Open Access Repositories and the freedom of Information

By Roxana Theodorou

It's a well known fact that for the last two decades at least, academic (and not only) libraries are facing serious economic problems that do not allow them to serve their purpose to the fullest. Additionally to that, subscriptions to scientific journals have been rising to extraordinary levels, making the situation even more difficult. In fact, between 1975 -1995 the prices of subscription based journals have risen to 300% above inflation. After 1995 and until today, subscriptions continue to rise rapidly, although, to be fair, not as much as previously. It is, however clear, that library budgets cannot meet these financial demands. So, they try to face their problems, mainly by forming consortia that negotiate the prices with the publishers and try hard to address the needs of their users as best as they can.

But how did the publishing world became so tough? Until WW2 academic publishing was mainly in the hands of universities and learning societies. With the end of WW2 there was an ever rising interest in science and scientific information. Commercial publishers saw an opportunity there and were very keen and quick to invest, buying out small publishers or simply by founding new journals. Through time and after many mergers and acquisitions, today, scientific publishing is in the hands of a few, very important and very fierce publishers. And because of this new found oligopoly they have the power to decide and enforce the rules of the game.

Additionally, Academic publishing belongs to a special kind of market; it's a two-sided market. Two-sided markets, also called two-sided networks, are economic networks having two distinct user groups that provide each other with network benefits. Example markets include credit cards, composed of cardholders and merchants; HMOs (patients and doctors); operating systems (end-users and developers), travel reservation services (travelers and airlines); yellow pages (advertisers and consumers); video games (gamers and game developers); and communication networks, such as the Internet. Benefits to each group exhibit demand economies of scale. Consumers, for example, prefer credit cards honored by more merchants, while merchants prefer cards carried by more consumers. In two sided markets the act of creating a good and giving it to consumers cannot be separated. That, in fact, means that the intermediate has an additional power in deciding the price of the product.

Another paradox about academic publishing is the fact that the end user (namely the reader) rarely ever is burdened with the cost of the product. Libraries are those that have to pay for scientific journals and ensure that the users will have continuous and unhindered access.

Also, the authors of scientific content almost in whole belong to the institutions that buy their works in the form of published articles. But the fact is that part of the authors' salaries is for that purpose exactly: to research, write and publish their conclusions. So the institution ends up paying twice for the same product, which partly belonged to them in the first place.

In order for libraries to cope with the increasing prices of subscriptions they have formed coalitions that negotiate with the publishers in order to ensure lower rates and better terms of access. And although, at first there was some success, the prices are still too high, and with the help of secret deals with some of the partners of coalitions publishers have managed to prevail, once again.

Open access is a very popular term nowadays. There are many implementations of open access with the two most important being open access journals and open access repositories. Each implementation has its own advantages and disadvantages that academic libraries and institutions can examine and try and find the most efficient way to publish scientific information.

The goal is always to create a product that is cheap (or at least cheaper), easy to use, reliable and of high quality. Using the experience of the past and combining partnerships through coalitions and open access implementations, we now propose a new publishing system that could take libraries and institutions out of their very difficult situation.

The business model of coalition publishing is based on the idea of "returning" the dissemination and administration of scientific publishing to the hands of academic institutions and their corresponding libraries. The basic idea is this; institutions of the same scientific interest form publishing coalitions and create and maintain cross institutional repositories in which they publish (in electronic form only) all the scientific production of their members (and anyone else interested, as long as the works cover the given subject). But, in order for these repositories to be able to compete against subscription based journals their contents should be of high quality. That means that they should undergo some kind of selection policy, the best till now being, peer review.

Half of the existing open access repositories are subject based. And only 13% of those are cross institutional. But, till now, there are no certain selection policies enforced that ensure the quality of the contents. And there is no certain way to tell either, what *kind* of works are accumulated in them (theses, articles, preprints or post prints, teaching material etc). It would then be fair to say that till now, the contents of institutional repositories are not considered equally useful and credible as that of scientific journals.

Most academic and scientific libraries in their effort to develop and offer e-services to their users come to face a phenomenon known as competitive convergence. This actually means that competitive establishments, in their effort to obtain competitive advantage, come to use the same techniques and offer the same services in a similar way. The only way to break this vicious cycle is for a business to try and find the functions that make it unique in its market and try and use them to its benefit. In the case of academic libraries that could mean that they have to manipulate and use some the practices of their competitors.

A subject based repository has a starting advantage. Its content is fairly obvious. The user can automatically know which science is covered in any particular repository and not lose time and effort looking among information irrelevant to her research. But of course, thematic relevance is not enough. In order for a repository to be really of use to its users has to be exhaustive, inclusive, valid, and up to date. These characteristics can only be achieved through the active participation of as many authors as possible. But authors also need some incentives in order to submit their works in a repository, instead of an established scientific journal. The repositories should provide that:

- In the coalition take part scientific/research and teaching institutions acclaimed internationally at their respective fields. This will draw to the coalition smaller institutions which wouldn't otherwise attempt such an ambitious project. It will also give the guaranty needed to the authors for the credibility and longevity of the repository.
- The repository should include as many file formats and genres of works as possible. On the same time though there must be a strict selection policy enforced, that will clearly state what should and should not to be included in the repository. This two should be carefully balanced so that they do not to mutually cancel one another.

Much like traditional library coalitions, cooperative subject repositories have to be centrally controlled in order to function properly. But on the same time, the goal has to be the

creation of a flexible and effective cooperative union in order to survive in the world of scientific publishing.

Probably the most effective way to go about it would be to create a central management system that would encompass the basic logic of a P2P network. That way all of the participants will have equal access to and share resources, control and responsibility of procedures. P2P networks can be separated into 3 different categories.

1. Centralized (1st generation) networks. There is a centralized index server on which is stored information about the contents of the files that the users wish to share. Users search the index server and when the desired document is fount a link opens between the user and the owner of the file. (Napster and DC++ are examples of such networks)

2. Decentralized (2nd generation) networks. The philosophy here is completely different. Every system that participates in the network serves both as client and as server. As long as someone is connected to the network with the use of the appropriate software his existence is made known to a small group of connected computers, which in turn, make their existence known to a larger group of computers etc. The user can, this way, search for any information on the shared files between the connected computers.

3. 3rd generation. These networks have mainly characteristics of anonymity, like Freenet and Entropy. They are decentralized, and their philosophy is based, besides anonymity, on high viability, constant file sharing, and encoding, so that no one ever can take absolute charge of the network. This kind of networks is still evolving and has been characterized as small global networks.

Creating small and closed off repositories will not contribute much to scientific information. One main reason why repositories exit is that they function as a vivid, living, growing advertisement of the institution they belong to and an easy way to manage their scientific production. But a repository could not be only that, otherwise it is damned to wither and die.

It would be difficult to describe in such a short time the whole organization of the proposed model. But, in rough terms the concept is this:

1. Following the necessary negations institutions of the same orientation agree to form a publishing coalition. In order to maintain the cost low, publication should be only in electronic form.

2. The definitive structure of the coalition will not be discussed at this presentation but it would certainly have a basic managerial department that organizes the whole business, and two basic departments one for administrative and another for economic affairs.

3. The most difficult part of this project would be, of course, the funding. Funds could come from a variety of places:

    a. Advertisements, if the coalition desires it, there could be some advertisement space on the site of the repository that could in fact generate a steady income.

    b. Author payments. Authors pay a small fee in order to publish their works. Of course in order for author to be *willing* to pay, the repository should ensure that the quality of its services is very high.

    c. Member fees. All members of the coalition pay annual fees in order to sustain the repository. The fees depend on the size of each institution.

    d. Added value products. Although the publications of the coalition are strictly in electronic form, in order to lower costs, if a user wishes to obtain a printed and bind volume of, let's say, a collection of works, this service could be charged extra. Almost all institutions have a publications office that is already equipped to do such works, so it wouldn't be really difficult to organize such services.

    e. Public funds. Most scientific institutions are already funded (partly or in whole) by the governments of their countries. A small part of these funds could be reallocated to cover some of the expenses of the publishing coalition. If the project meets success, then the coalition could ask for additional funds (taking into account that subscriptions to scientific journal will be a lot less).

    f. Private funds. The coalition could ask for donations from private donors.

4. Depending on the structure of the coalition, it should be decided where and how the servers and the services are based.

5. Finally, comes the organization of the peer review process. In the case of cooperative subject repositories peer review can be enforced in two distinct stages,

before and after publication, intensifying the participatory process and upgrading the role of the reader to reviewer. More precisely, at first, groups of 2 or 3 reviewers (that come from the institutions that take part in the coalition) assess in a blind process the submitted works and decide whether it gets published or not, if there is any need for editing etc. After publication the review process is open to the public and the readers can place their comments on the site (of course the moderator of the site, always, retains the right to delete any comments that are insulting, irrelevant or in any way inappropriate). The process of open review can quickly transform itself into a need kind of citation method, especially if the users start to interlink relevant works and comments to the original work, making this way, new research more visible and creating a more intense impact for any published work.

But, what comes to everybody's mind when we talk about publishing (or any other kind of business for that matter) is how much does it cost? Publishing in a repository, even if we take into account peer review, is very much cost effective.

The cost can be broken down into many categories. There is the cost of the installation of the software, the cost of any customization needed, salaries for the staff, functional costs etc.

According to S. Gibbons (Gibbons, 2005), installation and maintenance cost of a repository can vary dramatically, depending on its size and orientation. It depends on the software chosen (whether it is open source or commercial), the number of people working for the repository, the equipment that will be used etc. If the institution chooses to join SHERPA then the cost of mere installation (the cost of the server and man-hours) comes up to €4.300. Queen's University, QSpace, cost for the organization and setting up, €37.000 (including, equipment, customization and salaries). It also costs €37.000 / p.a. to maintain the repository.

Presumably, the repository with the largest setting up and maintenance cost in that of MIT, which is considered the most advanced and complicated of its kind. Setting up MIT's DSpace cost €1.307.000, but we should take into account that this figure includes the cost of the development of DSpace, an open source software used globally by a large number of users and institutions. The annual maintenance cost of the repository is a lot less, of course, and it totals €207.000. This includes all the

costs of the repository (hardware, software, functional costs) and all the salaries and insurance costs of the staff.

The repository of the University of Rochester cost around €145.000 to set up, with major customizations. At this point, the repository has more than 50 collections depending on the subject and the department they serve.

Houghton, on his JISC repost on 2009 determined the total cost of an article from the time the author captures the concept until it gets processed from a library and is finally available to the users. Houghton calculated this cost to be €10.600. From that figure only a small part burdens the publisher. And depending the form of publication the publishing cost may vary. In the case of subscription based journals the publishing cost is €2.580 (for e-only publishing), and for open access journals the publishing cost is €1.682. In the case of open access repositories (and if a paper is submitted for publication only once) the publishing cost is even lower, although it very hard to determine exactly.

But what can be determined is the cost per article for the system of higher education depending on the form of publication. For subscription based journals the cost for HE is €9.160 / per article, for open access journals is €8.262 and for OA repositories €8.262. It should be noted that this figure includes the peer review process. The cost differences are very big and should be taken seriously into account.

But a project as ambitious as this could face a number of problems:

1. there could be difficulties for different institutions to reach agreements
2. There could be lack of funds for the initial stages of the project.
3. Members of the coalition may disagree in the use of standards and procedures.
4. The larger and more robust members of the coalition may try to enforce their priorities to the other members.
5. If there is no robust business plan and clear priorities the project could reduce itself in mere vision.
6. Commercial publishers will oppose to such an effort and try to boycott it in any way possible.
7. Lack of cross-functionality between different applications.

But there is also some true potential to such a project:

A cross institutional repository is economically viable and cost effective. It is not just easy to maintain but also, if it functions properly and grows in time, it could save serious money from cancelled subscriptions.

Researchers themselves seem to be every day more eager to participate in open access mediums. Open access journals are growing and repositories are being constantly set up in all kinds of institutions. On January 2008 there was an unprecedented movement of support of Open Access. The European research council and the US National Institute of Health officially adopted their open access mandates. But that was not the end of it. 8 other very important organizations actively supported open access. Of course, these movements are not the only ones. All around the world, publicly funded research bodies are supporting open access, either by setting up repositories, either by encouraging researchers to submit their works in open access mediums or by adopting open access mandates.

The advantages of OA are many, especially when high standards of quality are achieved. Access to scientific results becomes easier and quicker. This saves time and money not only to researchers themselves, but also for their funding bodies, institutions and libraries. On the other hand, open access articles are more likely to be cited compared to the subscription based model. This means increased visibility for the work and the authors and their institutions.

Maybe it is still too early to come to any definite conclusion as to how open access is going to be in the future. But for the time being it seems that repositories will play an important part in scientific communication. Open access is not going to be the only publishing alternative in the near future. But its popularity can be a force for changes. Coalition publishing cannot completely overturn the present status quo, which exists and serves scientists and scientific information for nearly 350 years. But, it can, and probably will be, the excuse and means needed in order to improve the system to the benefit of society.