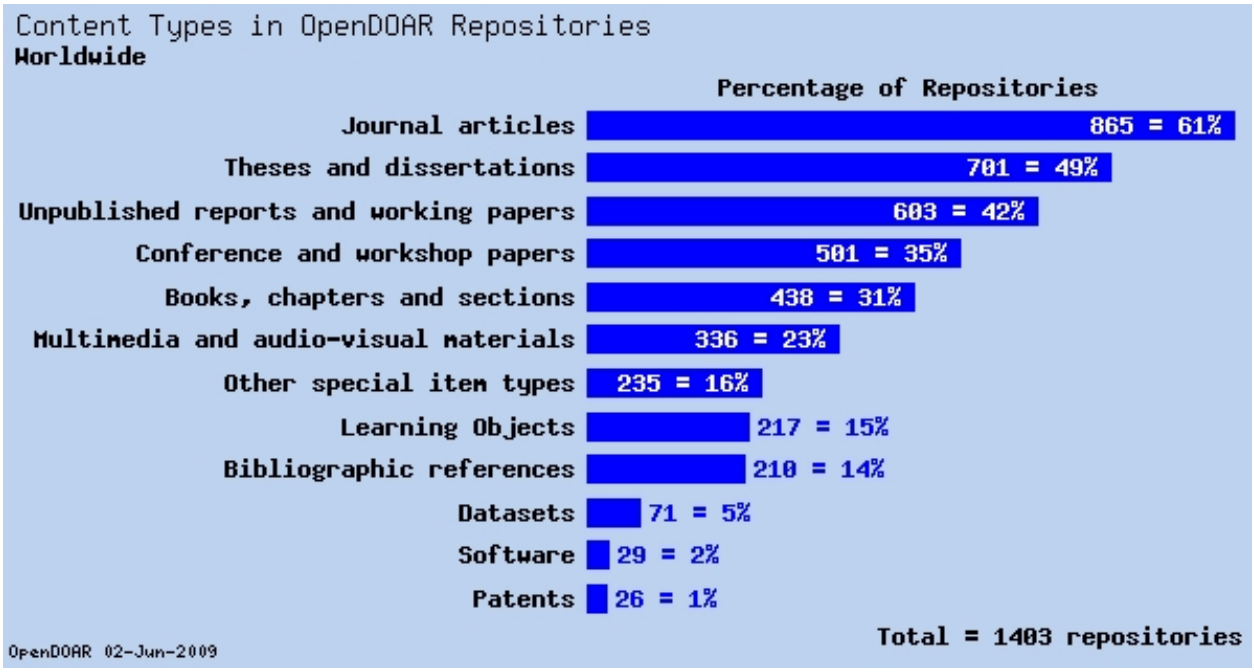
Kwasik and Fulda (2005) provides the following representation of traditional versus open access publishing:




http://southernlibrarianship.icaap.org/content/v09n03/mcguigan_g01.html




6.5. What kind of content is found in IR's



6.6. Geographic distribution of IR's?

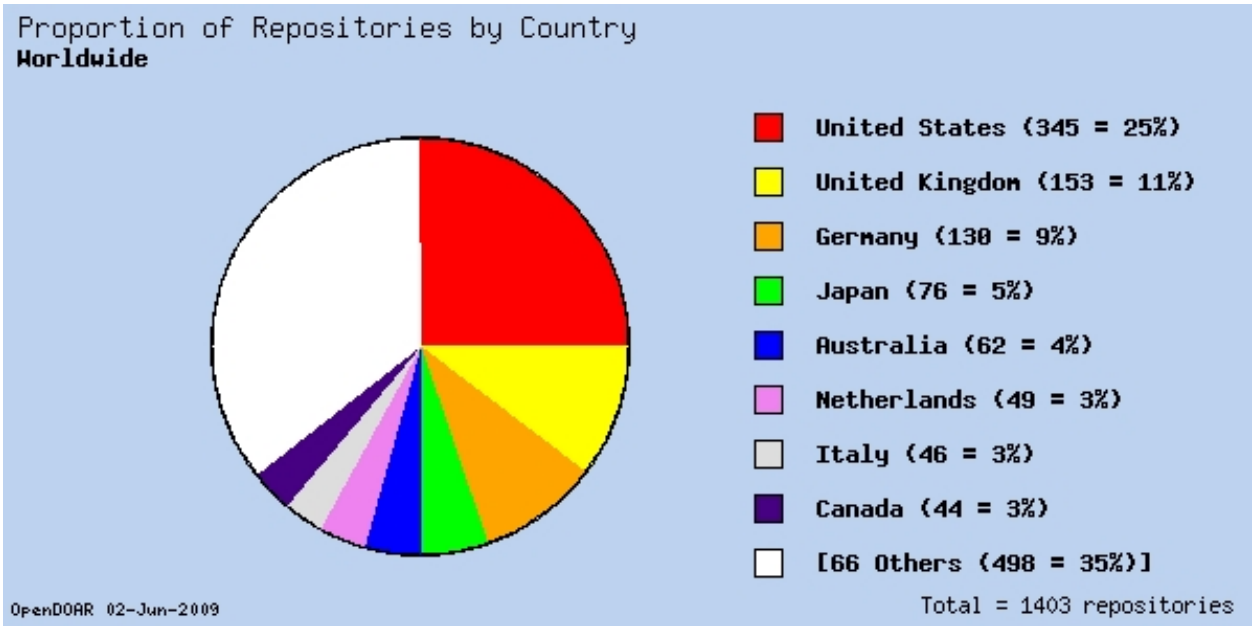
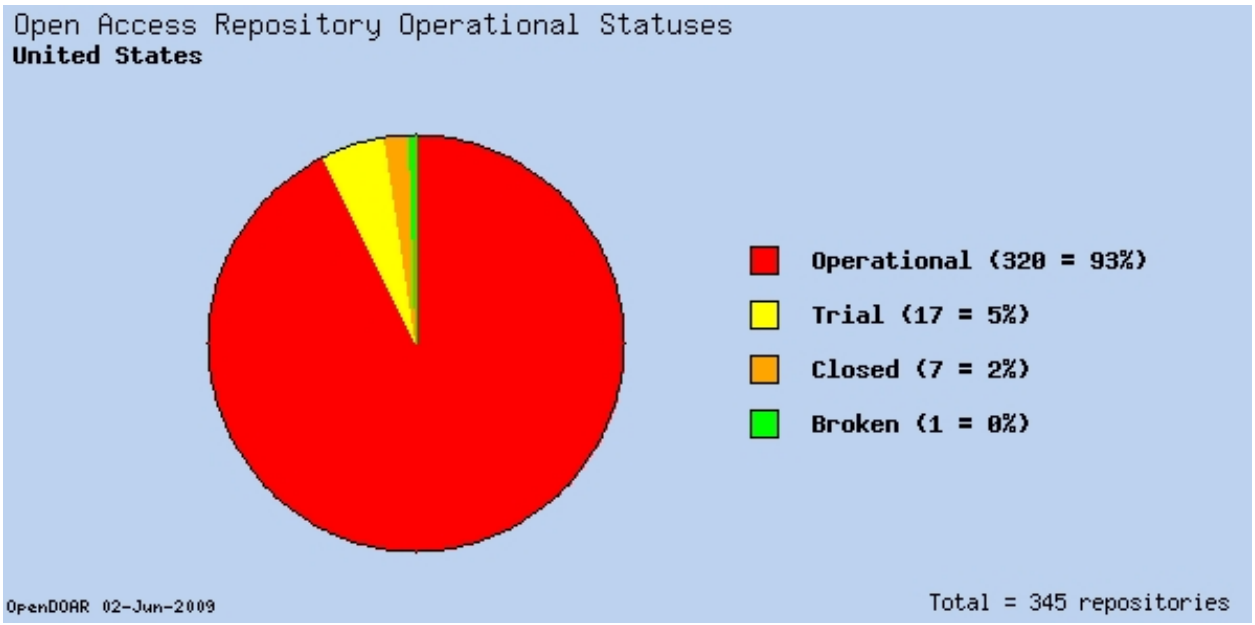



Figure 10: Geographic distribution of IR's (Directory of Open Access Repositories, 2009)




6.7. The future of institutional repositories

The Pew Internet study indicates that 28% of teenagers will contribute online content by blogging and 55% by creating a social networking profile for Gen Y individuals the respective percentages are 20% and 60%. These numbers are in sharp contrast to the 6%

and 16% respectively for young boomers (Jones & Fox, 2009). This information suggests that being comfortable with contributing online content is a generational issue and promises a bright future for IR's as the older generations vacate their places at institutions of higher education.

## 7. Conclusion

Lynch (2003) is of the opinion that IR's are not a challenge or alternative to disciplinary repositories; rather, they complement them, just as they can complement existing venues of scholarly publication. The bottom line is that there are strong indications that IR's are now mainstream and has become part of the infrastructure many academics expect from their institutions.

## Bibliography


Cervone, H.F. (2004). The Repository Adventure: On the way to changing scholarly communication, libraries may end up changing themselves. *Library Journal.* 129 (10), 44 – 46.

Davis, F. D. et al (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. In *Management Science*. 35, 982-1003.

*Directory of Open Access Repositories*. (2009). Retrieved May 25, 2009, from http://www.opendoar.org/index.html

Gartner Group (2004). Understanding Hype Cycles. Retrieved May 25, 2009, from http://www.gartner.com/hc/asset50595.jsp

*Institutional Archives Registry*. (2005). Retrieved April 25, 2005, from http://archives.eprints.org/

Jones, S & Fox, S. (2009) Pew Internet & American Life Project: Generations Online in 2009. Retrieved May 22, 2009, from http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx

Kawisk, H. & Fulda, P.O. (2007) Open Access and Scholarly Communication -- A Selection of Key Web Sites. Retrieved May 22, 2009 from http://www.istl.org/05-summer/internet.html

Lynch, Clifford A. (2003) Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL, no. 226 (February 2003): 1-7.

Mark Ware Consulting (2006). Scientific publishing in transition: an overview of current developments. Retrieved May 15, 2009 from, http://www.stm-assoc.org

Markey, K et al. (2007) Census of Institutional Repositories in the United States MIRACLE Project Research Findings. Retrieved May 15, 2009 from http://www.clir.org/pubs/abstract/pub140abst.html.

Moore, G. A. (1991). *Crossing the Chasm: Marketing and Selling Technology Products to Mainstream Customers.* Harper Collins: New York.

Product Life Cycle. (2004). Retrieved May 1, 2009, from http://www.witiger.com/marketing/PLC.htm

*Registry of Open Access Repositories*. (2009). Retrieved May 12, 2009, from http://roar.eprints.org/index.php

Rogers, E. M. (1995). *Diffusion of Innovations*, 4th ed., The Free Press: New York.

Scholarly Publishing and Academic Resources Coalition (2009). Retrieved May 15, 2009, from http://www.arl.org/sparc/about/index.shtml

Sherpa/Romeo. (2009) Publishers copyright policies and self-archiving. Retrieved May 22, 2009, from http://www.sherpa.ac.uk/romeo.php?stats=yes

Smith, K. (2008). Institutional Repositories and E-Journal Archiving: What Are We Learning? Retrieved May 16, 2009 from http://dx.doi.org/10.3998/3336451.0011.107

Swanepoel, M. (2005). Digital repositories: all hype and no substance. *New Review of Information Networking*. 11 (1), 13 – 25.

Twist, J. (2004). Peering beyond the technology Hype. Retrieved May 15, 2009, from http://news.bbc.co.uk/1/hi/technology/3577746.stm

Vallerand, R. J. (1997). Towards a Hierarchical Model of Intrinsic and Extrinsic Motivation. In *Advances in Experimental Social Psychology.* 29, 271-360.

Waller, A. (2009, April 23). E-LIS now has over 9,000 documents. Message posted to http://oalibrarian.blogspot.com/2009/04/e-lis-now-has-over-9000-documents.html