Institutional Repositories

in scholarly communication:

a literature review on models,

issues and current trends

Pietro Gozetti

SUMMARY

1 AIMS, OBJECTIVES AND METHODOLOGICAL APPROACH

The main aims of this literature review are:
- Report on relevant sources about IR, and some references about the environment they came from.
- Give an overview concerning causes, consequences and impact of IR application in the scholarly communication channel.
- Understand current trends in changing scholarly communication models through IR
- Provide a critical overview about benefits, but also obstacles, problems and issues that need to be faced in developing IR.
- Earn deeper understanding on the role librarians play in the implementation, management and advocacy of IRs.
- Draw a scheme of possible areas for further studies in this field.

1.1 Scope

This work takes into account the relevant English-spoken literature produced at all 2005, and especially in the period 2002/2005, since discussion about IR came out after open access movement and the release of a technical framework concerning metadata, harvesting, networking etc.

This work is focused on IRs for scholarly communication, so discipline-based repositories or IRs application in other fields such as learning are not included into this literature review. This report doesn't take into account specific applications, softwares or project, since their analysis and comparing would have been too ambitious for the economy of this work: significant projects were considered where they provided relevant contribution to development and establishing broader and applicable models.

The literature review is structured according to major themes discussed in selected contributions: after a general introduction about the background environment, discussion will be made about definition of an IR; benefits, possible changes in scholarly communication; models/management/issues (marketing the IR to the authors, encouraging the authors to participate by adding their materials, and addressing costs associated with keeping the IR running); IR outcomes/impact; librarian's role: involved in all previous areas.

2   INTRODUCTION – THE SCENARIO

In the last 30 years, scholarly communication has been affected by the so-called 'serial crisis': scholarly journals' prices have increased more rapidly than research institutions' budgets, forcing libraries to cancel subscriptions, reducing access, readership and circulation of scholarly output, which caused further prices increase. As reported by Carlson. ARL statistics estimate that journal subscription rates have gone up an average of 8.5 percent per year since 1986, while library budgets have increased 5.6 percent per year (Carlson, 2002).

Deals with providers and library consortia checked this vicious circle but didn't erase the crisis. Instead, dissemination of technology, networking systems, which seemed to lighten costs of managing, storing and disseminating scholarly content, didn't resolve the situation, since publishers retained a monopolistic position and libraries had to dedicate increasing budgets for managing and purchasing access.

This situation is well described by Peter Suber, who talks about 'permission crisis' (Suber, 2003), relating to financial and technical barriers to access (such as subscription fees and login/password requirements) risen by commercial publishers. A complex debate and ferment developed as in the meantime open access movement were spreading as well. Open access literature demonstrates consciousness of need for a distributed model in order to achieve dissemination, since in a world where information grows very much, no one institution could grant storing and access to all content.

We won't pay much attention on the open access movement itself, since it's not the core focus of this study, but we recommend a giant and up-to date bibliography by Charles Bailey[1](Bailey, 2005a). What we are interested in stressing is that this state of mind influenced governments and international research institutions' consciousness for alternative dissemination models in scholarly and research output: key texts as the Santa Fe Convention on the Open Archives Initiative (October 21-22, 1999) (Van de Sompel and Lagoze, 2000), the Budapest Open Access Initiative, the Bethesda Statement on Open Access Publishing (April 11, 2003) and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (20-22 October 2003, Berlin), which establish the will to promote free dissemination and availability of knowledge, are indeed milestones of open access philosophy, and a ideal basis that helped in developing a completely new paradigm.

Authors found out alternative (although not substitutive or competing with publishers) ways to disseminate their research output among their communities, through at least 3 main models which today seem to be stable and acknowledged:

- open access journals,
- discipline-based repositories and
- institutional repositories.

The first two models owe a lot to the past: open access journals are the free version of e-journals (an attempt to bypass editorial costs transposing scholarly content into a new medium), while subject-based repositories emerged in communities already accustomed to fastly share preprints even prior to the Internet (Bjork, 2004), especially in the scientific and technical sectors (Tennant, 2002): arXiv, the open archive for physics, was born in 1991, while CogPrints, the Cognitive Science Eprints Archive, was launched in 1994 (Yiotis, 2005).

New paradigms raised as people learned how to use the available technology to create some completely new ways to perform communication, in order to create networks with existent

---

[1] The Open Access Bibliography is also integrated by the Open Access Webliography, which lists several open access related websites.

resources: Waaijers talks about "libratories", indicating an environment which combines library, repository and collaboratory initiatives (Waaijers, 2005).

Several initiatives tried to find a solution to the above-mentioned gap, promoting alternative publishing models, the first of them, SPARC (Scholarly Publishing and Academic Resources Coalition), was promoted by ARL in 1998. SPARC, an alliance of universities, research libraries, and organizations which coordinates projects devoted to relevant information dissemination in a networked system, endorsed and relied on open access journals and more recently on IR as a way to liberate peer-reviewed content through the Internet (Prosser, 2003).

3 INSTITUTIONAL REPOSITORIES: DEFINITION AND OBJECTIVES.

3.1 Definition

Institutional repositories made their first appearance in literature in 2002 with the publication of the SPARC position paper on Institutional Repositories, in which Crow defined institutional repositories as

"digital collections capturing and preserving the intellectual output of a single or multi-university community" (Crow, 2002).

This definition, broader and highly applicable, is greatly acknowledged by other authors which almost always refer to Crow (Johnson, 2002). A narrower but valuable definition is provided by Lynch, which defines IRs as

"a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institutions and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution." (Lynch, 2003).

This last definition tends to exclude IR for other purposes but stresses the role of an IR on dissemination of content, and, as most of the authors, seem to emphasize preservation, as the first step to guarantee dissemination through time. Clearly Lynch don't look at an IR as to a simple collection, but as a complex system of services that provide added value to the collection itself.

Key elements and requirements to define an IR are recurrent in the literature: even if in literature reading emerged that different authors use different names and approaches, depending on their background, who writes here recognized in several papers the same distinctive characteristics as they were introduced by Crow, Prosser, Johnson and Ware, who first wrote about IR. Here in brief are explained the essential elements of an IR:

- Institutionally defined: it means that an IR is the natural extention of history, research, prestige and impact of an institution or a group of institutions, as consortia.
- Scholarly content: Bailey (Bailey, 2005a) pays attention to variety of materials that IRs can contain:

   "An institutional repository includes a variety of materials produced by scholars from many units, such as e-prints, technical reports, theses and dissertations, data sets, and teaching materials. Some institutional repositories are also being used as electronic presses, publishing e-books and e-journals" (Bailey, 2005a).

Cervone and Gelfand also remarks that much of data contained in IRs is high value grey literature, that otherwise couldn't be published. A content management and document version system infrastructure, including flexible policies to contribute, approve, access data, is the framework which content is set in.

- Cumulative and perpetual: data can't be cancelled (if not illegal or bad science materials) and a policy for preservation/access must be undertaken.

- open and interoperable: IR supports standards and protocols for managing metadata, with few/none barriers to access.(Johnson, 2002).

3.2 Aims and objectives of IRs

Crow also gives a good definition about aims and objectives of IR, while examining the strategic roles IR have for research institutes (Crow, 2002), claiming they

> "provide a critical component in reforming the system of scholarly communication-a component that expands access to research, reasserts control over scholarship by the academy, increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to the institutions and libraries that support them".

According to Johnson, while traditional publishing model limits readership, obscures institutional origin, costs much, the new model implies no monopoly, increase of output, awareness (Johnson, 2002).

This aim also complies with the Freedom of Information Act, which establishes the right for everyone to access information held by a public authority and implies for each university to have an up to date publication scheme and a "digital records management system" (according to Barton et al., 2003) which allows universities to have knowledge and control on their whole publications and make them available: as Johnsons and Bailey (Bailey, 2005) remark as well, literature before published on personal or departmental sites can now be hosted on permanent repositories, since metadata standards, networking technologies and interoperability protocols are now enough widespread and reliable. The new scholarly communication model drawn by faculty members is horizontal, disaggregated and unbundles different functions of scholarly communication model (four functions indicated by Crow in 2002 and Prosser in 2004: registration, certification, awareness, archive).

Both Prosser and Crow (as many authors which cite them) retain IR can serve as

> "indicators of a university's quality […] to demonstrate the scientific, societal, and economic relevance of its research activities, thus increasing the institution's visibility, status, and public value".

The author stresses that these objectives can be achieved without competition, rather in a complementary way, with traditional commercial publishing and that this disaggregated model can be very economic for those institutions which can't afford huge technical investments, if self-archiving, effective copyright policies and encouragement to improve and increase research outputs from faculties are well set up. As a result, sensible changes can be gained without altering the financial or technical resources, but reallocating and reorganizing them maintaining advantages, as retaining intellectual property for authors and increasing research output's use and prestige without paying subscriptions.

One of the aims for developing IRs, as emerges from literature and especially in the clear paper by Ginsparg, is the will to create a network of collections which represent the best output of different institution in a distributed model (Ginsparg, 2000). In order to create a huge critical mass of content and a network of institutional repositories, interoperability is a key issue, since it allows disaggregated repositories to connect in a network of resources which maintain their originality and peculiarities, as Johnson claims. Crow also makes a smart analysis considering IRs as an application

of the role of creator of information by academic institutions and on the other hand as a propulsive initiative to significantly change future dynamics in scholarly communication.

Stevan Harnad, one of the most passionate advocates of open access, has outlined five broad aims of institutional repositories: self-archiving, management of digital collections, preservation of digital materials, housing of teaching materials, and electronic publishing of journals and books. In literature, however, self archiving is perceived as a way to obtain content recruitment (as we'll see later), while preservation has supporters and detractors (and as self-archiving, is perceived as a way to grant long-term access and provide value added service which only publishers at present can guarantee, rather than an aim itself).

3.3   Benefits

Crow remarks how progress in research relies largely on the amount of available information: retrieve more (and more quickly) relevant research means improved scholarly communication, research and teaching (Johnson, 2002). Traditional advocates, as Crow and Prosser (Prosser, 2005), stress the contribute of IRs in creating a new scholarly publishing paradigm which reveals weaknesses of current model and unbundling its traditional components will gain more economic advantages. Below we synthesize how publication elements will be redefined according to these authors:

- Registration (claiming the intellectual priority of a research) will be addressed by authors instead of publishers;
- Certification (quality evaluation and inner validity) will still be addressed by academic referees but sponsored by institutions;
- Awareness (dissemination and accessibility of research) will be addressed by librarians through IRs and related search tools instead of advertising and research tools sponsored by publishers;
- Archiving (preservation) will remain a library/institutional duty;

Prosser and Johnson stress the benefits of gaining broad access, dissemination and federated research for each institution, better impact for research, visibility for institution, advertisement, funding, measurement of impact for each institution through centralization of content (Prosser, 2003), (Johnson, 2002). These authors, endorsed by Gelfand, stress how centralization of institutional content can add "prestige and visibility to resources that without this institutional affiliation may not have peer review, be available digitally and thus remotely, and have perpetual access" (Gelfand, 2005). Centralization empowers dissemination, which emerges as strictly interconnected with interoperability: IRs are searchable through search engines, since digital objects are described by metadata, which can be harvested by external system through the OAI-Protocol for Metadata Harvesting[2]. Interoperability is a recurrent theme in the literature, and almost all the authors which talk about it (above all, Harnad) refer to Open Archival Information System (OAIS) Reference Model, and specifically to OAI-PMH as the protocol for making IRs connected [3].

---

[2]   At present best known OAI-PMH search system is the cleverly named OAIster: http://oaister.umdl.umich.edu/o/oaister/

[3]   Another known standard is the Metadata Encoding and Transmission Standard (METS) [http://www.loc.gov/standards/mets]. Other organizations involved in standards and repository design and operations include the Digital Library Federation [http://www.dlf.org], Coalition for Networked Information [http://www.cni.org], OCLC [http://www.OCLC.org], RLG [http://www.rlg.org], the electronic theses and dissertations program at Virginia Tech [http://scholar.lib.vt.edu/theses; http://www.thesis.org/standards/metadata/current.html], and Creative Commons [http://www.creativecommons.org].

Broader access means lower barriers to retrieve information: this doesn't mean that licensing and copyright agreements don't play a relevant role in managing IRs' content; Bailey clarifies the relationship between IRs and open access, relating to content and licensing issues:

> "IRs (as well as other digital repositories) are not typically pure "open access" repositories. Rather, they contain digital materials that have a mixed bag of copyright or license terms, and, generally, there is free and unrestricted access to these materials [...] Open access advocate Harnad […] has argued that the free vs open access distinction is "both spurious and a retardant on progress toward free/open access" and that open access should simply be defined as "free, immediate, permanent access to refereed-article full-texts online"" (Bailey, 2005a).

One of the major benefits perceived by a strong percentage of authors is that IRs, accordingly with their own policies, can hold several kinds of material, from peer-reviewed articles to reports, datasets, audio, video, learning objects, raw data, since technology allows it. IRs give scholars the opportunity to share not only formal publications (for example peer-reviewed articles), but also all that contributions which are gathered under the name of grey literature and which are the most difficult to find and preserve over the long period: Gelfand reports that grey literature can earn legitimacy, even if literature hasn't treated deeply this issue until now.

Grey Literature 1997 Luxembourg Conference defined grey literature as "that which is produced on all levels of government, academics, business, and industry in print and electronic formats, but which is not controlled by commercial publishers" (Gelfand, 2005). Gelfand remarks that "grey literature differs from commercial publications in that it is not based solely or even principally on an economic model, but rather on a communication model which we also now describe as scholarly communications" and that its uncertainty in identification, bibliographical description and retrieval justify why it brings several problems in collection development and management (Gelfand, 2005). Grey literature have great consideration in the new scholarly communication paradigm, since it covers almost 60% of content in IRs, according to Ware (2004).

Crow and Johnson (Johnson, 2002) agree in stating that IRs, centralizing all output by the same institution instead of disseminating it through several journals, can also serve as measurable indicators for the inner institution's academic value regardless in which the form the research is presented in, making fundraising more easy. Within the DAEDALUS project, as an example, scholars were able to list their publications for curricula simply downloading their deposited researches records into a simple reference manager (Ashworth et al., 2004). Quality control is indeed one of the most debated arguments, and many authors think that only in the future will be possible to find a rigorous alternative to peer review: but if new publishing paradigm doesn't rely on subscription, who will pay for reviewing materials? Some think it can be done on a voluntary basis (authors must be reviewed by other institutions' scholars to maintain equity), some retain that citation ranking will be the future of peer review, considering that open access literature has more chances to be read and cited by scholars than peer-reviewed articles rather than relying on prestige of journals, scholars, institutions and research can gain much more benefits by considering real use and appreciation of contributes even if they're not published on prestigious papers. As Crow claims:

> "Research has demonstrated that, with appropriate indexing and search mechanisms in place, open access online articles have appreciably higher citation rates than traditionally published articles. This type of visibility and awareness bodes well for both the individual author and for the author's host institution. Additionally, value-added services such as enhanced citation indexing and name

authority control will allow a more robust qualitative analysis of faculty performance where impact on one's field is a measurement. The aggregating mechanisms that enable the overall assessment of the qualitative impact of a scholar's body of work will make it easier for academic institutions to emphasize the quality, and de-emphasize the quantity, of an author's work.[53] This will weaken the quantity-driven rationale for the superfluous splintering of research into multiple publication submissions. The ability to gauge a faculty member's publishing performance on qualitative rather than quantitative terms should benefit both faculty and their host institutions" (Crow, 2002).

A undeniable benefit of IRs, according to Crow, is the opportunity of cancelling subscriptions for journals an institutions produces for, since scholarly output would be available in the repository. This also means that authors shouldn't let go their copyright to publishers and universities shouldn't pay anymore for obtaining from publishers the research output they previously funded. Rowland claims, however, that decreasing subsciptions will be a long and not catastrophic process, as IRs and disciplinary repositories haven't yet replaced traditional publishers' journals in ten years (Rowland, 2005).

Bjork states that IRs can be devoted to find alternative marketing channels for universities (Bjork, 2004) and for example, according to Shearer, IR can be useful for helping developing an effective and economic scholarly communication in the developing countries.

Harnad remarks also how publication can be a more work-in-progress process, interconnecting pre-prints with their corrections, revisions and updates: this means progress and ameliorating productivity. Relevance of research will be defined by new "scientometric" indicators, which are based on use and citations. Costs are lowered, due to open source and customizable softwares (Harnad, 2001a).

4   INSTITUTIONAL REPOSITORIES MANAGEMENT AND ISSUES

4.1   Costs

Crow remarks how implementations, projects and initiatives have been diverse, making difficult to outline a universal economic model: costs for developing and running can vary from no cost (for institution which reallocate resources from other previous services), to hundreds of thousands of dollars (for institutions which invest in technical and staff resources). IRs require a huge effort, both in the management/policy system (choosing and implementing content management/recruitment, staff to dedicate, training staff, marketing the IR and facing licensing agreements), and in the technical environment (choosing metadata, infrastructure, software, customization) (Crow, 2002).

Bailey and Gibbons warn that open access doesn't mean that IRs are costless: some open access advocates, focusing on technical costs of IRs, can perceive IRs as cheap to support and quick to implement while librarians may tend to consider additional costs, such as staff and user training and support, IR advocacy and promotion, metadata creation and maintenance and long-term digital preservation (Gibbons, 2004b; Bailey, 2005b). Harnad also points out how costs haven't vanished, especially for quality control and validation, but that this cost takes about 10% of traditional publishing process' total cost.

Drake points out the long-term perspective which involves technology and policy sustainability:

> "Maintenance of content, software, and accessibility can change. IT staff and librarians need to know the consequences of changes in hardware, software, and standards and be able to adjust accordingly. Repositories cannot be sustained without long-term infusions of funds. Everyone involved in a repository needs to understand that the project has become part of their everyday lives and will require attention and funding in perpetuity. Too often managers in corporations seem unable to look beyond the quarter's bottom line and shy away from long term commitments. Their reluctance to commit funds is exacerbated in an uncertain economy. Many managers in academe emulate their corporate colleagues through their reluctance to raise and dedicate enough money to ensure that the repository is funded at an appropriate level forever" (Drake, 2004).

4.2   Implementation

Literature was full with implementation models and case studies about development and outcomes of local IRs: it would have been too difficult and expensive to compare all practical initiatives, which are not the object of this study of course. Different softwares, policies, kinds of institutions, user needs make repository's environment very intricate and scientifically incomparable: a study in this direction is inappropriate in writer's opinion, since the literature clearly reveals that IRs have more to do with creativity rather than with an impersonal model to be adapted aprioristically.

At this time, best practices seem to guide the implementation of repositories: experiences are all different and rooted in each institution's environment, so it's difficult to outline a universal

model. Each institution can choose its own policies, taking advance and reusing other repositories' experiences and outcomes. As Dill et al. report, implementation requires several stages of inquiry that have influence on choices and decision making: technological issues are for example choosing a repository platform, assessing technical needs, establish metadata and vocabulary principles, while political issues can be summarized in evaluating requirements of staff (and its training), promotion, and marketing challenges (Dill and Palmer, 2005).

Most successful implementations, highly recognized in the literature, are two open source softwares, Eprints, developed by the University of Southampton in late 2000, and MIT's DSpace in 2002. Both systems now have numerous installations. Ware, in his valuable paper, measures and compares increasing widespreading of IRs all over he world (Ware, 2004). According to Cervone (2004), DSpace and EPrints[4] are the most common softwares, due to the fact that are open source and highly customizable softwares, even if commercial software surfaced, and are useful for those institutions who can't provide technical staff for customization and maintenance: commercial products generally are more interoperable with previous library services (catalogues and databases) and have best developed modules for access, authorization and rights management, but are less modifiable. Choosing the software depends on aims and goals to address: EPrints, the oldest and most widespread, reproduces traditional text-based scholarship in the form of preprints and postprints; DSpace, instead, allows the deposit for a wide range of digital material types: audio, video, datasets and programs (Cervone, 2004).

Tennant recognizes that building an IR means for the single institution to investigate requirements and characteristics in order to establish an implementation model: he claims that more than choosing a software, stakeholders must concentrate in designing a strategy to meet users' needs in a system which is compliant with existing resources and single institution's peculiarities. The author retains there can be distributed, semi distributed or more or less centralized models according to kinds of documents and community needs: so it's very improbable one model will provide answers for different institutions (Tennant, 2002).

Every kind of institution can create an IR, without care of kinds of information and its provenance (Johnson, 2002), that's why some have wished the introduction of IR in other contexts, such industry (Gallagher, 2005), public libraries, historical societies, museums, and other cultural institutions (Cervone, 2004), corporations, government agencies and no-profit institutions (Drake, 2004), in order to store and manage historical and administrative documentation in a flexible and economic way. Three authors (Shearer, 2003, Chan & Costa, 2005 and Anuradha, 2005), also point out that the economic model under IR can be highly sustainable even in developing countries and can contribute to promoting a better dissemination of information and in some cases to start dissemination where it's absent[5].

In the literature, main variables and decisions are to be taken into account while implementing a repository: input activity, context and disciplines, purposes, advocacy activities, archiving policies, copyright policies (can discourage uploading, according to Crow), content type (more variety= less use?), staff support, quality control policies, access levels, software, use (Crow, 2002), (Drake, 2004).

Model design includes recognizing kinds of publication: some institution can opt for peer reviewed materials, other for grey literature, other again for non evaluated papers (or a mix between these models) (Young, 2002). Critical mass is important both for community of users and for institutions, as Gibbons points out that the best way to demonstrate the persistent value of an IR, in

---

[4] Eprints was available from the Joint Information Systems Committee (JISC) in the UK and the National Science Foundation (NSF) in the United States. DSpace emerged as a joint program between the MIT Library and Hewlett-Packard. FEDORA (Flexible Extensible Digital Object and Repository Architecture), which doesn't provide a user interface out of the box, but allows for a wide range of digital material types, has been developed by University of Virginia and Cornell University.

[5] The second author also describes how the Indian Institute of Science of Bangalore, India, developed its personal strategies to start an IR, see Anuradha (2005).

order to keep alive institutional commitment, is to populate IR with content (Gibbons, 2004b). Prosser claims that peer-reviewed journals can live beside IR, since a preprint can be stored in an IR and the definitive version can be published: but with high value content made freely available, commercial journals couldn't charge subscribers for access their own content, so this should bring to conversion into open access journals by all journals. This implies that publishers should find alternative sources for funding. According to Prosser, the new model raising is a pay for dissemination model (the author pays) rather than a traditional pay for access one (the reader pays) (Prosser, 2003).

A huge opportunity and stake is indeed felt by several authors: management requires continuous communication/discussion among stakeholders involved in the process (Johnson, 2002). According to Drake, collaboration, trust and harmony with staff and managers will encourage faculty to feed and use the repository, while group working and sharing of a common vision among librarians, IT staff and information professionals is necessary (Drake, 2004).

There's high awareness among authors about the key issue of stating clear policies for content recruitment: according to Genoni, this commitment is strictly connected with collection development and management traditionally addressed by librarians. The author, starting by the declaration by OCLC, which considered the most hard challenge not a technical but a cultural one, i.e. make clear "what an institutional repository is, what it contains and what its governance structure should be" (OCLC, 2004), takes into account different kinds of content to be hosted in an IR. He agrees substantially with Lynch, who thinks that a repository should hold all kinds of materials produced by a campus, including lectures and meeting papers (Lynch, 2003), and states that maybe certain kinds of content will come out from best practices, but none is allowed to establish what content IRs should contain, except the institution itself.

Much of literature still focuses on advocacy and promotion of IRs, sometimes replicating themes and divulgating aspects of previous literature without adding something new to aquired knowledge, only more recently few authors expressed concerns about open access (mainly representative of publishers): what still is missing is a strong measurement and critical evaluation of benefits and outputs in the IR environment. Richardson points out the need of an evidence-based approach to really understand what has been done and what needs improvement. He states that policies that seem to change radically scholarly communication are announced frequently, but that business in publishing is a complex process: IRs must be planned and evaluated through practical evidence of measurable and comparable pros and cons. Any policy must be backed with evidence of benefits to the community and analysis of potential impact on current publishing model. He also reflects on the fact that subscription revenue is vital to allow publishers fund the process of peer review, where other models aren't available (Richardson, 2005). The importance of establishing comparable measures for repositories also emerges from Shearer (Shearer, 2003), who designs 3 parameters for judging success of an IR: accessibility (no fees), satisfaction (degree of meeting user needs by a system) and usefulness, the last two related to input activity (a scholar will use more likely an IR with much and highly accessed content).

4.3 Self-archiving

Self-archiving, i.e. the autonomous uploading of documents and related metadata by authors in an IR (Harnad, 2001b), is one of the most debated themes in literature. The importance of self-archiving is acknowledged, because it's the authors' habit to upload or not their papers, thus the development of critical mass of content, to mark the success of IR in terms of usage and quality (Crow, 2002b; Wheatley, 2004).(Allard et al., 2005).

Some authors think archiving as an operation librarians are committed to do, standing the fact that authors are busy and unprepared, perceiving it as more work discouraging them from

depositing[6], others think that none can perform this action better than the creator of a paper, since faculty practice of posting research online (web sites, departmental sites, disciplinary repositories) is documented (Johnson, 2002), others recommend mutual support (Pinfield). Self-archiving is a critical point, since participation or not by authors have huge consequences on content recruitment: several authors retain that, by self-archiving, the author is populating the IR with content and is therefore an key partner in collection development (Chan, 2004; Harnad, 2001; Johnson, 2002, (Prosser, 2005).

Self-archiving is supported by the majority of authors, many of them also believe self-archiving should be mandatory: several initiatives applied a mandatory policy (not least PubmedCentral, which forces authors to deposit public-funded research on the repository). Several institution already apply a mandatory policy, for instance University of Glasgow, where this service is provided by librarians (Ashworth, 2004). In 2004 the House of Commons Science and Technology Committee report, recognizing the easyness of use and economic and research benefits of IRs, recommended mandatory self-archiving as a way to increase crtical mass of content in IRS (Great Britain, 2004), (Chan et al., 2005) and (Gibson, 2005). The same year, the Wellcome Trust, funding body for biomedical research in the UK, announced a similar policy for papers within six months of publication, while the US National Institutes of Health and Australia's National Scholarly Communications Forum followed with similar announcements (Chan et al., 2005). Nevertheless, in its response to the Science and Technology Committee's report, the Government's reaction wasn't very enthusiastic, since it preferred to stick to a "market driven" approach for OA publishing, leaving to individual institutions the choice to set up their own policies(Chillingworth, 2005). Harris criticizs that government (in UK) should take more seriously the challenge of being a proactive agent in changing scholarly communication: he claims decisions by government were too warm and not cutting edge with past (with the risk to promote involution) (Harris, 2005).

Pinfield (Pinfield, 2005) explains that mandatory self-archiving would accelerate change and make benefits more evident, even for those disciplines which have less familiarity with self-archiving, but the risk is that many scholars don't like to be forced doing anything: funding bodies should apply mandatory deposition as a condition to obtain grants. Harnad is one of the most passionate advocates of mandatory self-archiving as a strategy to deliver content recruitment: since technical or financial barriers are low, this could be achieved in a relatively short time, he says.

However, mandatory policies found also violent criticism: Peek reports about negative reactions against the Research Council of the UK by the Association of Learned and Professional Publishers (ALPSP), which claimed that the

> "consequences of the destruction of journals' viability are very serious. Not
>
> only will it become impossible to support the whole process of quality control,
>
> including (but not limited to) peer review, but in addition, the research community
>
> will lose all the other value and prestige which is added, for both author and reader,
>
> through inclusion in a highly rated journal with a clearly understood audience and
>
> rich online function" (Peek, 2005).

Richardson criticizes that self-archiving can lead to multiple copies of articles being available through the internet (with different degrees of authority), and that this behavior can bring to an increase of publication costs (Richardson, 2005).

Either mandatory or not, self-archiving is recommended both by Harnad and Bailey, since scholars write for research progress and to be cited: it means that uploading document which have

---

[6] We talk elsewhere in this brief study about relationship between faculty and librarians and about the role of librarian as a recruiter for content. The focus of this section is faculty and self-archiving itself.

high rates of reading translates into higher citations rates. Difficulties reside in inertia, of course, but also in the career building system, which is largely shaped on publications on journals.

Many scholars are worried to infringe copyright, even when they have complete rights to self-archive, as reported by Greig et al. (2005): this makes them reluctant to use the IR. Cranfield University, University of Glasgow and University of Kansas started content recruitment management through the SHERPA/ RoMEO Publisher Copyright Policies & Self-Archiving database, which identifies publishers' self-archiving policies. Indeed, recruitment of content remains difficult until uncertainty about copyright terms and conditions in deals will persist among authors: Greig describes several activities project DAEDALUS went through and concludes that deeper knowledge is needed about authors attitudes to copyright, copyright checking process, relationship and agreements between authors and publishers, distinction that may arise among publishers between pre-print and post-print versions (Greig, 2005, Greig and Nixon, 2005).

4.4   Preservation

The issue about addressing long-term access to content is very strong in the literature and preservation is perceived as one of the key issues that can determinate the success of IR. Clearly preservation requires proactive management, long term planning and huge efforts that can't be faced by single institutions, as Cervone, Pinfield and Barton et al. say, thus strategies must be organized and funded only on an organisational-based strategy. Stanescu also remarks how implementing models for achieving preservation must be a priority: he claims that preservation plans should be based on objective analysis of risk trends (for file types, softwares and hardware) rather than on individuals' opinions and experiences (Stanescu, 2005).

But who should be responsible of this duty? Cervone (2004) raises two stakeholders: publishers and librarians, but many authors remain sceptical of the publishers' suitability for this task given their short-term perspective. Indeed IRs, being built on institutional or multi-institutional basis, can afford it and serve as attracting points for further research (and funding) in preservation strategies (Barton et al., 2003); Pinfield and James propose to create consortia to afford an activity which is also difficult to quantify in terms of financial resources: he looks at the SHERPA project as a model for further planning. Crow proposes to lay the task in the hands of librarians, those professionally prepared (Crow, 2002). Many initiatives have been taken to solve preservation issues: the JISC Digital Preservation and Records Management Programme launched the Digital Curation Centre and supports the Digital Preservation Coalition, which aim to develop different approaches to achieve long-term preservation (Carpenter, 2005). Stanford University developed the LOCKSS (Lots of Copies Keeps Stuff Safe) initiative, an OAI compliant software system that exploits peer-to-peer networking technology to maintain electronic publications' copies (Eaton, 2005).

Nevertheless, from an analysis of the available literature emerges that there are no universally-accepted archival standards for ensuring preservation yet, and authors substantially agree in relying on technical best practices to achieve preservation, though agreement about which practice is the most suitable hasn't been apparently reached yet. OAIS and RLG/OCLC efforts on Trusted Digital Repositories provided a framework for terminology, key elements for IR design and requirements for preservation, but still preservation practices owe much more to practical experiences by developers.

Some papers show more interest in developing access and dissemination now, and some prioritize protecting that access for the future. Stevan Harnad (2001) supports the idea that access and filling IRs with content are the major priorities and that preservation could be a distraction right now, stating that ArXiv (the first discipline-based repository) was still granting access ten years after its launch. Pinfield and James follow partially this viewpoint, admitting that for e-prints the problem can be simpler, but, considering the high variety of formats available through IRs,

preservation can't be considered only a technical issue (Pinfield and James, 2003). Some implementation softwares, as DSpace, care much more for preservation than others, as EPrints, but their aims are different. Repositories with scholarly output generally care more for perpetual access, and this kind of repository is the focus of this literature review and definitely the most common: Pinfield and James, as Carpenter, recommend to select, at a management level, kinds of materials to be preserved, and to choose where risks of loss are greater (Pinfield and James, 2003). According to Wheatley (Wheatley, 2004) security and authenticity, verification and storage are well addressed by current softwares; he claims that as digital obsolescence implies updating tools or migration of formats (which are expensive activities), the best strategy rather consists in storing digital objects as bitstreams, independent from medium: then raw data would be interpreted by users once they access the file through metadata. This implies that metadata are extracted during ingest, stored in a framework, constantly monitored (to grant access even when technology changes), changed when necessary (with a system that allows keeping path of changes), while digital objects must be rendered through a displaying process, to make sense of them. Preservation starts with ingest, and involves every stage of a digital object's life. Cervone instead believes that migration is the best solution: in his opinion providing compatible retrieval and rendering technologies for digital material is required, while IRs provide mechanisms to identify material to simplify future migration activities. He pays strong attention on standards and protocols which are needed to ensure continuous access to information, recognizing the Open Archival Information System (OAIS) Reference Model as the most economic and widespread framework for further standard's development and economic exporting for migration. He lines up with other authors claiming that:

> "Planning for preservation is tricky. It is difficult to predict the many critical aspects of the preservation puzzle […]. Preservation must be integral to the planning, design, and budgetary process for repositories if institutions don't want commitments to exceed resources" (Cervone, 2004).

4.5 Identification

Though it's barely treated in the literature, as a mere technical aspect, identification is a problem also: identifying and locating content is a key issue in repository systems: to grant long-term access, each object should have a unique and persistent identifier independent from the software which is in use, that must remain valid even if the content migrates to a new system or if the management responsibility of the institutional repository changes (Cervone, 2004). Various standards (DOI, ARK, URN etc.) are used by different institutions, while interoperability is only gained through an integrated system of identification. DPC and DCC are in charge of drawing a regular standard.

5   CHALLENGES IN CURRENT TRENDS

5.1 *Impact*

IRs developed quickly in different institutions all over the world through different systems, implementations and structures, and the development of a supporting series of tools means that content is increasing and users are interested in searching through value added systems. JISC (Joint Information Systems Committee), CURL (Consortium of Research Libraries) and SPARCEurope, together with the OSI (Open Society Institute) have launched OpenDOAR, the Directory of Open Access Repositories, in order to monitor the growth and variety of IRs and also offer a research tool to search for IRs given some parameters (2005a). Another tool is OAIster, a search engine which scans content of thousands of IRs all over the world. The RoMEO project, which lists all major publishers' policies on copyright and reproduction of materials, is also increasing and supporting authors and librarians in content management organization (Gadd et al., 2003b) and (Gadd et al., 2003a).

Commercial and institutional providers and publishers have understood the opportunity (and somehow the threat) of IR and tried to get involved in more or less open access initiatives: that's what Peek calls "the cascading effect" (Peek, 2004) (Elsevier's policy on self-archiving which we're about to mention is a clear example). Strategic deals can provide benefits for both partners: visibility for the institution and comprehensiveness for providers.

In June 2005 Elsevier, collaborating with T-Space, the University of Toronto's IR, launched Scirus Repository Search, a new service that indexed the full-text of T-Space's repository providing additional search capabilities at no cost (2005b). Ebrary developed a product to create online institutional repositories for libraries, integrated both with library resources and Ebrary commercial tools (2003). Repositories had a huge impact on public research institution as well, the most prestigious example is the Open Repository started by BioMed Central, (on commitment of National Institute of Health) which aims to help institutions in building and maintaining their own institutional repositories. This is one of the early alternatives for those institution which can't afford launching their own IRs with their technical, financial or human resources (Chillingworth, 2004), (2004).

Views and perspectives among publishers are very variegated, since Blackwell remains sceptical about moving towards alternative publishing models, especially if undertaken by academic institutions and retains unlikely a revolution in the brief term (Blackwell et al., 2004). An hybrid position was taken by Springer, which started some open access initiatives, but testified the publisher's journal as authoritative source of quality (Peek, 2004).

Traditional print model is integrated: refereeing, editing of standards, dissemination and marketing are part of a unique process (vedi anche Prosser 2004 e Crow). Some say that it justifies high costs of journal (but apparently not the huge annual increase), other think that alternative models, unbundling those operations, can show the weakness of the previous statement.

IRs and open access publishing model had a huge impact even on self archiving agreements with publishers: in 2004 Elsevier allowed its authors to self-archive their papers in IRs (Peek, 2004), while in 2005 authors of the Nature Publishing Group could also deposit their manuscripts into their own IRs six months after the original publication.

These experiences mean that a complementary and hybrid environment is emerging, as Hubbard agrees, when he talks about hybrid approaches in implementing IRs: they aren't built for replacing traditional scholarly publishing model, but to change it, even radically, in order to obtain best performance. Hybrid approach means that scholarly content will be available through different channels, with a significant and necessary quality control (Hubbard, 2003). Oppenheim, agreeing with many others about IRs not being a substitute for traditional scholarly journals, adds that open

access publishing hasn't yet found a clear and stable business model, but it's process which has started and is not meant to be erased (Oppenheim, 2005).

A study by Graham et al. demonstrates another influence IRs can bring to traditional relationship inside institutions: a repository can be an economic way to establish relationship between libraries and isolated departments/communities which previously had little or no involvement with library services (Graham et al., 2005).

5.2 Barriers and obstacles

Concerns about the realization of an ideal landscape where scholarly communication goes smoothly are well synthesized from one of IRs advocates, Crow:

> "Altering the structure of the scholarly publishing model will be neither simple nor immediate. The stakes are high for all the well-entrenched participants in the system- faculty, librarians, and publishers-and the inertia of the traditional publishing paradigm is immense. In the near-term, large journal publishers have both the power and the incentive to maintain the status quo: the prestigious journals they control appear integral to the very structure of academic professional advancement. However, digital publishing and networking technologies, harnessed by an increasingly dissatisfied library market-as well as by authors themselves-are now driving fundamental changes to this publishing model at an accelerating pace. And new communications paradigms, especially when constructed by the scholars themselves, can eliminate seemingly insurmountable publisher advantages in relatively short order" (Crow, 2002).

Many authors, as Young (Young, 2002) and Cervone, states that scholarly journal system can be an obstacle to free sharing of content and to advances in sciences, since it strongly influences faculty members' habits. IR supporters recognize the difficulty of changing authors' habits about uploading and evaluating the content stored in IR. MacKenzie Smith, quoted by Young, says that "professors are busy, and they may not use the repository if they perceive it as more work, even if they like it in principle", while Cervone fear the same perception for staff members.

Implementing new models for scholarly communication and renew the scholarly communication will not be immediate: Johnson, as Cervone, retains the hugest obstacle in implementing the IR model is definitely "inertia of the traditional publishing paradigm"; he says

> "large journal publishers have both the power and the incentive to maintain the status quo: the prestigious journals they control appear integral to the very structure of academic professional advancement" (Johnson, 2002).

A well recognized challenge is defying scholars' and stakeholders' inertia against change: faculty must understand that open access articles can be more cited (that's the greatest benefit for them) as reported by Rowland (Rowland, 2005), while publishers want to maintain the status quo, for prestige, since journals are strictly integrated with academic professional advancement system (Johnson, 2002), (Crow, 2002). Scholars must be aware of what content they can give away for free, so librarians' role is necessary in advocacy and training about copyright, licensing and relationship with publishers' agreements (Chan et al., 2005): as we say elsewhere in this study,

Who is in charge of adding metadata, indexing and describing content is also a debated obstacle, as in the literature there's no agreement and no universal policies models: Young believes it can be an user activity(Young, 2002) but other researchers think librarians are committed to do it, as we said previously.

Inertia is a huge obstacle too, but it's also justified by worries about copyright and plagiarism, especially in humanities (Young, 2002).

Inertia can reside also in cultural and academic barriers which are well described by Jenkins (2005), who reports that getting faculty members inside IRs required a shift in habits, values (and habits are diehard). Librarians' advocacy must increase positive perception by faculty members/administrators in order to recruit/make available more content. As this author say,

> "authors who submit material to the IR are, in a sense, risk-takers and academia is a risk-averse environment. The traditional culture of academic publishing accounts for some of the resistance to IRs; journal and monograph publication processes are deeply embedded in the scholarly process. Across disciplines, publishing in journals and monographs has been the standard for over 100 years, and integrating a new genre into scholarly communication is a significant challenge. Faculty depend on the traditional genres of communication not only to disseminate research, but also to get tenure and establish themselves in their field"
> (Jenkins et al., 2005).

As emerges from literature analysis, faculty perceptions and approaches to scholarly publishing changes may vary depending on academic discipline (Jenkins et al., 2005), (Crow, 2002): some disciplines were more receptive in experimenting repositories and sharing of work-in-progress content, while other ones, such Humanities and Social Sciences, didn't enjoy such a tradition in communication skills and faculty perceive their research as proprietary. Jenkins adds that some disciplines don't prioritize change in scholarly publishing, and in some cases journals are controlled and owned by a few publishers.

Jenkins also identifies other critical barriers and resistance: fear of disrupting existing relationships with publishers and generally ignorance of copyright law, concerns about not being able to publish in traditional journal after publishing in IRs,. reluctance to give away research outputs without traditional validation (the perception that repository content, which may not go through peer-review, is of lower quality than traditional scholarly journals, is also mentioned by Cervone (Cervone, 2004), reluctance to share work-in-progress research, reluctance/lack of time to modify bureaucratic processes and learn new procedures, reluctance to have their research output marked under institutional name, mistrust of the long-term access of digital content.

Nixon (2002) provides an additional perspective on this issue:

> "The challenge, ultimately will not be the technical implementation of an e-prints service but rather the cultural change necessary for it to become embedded and commonplace in the activities of the institution" (Nixon, 2002).

Crow retains that one of the biggest challenges is indeed marketing IRs to make them understandable and addressing advocacy to encourage voluntary involvement by faculty members:

> "While gaining the participation of faculty authors is essential to effecting an evolutionary change in the structure of scholarly publishing, early experience

suggests better success when positioning the repository as a complement to, rather
than as a replacement for, traditional print journals"

5.3   The publishers's position

Crow analyses carefully and in the deepest way (in comparison with others) all consequences for commercial, society and academic publishers. Commercial publishers offer high value services as peer-review, citation linking, controlled vocabularies, which justify high prices subscriptions. IRs disturb this system, and many publishers are fear the consequence of this alternative model: to synthesize Crow's position, publishers can choose between being inflexible or face the loss of channel exclusivity and adapt/appeal to a new market. Surely their role is not doomed, simply changed, especially if we take into account some discipline-based repositories, running since early nineties, which didn't erase traditional scholarly publishing.

Ware (2004) claims that the major challenge for publishers will be how to respond to more liberal copyright agreements: the RoMEO project revealed that publishers can agree to self-archiving of pre-prints, but not always to free access for peer-reviewed articles, because they feel the distinction between these two kinds of materials. The author claims that publishers could contribute to OA communication making available their bibliographical data and harvest OAI metadata (as Elsevier's Scirus does).

Learned society publishers are less aggressive in exploiting their monopolies than commercial publishers, but contribute strongly to research institutions' expenses and have great power, which will not let down simply for the goods of research. Their role must be redefined, but the author still doesn't know which shape this redefinition will take (and other authors don't mention learned society publishers): since worthy literature is much more than in a print environment with page constraints, new metrics for quality and validity of research must be provided and learned society publishers have the financial and political credentials to play a role in new peer review and certification activities. Societies could also collaborate with libraries to develop author/document authority control. Learned societies have long-standing relationships with their members and they should be able to act as focal points for the research communities they represent.

University presses can also dive into the new market, by allowing electronic versions of press monographs to be made available in IRs, and maybe they will be included in IRs, which can be a new publishing body.

Government Agencies and Other Funding Sources are interested in dissemination of research outputs they funded: deals for grants could include the mandate to deposit outputs in IRs (as actually happened later Crow's report), or subsidies for authors/departments supporting open access repositories.

Prosser (Prosser, 2003) designes some steps to be taken in the near future: publishers should be more flexible in agreements with authors, allowing deposit (RoMEO), should consider to convert their journals to open access and founding new oa journals.

6   LIBRARIANS AS FUNDAMENTAL STAKEHOLDERS IN IRS MANAGEMENT

The evidence of considered literature seem to stress the importance of librarians as key stakeholders: their traditional roles are extended, but there are also new commitments. However, their roles and training has not been so deeply investigated until the last two years. Early literature seemed to give more emphasis to implementation models, including roles as publishers, managers, faculty members, as reported by Johnson (Johnson, 2002) and to explanation of what IR are, more recently there have been studies about the human element. According to Prosser (2003), librarians should establish IR, help faculty in archiving, digitizing, training faculty members on how to find oa resources, cancel subscriptions to journals that cover the same areas and have the same quality of oa archives.

A valuable contribute comes from Suzie Allard, which scans the IR literature to understand possible roles for librarians in IR projects and processes. At the time of the article, little discussion had been made about the librarian's role, but Allard identifies these ones: understanding software, project planning and management, collection definition/developing, metadata guidance, submission review, and author training. She claims that literature is not providing librarians with the adequate knowledge to provide services in an IR environment.(Allard et al., 2005)

Rockman claims that reference librarians are "natural partners to be involved with institutional repositories" because of "their service orientation, subject experience as knowledge managers, and communication skills". The overview study by Horwood et al. places librarians in a role which is continuation of past tasks but also a innovation, as requires a proactive approach towards several aspects of IRs policy and advocacy (Horwood et al., 2004): traditionally, librarians are experienced in selecting, describing, storing, and managing information content (Chan et al., 2005) (Jenkins et al., 2005), but general agreement on proactive role for librarians in the IR environment seems clear for major open access advocates as well, as reported by Caldwell (Caldwell, 2004). Bailey points out the value of the "human element" in the system, stressing traditional skill that are applicable to the new environment:

> "Reference librarians are a library's eyes and ears. They understand user needs and perceptions. They know what is working and what is not. When they act as subject selectors, they are the library's primary liaison with faculty in their subject areas and its most visible representatives. They know how to help, inform, persuade, and teach users. For an IR to succeed, it is essential that they be involved in its planning, implementation, and operation" (Bailey, 2005b).

They can effectively collaborate with faculty and students to encourage them to add content and, exploiting their experience with interfaces, content development policies and teaching skills, serve as a key human element for development, management, dissemination, sustainability and success of institutional repositories. The author stresses the importance of collaborating. Their main activities can be related to teaching, learning, and research processes on their campuses. (Rockman, 2005). Collaboration among staff with different expertise from all parts of the library, as strong administrative support, are essential, according to Rockman and Jenkins (Jenkins et al., 2005).

In the literature I identified these core activities for librarians, which come from their knowledge and confidence with users and from their traditional expertise:
- Helping in creating IR policies and procedures for content management (kinds of materials to be included, as also Chan points out (Chan et al., 2005). I didn't find any mention to a vertical approach with managers and directors in order to establish a stable role in this area. I also didn't find strong reference to evaluation of collection

performance, and decision making relating to access, conservation, and preservation, quickly mentioned only by one author (Genoni, 2004).

- Assisting IT staff and faculty members in designing the IR user interface.
- Helping to identify current self-archiving activity on campus to help, encourage and develop content recruitment initiatives for academic authors. This proactive duty is specially recommended by Jenkins (Jenkins et al., 2005) and Mackie (Mackie, 2004). Chan claims that librarians should contact publishers to investigate their policies and license agreements where authors refuse or fail to do it (Chan et al., 2005).
- Marketing and promoting the IR to faculty in campuses, using a terminology which is familiar for them and focusing on benefits for individuals or departments, through guides, handouts. As Gibbons particularly remarks, the challenge lies not much in the technical issue, as in recruiting content, i.e. developing a change of mindset among academic authors(Gibbons, 2004a): librarians can build trust in institutional repositories by faculty members, explaining copyright policies, licenses (as Creative Commons) and publisher e-print policies, making self-archiving an usual part of research processes (Chan et al., 2005). This role expands towards students and graduated, to shape their habits: the same author believes that promotion must be undertaken at all levels, from university administrators to department heads, canter directors, researchers, and graduate students, through newsletters, local newspapers and conferences. Faculty members must perceive institutional repositories as complementary to traditional publishing, especially in disciplines that lack a tradition of collaboration and research output sharing: once "faculty members realize the impact and benefits of putting their papers on an open access platform, they will be motivated to take action" (Chan et al., 2005). Some authors push the topic further, thinking the librarian as an agent which can do some activities (self-archiving and metadata providing, contact with publishers, organization of collections) on behalf of authors (when required) in addition to verifying that users are comfortable with this new environment (Jenkins et al., 2005). Marketing role is indeed the most recognized and agreed role for librarians, since it has a strong weight in case studies and papers (Buehler and Boateng, 2005), (Phillips et al., 2005), (Graham et al., 2005).
- Providing IR metadata, such as local controlled vocabularies, to be negotiated with users. Genoni also suggests that librarians should evaluate the performance of the collection and monitor the process. There poor evidence in the literature about which shape will take the metadata supporting by librarians, the discussion is about letting authors doing on their own (with help) or being in charge of providing metadata. This is a problem that is being faced by the technological point of view from single application software (it's the ingest, uploading workflow process, which guides the user step by step), but not yet from a "human" viewpoint: as Beall says, there are lots of errors which can occur in metadata providing (that can stop access to digital content, even if it's available) (Beall, 2005) so the need for vocabularies, even local institutional vocabularies and thesauri, is high, especially for users which aren't accustomed in indexing. The impression I had reading materials is that elaborating indexes should be a mutual operation established through dialogue between librarians and researchers as specialists in their disciplines. None paper yet analyses how this should be done.
- Training users in IR deposit and searching procedures. Reference is a consistent activity even after the launch and training for IRs: Bailey and Jenkins recommend assistance in IR use, to help authors evaluating and trusting search results in this area. Jenkins also advise to integrate IRs into users' research vocabulary, clarifying their content (Jenkins et al., 2005). An interesting case study revealed that faculty members have been slow to put their content into the IR, mainly because they have not

understood how they would benefit from doing so (Bell et al., 2005). Chan (2005) reports that once IRs are federated and cross-searchable can form a universe of information, and librarians have the task to teach research strategies for this tool.

Case studies are also interesting to see how different institutions, with different financial-human resources and expertise have managed to make an IR work[7]: what really seems to lack is a study on all those single cases, in order to establish one or more best practices in this area. From the available literature seems clear that differences between institutions, and obviously different levels of experience by librarians, prevent from putting together something more complex than few advices or practical examples. Where faculty research habits were surveyed, it was more easy for librarians to appeal to users, delivering knowledge and effective advocacy about IRs. Bell reports that

> "The key to success is to provide adequate training to Library Liaisons so
>
> that they will speak confidently to faculty members in understandable terms, and are
>
> prepared to answer any questions that might arise" (Bell et al., 2005).

Caldwell, citing Peter Suber, retains that librarians shouldn't be alone in training faculty members, especially in self-archiving procedures. Training faculty means: understanding of the IR and its motivations, critical information in order to answer questions the faculty may ask and knowledge on how deposit process works. Advocacy must be undertaken by deep understanding and dialogue with faculty members, as above mentioned authors claim, and as also Foster and Gibbons (Foster and Gibbons, 2005); talking in faculty language, answering questions confidently, especially about copyright, which results to be the greatest concern for scholars, as reported in different papers (Ashworth et al., 2004) are also a key abilities to be earned. Faculty members will want to use the IR once they know that others are finding, using, and citing the work that they place there. To achieve impact on faculty and other stakeholders, librarians must be trained as well: among others, Chang claims that libraries need to recruit librarians who have updated skills in digital collection management and Open Archive Information System (OAIS) management, and possess the leading charisma to involve people in IRs policies, both in vertical and horizontal directions (Chang, 2003).

Drake remarks that librarians are taking leadership roles in planning and building these repositories, fulfilling their roles as experts in collecting, describing, preserving, and providing stewardship for documents and digital information (Drake, 2004), while, according to Genoni, they can apply their skills in content development as well, as we reported above (Genoni, 2004).

---

[7] Some of the case studies taken into account were University of Oregon, University of Rochester, Hong Kong University of Science and Technology.

7   CONCLUSION

Literature concerning IRs is definetly strong and various. It demonstrates that IRs are perceived as a proactive policy for radically change scholarly communication models. Considering criticism and thoughts from advocated and detractors gave the idea that change is not supposed to be nor immediate or traumatic.

Many efforts are documented, in order to achieve dissemination: from content recruitment to worries for guaranteeing access for this content, from self-archiving discussions to advocacy and marketing roles. IRs have now earned respect as an aconomic way for changing scholarly communication into a more economic and user-oriented channel, as government and big publishers paid attention to them.

7.1   Strenghts, weaknesses and further research suggestions

Several strenghts were highlighted in the scanned literature:

- papers reveal high consciousness of the importance of IRs as a new scholarly communication channel, citation rates for open access papers are higher and the use is increasing.
- Standards are stable and highly recognized, which makes IRs interoperable and the creation of a world digital library nearer.
- The number and entity of research projects promoted by specific institutional bodies are increasing, running and offering relevant output and data to reflect on.
- Practical initiative are different and may serve as examples for institutions which wish to start IRs.
- There's a growing attention to the importance of librarians and increasing literature about how librarians can face different stakes and challenges.
- Efforts and consortia for specific or technical issues as preservation are taking place, exploring alternatives which can be useful even outside IRs.

Weak points outlined were the following:

- Lack of measurements, comprative studies for specific services or technical aspects, there's almost complete lack of an evidence based approach.
- Studies on citations, alternative peer review processes and quality assessment haven't been undertaken seriously yet.
- There's a lot of discussion and disagreement about impact on publishers and the new role for them.
- Literature lacks studies about how things must be done, especially surveys on user needs and habits, costs, metadata management: stakeholders may have a good knowledge of open access but not know how practically to run an IR.
- There's still scarce reference about how libraries can reallocate financial and economic resources in order to run IRs while maintaining traditional services.

We have outlined other big opportunities which emerge from literature and that should be investigated in the future.They refer to the following areas:

- Developing IRs for dissemination of information in the developing countries.
- Use of IRs for publication as if they were University Presses.

- Creation of deals and redefining the role of publishers.
- Involvement of all key stakeholders.
- Definition and development of value added services which make repositories competitive with other information providers.
- Exploiting exsisting technology to improve services and creativity, customization. How to use open source softwares and adapt them in relation to user needs.

The most acknowledged threats for the future involve key elements that can estabilish or not the success of IRs:

- Lack of use by scholars: the vicious circle (lack of depositing activity>less content>less use>decading of IRs model) is indeed rooted in scholars' habits, and advocacy must be massive to avoid it.
- Tightening of publishers' position is also a huge threat, if they react to IRs and open access as a enemy and put barriers to access, not permitting post-prints publication or imposing to upload after some time.
- Lack of interest and long term commitment by faculty administration can be disastrous for IRs policies: people and institutional interest must be constant and effective.
- Decreasing quality of library services due to amount of work committed to librarians: IRs could turn out to increase costs for libraries instead of relieving them, if resource management and planning is not set up properly.
- Training staff become necessary, staff must be trained to train other faculty members.

8   CRITICAL ACCOUNT

8.1   Outcomes and self-criticism

Findings and perspectives outlined in this work, even if limited, can provide wide knowledge about what IRs are, how they are supposed to change current publishing model and how different stakeholders take into account the impact of IRs in their own environment. I think also this small work can give a general landscape of theoretical and practical issues related to IRs management and advocacy.

The major challenge for me was facing the high amount of information I gained from literature scanning. Many decisions and choiches had to be made in order to balance depth with width. IRs cross many open access related aspects, so I tried to choose few relevant sources for each of those aspects. In addition, a high percentage of literature (which later was discarded), aimed at divulgating IRs concepts, without adding new ideas or effective scientific contribution to the matter. Many articles also repeated old concepts in their first half, while concentrating on new contribution in the second one.

Elements of validity for this work lie in the following areas:

- The width of research scope: this literature review analyses briefly major aspects and issues related to IRs, from the technical/political viewpoint, as implementation and technological framework setting to the development/management viewpoint, regarding policies for content recruitment, scholars' research habits change and not least the role of librarians as change agents.
- The consideration of both no-profit and commercial approach for developing IRs infrastructure: the hybrid market is costantly present in considering policies and development plans and competitiveness is one of the propulsive elements for ameliorating digital libraries.
- This work tries to group issues and topics which in literature appear very fragmented: a huge work was undertaken in extracting single authors' views and opinions in different areas and compare them into an integral framework.

On the other hands, limits to the research were the following:

- Lack of depth and detailed information. Each aspect would have required a dedicated literature review: an in depth analysis of all issues were impossible to make, since theoretical literature and case studies were abundant and variegated in approaches, methodologies and scopes.
- Partial information: only literature in English was taken into account, since English and U.S. projects were leading the trend: much literature in other languages replicated English literature for divulgation in other countries. However his doesn't mean this literature could be considered as completely irrelevant.
- Lack of updating: literature research was undertaken at the beginning of 2006, while lots of time were spent in reading, selecting and choosing information to be used and which structure to give it.

8.2   Research path and strategy

The first step was creating a conceptual map of the topic, based on previous knowledge about it. I wanted to explore more and more variegated resources in order to understand deeply and broadly some various aspects I've been interested in by previous researches I did.

Secondly, I compiled a list of keywords and phrases to be used effectively in information resources scanning.

I scanned all selected resources in order of authority: OPAC, databases, grey literature, websites and conference sites: all quotable literature was managed through Refworks (the web-based tool was useful could be accessed even when I hadn't my pc) and EndNote as reference manager tools, and all references were then imported in EndNote when writing started. The retrieved literature, cleaned by duplicates through the reference manager, was certainly overwhelming (over 850 records), since the topic could be related with several other broader topics. I chose to start reading the literature I retained more compliant with my interests.

I started skimming through the literature, reading abstracts, finding main topics inside every work and recurring authors. This was the hardest and time-consuming activity. Retrieving the literature was not hard at all, because most of the content was available online and free. Elsewhere I recurred to document delivery supply. Since the topic has not yet penetrated in depth into books, I strongly referred to articles. I tried to group literature in relation to main topics, but I soon recognized that many articles tried to embed many topics so the same article could refer to different sections of my previous conceptual map.

Then I redefined the starting conceptual map and related questions/hypotheses under the light of new materials. This action was hard as well, since the major problem for me is to stick to the narrow topic I have in mind, without going further. In this phase my research found a stop as trying to organise knowledge in the most appropriate way was an activity that really overwhelmed me. Finding lots of irrelevant literature, which superficially repeated other authors' concepts, was certainly another discouraging aspect, since my time was very little and my schedule too ambitious to be embarked as it was initially planned. The map has been shortened a lot, so the complex of this work can seem incomplete or uncovering some aspect that are however mentioned but not investigated in depth how as it was initially intended.

I started writing considerations and comparisons among authors; each section (which later became the chapters of the final report) was written separately, trying to include within each one all the distinctions I made in the conceptual map.

8.3 Information sources used

I tried to search methodically through research tool in order of relevance:
- OPACs, such as Library of Congress, COPAC, Northumbria University Catalogue, SBN (Italy) and metaOPACs (MAI, MetaOPAC Azalai).
- Databases[8]: LISA (CSA), ZETOC, Emerald Full-Text, Ebsco, Eric. The research was also repeated with he launch of Northumbria University NORA, and duplicated or previously literature judged as irrelevant discarded. This information, as expected) was the most relevant in terms of quality and quantity.
- Bibliographies by C.W. Bailey (cited in the references) were also a huge help in finding materials and their relevance.
- Open access resources: University of Michigan's OAIster
- Grey literature resources, as Google Scholar, Elsevier's Scirus, ISI Web of Knowledge. Google Scholar was customized to export references in Refworks.
- Search engines as Google were used only to find some specific information, that weren't unavailable through the above mentioned channels, or additional environmental knowledge that isn't documented in the literature (names of authors, curricula, personal sites etc.).

---

[8] I decided not to refer to single e-journals because the topic touches several aspects and each of them had a strong number of related journals, I thought indexing services as databases and search engines for journals had a sufficient covering.

BIBLIOGRAPHY

(2003) Ebrary announces institutional repository program. *Advanced Technology Libraries,* 32(6)**,** p. 3.

(2004) BioMed Central launches repository service. *Advanced Technology Libraries,* 33(10)**,** p. 9.

(2005a) OpenDOAR or directory of open access repositories. *Information Services & Use,* 25(2)**,** pp. 109-111.

(2005b) Scirus launches repository search service. *Advanced Technology Libraries,* 34(9)**,** pp. 7-8.

ALLARD, S., MACK, T. R. & FELTNER-REICHERT, M. (2005) The librarian's role in institutional repositories: a content analysis of the literature. *Reference Services Review,* 33(3)**,** pp. 325-336.

ANURADHA, K. T. (2005) Design and development of institutional repositories: a case study. *International Information and Library Review,* 37(3)**,** pp. 169-178.

ASHWORTH, S., MACKIE, M. & NIXON, W. J. (2004) The DAEDALUS project, developing institutional repositories at Glasgow University: the story so far. *Library Review,* 53(5)**,** pp. 259-264.

BAILEY, C. W. (2005a) *Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals,* Washington, D.C., Association of Research Libraries. Available at: http://www.digital-scholarship.com/oab/oab.pdf

BAILEY, C. W. (2005b) The role of reference librarians in institutional repositories. *Reference Services Review,* 33(3)**,** 259-267.

BAILEY, C. W. & HO, A. K. (2005b) *Open Access Webliography.* Available at: http://www.digital-scholarship.com/cwb/oaw.htm

BARTON, M., HOLLEY, R. & MURRAY, A. (2003) Planning for institutional repositories in the United Kingdom. *SCONUL Newsletter,* (29)**,** pp. 56-59. Available at: http://www.sconul.ac.uk/publications/newsletter/29/15.PDF

BEALL, J. (2005) Metadata and data quality problems in the digital library. Available at: http://jodi.tamu.edu/Articles/v06/i03/Beall/Beall.pdf

BELL, S., FOSTER, N. F. & GIBBONS, S. (2005) Reference librarians and the success of institutional repositories. *Reference Services Review,* 33(3)**,** pp. 283-290.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Available at: http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

Bethesda Statement on Open Access Publishing. Available at: http://www.earlham.edu/~peters/fos/bethesda.htm

BJORK, B.-C. (2004) Open access to scientific publications: an analysis of the barriers to change. *Information Research,* 9(2)**,** No page numbers.

Available at: http://informationr.net/ir/9-2/paper170.html

BLACKWELL, P., DUDMAN, J. & OLIVIERI, R. (2004) Family values. *Information World Review,* (206)**,** pp. 18-19.

Budapest Open Access Initiative. Available at: http://www.soros.org/openaccess/

BUEHLER, M. A. & BOATENG, A. (2005) The evolving impact of institutional repositories on reference librarians. *Reference Services Review,* 33(3)**,** pp. 291-300.

CALDWELL, T. (2004) Libraries take a role in archives. *Information World Review,* (205)**,** pp. 18-19.

CARLSON, S. (2002) Scholarly publishers aim to woo librarians away from self-published research. *The Chronicle of Higher Education,* November 7. Available at:

http://chronicle.com/free/2002/11/2002110701t.htm

CARPENTER, L. (2005) Supporting digital preservation and asset management in institutions. *Ariadne,* (43). Available at: http://www.ariadne.ac.uk/issue43/carpenter/

CERVONE, F. (2004) The repository adventure. *Library Journal,* 129(10)**,** pp. 44-46.

CHAN, D. L. H., KWOK, C. S. Y. & YIP, S. K. F. (2005) Changing roles of reference librarians: the case of the HKUST Institutional Repository. *Reference Services Review,* 33(3)**,** pp. 268-282.

CHAN, L. & COSTA, S. (2005) Participation in the global knowledge commons: Challenges and opportunities for research dissemination in developing countries. *New Library World,* 106(3/4)**,** pp. 141-163.

CHANG, S.-H. (2003) Institutional repositories: the library's new role. *OCLC Systems and Services,* 19(3)**,** pp. 77-79.

CHILLINGWORTH, M. (2004) BioMed Central launches its new Open Repository service. *Information World Review,* 206)**,** p. 11.

CHILLINGWORTH, M. (2005) Government sticks by its OA policy. *Information World Review,* (210)**,** 1. Available at:

http://www.iwr.co.uk/information-world-review/news/2083924/government-sticks-oa-policy

CROW, R. (2002) The case for institutional repositories: a SPARC position paper. *URL: http://www.arl.org/sparc/IR/ir.html* and www.arl.org/sparc/IR/IR_Final_Release_102.pdf

DILL, E. & PALMER, K. L. (2005) What's the Big IDeA? Considerations for Implementing an Institutional Repository. *Library Hi Tech News,* 22(6)**,** pp. 11-14.

DRAKE, M. A. (2004) Institutional repositories: hidden treasures. *Searcher,* 12(5)**,** pp. 41-45.

EATON, J. (2005) Volatility and the long term archive. *Managing Information,* 12(10)**,** p. 10.

FOSTER, N. F. & GIBBONS, S. (2005) Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine,* 11(1)**,** No page numbers. Available at: http://www.dlib.org//dlib/january05/foster/01foster.html

GADD, E., OPPENHEIM, C. & PROBETS, S. (2003a) The intellectual property rights issues facing self-archiving: key findings of the RoMEO Project. *D-Lib Magazine,* 9)**,** No page numbers. Available at: http://www.dlib.org/dlib/september03/gadd/09gadd.html

GADD, E., OPPENHEIM, C. & PROBETS, S. (2003b) Self-archiving: the right thing? An introduction to the RoMEO Project. *SCONUL Newsletter,* (29)**,** pp. 34-37.

GALLAGHER, R. (2005) Why we need institutional repositories. *The Scientist,* 19(19)**,** p. 8.

GELFAND, J. (2005) `Knock, Knock': Are Institutional Repositories a Home for Grey Literature? in FARACE, D. J. J. F. (Ed.) *GL6 Conference Proceedings: Sixth International Conference on Grey Literature: Work on Grey in Progress 6-7 December 2004.* Amsterdam, TextRelease.

GENONI, P. (2004) Content in institutional repositories: a collection management issue. *Library Management,* 25(6)**,** pp. 300-306.

GIBBONS, S. (2004a) Benefits of an institutional repository. *Library Technology Reports,* 40(4)**,** pp. 11-16.

GIBBONS, S. (2004b) Estabilishing an institutional repository. *Library Technology Reports,* 40(4)**,** pp. 54-56.

GIBSON, I. (2005) Overview of the House of Commons science and technology select committee inquiry into scientific publications. *Serials,* 18(1)**,** pp. 10-12.

GINSPARG, P. (2000) Creating a global knowledge network. *Freedom of Information Conference, 2000.* Available at: http://people.ccmr.cornell.edu/~ginsparg/blurb/pg01unesco.html

GRAHAM, J.-B., SKAGGS, B. L. & STEVENS, K. W. (2005) Digitizing a gap: a state-wide institutional repository project. *Reference Services Review,* 33(3)**,** pp. 337-345.

GREAT BRITAIN, H. O. C., SCIENCE AND TECHNOLOGY COMMITTEE (2004) Scientific Publications: Free for all? Tenth Report of Session 2003-04 HC 399- I & II. Available at: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm

GREIG, M. (2005) Repositories and copyright: experiences from the DAEDALUS project. *ALISS quarterly,* 1(1)**,** pp. 24-27.

GREIG, M. & NIXON, W. J. (2005) DAEDALUS: delivering the Glasgow ePrints Service. *Ariadne,* 45)**,** [np]. Available at: http://www.ariadne.ac.uk/issue45/greig-nixon/

HARNAD, S. (2001a) For whom the gate tolls? How and why to free the refereed research literature online through author/institution self-archiving, now. Available at: http://www.cogsci.soton.ac.uk/~harnad/Tp/resolution.htm

HARNAD, S. (2001b) The self-archiving initiative: freeing the refereed research literature online. *Nature,* (410), pp.1024-5.

HARRIS, E. (2005) Institutional repositories: is the open access door half open or half shut? *Learned Publishing,* 18(2)**,** pp. 85-89.

HORWOOD, L., SULLIVAN, S., YOUNG, E. & GARNER, J. (2004) OAI compliant institutional repositories and the role of library staff. *Library Management,* 25(4-5)**,** pp. 170-176.

HUBBARD, B. (2003) SHERPA and institutional repositories. *Serials,* 16(3)**,** pp. 243-247.

JENKINS, B., BREAKSTONE, E. & HIXSON, C. (2005) Content in, content out: the dual roles of the reference librarian in institutional repositories. *Reference Services Review,* 33(3)**,** pp. 312-324.

JOHNSON, R. K. (2002) Institutional repositories: partnering with faculty to enhance scholarly communication. *D-Lib Magazine,* 8(11)**,** No page numbers. Available at: http://www.dlib.org/dlib/november02/johnson/11johnson.html

LYNCH, C. A. (2003) Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL,* 226(February 2003)**,** pp. 1-7.

MACKIE, M. (2004) Filling institutional repositories: practical strategies from the DAEDALUS Project. *Ariadne,* (39). Available at: www.ariadne.ac.uk/issue39/mackie/

NIXON, W. J. (2002) The evolution of an institutional e-prints archive at the University of Glasgow. *Ariadne,* (32). Available at: www.ariadne.ac.uk/issue32/eprint-archives/

OCLC (2004) *The 2003 OCLC environmental scan: pattern recognition: a report to the OCLC membership,* Dublin, OCLC.

OPPENHEIM, C. (2005) Open access and the UK Science and Technology Select Committee Report Free for All? *Journal of Librarianship and Information Science,* 37(1)**,** pp. 3-6.

PEEK, R. (2004) The Cascading Effect. *Information Today,* 21(8)**,** pp. 17-18.

PEEK, R. (2005) Open debate on RCUK. *Information Today,* 22(9)**,** pp. 17-18.

PHILLIPS, H., CARR, R. & TEAL, J. (2005) Leading roles for reference librarians in institutional repositories: one library's experience. *Reference Services Review,* 33(3)**,** pp. 301-311.

PINFIELD, S. (2005) A mandate to self archive? The role of open access institutional repositories. *Serials,* 18(1)**,** pp. 30-34.

PINFIELD, S. & JAMES, H. (2003) The digital preservation of e-Prints. *D-Lib Magazine,* 9(9). Available at: http://www.dlib.org/dlib/september03/pinfield/09pinfield.html#1

PROSSER, D. C. (2003) Scholarly communication in the 21st century - the impact of new technologies and models. *Serials,* 16(2)**,** pp. 163-167.

PROSSER, D. C. (2005) Open access: the future of scholarly communication? *ALISS quarterly,* 1(1)**,** pp. 8-11.

RICHARDSON, M. (2005) Open access and institutional repositories: an evidence-based approach. *Serials,* 18(2)**,** pp. 98-103.

ROCKMAN, I. F. (2005) Distinct and expanded roles for reference librarians. *Reference Services Review,* 33(3)**,** pp. 257-258.

ROWLAND, F. (2005) How do we provide access to the content of scholarly research information? *Serials,* 18(3)**,** pp. 218-229.

Santa Fe Convention for the Open Archives Initiative. Available at: http://www.openarchives.org/sfc/sfc_entry.htm

SHEARER, M. K. (2003) Institutional repositories: towards the identification of critical success factors. *Canadian Journal of Information and Library Science,* 27(3)**,** pp. 89-108.

SPARC. Available at: http://www.arl.org/sparc/

STANESCU, A. (2005) Assessing the durability of formats in a digital preservation environment: The INFORM methodology. *OCLC Systems and Services,* 21(1)**,** pp. 61-81.

SUBER, P. (2003) Removing the Barriers to Research: An Introduction to Open Access for Librarians. *C&RL News,* 64(2). http://www.earlham.edu/~peters/writing/acrl.htm

TENNANT, R. (2002) Institutional Repositories. *Library Journal*, (September 15). Available at: http://www.libraryjournal.com/index.asp?layout=article&articleid=CA242297&publication=libraryjournal

VAN DE SOMPEL, H. & LAGOZE, C. (2000) The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine,* 6(2). Available at: http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

WAAIJERS, L. (2005) From libraries to "libratories". *First Monday,* 10(12)**,** [np]. Available at: http://www.firstmonday.org/issues/issue10_12/waaijers/index.html

WARE, M. (2004) Institutional repositories and scholarly publishing. *Learned Publishing,* 17(2)**,** pp. 115-124.

WHEATLEY, P. (2004) Institutional repositories in the context of digital preservation. *Microform and Imaging Review,* 33(3)**,** pp. 135-146.

YIOTIS, K. (2005) The Open Archives Initiative and Eprints Repositories. *b/ITe,* 22(2 supplement). Available at: http://www.sla.org/division/dite/bite/julaug2005/bitesupp07082005.pdf

YOUNG, J. R. (2002) 'Superarchives' Could Hold All Scholarly Output. *Chronicle of Higher Education*, 48(43), pp. A29-A30 Available at: http://chronicle.com/free/v48/i43/43a02901.htm

YOUNG, J. R. (2002) 'Superarchives' Could Hold All Scholarly Output. *Chronicle of Higher Education*, 48(43), pp. A29-A30 Available at: http://chronicle.com/free/v48/i43/43a02901.htm