McKay, Dana. (2007). Institutional repositories and their 'other' users: usability beyond authors. *Ariadne.52.*

Copyright © 2007 Dana McKay.

This is the author's version of the work. It is posted here with permission of the publisher for your personal use. No further distribution is permitted. The definitive version can be found at [http://www.ariadne.ac.uk/issue52/mckay/](http://www.ariadne.ac.uk/issue52/mckay/)
Institutional repositories' 'other' users: Usability beyond authors.

Dana McKay summarises the literature on the usability of institutional repositories, and points to directions for future work.

Introduction

If institutional repositories (IRs) were all that their proponents could have hoped, they would be providing researchers with better access to research, improving institutional prestige, and assisting with formal research assessment [1][2][3]. The reality, though, is that IRs are less frequently implemented, harder to fill, and less visible than their advocates would hope or expect [4][5][6].

While technical platforms for IRs, such as DSpace [7] and ePrints [8] have seen an abundance of research, little is known about the users of IRs— not how they use IR software, nor how usable it is for them. IR users can be split into three main groups: authors, information seekers, and data creators/maintainers; while authors are reasonably well understood, the latter groups are particularly under-studied.

Authors

Authors are better studied than any other users of IRs, perhaps because the first barrier to IR use is content recruitment, and authors are vital to content recruitment, even in IRs where deposit is performed by a third party. Several strategies to improve author involvement in IRs have been identified in the literature, and are summarised in Mark and Shearer's excellent review [9]. The literature on authors will not be further reviewed here.

Information seekers

Information seekers are the end users of any IR, and while there may be authors in this group (indeed, they may constitute the majority of this group), the goals and concerns of this group are very different from the goals and concerns of authors in their authorship role. Research shows that information seekers generally want to find information quickly and with a minimum of fuss [10][11][12], (though authors are differentiated by placing high value on peer-reviewed work [13]). It is clear that where information is freely available, information seekers are willing to use it, and trust it just as much as for-fee information [12]. Even authors are visibly willing to use 'free' published work; 88 percent report having used self-archived materials [14]. While IRs clearly have potential users, even researchers at a given institution are unlikely to know whether their institution has an IR [5][13]; hence IRs are likely to be largely unknown to researchers outside their host institutions (and certainly they must be unknown to the general public, who are among the purported beneficiaries of IRs [3]).
The first usability problem for IRs, then, is visibility; for IRs to be useful they must be seen, and it would appear at present that authors are quite right in perceiving them as 'islands' of information, set apart from those who might be interested in them [15]. This is a problem that can be addressed by search-engine harvesting of IRs, not just by Google Scholar (which attracts some use by academics, but is not usually the first information source they consult[13]) but by Google’s main service, which is the first stop for information for academics and the public alike [11][12][13]. This is not to say that other commercial search engines should be excluded; Google is mentioned by name here simply because it is the most popular [12][13]. Reinforcing the importance of search-engine harvesting is Nicholas' surprising conclusion that search-engine indexing of a journal is at least as important as making that journal open access in improving the journal's visitor numbers [16].

Other than this visibility problem, little is known about the usability of IRs for information seekers. A limited number of usability studies of IRs have been conducted (more on this below), but at the time of writing there are no known reports of actual usage of any IR. This dearth of usage data means we do not know: whether typical IR users are local or from outside the hosting institution; whether they find the IR via the institutional homepage or via search engine referrals; we do not know what kind of information they look for and use or how they use the functionality offered by IRs. While studies of IR usage would also be valuable, we can certainly learn from the usability studies of IRs and from the wealth of research about information seeking in other contexts (this, in particular, is likely to advance our knowledge of how best to design IRs [17]).

Usability studies of IR software

The work in this field is very limited; at the time of writing, only one complete report of a usability study of any IR with a focus on users could be found. That report is a comparative analysis of two of the big software players in the IR field, e-Prints and DSpace[18]. In this study, Kim performed a heuristic analysis of e-Prints and DSpace for a number of tasks (most of which involved searching for a known item), and then ran a between-groups study of users performing the same tasks. Kim predicted from the heuristic analysis that users would be faster and error rates would be lower for most tasks using DSpace (the reasons for this are analysed in depth in the paper); these predictions were proved accurate in the user studies. Despite the consistency of Kim's results, they are in contrast to Ottaviani's findings, which show problems with interface terminology and context indicators in DSpace in real-world use [19]. Kim's findings also differ from Atkinson's experience, in which researchers found it very difficult to perform one of their common real-life tasks using DSpace [20].

The implications of this work are not software-specific; we can see that heuristic analysis can give good usability predictions, and that usability studies of specific tasks can tell us about software performance for those tasks. However, when the results are contradicted by studies of users attempting real-world tasks, we see that understanding what users would like to do with software, and ensuring that these tasks are both possible and simple to do, should be a priority when developing IRs.
Usage studies of online research resources

Despite the lack of IR usage studies, we can gain some understanding of how users are likely to use IRs for information seeking by looking to usage studies of journal databases and open access research repositories.

Recent studies of large journal databases show that users read and download an unexpectedly wide range of material in comparison with the range of papers actually cited. Obsolescence is not nearly as pronounced in downloads as it is in citations; while there is some recency effect (particularly in the sciences) older articles are downloaded much more frequently than they are cited [21]. It is suggested that this may be a result of search engine use [21]. Article popularity is also not as clear cut as might be expected. In a one month study of a 'big deal' [22] nearly all available journals were accessed at least once, although the top half of the journals accounted for over 90 percent of the usage and three quarters of articles were viewed only once. Fewer than 1 percent of articles were downloaded more than ten times.

Studies of what people actually do with journal databases show that the typical user visits infrequently, views articles from just a single subject area, and views only a small number of articles [22][23][24]. There is a correlation between the number of items viewed and the frequency with which users returned, suggesting that there is a small but significant minority of 'dedicated researchers'. There is also a significant level of browsing, particularly of journals’ tables of contents, observed in the statistics, [23].

Usage studies of open access research collections in computing confirm the pattern of infrequent visitors who view and download only a small number of articles [25][26][27]. Moreover, users typically type in short queries (1–3 words), do not change default search options, and view only the first page of search results. Spelling errors are not often observed, but a number of queries returning few or no results are a result of less-popular local spelling variants (for example 'optimisation' versus 'optimization') [25].

While these studies do not investigate IR usage, the systems they describe are similar in purpose to IRs and we can reasonably expect information seekers to use them in similar ways. Given that assumption, we can infer that typical IR users will visit infrequently, download only a few articles at a time, perform very simple searches, and use results from the top of the results list (though they will browse widely in other ways if offered the chance). Conversely, as a group, IR users are likely to use a wide range of articles, not just those that are new or popular, because their searches will return a wide range of articles, not just the most recent or popular articles. This picture of users suggests search mechanisms should be easy to use, that search defaults should produce a wide range of results, and that results should be displayed with the best possible relevance rankings. IRs should also facilitate browsing (preferably of the whole collection, as well as search results), and provide the widest possible range of articles.

Information seeking behaviour

When looking for information, users do more that just use a search box, particularly if it is not clear to them exactly what they are looking for. Instead, they engage in a process that continues until they find what they want, find something close enough (otherwise known as 'satisficing ' [10]), or just give up [28]. This process has been
described in a number of models (see for example [29][30]), but the models are broadly similar and can be abstracted to six steps (though the process is often iterative). Those six steps are:

1. Perceiving a need for information
2. Investigating the ways in which the information need might be met, including assessing available sources of information, and possible searching and browsing for preliminary results
3. Clarifying the information need to a small number of specific questions, based on the available resources and personal interest
4. Querying information sources to meet the need
5. Browsing and assessing results
6. Assimilating results and refining queries if the information need remains unmet.

It is important to recognise this process when designing an information system, and to support it as much as possible. An example of an 'information system' that supports this process well is a library reference desk, where a librarian facilitates information seeking [28][31]. In terms of IR usability, this process suggests we should include browsing functionality (to help users with assessing the information source, and clarifying their information need), and that we should allow users to interleave searching and browsing (this reflects the iterative nature of this process, and is supported by [32] which shows that users of a journal database do interleave searching and browsing).

**Data creators/maintainers**

An IR's data creators/maintainers (henceforth referred to as 'data maintainers') are those who create metadata, upload documents, and generally contribute to or oversee the IR's document collection. Data maintainers may be librarians; the group may also include authors at institutions where author self-deposit is used. Very little research attention has been paid to this group, particularly in the usability field, yet they are vitally important to the creation and maintenance of any IR. Moreover, data maintainers are engaged in an entirely new role. This role is likely to require some combination of technical expertise, an understanding of metadata and metadata standards, copyright knowledge and the inclination to collate research publications[33]. There is no comparable role within any other information system, particularly when it comes to author self-submission (authors who deposit in subject archives such as arXiv.org are sufficiently highly motivated to ignore usability problems[34], while it is difficult to motivate authors to submit to IRs at all[4]); thus there is no other research we can draw on to bolster our limited understanding of this role.

**Leadership**

Librarians have demonstrated leadership in the IR field [4][6][35][36], and creating IRs [37][38] and encouraging OA mandates [37] is seen as a way forward for libraries in an age of digital information. There are a number of reports describing how well-suited librarians are to IR involvement [11][39][40][41][42], and how they provide a wide range of necessary functions, including overcoming publisher and academic
resistance [38], providing good metadata standards, and pushing for inclusion in external search services [41]. Not only do librarians possess the necessary skills to provide leadership in the establishment of IRs, Carver [40] posits that libraries are best placed to use these skills, being at the nexus between published work, academics, and information access.

The benefits of librarians' involvement in IRs do not all flow one way, however; documented benefits to libraries include greater visibility within their research communities[37], opportunities to provide more tailored services for their patrons[41], and improved research collaboration with other libraries[43].

Despite all the potential benefits, it would be foolish to suggest that IR leadership never has any negative impact on libraries. Those with experience caution about the amount of staff time that may be absorbed by IRs [36][38]; moreover, Piroun warns about the high level of technical expertise required by some systems (expertise not readily available in all libraries) [36]. There are a number of reports about the inertia of academics and their resistance to involving themselves in IRs [4][15][40][44][45], meaning that librarians must provide leadership, because they are the only people suited and available to do so. Finally, despite resolute predictions that self-archiving would make research literature free for the taking [1], not one of the library publications about IRs mentions a reduction in journal subscriptions (and hence cost) as a benefit.

Usability for data creators and maintainers

Unfortunately, as with information seekers, there are only a limited number of reports of IR software usability for data maintainers. Despite assertions that authors can be easily trained to submit their own work[46], usability reports about IR software are predominantly negative, both in terms of what users can do with the software and how the software appears and behaves (though it should be noted that the research to date only covers DSpace and ePrints).

The main problems for data maintainers reported in the literature are:

- **Terminology:** The terminology used in the deposit and management interfaces of DSpace and ePrints was confusing and inappropriate for both authors and librarians (though in different ways) [19][20][47].

- **Process:** to deposit an item or to update its metadata, users were required to click through a number of screens. Often users needed to see only one or two of these screens, but they were nonetheless forced to follow the same linear progression through the screens. Even in early trials, users saw this as tedious and frustrating [19][20][47].

- **Metadata requirements:** In one study, the detailed metadata input fields displayed by ePrints and DSpace in their document deposit interfaces were daunting to both academic staff and librarians. Both groups complained that they often did not have all the metadata, and that it was not clear from either system which elements were required and which were optional [47]. It is worth noting that the requirement of detailed metadata at the time of deposit is also raised as a problem in the design stage of a image repository system [48].
- **Detail suppression:** One group studied reported having difficulty making sections of a record private when using DSpace. This limited what they were able to put in the database [19].

- **Formatting and authority control:** Examples of formatting and authority control problems include, whether record titles were entered in title case or lower case, and whether author names were entered first name or last name first [47]. Authority control for author names in particular has also been mentioned by bloggers working in the IR field [49][50], though requiring depositors to create authority versions of names during the deposit process has been shown to be unwieldy (at least for an image repository [48]). This issue is not only frustrating for librarians or authors who must enter details of a work, but may also cause problems for information seekers who type in the 'wrong' variant of a name. (It should be noted that ePrints version 3 has an 'auto-complete feature' that will help automate authority control [51]. This version of ePrints was not the subject of any of the reports reviewed here).

- **Research reporting:** A group of librarians who wished to use DSpace for research reporting (not an uncommon task for IRs [2][35]) found the search and browse functionalities 'weak', and the search results display awkward [20].

It should be noted that while DSpace is more heavily criticised than ePrints, it is also more widely tested and thus may not be any less usable for data maintainers (though in a comparative test between DSpace and ePrints, a slight preference for ePrints emerged among both librarians and author-depositors [47]).

This literature reflects serious usability issues that may engender resentment among librarians taking time out of other duties to maintain IRs, and may also discourage authors who are ambivalent about self-deposit at best. However, at present it is impossible to calculate the real impact of these usability problems because we do not know how data creation and maintenance fits into the work practices of the authors and librarians involved (despite claims that self-archiving should take authors less than one hour per year each [46], and criticism of authors for not doing it [52]).

Libraries and librarians currently display a high level of commitment to IR data creation and maintenance; if this level of commitment is to be retained, it is necessary to pay attention to the needs and experiences of librarians. Conversely, if the commitment of authors is to be increased, it is necessary to make their initial experiences at a minimum not frustrating and daunting. This means not only improving the usability of data creator and maintainer interfaces, but also understanding how the work involved in data creation and maintenance fits into the way people involved do their jobs, and make this fit as streamlined as possible (for an example of how work practices of users were taken into account during the design of an image repository to great benefit, see [48]).

**Conclusions and future directions**

In reviewing the literature about IR use and usability, we see that authors are well studied, and that there are a number of proven methods of engaging them. However, there are two other user groups for IRs that have not attracted nearly so much attention thus far—information seekers (or 'end users') and data creators and maintainers.
Information seekers, while they are not closely studied with respect to IRs, are well studied in general, and by understanding both the information seeking process and the behaviour of this group in similar systems, we can make predictions about how they may use IRs. Their visits are likely to be short, with short searches, and they are likely to view only a few articles. They will make use of browsing features, if they are provided, and this could lead to better information seeking. Collectively, they will use a wide range of articles spread over a long time period. Even academic information seekers use Google (or other commercial search engines) first; IRs that are harvested by search engines will see a higher level of use than those that are not. All that we know about information seekers should be incorporated into our design of information seeker interfaces within IRs, but the few usability studies available suggest this has not been the case.

Data creators and maintainers have been largely ignored in the literature; despite their role being completely new, we know very little about how it fits into other job responsibilities and expectations. The suitability of data maintenance interfaces to their users is also largely unknown, though the available literature suggests that improvements are needed in this area.

Future directions

For IRs to attract the level of use they need to revolutionise digital scholarship, they must be both useful and usable.

For IRs to be useful, they must first have information in them, and little is known about their usability for the group (data creators and maintainers) who create the information. We know that this group is made up of librarians and authors, for the most part, but we do not know how the work of populating an IR fits into their workflow. Thus far, usability reports have been largely negative, and we can make suggestions as to how to avoid these mistakes, but we cannot make general suggestions for good design of IRs from a data maintainer's perspective. Observational studies of data maintainers would provide an understanding of the tasks IRs are used for and the way they fit into data maintainers' larger work roles, and also suggest ways of improving the fit of IR software to the tasks for which it is used. More formal usability testing could then be used for fine-tuning the design of data maintainers' IR interfaces.

 Virtually nothing is known about IR end users. We do not know how many people are using IRs, whether they are academics or laypeople, or how they most often find IRs, though it is reasonable to suspect they may find IRs with Google, given that this is the starting point for most information seekers. Studies of usage logs from a well-populated IR could answer these questions, and provide avenues for further investigation into how we might improve information seekers' IR experiences.

Acknowledgements

The work to prepare this review was funded by the ARROW project (Australian Research Repositories Online to the World, http://www.arrow.edu.au). The ARROW project is funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education.
References

Author Details

Dana McKay
Usability Researcher
Swinburne University of Technology,
Melbourne, Australia

Email: dmckay@swin.edu.au

___________________________________________________________

Keywords: please list in order of importance the keywords associated with the content of your article (usually up to 8). These will not be made visible in the text of your article but may be added to the embedded metadata:

1. Usability
2. Institutional repositories
3. Information seeking
4. Libraries
5. Authors
6.
7.
8.
9.
10.
11.


