PEER D4.1 Behavioural Research: Authors and Users vis-à-vis Journals and Repositories - Baseline report

Jenny Fry, Charles Oppenheim, Steve Probets, Claire Creaser, Helen Greenwood, Valérie Spezi, Sonya White

PEER Behavioural Research:

Authors and Users vis-à-vis Journals and Repositories

Baseline report

Jenny Fry, Charles Oppenheim, Steve Probets
Department of Information Science

Claire Creaser, Helen Greenwood, Valérie Spezi, Sonya White
LISU

September 2009

LISU Research & consultancy for performance management
Information, cultural & academic services

Loughborough University

Contents

1 Introduction

Ways in which scientific and scholarly knowledge is created and disseminated are undergoing radical change in light of new digital technologies, though the extent and pace of this change is not uniform across disciplines (Tenopir, 2003; Harley *et al*, 2008). Something akin to a paradigm shift in the scholarly communication system has also been underpinned by initiatives such as open access, open science and data sharing (Borgman, 2007). Increasingly, funding agencies are implementing mandatory deposit of the outputs from publicly funded research, and whilst knowledge as a national public good is a compelling case for open access, in practice it does not always sit well with institutional structures, disciplinary practices or incentive systems. Even where mandates do not exist, funding agencies, and increasingly higher education institutions, have open access policies based on voluntary self-archiving. PEER has identified a lack of evidence of the potential impact that the systematic archiving of research outputs in open access repositories might have on the ecology of European research and the most efficacious model for doing so.

The Behavioural research: Authors and users vis-à-vis journals and repositories project was commissioned by PEER in April 2009 as part of a broader initiative to investigate the effects of the large-scale, systematic deposit of authors' final peer-reviewed manuscripts (also called stage-two research outputs) on reader access, author visibility, and journal viability, as well as on the broader ecology of European research. The specific aim of the behavioural research is to understand the extent to which authors and users are aware of open access (OA), the different ways of achieving it, and the (de)motivating factors that influence its uptake.

Below we briefly set the context for the behavioural research, indicating current understanding of new forms of scholarly communication, information behaviours in digital information environments, repository development, perspectives on open access to scientific and scholarly knowledge, and authors' attitudes to open access. This context is based on a preliminary literature review that informed the data gathering and analysis. For the purpose of presenting initial analysis of our empirical data the related literature is used in an illustrative way, and is not intended to be read as a comprehensive review.

1.1 Scholarly communication

Scholarly communication is commonly defined as the process by which information is exchanged between scholars and contributions to knowledge are attributed to specific individuals, validated by peers, disseminated throughout the discipline and beyond, and archived for the scholarly/scientific record (Mabe, 2006; Pinfield, 2009; Borgman, 2007). Sociological definitions of scholarly communication take reward structures (e.g., peer recognition, career advancement, research evaluation, research funding) into account in defining its characteristics (Hagstrom, 1965; Prosser, 2005). Traditionally, scholarly communication in the sciences has been supported by peer-reviewed journals, although communication may also take place through many other channels, such as books, monographs, conference proceedings, reports and manuscripts. Communication can also take place via less formal channels, such as emails and seminars as well as novel forms of dissemination such as Web pages and e-print repositories (Borgman, 2007).

The scholarly communication system is currently in a state of flux. The advent of electronic publishing, the Internet and social networking tools mean that the choices available to a scholar for communicating both preliminary and final research results to the scholarly/scientific community, and, potentially, to the public at large, have dramatically changed. The traditional methods of preliminary discussions with colleagues, presentations at conferences and submission of manuscripts to journals or for monograph publishing have expanded significantly (Oppenheim, 2008). Electronic publishing and open access have challenged publishers' business models (Toledano, 2003), which have been traditionally based on what Hagstrom (1965, p.12) describes as 'gift-giving', whereby authors make a contribution to a journal in return for the social recognition it engenders, rather than financial payment or royalties. Some argue that this practice of 'gift-giving' is indicative that scholarly communication has always rested upon the concept of open science, which Borgman defines as *"the principle that a free flow of scholarly communication benefits both scholars and the society at large"* (Borgman, 2007, p.65).

By-products of research, such as datasets and methodologies to process large-scale datasets, are increasingly recognised as research outputs, most notably within an e-research framework, but evaluation of researchers' productivity and research impact still relies on peer-reviewed journals and traditional metrics (Fry and Thelwall, 2008). The impact of electronic publishing on dissemination has been twofold; firstly, electronic publishing has helped speed up the diffusion of knowledge and thus drastically reduced lengthy publication lags in some disciplines, thereby enabling faster access to articles. Secondly, it has allowed the storage and retrieval of scholarly output to be made much easier. Whilst acknowledging the influence of new information technologies on the way scholars locate and disseminate material, Borgman (2007) also notes that, overall, the main functions of the scholarly communication system have changed very little and there is a notable degree of continuity in scholarly publishing.

1.2 Information behaviours

Networked digital resources are not isolated entities, they exist in a broader information landscape of competing/complementary resources and large-scale access resources, such as *Web of Knowledge, JCE (Journal of Chemical Education) Online* or Institutional repositories, such as MIT's DSpace (Palmer, 2005). Being digital typically infers being interlinked and these inter-linkages form dynamic "connective structures" (Palmer, 2005) that are shaped by multiple stakeholders in the scholarly communication process. It is likely that the location of an existing or emergent information resource within a discipline's connective structure will influence its uptake and use.

Information seeking serves a number of functions in the creation of scientific and scholarly knowledge. For example, Palmer and Neumann (2002) found that, in interdisciplinary fields, there is a need for translation work and boundary crossing across disciplinary information landscapes. This is where connective structures might be most influential in how researchers locate sources, since variation in language systems across disciplines can lead to what Palmer and Neumann (2002) describe as "excavating" work - the tracing of intellectual paths through resources and sources. The need for excavation work is compounded in those intellectual fields where relevant material is scattered across diverse disciplines and resources. To give an example of how this plays out across broad disciplinary groupings, Vakkari and Talja (2005) found that in medicine 52% of faculty used publications mainly from their own field, in

engineering this was 40%, and in the social sciences only 21% mainly used publications in their own field.

Extending earlier work on interdisciplinary and humanities scholars, Palmer (2005) compared information behaviours between the humanities and life sciences and identified two interrelated types of search: confirmation and discovery searching. “Confirmation searching” is employed when a researcher wishes to confirm their existing ideas or rejuvenate memories of related work. This often involves known item searching or performing basic current awareness in order to refresh and affirm ideas. Whilst Palmer (2005) found this type of search was more common amongst humanities scholars, life science scholars also performed confirmation searching, but searches of this nature were typically directed, using well-established key words and searching core disciplinary resources. “Discovery searching” was commonly employed by interdisciplinary scholars and scientists to probe outside of their main research area. Hence, discovery searching involves searching “intellectually distant” or unknown resources, typically non-specialist such as aggregated literature databases or Internet search engines.

Information seeking and reading behaviours of researchers have been the focus of many studies over the last few decades, notably since the advent of electronic publishing and networked digital resources, and there is now a substantial body of work concerned with information search patterns. With researchers generally heavily reliant on digital information resources to locate and access research, scholarly communication systems have become highly complex, with a mass of digital means by which content can be retrieved and accessed (Palmer, 2005). In most disciplines the choice of digitally accessible resources is rich, increasing both the reach and accessibility of scholarly communication, and, at the same time, making it more complex, as boundaries between unpublished/published and different producers become blurred. The choice includes commercial aggregators, such as Web of Knowledge or ScienceDirect; collaborative initiatives from various stakeholders in the scholarly communication system, such as BioOne; institutional initiatives, such as institutional repositories, subject-based repositories; and digital access resources produced by researchers for the benefit of the research communities to which they belong (Palmer, 2005).

Given the increased competition between information gatekeepers and the limited time available to researchers working under increasing institutional pressures, there has been a great deal of concern regarding the uptake of networked digital resources, particularly what Duranceau (2008) refers to as non-market resources such as institutional and subject-based repositories. For example, in a French nationwide survey of mathematicians and computer scientists, Wojciechowska (2007) was specifically concerned with ways in which researchers became aware of publicly accessible repositories and found that colleagues were the most frequent source of information (42%), with universities being the least frequent (0.8%). In terms of searching the web for full-text research articles Wojciechowska (2007) found Google (66%) and arXiv (66%) to be the most popular, with personal web pages also being quite important. The respective roles of traditional/emergent and market/non-market gatekeepers varies across disciplines (Fry, Virkar, and Schroeder, 2008) and in a scoping study of the digital information landscape within high-energy physics Gentil-Beccot *et al* (2009) illustrate that in some disciplines market-based resources are only minimally used to locate information sources and instead subject-based repositories are an integral part of the daily workflow. Regarding how researchers locate open access materials more generally, King *et al.* (2009) confirm the findings reported above that scholars generally identify and locate open access publications following discussion with peers or through online searching.

Studies have shown that the use of digital resources such as e-journals, e-books and databases may significantly impact on scholarly work and productivity. Thus Vakkari (2008) found that, in the case of digital libraries, the improved accessibility, availability and range of electronic literature was perceived by researchers as having a definite positive impact on their research, with variations across disciplines. He also found that increased access to electronic literature was perceived by researchers as having a positive impact on their total number of publications as well as the number of international publications, again with variations across disciplines and research experience. Vakkari's findings corroborates results from earlier studies (Hesse *et al.,* 1993; Barjak, 2006; Brown *et al.*, 2007) and thus may be understood as part of a more general trend. Academics' perceptions of increased productivity linked to digital access to resources have been recently validated by Rowlands *et al.* (2009) who, through their extensive analysis of log files collected in 10 UK higher education and government institutions and covering about 1500 journals, found that for each institution studied there exists a strong positive correlation between academics' use of e-journals (including number of pages viewed) and their productivity in terms of number of articles published.

Whilst Vakkari (2008) and Rowlands *et al.*(2009) focused on the positive impact of digital resources on scholarship, Evans (2008) has claimed that scholars' information seeking behaviours in networked digital environments may have consequences for narrowing down the breadth of knowledge and scholarship, due to the efficiency of online searching and indexing afforded by aggregated access resources. In other words, the fact that browsing is replaced by advanced search capabilities will influence how scholars locate sources, the type of sources they locate, relevance judgements and potentially citation patterns. Whilst it would be intuitive to think that the greater online availability and accessibility of journal articles (notably backfiles) and journal titles would lead to an increased number of articles being read and a wider range of journal titles being cited, Evans (2008) has observed the opposite trend with regard to citation practices. Evans (2008) found that the more back issues are made available online, the more recent issues are cited, and the more journals titles/articles are made available online, the less journal titles/articles are actually referenced. This finding contrasts with studies of reader behaviours reported below, which indicate an increase in the number of journal articles read year on year, so it may be the case that it is not feasible to extrapolate from number of sources cited to number of sources read. According to Evans (2008), online searching of digital resources is more likely to bring scholars to the *'prevailing opinion',* and so narrowing down the breadth of Science, as non-consensual opinions and ideas get forgotten more quickly in such an information landscape. Indeed, as Evans puts it, "by enabling scientists to quickly reach and converge with prevailing opinion, electronic journals hasten scientific consensus" (Evans, 2008, p.398). It should be noted that Evans' conclusions have been challenged by open access advocates. Looked at in the context of search preferences identified by Palmer (2005), with scientists tending to adopt directed search strategies, Evans' argument may be indicative of a trend towards increasingly specialised search strategies.

In relation to digital resources, Tenopir and colleagues have extensively studied the information seeking, reading and use patterns of a large number of research communities spread over many disciplines, both within the university sector and beyond. With journal articles at the heart of scholarly communication, electronic publishing has affected the way researchers search, identify and read journal articles. Tenopir and colleagues have found that there has been a change in the way readers locate information and access articles. Information retrieval is today mainly done through online searching, whether it be using search engines or citations, abstracts and full-text databases. Tenopir *et al.* (2009a) have argued that journals are by and large the

first source of information used by readers, with 93.9% of researchers' readings being of journal articles. For roughly half of the journal articles read, the information held in these articles was already known by readers, and they had known about it mostly via other journal articles. Tenopir *et al.* (2009a) also found that there has been an important increase in article reading over the last decade, coinciding with a greater availability and increase of electronic journals, and the need for scholars *'to read more merely to keep up with the same proportion of the literature'* (Tenopir *et al.,* 2009a, p.27). Their findings show that, on average, readers read about 280 articles per year, most of them provided by their library, against 216 in 2000-02, and 150 in 1977, and that 48.5% of their readings are used in their research, and a further 10.8% are used in the writing of articles, proposals, reports etc. Based on the time spent on reading and the purposes of reading, Tenopir *et al.* (2009a) conclude that journal articles remain the main channel for scholarly communication. Whilst journal articles undeniably play an important role in the research life-cycle and scholarship, there however exist significant variations in reading patterns, and information seeking patterns, across faculty. Such variations in reading patterns may be attributed to specific readers characteristics such as subject discipline and work responsibilities, and to a lesser extent, to age, productivity and purpose of the reading (Tenopir *et al.*, 2009b). For example, readings by faculty in Engineering and Sciences are undertaken for the purpose of research whilst reading patterns of Social Sciences and Humanities faculty tend to be primarily teaching-driven. With regards to work responsibilities, Tenopir *et al.* (2009b) have shown that, while reading times per article are similar across faculty, the amount of reading is higher for research-active faculty compared to faculty primarily involved in teaching.

1.3 Open Access

The concept of 'open access' has become much more prominent with the development of electronic scholarly communication tools. Open access was never realistic in a print-based scholarly communication system because there was no sustainable business model for such an approach. Open access is defined as the free and unrestricted access to online digital literature by anyone with a networked computer from anywhere (Budapest Open Access Initiative [BOAI], 2002). Further definitions have been put forward by the *Bethesda Statement on Open access Publishing* (2003) and the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (2003), but the BOAI definition is sufficient for all purposes. The other definitions specify further criteria that apply to open access materials.

There are two types of open access, the 'open access publishing' strategy, also called the 'gold' road, as open access is reached at the level of publication without further action from stakeholders, and the 'self-archiving' strategy, referred to as the 'green' road, which requires action from one of the stakeholders (authors, publishers, institutions, etc.) at some point in an article's life cycle (pre-print or authors' final version or publishers' edited file) to deposit the article in an appropriate repository. It is very unfortunate that some commentators equate the gold road to open access (Oppenheim, 2008).

Open access journals follow the traditional model of scholarly publishing. Articles are submitted, (normally) peer-reviewed and published in the same way as for subscription journals, the only difference being that articles are made freely available at the point of access. Open access journals are always available in electronic form, and sometimes also in print form. There are a number of different business models for open access publishing, and so far there is not a consensus approach. In some cases, the open access journal is run on volunteer effort, with

few if any explicit costs and revenues involved; in other cases, the journal is intended to break even, or make a profit, with income coming from the authors or someone acting on behalf of the authors. That income might be at submission stage, at acceptance stage, or at both stages. This latter model is commonly called the 'author-pays' model, although payments are typically made by the authors' funding bodies or employing institutions rather than the authors themselves (Oppenheim, 2008). Some traditional subscription-based journals also offer authors a choice whereby they can make their articles available via open access without an embargo period, for a fee. This is often referred to as the 'hybrid' model.

The green model for open access is generally achieved through authors' deposit of a copy of their accepted peer-reviewed articles in a repository, be it an institutional repository or a subject-based repository. Authors generally post the final version of their peer-reviewed articles, i.e. the edited version after peer-review, also called stage-two manuscript. Some journal publishers also allow their authors to post the publisher's version, i.e. the PDF file which has been copy-edited, typeset and page numbered. Researchers within economics and physics have been early adopters of self-archiving as a complementary route to accessing scholarly works. It is common practice in these disciplines to upload pre-prints (papers not yet peer-reviewed) whilst submitting them to journals for consideration.

A particularly contentious matter of policy relating to open access is that of mandates. Mandates are being imposed by some funding agencies and by some Higher Education Institutions on scholars, requiring the scholar to make outputs open access, whether through the gold road or the green road. To date, anecdotal evidence shows that a majority of authors do not seem willing to go down the open access route of publishing voluntarily, although Swan and Brown (2005) found in their survey that 81% of authors would comply '*willingly*' if they were required to deposit a copy of their publications in their institutional repository. This may be for a number of technical and social reasons, or simply because of inertia. Whatever the reason, proponents of open access have vigorously promoted the concept of mandates, and indeed, have succeeded in changing the law in the USA so that outputs from research funded by the National Institutes of Health has to appear in open access form, after an embargo if need be.

There is some evidence that systematic deposit of authors' final versions (i.e., including amendments after peer review – stage-two manuscripts) might eventually lead libraries to cancel some of their journal subscriptions (Beckett and Inger, 2007), as librarians might choose to attach little importance to whether it is the publisher's version that is accessed by their readers or the author's final version. They have supposedly the same content, but critical differences may occur, as pointed out by Wates and Campbell (2007). However, it should be stressed that currently there is no evidence of librarians cancelling subscription-based journals due to the ready availability of open access materials.

In any case, authors such as Pinfield, believe open access repositories can co-exist harmoniously with scholarly publishing, as *"it has recently become apparent that there is a potential for repositories and journals to interact with each other on an ongoing basis and between them form a coherent OA scholarly communication system"* (Pinfield, 2009, p.165). Although publishers claim that their dissemination role often eclipses other value-added functions they perform in the article lifecycle and scholarly communication more generally, the impact of electronic publishing and open access has nevertheless disrupted publishers' monopoly of dissemination of scholarly works. As Velterop pointed out *"dissemination is just*

one of the many functions performed. Dissemination, however, is the one on which the economic value of a publisher is built" (Velterop, 2008, p.119).

With the advent of electronic publishing, some have questioned the legitimacy of the significant increase in subscription prices charged by publishers (the so-called 'serials crisis'), and raised the question of whether electronic publications should be cheaper or even free. Against such claims, Mabe (2006) argues that electronic publishing has not reduced publishing costs but only changed their structure. Publishing costs can be separated into 'fixed costs' and 'variable costs'. Fixed costs are incurred by the creation of the first copy, involving activities such as copy-editing, typesetting etc., whereas variable costs occur in activities such as reproduction and dissemination; Mabe concludes that savings made in variable costs by electronic publication have been offset by the emergence of new costs pertaining to the acquisition and maintenance of the digital technology and related processes, as well as the creation of new value-added services. Mabe also points out that publishing costs have generally increased for publishers maintaining both print and online media for the dissemination of scholarly journal articles.

The following advantages for open access are often cited in the literature:

- greater access to scientific literature for all research communities, especially researchers in developing countries
- greater speed of dissemination of knowledge
- citation advantage
- the general public's right to access publications based on public-funded research (as well as greater accountability of public money)
- advancement of Science in its most general sense

Publishers, however, can point to the significant added value they provide to the scholarly communication process and the many initiatives that make large numbers of scholarly articles available at a very reasonable price, often with special discounts for those in developing countries.

It is worth looking at one of the alleged advantages of open access – the so-called citation advantage – in more detail, because it is so controversial. Citation studies have attracted considerable attention in the last few years, with their use as a measurement tool to respond to increasing pressures to measure the research impact of individual scholars or research groups. However, no consensus has yet been reached on whether there exists an open access citation effect, and the extent and nature of such an effect if it exists. Research on open access citation advantage is still a developing area, with the majority of studies showing such an advantage and a minority showing none. In light of the many different conclusions put forward in the numerous open access/citation counts studies found in the literature (some of which are detailed in the next paragraph), evidence of an open access advantage in terms of citation counts remains to be clearly identified and spelt out. While acknowledging the existence of a possible citation effect, Velterop (2008) argues that any citation effect can be only a temporary advantage that is bound to disappear as the volume of open access material grows towards 100%. Generally, making articles available via open access leads their authors to receive more citations, which, it is believed, reflects the impact of those authors on their discipline and thus the quality of their work. There could, in fact, be a number of reasons for the advantage; it may

be due to the fact that authors tend to put their best work in open access form; it may be because open access materials, whether green or gold, are easily picked up by Google and Google Scholar, which are the search engines of choice for many scholars. It may be that open access articles appear earlier than subscription-based articles and so have more time to accumulate citations.

The positive correlation observed between open access and increased citation levels has sometimes been interpreted as a direct causal link (Harnad and Brody, 2004). Others, for example Kurtz *et al.* (2005), Davis and Fromerth (2007) and Moed (2007), have proposed other reasons to explain the observed citation count differential, such as a subconscious selection from authors to self-archive their best quality papers or the availability of pre-prints which then attract earlier citations. In their review of the literature on open access and citation impact, including early studies from Harnad and Brody (2004), Antelman (2004), and Hajjem, Harnad and Gringas (2005) to more recent studies (Moed, 2007), Craig *et al.* (2007) have scrutinised both the conclusions and the methodologies supporting those conclusions. They concluded that differences in citation counts between open access articles and non open access articles could often be ascribed to methodological biases both in selection and citation analysis, notably in earlier studies, and that a causal link between open access and citation impact remained to be demonstrated. In a recent study conducted by Norris, Oppenheim and Rowland (2008), it has been found that open access articles attracted more citations than those kept behind subscription barriers: 49% of the articles selected had an open access copy available and those open access articles had, on average, a citation count 157% higher than toll access only articles. The study also confirmed earlier results showing that the open access advantage varied greatly across disciplines (Antelman, 2004; Hajjem, Harnad and Gringas, 2005). Norris, Oppenheim and Rowland (2008), however, concluded that, despite evidence of an open access advantage across disciplines, the underlying cause for the association between open access and a greater number of citations had not been identified. Other studies, such as that of Davis *et al.* (2008), have shown that whilst it is true that open access articles do attract a greater readership, the assertion that open access articles attract more citations than subscription-based articles is not validated, at least in the first year following publication which was the focus of their study. Evans and Reimer (2009), in yet another extensive citation counts study, have reported that the actual open access advantage may be much less than had been suggested in earlier studies. They have found, for instance, that recent publications attract a small open access advantage (circa 8%). They have also reported great variations in the impact of open access across disciplines with some disciplines, such as multidisciplinary research in the natural sciences, benefiting greatly from open access while the effect of open access on other disciplines with a strong established culture of sharing pre-prints, like physics or the social sciences, is negligible or nonexistent. Evans and Reimer (2009) argue that the benefits of open access are, by and large, not confined to variable increases in citation counts but rather lie in the opportunities that open access offers for wider participation in science.

As pointed out by Tenopir (2007), in any case, citation counting has flaws and does not always reflect the quality of the work of individual scholars. It would be fairer to say that citation counts reflect impact, and this is not necessarily synonymous with quality, although often it is. Other measurement tools, such as the h-index or other multi-dimensional metrics, are being developed to help measure better the quality of individual scholars as pressure to evaluate scholars' work grows.

Another controversial matter in the open access debate is cost savings. Claimed potential savings offered by open access publishing are of interest to higher education institutions wishing to mitigate the concurrent effects of increasing[1] journal subscription prices and library budget cuts. However, whilst it is well acknowledged that open access may have a global positive impact on research (and funding of research) worldwide, it remains to be seen how such global positive impact would spread at national and institutional levels. Indeed, it is yet unclear how cost savings in an open access environment would play out at national and institutional levels. Moreover, it has been argued that the 'societal' benefits of open access must be checked against the costs of switching to an open access environment, to determine whether or not it is beneficial to society as a whole (Anderson, 2007). Studies of cost savings in an open access environment have so far provided us with some significantly different conclusions. A 2008 RIN study (RIN, 2008) has found that, whilst the author-pays model may achieve significant cost savings globally in terms of production, distribution and access, it remains unclear, for instance, whether the UK higher education sector would directly benefit from this model as research-intensive institutions could see their contribution towards author-pay fees increase significantly, thus cancelling research libraries' savings achieved through the diminution of subscription costs made possible with electronic-only access. On the other hand, a second study, carried out in 2008 for the JISC, compared the costs and benefits of various scholarly publishing models and estimated that a shift from subscription journal publishing to open access journal publishing could have brought the UK higher education sector as a whole savings of up to £80 million per annum in 2007, whilst a shift from subscription journal publishing to open access with overlay services could have brought savings up to £116 million (Houghton *et al.*, 2009, p.224 225). Despite such an apparently compelling case, issues remain surrounding the costs associated with moving to an open access only publishing model, including the question of who is going to pay for authors to publish in open access journals, as well as the costs for libraries to maintain journal subscriptions whilst the transition to open access takes place. Furthermore, initial responses to this report were polarised, with critics concerned about certain costs not being taken into account in the model, and about the accuracy of some of the figures input into the model. Further research in this area would help to validate, or invalidate, these preliminary results, and would also help develop the understanding of the economics of open access publishing.

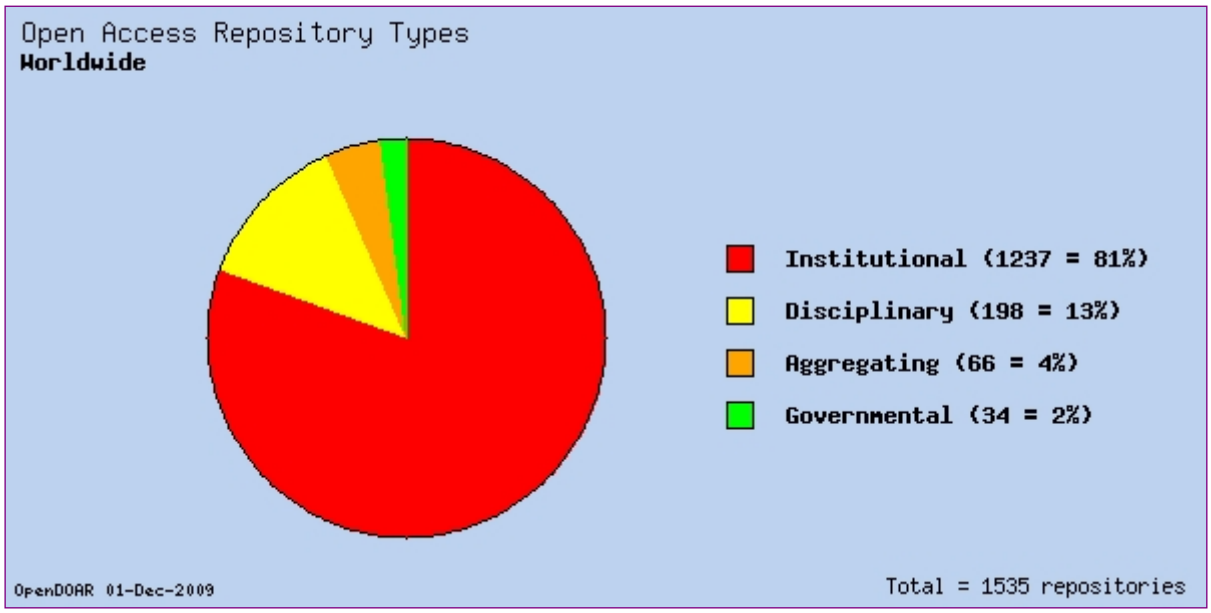
1.4 Repositories

As indicated earlier, the development of networked digital resources has introduced new opportunities for scholars to disseminate their work in ways that complement traditional scholarly communication systems. Publicly accessible repositories are the direct application of new technologies to enhance both informal and formal scholarly communication. Repositories can help mitigate against some of the perceived imperfections of the traditional scholarly communication system, for example they can reduce publication lags by enabling authors to deposit a version of an article as soon as it is complete, as well as reducing restrictions on access to the scientific and scholarly literature (in terms of price and availability), by opening up research outputs to greater readership and enabling widespread dissemination of scholarly works. Simpson and Hey (2006) claim that many types of repositories exist, and that researchers eager to make their research outputs available via the repository route could be faced with the 'problem of repository choice' (Simpson and Hey, 2006, p.226). Despite this

[1] Though there is a clear trend towards small or even zero increases in prices for 2009-2010, no doubt caused by the recession

claim many common repositories can be considered to be either subject-based (or disciplinary) repositories or institutional repositories.

Figure 1.1 Open access repository types



Source: OpenDOAR (1 December 2009)2

As shown in Figure 1.1, publicly accessible repositories, interchangeably called open access repositories, are mainly of two kinds, subject-based and institutional repositories, and primarily rest upon the principle of self-archiving (as opposed to harvesting repositories).

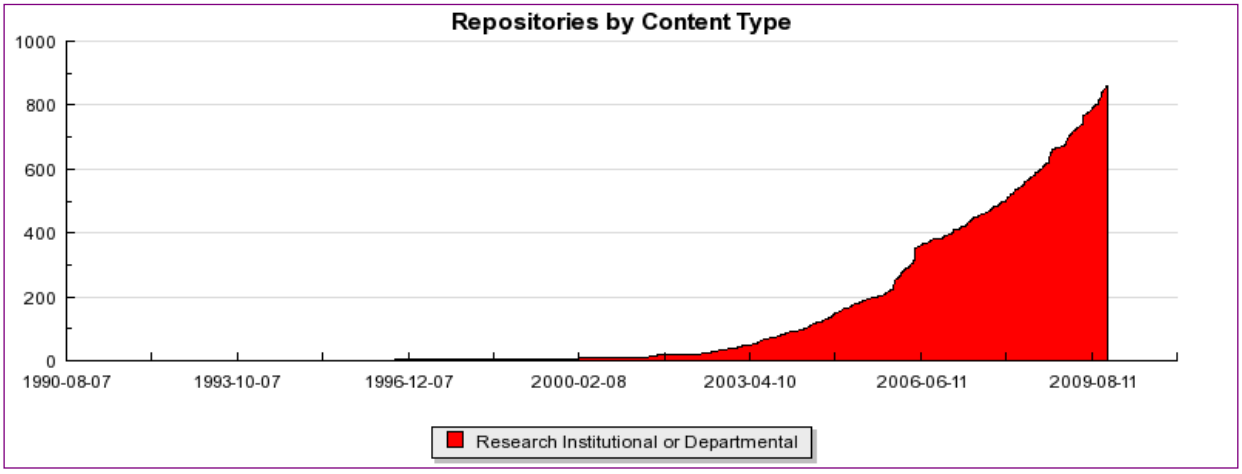
Subject-based repositories were pioneered in the early 1990s within specific disciplines as a response to a desire to speed up and facilitate communication between research groups and research communities. The computer sciences and high-energy physics communities have been at the forefront of the movement in subject-based repositories by making their research outputs publicly available. To a certain extent, it could be argued that this movement has grown out of the practice of making papers and other types of output available via personal or research group web pages, particularly in the computer sciences. In many disciplines this is still a norm, and researchers perceive limited benefits to changing this practice and adopting the use of subject-based or institutional repositories instead (Duranceau, 2008). The use of subject-based repositories within the high-energy physics community has grown out of a slightly different tradition whereby, prior to the web, typed-up pre-prints (different from the working paper series that many computer science and economics departments adapted to disseminate via the web and which Kling, Spector and McKim (2002) termed the 'Guild Model' of publishing) would be disseminated to key members and departments of the community. The creation of the arXiv repository, by Paul Ginsparg at Los Alamos National Research Laboratory in 1991, both to disseminate research findings and make them easily accessible in a timely way to the research community was a natural development of the existing pre-print tradition in the High-Energy Physics community. To date, the success of arXiv, now maintained by Cornell University, has been replicated in only a few other disciplines, such as the economics community which has developed the RePEc repository of working papers underpinned by the same principles.

[2] Note that data collected from OpenDOAR (1535 repositories) also include non-research/departmental repositories (for example, learning objects repositories).

Content in subject-based repositories is generally deposited by authors themselves (self-archiving), but there also exist some subject-based repositories, such as PubMed Central for the biomedical and life sciences, for which deposit of scholarly publications is also the responsibility of those publishers who have volunteered to be part of the repository. The US National Institutes of Health (NIH) have implemented, in April 2008, a funder mandate requiring authors conducting NIH-funded research to deposit a copy of their last version of their peer-reviewed published articles in PubMed Central. PubMed Central operates as a hybrid repository whereby individual authors, subject to funder mandates such as the NIH or Wellcome Trust mandates, can deposit a copy of their publications and participating publishers agree to make part or all of their content freely available after an agreed embargo period.

In less than a decade, higher education institutions around the world have embraced the concept of institutional repositories. There have been national differences in approach, with some countries developing policy and initiatives for publicly accessible repositories more rapidly than others. Literature on repositories usually shows that UK higher education institutions are today in the top tier in terms of repository infrastructure, with a well-developed network of institutional repositories. The Registry of Open Access Repositories currently reports 860 institutional or departmental repositories worldwide *(Figure 1.1)* of which 172 are in the US, 97in the UK, 63 in Germany, and 57 in Japan *(Figure 1.3).* These figures must be normalised against the number of higher education institutions in order to gauge the extent to which institutional repositories have diffused. In the UK, the representative body for the higher education sector, Universities UK[3], lists 116 universities, which means that 82.75% of institutions reported having an institutional repository. A recent European survey conducted in 2008 as part of the DRIVER project indicated that it is estimated that nearly half of European universities (spread over 22 countries) now have an institutional repository (Vernooy-Gerritsen *et al.,* 2009). As an increasing amount of government funding is now directed to open access publishing, this may foster the development of institutional repositories worldwide even more. For example, in the UK, for 2007-8, a total funding of £1.2 billion was provided by the UK research councils, of which £0.76 billion was funded under an open access policy (Sherpa-Juliet, 2009).

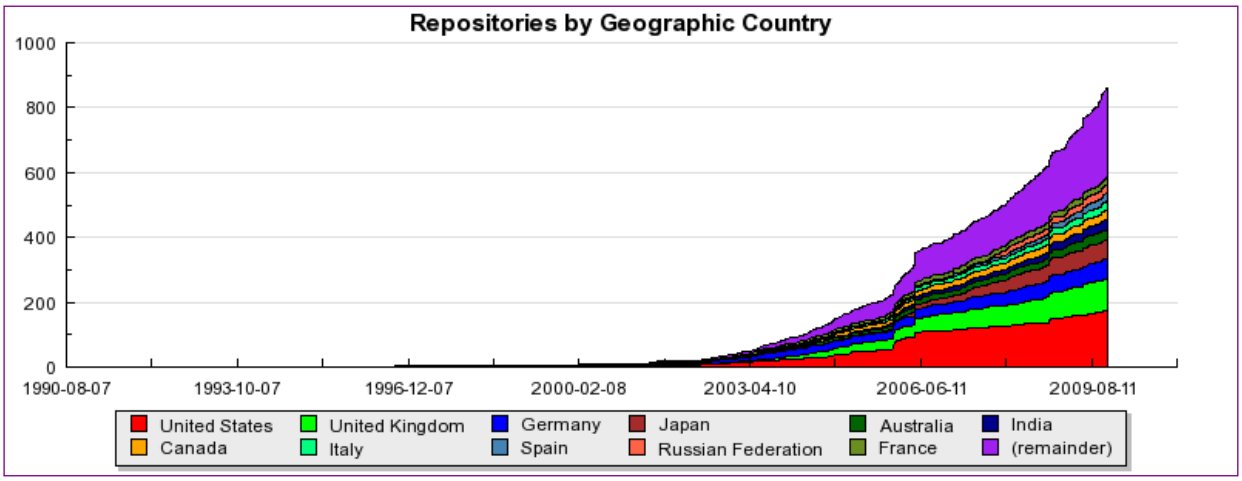
Figure 1.2 Growth of institutional research repositories (institutional or departmental)



Source: ROAR (1 December 2009)

[3] http://www.universitiesuk.ac.uk/Pages/Default.aspx

Figure 1.3 Growth of institutional research repositories by countries



Source: ROAR (1 December 2009)

The reasons underlying the creation of institutional repositories are multiple and diverse, from the adherence to the general principle of opening up access to scholarly publications, via the increase of the visibility (and possibly usage and citation counts) of scholars' research outputs, and hence those of the institution, to showcasing institutional research outputs. The rapid growth of institutional repositories, at least in the UK, has certainly also been driven by an increasing availability of public funds under the aegis of the JISC. Similarly, the EU-sponsored DRIVER project fostered the development of a European repository infrastructure. Such initiatives enable institutions to raise their profile in what has become a highly competitive higher education sector, while ensuring that knowledge as a public good reaches the relevant research communities, as well as the general public, for the benefit of the society as a whole.

National repositories for e-prints are also starting to be developed (these need to be understood in the broader information context whereby national repositories are also being developed for other types of research outputs, such as data archives and centres). In the UK, researchers at institutions without an institutional repository can deposit material in The Depot (http://depot.edina.ac.uk/). The Depot functions as a national repository and ensures that the entire UK research community has a place to deposit material. The Depot has not been developed as a long-term solution and it is the intention that material be transferred into permanent institutional repositories as they are developed.

Institutional repositories can also support subject-based services. Most repositories support protocols such as OAI-PMH by which metadata can be harvested. This allows subject-based services to be built based on metadata harvested from institutional repositories.
Rowland *et al* (2004) recommend a model in which full text articles remain in distributed archives (either institutional, subject-based or national), but that subject-based services are developed based on metadata harvested from these archives.

A fast-growing corpus of literature, which can be found in the extensive (but non exhaustive) bibliography edited by Charles Bailey (2009), is now looking into open access repositories and related issues such as authors' engagement, repository management, and how repositories fit into the scholarly communication system and publishing industry. Despite the infrastructure for institutional repositories being in place, the acquisition of content has been slow and consequently current uptake by readers has also been slow. As Borgman (2007) argues,

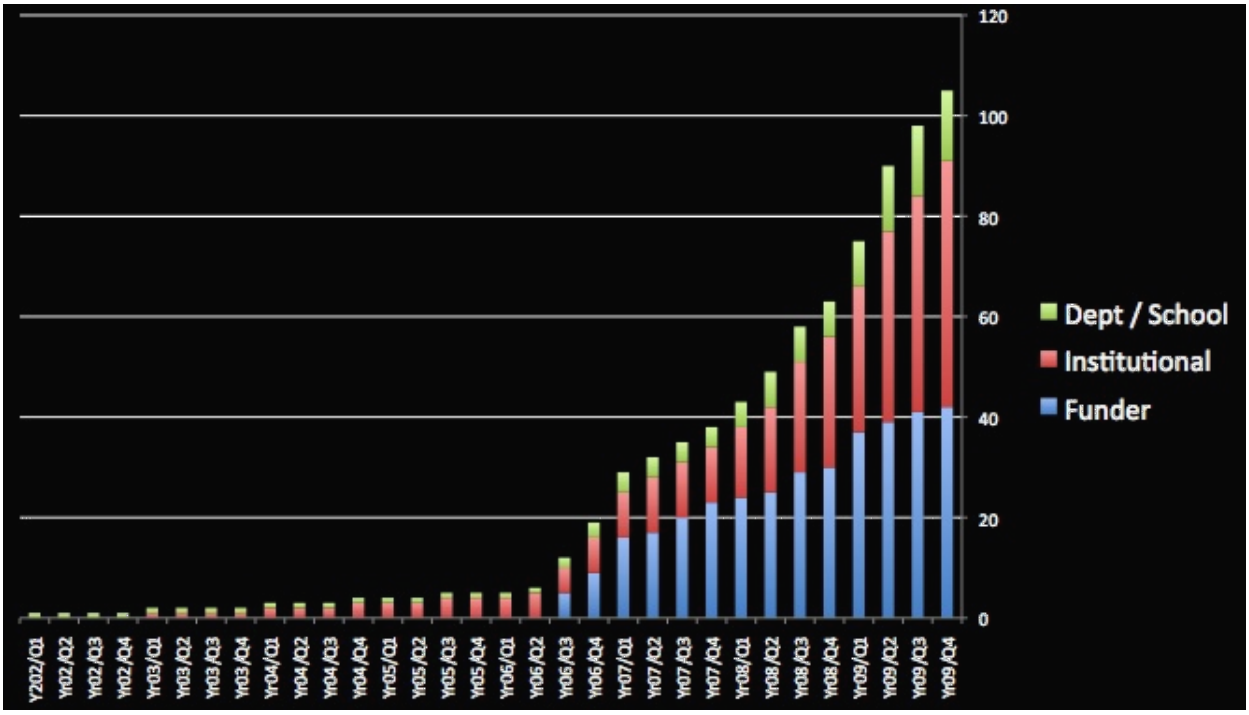
content is the most important part of any information infrastructure and it is not uncommon for there to be a long lead in terms of adoption. Based on the date of registration to the Registry of Open Access Repositories (ROAR), assuming registration is made immediately after the creation of the repository, half of institutional repositories have been created over the last three years *(Figure 1.2).* This means that acquisition of research content is still in its early stages and most institutions seem to be in the process of devising strategies for content recruitment, rather than operating at full capacity. The difficulty in populating institutional repositories with research outputs (Mackie, 2004; Foster and Gibbons, 2005; Van Westreinen and Lynch, 2005) has raised two interconnected issues, the realization that there exist different author self-archiving patterns, and the controversial policy to mandate deposit of research outputs by institutions, which is gaining momentum worldwide. To date (December 2009), ROARMAP, which has become an authoritative source listing the development of open access mandates, lists 176 mandates worldwide, of which 77 are institutional mandates, 18 are departmental mandates, 42 are funder mandates and 39 are thesis mandates (ROARMAP, 2009).

Repository awareness, users' engagement and authors' self-archiving are some of the complex issues with which repository managers are faced. There is some evidence that authors' rates of self-archiving are still very low, with most deposits of research material being made by intermediaries such as library staff or administrative/departmental staff, and that a relatively small portion of research articles held in institutional repositories are actually made available in full-text (Xia and Liu, 2007). Moreover, some studies have suggested that significant proportions of those scholars who have material held in their institutional repository are not necessarily fully aware of the existence of that institutional repository (Kim, 2007). This may be explained by the intermediation role played by library staff, whereby authors have a somewhat distant engagement with the whole self-archiving process. Furthermore, it is often reported in the literature that self-archiving practices can be closely associated with disciplinary cultures, notably with physicists more inclined than any other disciplines to self-archive in repositories due to their tradition of sharing pre-prints. Such relation between disciplinary culture and self-archiving practice has often underpinned discipline-specific repository advocacy campaigns. Some, however, have questioned the validity of such assertions by testing whether there is indeed any correlation between self-archiving practices and disciplinary cultures. Preliminary studies on the topic have brought interesting results. In their comparison of deposits in the subject repository arXiv and the institutional repository e-Prints Soton, Xia (2008) has found no correlation between self-archiving practice and disciplinary culture for the group of physicists studied. In other words, Xia's findings show that scholars depositing regularly in arXiv are not necessarily depositing their research materials for inclusion in their institutional repository. This validates earlier findings (Xia, 2007) suggesting that familiarity with self-archiving practice and experience with self-archiving processes (notably the ability to make deposits more quickly as familiarity with the process grows) do not necessarily play an important role in authors' choice to place research outputs in their institutional repository. Moreover, Xia (2008) suggests that the very fact that authors have already deposited their research materials in arXiv in the first place might deter them from placing the same materials in their institutional repository. The duplication of effort for authors who are already part of a subject-based repository community is seen as a major barrier to their engagement with their institutional repository.

As research evaluation grows in some countries, it is likely that it will foster the institutional collection of research outputs produced by faculties via institutional repositories (Van Westrienen and Lynch, 2005), thus enabling institutions both to better understand the type

and quality of research undertaken under its aegis, and hence its research impact, and to manage better the administrative process of identifying, collecting and submitting the research outputs to be entered in the research evaluation exercise. The rapid growth of institutional mandates observed worldwide *(see Figure 1.4)* seems to indicate that advocacy campaigns aimed at raising repository awareness and engaging research communities in self-archiving have not delivered the expected volume of research content to institutional repositories. It is often reported in repository circles that the baseline rate for spontaneous self-archiving (i.e. not required by a mandate) peaks at 15% worlwide, although the origins of this figure are not clear. Most higher education institutions currently offer their researchers a mediated deposit service whereby libraries manage the deposit of research outputs and associated tasks such as metadata creation, copyright checks, licence and preservation management. In this context, the only requirement from authors is to send the bibliographic details of their research outputs, together with their last version of the documents, to library staff. Ideally, it would be expected that researchers deposit their outputs themselves, although some have voiced their concerns about the impact this would have on quality standards, such as metadata creation or file preservation (Joint, 2006).

Figure 1.4 Growth of mandates in higher education



Source: Enabling Open Scholarship

Whilst there is some scope for a repository architecture whereby a distributed approach of institutional repositories could support both subject-based repositories and subject-based services, this approach has not received widespread acceptance. Moreover, the very fact that subject-repositories and institutional repositories co-exist is still stirring debate in repository circles, as regularly witnessed on the JISC-repositories discussion forum[4]. Whilst subject-based repositories bring to subject communities valuable advantages such as relevance of the audience, the success of subject-based repositories such as arXiv, E-LIS, RePEc or SSRN has not been replicated in other disciplines. Furthermore, the issue of sustainability of some of the

[4] https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=JISC-REPOSITORIES

subject-based repositories has been raised and concerns have been voiced over the long-term management of the collections of digital resources held in these repositories (Simpson and Hey, 2006). In this context, it is often argued that institutions are in the best position to ingest, manage, expose and preserve their employees' research outputs along with numerous other institutional digital assets (Simpson and Hey, 2006). Moreover, whilst open access publishing business models are still being debated, authors' funding for open access publishing is not always available and open access journals are still striving for high impact levels. In this context Swan indicates that institutional repositories may offer an immediate and alternative solution (Swan, 2007). In light of the amount of funding and human input currently deployed to set up, populate and maintain institutional repositories, it could be suggested that the return on investment in institutional repositories remains low, however.

Some observers have suggested that the lack of user engagement towards institutional repositories can be attributed to a lack of effective promotion by institutions to their research communities. Indeed, whereas scholars often place inter-disciplinary and/or inter-institutional collaboration mechanisms, facilitating co-authoring (versioning control etc.) and preservation (notably data preservation), high in the list of motivating factors to use institutional repositories, institutional promotion of repositories has rather been focused on opportunities for better institutional asset management and greater dissemination (Aschenbrenner *et al.*, 2008). Scholars generally tend to consider that dissemination via journal publishing meets their needs in terms of diffusion of knowledge and peer recognition, and thus do not regard repositories as a 'much needed' dissemination tool in the knowledge dissemination landscape. In other words, scholars do not perceive dissemination of scholarly outputs as an issue. According to Aschenbrenner *et al.* (2008), for users to engage with repositories, repositories must strengthen their focus on providing user services and should aim to embed repositories in researchers' workflows, and thus gain users' trust. However Aschenbrenner *et al.* (2008) also believe that this cannot be achieved locally by separate institutional repository entities and indicate that there is a need to develop a collaborative repository agenda based upon cross-repository interoperability.

1.5 Authors' attitudes in an open access context

Some studies suggest evidence of advantages associated with open access, however, despite this, researchers have not yet fully engaged with open access publishing or self-archiving. Swan and Brown's studies conducted successively in 2004 and 2005 have set milestones in the understanding of authors' perceptions of open access. These studies have shown that there is generally a low level of awareness of open access. Researchers generally have a very confused understanding of the 'open access' concept, its purpose and the means to achieve it. A 2004 survey of authors' attitudes towards open access showed that 82% knew 'nothing at all' or 'a little' about open access publishing (Rowlands *et al.*, 2004, p.22). The study also indicated that scholars showed a positive perception of open access publishing despite having some reservations about quality and preservation in such a model. However, concerns about the perceived low impact factor of open access journals have also been voiced in the open access literature, notably by authors subject to research audits such as the Research Assessment Exercise (RAE) in the UK, and was put forward as a reason for not choosing open access publishing for dissemination of their research (Swan and Brown, 2004; Watson, 2007).

Results from Swan and Brown's 2005 survey show that self-archiving practice remains limited amongst researchers, with only a minority of respondents making a copy of their articles

available on open access via institutional or personal Web pages or via institutional or subject-based repositories (Swan and Brown, 2005, p.32). It is often reported in the literature that the spontaneous self-archiving rate (i.e. free from any mandate/requirement pressure) peaks at 20% (Swan and Carr, 2008; Harnad, 2006). Furthermore, studies have shown that most researchers do not know whether their own institution has a repository. A 2007 UK study indicated that 72% of researchers did not know whether their own institution had a repository (Brown and Swan, 2007). However, this situation is changing as a large majority of higher education institutions have now set up their own digital repository and are engaged in advocacy campaigns targeting the academic community. A recent RCUK study indicated that only 43% of authors did not know that their institution had an institutional repository (SQW & LISU, 2009), thus suggesting an increased awareness of institutional repositories. A study conducted at Cranfield University (UK) has spelt out some interesting facts regarding the impact of repository advocacy in that knowing about the existence of one's institutional repository does mean that one understands its aim and purpose, and how this relates directly to scholars (both as users and authors), but does not mean that one will necessarily start placing research materials in there (Watson, 2007).

It is also expected that the recent impetus for institutional, departmental and funder open access mandates observed across the world will raise researcher awareness of open access repositories, but any such change in trends needs to be confirmed by further studies. It is also noted that repository awareness amongst scholars seems to be slightly higher in regard to subject-based repositories, particularly in disciplines such as physics and economics where there are long-standing traditions of depositing pre-prints and working papers. To date (December 2009), 77 institutional mandates, 18 departmental mandates and 42 funder mandates have been listed in ROARMAP (2009) worldwide. Note, however, that there is a time delay before such mandates are registered, so that these figures represent a lower bound to the number of mandates likely to be in place. A separate issue, not considered here, is the rigour with which such mandates are enforced.

A more recent study published in 2008 by Swan (2008) shows that very little progress has been made in the understanding of open access since Swan and Brown's seminal works. The following arguments are often cited in the open access literature as barriers to authors' self-archiving (Van Westrienen and Lynch, 2005; Swan and Brown, 2004, 2005; Harnad, 2006)

- Researchers often feel that open access material is of lower quality than toll access journal content, more precisely open access works are often associated with non-peer-reviewed materials.
- Copyright and publishers' self-archiving policies are perceived as a major barrier to self-archiving. Generally, authors do not know whether they can upload a copy of their article onto a repository or website, nor do they know which version (pre-print, author's final copy or publisher's copy) they can mount. Some studies tend, however, to moderate such claims. In her 2006 study, Antelman found that there exists no definite and specific correlation between publishers' policies and authors' deposit behaviour in the social sciences, and suggested that self-archiving behaviour is primarily affected by disciplinary norms.
- The deposit process itself is often cited as a barrier. Time pressure, technical difficulties and learning new processes (which can range from file upload to metadata input) are all

part of the deposit process and may all contribute to authors' concerns about depositing material in an open access repository.

- Authors' inertia is often reported as an obstacle to self-archiving practice. Swan and Brown (2005), Foster and Gibbons (2005) and McKay (2007) have reported that, despite acknowledging the open access benefits in principle, most authors generally do not take the principle into action and never get around to depositing their articles in a publicly accessible repository.

In summary, changes in the scholarly communication system can be linked back to several factors; the development of electronic publishing, increased availability of networked resources, digitisation of data and the tools used to manipulate data, and open access initiatives and policy. The trends to emerge have not been uniform across disciplines and whilst differences may have been identified at the coarse-grained level of the physical sciences, life sciences, social sciences and humanities little is known about the similarities or differences within and across these broad disciplinary groups. In particular, attitudes and behaviours towards open access, whether it be open access publishing or sharing data and other types of outputs in an open access manner, seem to vary greatly amongst scholarly and scientific communities, even within the same discipline. Despite the SHERPA/RoMEO service[5], there is often confusion amongst authors regarding their rights and responsibilities in relation to copyright agreements with publishers, and again this level of understanding appears to vary community by community. Despite changes in the way in which scholarly and scientific knowledge is created and disseminated researchers remain 'locked-in' to peer-reviewed published journal articles, both in terms of recognition and reward and information seeking. In terms of how researchers locate and access information, availability of online access to full-text articles is often conflated with open access. The location of open access resources and sources in the "connective structures" of digital information landscapes and the pathways that information seekers construct to access them is not well understood, although current understanding suggests that word of mouth and internet search engines are influential in this process. Although, such pathways are likely to be shaped by the type of search being conducted and the extent to which a specific topic spans disciplinary boundaries.

The literature on repositories is broad in scope and within the context of this report we have covered only a narrow range of that literature. What emerges, however, is that it is not clear what the best model for governing repositories will be e.g. institutional/subject – full-text/metadata. The debate on mandates, whether or not they are feasible or capable of achieving institutional goals is ongoing. In the meantime, some interim measures have been taken, such as The Depot, but these will need to be integrated into longer-term sustainable solutions. Given that the environment with regards to repositories is still uncertain this may impact negatively on researchers' willingness to contribute to them.

1.6 Research questions

This behavioural research aims to contribute to this evidence base by exploring perceptions, motivations and behaviours of authors and users towards journal publishing and depositing in repositories. The research will be longitudinal and will use mixed methods to address the following set of research questions:

[5] http://www.sherpa.ac.uk/romeo/

- In seeking information what choices do readers make in locating and selecting sources and in what ways do such choices influence the role played by repositories in information seeking behaviours?
- In publishing research, what choices do authors make in locating and selecting appropriate outlets, and what are the major influences on their choices? Where do repositories fit in the dissemination landscape?
- What common perceptions do readers have in relation to repositories, e.g. quality, authority of versions, and availability, and how do such perceptions influence information behaviours?
- What common perceptions do authors have in relation to repositories, e.g. visibility, impact, and recognition, and how do such perceptions influence publication and dissemination behaviours?
- Are there identifiable coarse-grained characteristics of authors and readers that influence their behaviour (e.g. institutional type, region, discipline, career status etc.)?
- How do social/institutional factors influence author and reader behaviours (e.g. mandates, embargoes, research cultures)?
- What tensions, if any, exist between institutional (e.g. employer/funder/publisher) policies and practice, and disciplinary norms and practices? In what ways do such tensions influence authors and readers?

This Baseline Report outlines findings from the first phase of the research and identifies the key themes to emerge. It also identifies priorities for further analysis and future work.

1.7 Methodology

The first phase of this study comprised two primary data gathering elements, in addition to the background research reported above. An extensive survey of European researchers gathered evidence on a broad basis covering many of the issues in the research questions. This was supported by a series of focus groups that ran concurrently and explored specific issues in greater depth. Individual telephone interviews with selected survey respondents remain to be carried out, to add to the richness of the data.

1.7.1 Survey of researchers

An electronic survey of researchers was conducted between June and August 2009. The questionnaire was designed to cover those issues addressed by the research questions above, in so far as a broad survey instrument was appropriate, and was informed by the background research. The survey included questions addressed to scholars both as writers of research outputs and as readers or consumers of those outputs, and focussed on research dissemination via journal articles. The survey was piloted with a selected number of researchers known to the research team and spread across European higher education institutions and research centres, and one researcher based outside Europe. Because the survey was being distributed by publishers across Europe, it was important to ascertain not only that the questions were relevant but also that non-English native speakers would understand the questions easily. Details of the pilot participants are given in Table 1.1.

Table 1.1 Pilot participants

Country	Institution type	Native language
Australia	HE institution	German
France	HE institution	Dutch
France	HE institution	French
Germany	Research centre	Greek
Netherlands	Research centre	German
UK	HE institution	Bengali
UK	HE institution	Italian

Invitations to complete the survey were distributed via the twelve publishers participating in the PEER Observatory. The intention was to restrict the distribution to EU-based corresponding authors, who had manuscripts published in those journals included in the PEER Observatory and the control group since the beginning of 2008. This time scale was set to reduce the potential number of currently inactive researchers who might be approached. Such a restriction of the circulation was not always possible, and the survey was distributed more widely by several publishers. Non-EU respondents were filtered out from the responses in the analysis. One publisher was not able to distribute the survey on our behalf; in this case the research team trawled the tables of contents of participating journals to extract the email addresses of EU based corresponding authors. An incentive prize draw for Amazon vouchers was offered to encourage response, and one respondent was selected at random from the completed responses received by the closing date.

Survey responses were checked for completeness, coded, and quantitative responses analysed using the SPSS software package. All respondents answering more than just the first section of the questionnaire were included in the analysis. Comments made by respondents remain to be analysed and will be included in the final report. A total of 3,139 valid responses from researchers in the European Union were received, although not all respondents answered every question.

As well as overall summaries of the responses, all questions have been analysed by broad discipline group, institution type and career stage of the respondent. Where sufficient data were available, differences between groups were identified using a χ^2 test. Results are noted in this report as being statistically significant when the probability that the observed differences between the groups occurred by pure chance, and that in the wider population no such differences exist, is less than 5% ($p<0.05$) or 1% ($p<0.01$).

There were generally insufficient responses from some institution types to perform a formal test of apparent differences for the majority of questions, although it may be possible to undertake further analyses by combining selected categories. Analyses by region have not been carried out at this stage, due to the small numbers of respondents from some European countries. The research team plan to consult over the most suitable geographic breakdown to investigate potential regional differences in behaviour.

1.7.1.1 Demographics of the sample

All countries in the EU were represented, with more than 100 responses received from each of seven countries (UK, Germany, Italy, Spain, France, the Netherlands and Sweden) *(Figure 1.5)*.

Figure 1.5 Survey response by country

The largest single group of respondents had been employed in research for between five and 14 years *(Figure 1.6)*. Seventy percent were based in universities or colleges, with a further 20% based in research institutes *(Figure 1.7)*. Respondents were asked to indicate the areas of their research by ticking one or more of 39 distinct subjects, grouped into four broad areas. More than half of all respondents were from the Physical sciences & mathematics *(Figure 1.8).* Respondents were allocated to an Interdisciplinary group if they ticked subjects from two or more of the broad areas included; the majority of these had indicated subjects in both the Medical sciences and Life sciences areas.

It is not clear at this stage whether the bias towards Physical sciences & mathematics was inherent in the initial distribution of the survey, or whether researchers in these disciplines were more likely to respond to the survey (see section 1.7.3 for further discussion of this issue).

Figure 1.6 Research experience

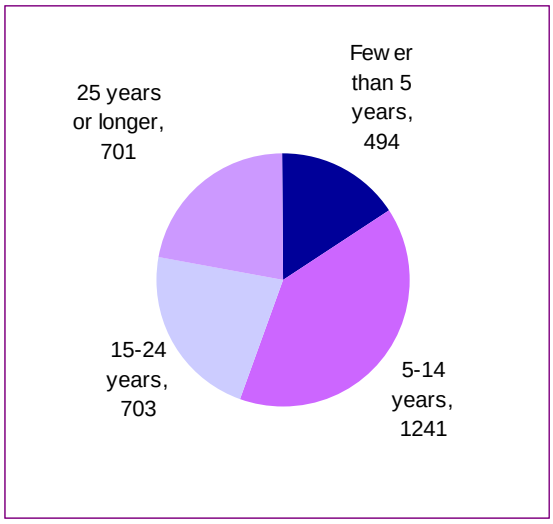


Figure 1.7 Institution type

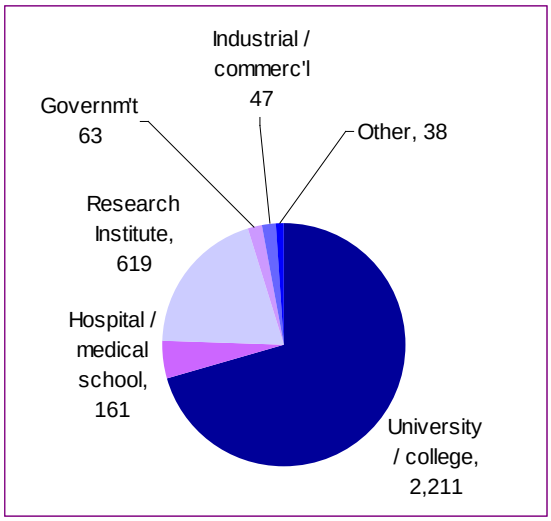
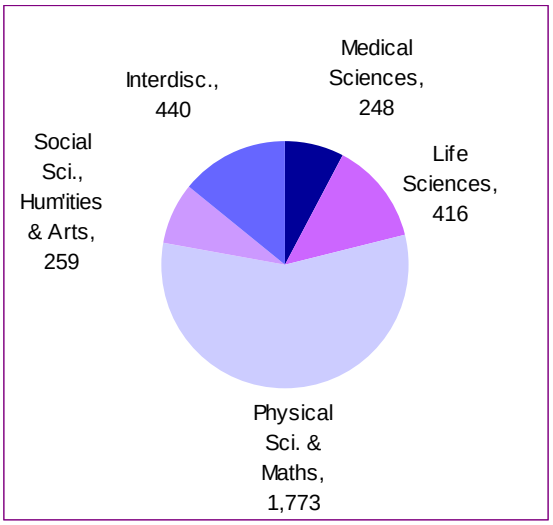


Figure 1.8 Subject area



1.7.2 Focus Groups

Four focus groups were conducted in parallel to the collection of quantitative data gathered through the web-based survey, in order to bring a qualitative perspective to the research. This allowed the research team to investigate in-depth scholars' attitudes and opinions about open access material and deposit of stage-two manuscripts in publicly accessible repositories.

1.7.2.1 Sampling and selection

The four broad disciplinary groups (Medical sciences; Social sciences, humanities & arts; Life sciences; Physical sciences & mathematics) represented by the participating journals in the PEER experiment were chosen for the focus groups in order to facilitate comparative analysis. A decision was made to hold the focus groups in various locations across Europe in order to build on a large spectrum of opinions and practices and identify potential regional and legal differences influencing researchers' behaviour towards open access and self-archiving. Focus groups were disciplinary-focused in order to explore the impact of research cultures on scholars' attitudes towards open access and deposit of stage-two manuscripts. Location of the venues for

the focus groups was agreed with the PEER Research Oversight Group, and appropriate venues were identified in London, Berlin, Budapest and Rome. It was considered important to avoid concentrating the focus groups in northern Europe in order to attract participants from a variety of open access contexts since there have been national differences in policy.

The intention was that venues and disciplinary groups would map onto one another in terms of availability of participants. Based on the research team's experience, it was anticipated that it would be particularly difficult to recruit medical scientists, and therefore holding the Medical sciences focus group in London was seen as advantageous due to the numerous medical institutions based there. There are a number of distinguished institutions in the social sciences and humanities in Berlin, so the Social sciences, humanities & arts focus group was held there. The three-year research evaluation[6], conducted in 2000-2003, in the Italian higher education system helped to identify areas of strength of the universities based in Rome. The disciplinary grouping of 'Physical sciences & mathematics' was identified as a strong match for Rome. Budapest has a strong Life sciences research community and was, therefore, ideal for holding the Life sciences focus group.

Researchers and academics from higher education institutions and research centres, within reasonable travelling time of each venue, were then identified and invited to participate in the focus group sessions. The selection process included a systematic, and extremely labour-intensive (notably when institutional web pages were not available in English), browsing of staff pages of relevant departments of selected universities and research institutions. Potential participants were identified and their contact details, as well as the position they held, were recorded in a database, which was later used for sending the email invitations to participate in the research. For the Budapest focus group, where it was particularly difficult to identify relevant individuals, additional invitations were sent to those Hungarian researchers publishing in the life sciences journals of the publisher for which the research team had compiled its own distribution list for the survey *(see 1.7.1)*.

A combination of the fact that focus group participation is notoriously low and that the focus groups were planned to be held in the two first weeks of July (potentially conflicting with the academic year e.g. conferences, exams and annual leave) necessitated the research team to over recruit for the focus groups. For each focus group we set a target of 10-12 participants, which meant that invitations needed to be sent out on a large-scale, as illustrated in Table 1.2. Even so, we were unable to reach our targets in the time available, and as not all those accepting the invitation actually attended, some focus groups had a lower attendance than we would have liked.

[6] The VTR [Valutazione Triennale della Ricerca] report was commissioned by the CIVR [Comitato di Indirizzo per la Valutazione della Ricerca] in order to review Italy's research policy and priorities as well as allocation of research funding. The report is available at: http://vtr2006.cineca.it/php5/vtr_rel_civr_index.php?sel_lingua=EN.

Table 1.2 Focus group participation

Location	Disciplines	Number of scholars invited	Number accepting the invitation	Number of attendees
London	Medical sciences	750	6	3
Berlin	Social sciences humanities & arts	677	7	5
Budapest	Life sciences	364	6	6
Rome	Physical sciences & mathematics	646	9	7

Although the focus groups were primarily disciplinary-focused, an exception was made for one Life sciences candidate who was keen to participate given their interest in open access, but who could not attend the appropriate focus group in Budapest (because s/he was visiting a German university at that time), therefore they were invited to join the Social sciences, humanities & arts group in Berlin instead.

1.7.2.2 Profile of the participants

Participants were at varying stages of their careers, although mid-career and senior researchers predominated overall. The Social sciences, humanities & arts focus group comprised relatively senior researchers from a broad range of disciplines, including law, sociology, gender studies and semiotics. The Medical sciences focus group comprised researchers who were generally at earlier stages of their careers in the fields of immunology, anatomy, ethno-botany and pharmacology. In the Physical sciences & mathematics focus group the majority of participants were physicists, but structural engineering and chemistry were also represented. Participants ranged from very senior members of the physics community to researchers at earlier stages of their careers. The Life sciences focus group involved participants from a broad range of disciplines, most of whom were mid-career to senior researchers. Details are given in Table 1.3.

Table 1.3 Focus group profiles

Broad disciplines	Specific research fields of participants	Institution spread	Research experience		Gender	
Medical sciences	Immunology and virology Anatomy Medical Ethno-botany, ethno-pharmacy, ethno-anthropology and pharmacology	3	Early career researchers Mid-career researchers Senior researchers	2 1 0	Female Male	2 1
Social sciences, humanities & arts	Sociology and gender studies Sociology and family policies, welfare policies, social assistance policies Linguistics and semiotics Consumer law, media information law and copyright law Plant nutrition and fertilisation	5	Early career researchers Mid-career researchers Senior researchers	1 1 3	Female Male	2 3
Life sciences	Organic plant breeding Molecular genetics Parasitological and infectious diseases Neuropsychology Human health risk assessment and occupational assessment Biomechanical mechanism of plant diseases	3	Early career researchers Mid-career researchers Senior researchers	1 2 3	Female Male	1 5
Physical sciences & mathematics	Structural mechanics Theory of molecular physics and chemical physics Theoretical physics and mathematical physics Condensed matter theory Civil engineering Chemistry	2	Early career researchers Mid-career researchers Senior researchers	0 3 4	Female Male	2 5

1.7.2.3 Focus group protocol and analysis

All focus groups were conducted in English and each was facilitated by three members of the research team. The focus groups were conducted through the use of scenarios to loosely guide the discussion on specific themes and encourage participants to present their views and opinions. Four scenarios in total were devised by the research team, of which two scenarios aimed to investigate researchers' views as authors and the other two aimed at exploring their views as readers. The scenarios enabled the researchers to ensure that specific themes were consistently addressed across each focus group. The research team was also sensitive to unexpected themes that emerged during the discussion as they provided additional valuable information on scholars' perceptions of open access repositories and self-archiving. Further details about the scenarios are provided in the focus group protocol in Appendix 2. Verbal permission to audio-record each focus group discussion was obtained and participants were informed that the data would be treated with confidentiality.

The focus groups were audio-recorded and later transcribed in full. The qualitative data were then coded and categorised, and relationships between codes were mapped out. The coding scheme was discussed by the research team and was checked against the research questions to ensure that the analysis of the data was closely mapping onto the research questions.

As part of the process of validating the focus group findings, each participant was sent a copy of the focus group write-up to check that their views and those of their group had not been distorted or misinterpreted. Participants were also invited to comment on the findings and provide additional thoughts arising from the reading of the write-up, if they wished to do so.

1.7.3 Limitations of the research

The survey analysis is inevitably weighted towards the views of authors by virtue of being distributed by publishers to their authors. Furthermore, the profile of respondents is heavily weighted towards physical sciences and mathematics disciplines, and researchers based within universities. The former may result both from a response bias and a sample bias. Indeed, It should be noted that the distribution of journals titles participating in the PEER experiment showed a slight bias towards the physical sciences *(Figure 1.9).* Whilst it is true that dissemination patterns collected from the survey show that physical sciences and mathematics disciplines tend to heavily rely on journal articles to disseminate their research outputs *(Figure 1.10),* thus reinforcing the probability for this group to be contacted by the publishers participating in the PEER experiment, the difference with the other disciplines covered in this study remains marginal. It should also be noted that, despite the research team requesting that all participating publishers send the survey out to their authors on a similar date, the survey was ultimately distributed at different times by each of the publishers, with the same fixed closing date[7], therefore authors in some fields may have had more opportunity to reply than others, which may also explain the high response from scholars in Physical sciences and mathematics. Alternatively, a response bias could come from individual awareness of open access. Intuitively, scholars knowing about open access may be more likely to participate in the survey than those who know little or nothing about it. The early adoption of the open access culture by the Physical sciences and mathematics communities (open distribution and sharing of pre-prints through open access repositories such as arXiv) means that such communities are already quite familiar with open access developments in scholarly communication, and thus are able to actively participate in the debate, whereas other communities less familiar with open access and repositories may have not developed a keen interest in the current debate.

[7] Only one publishers had a later closing date (by a week) owing to the fact that this specific publisher was unable to distribute the survey within the agreed timeframe.

Figure 1.9 Distribution of the journal titles participating in the PEER experiment

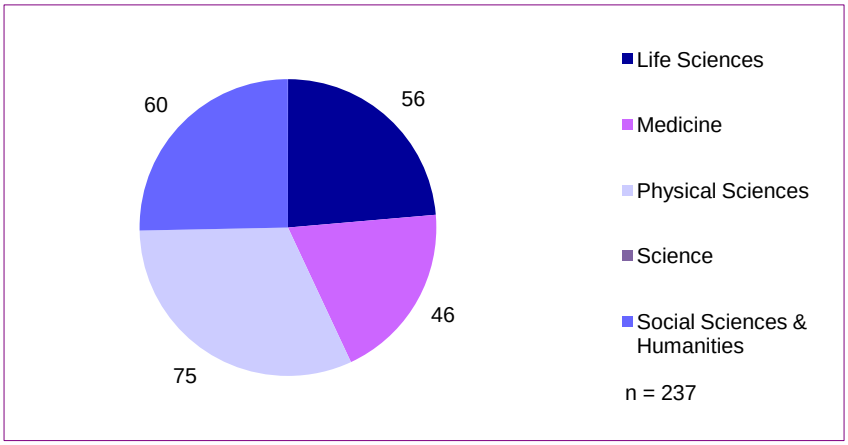
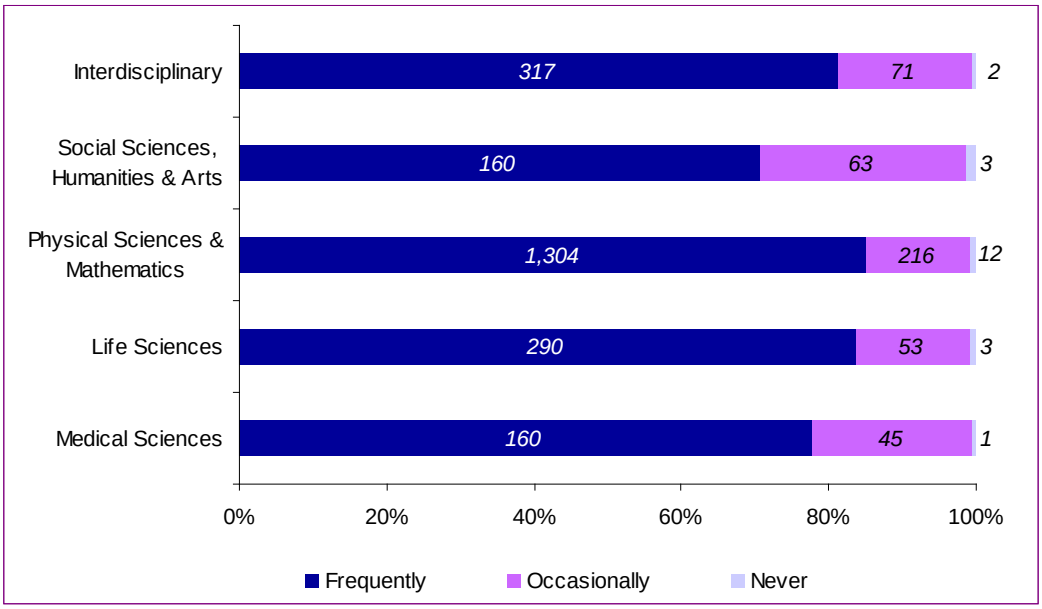


Figure 1.10 Frequency of publication in a peer-reviewed journal in the last five years



Focus groups are not meant to provide a basis for generalization. They explore the views and opinions of individuals, and hence bring to the research an in-depth analysis. In such context, focus group findings must be interpreted with caution. They do not provide a statistical representation of the views of the research communities - although occurrences and intensity are important factors to be accounted for, they only tell us about individual feelings and experience. They only provide pointers for a better understanding of researchers' attitudes. Moreover, the findings of the focus groups cannot also be regarded as representative given the small numbers of participants and their self-selecting nature. The focus groups inevitably comprised interested parties, which may go some way to explain the predominance of favourable views regarding open access. Conversely, some participants indicated at the beginning of the focus groups that they did not know much about open access and repositories and that they had come to the session to learn more about this area.

Ideally, the research team would have liked to have conducted the survey first, and then followed-up with the focus groups, in order to validate quantitative findings with qualitative data,

and expand on the survey findings. However, owing to time constraints, it was agreed with the PEER Oversight Group that the focus groups and the survey would run concurrently. The implication of this is that the qualitative analysis is not intended to validate the quantitative analysis, but rather to provide further information. However, work planned for the later stages of the project *(Section 4.3)*, in particular the workshop to be held in March 2010, the follow-up interviews, and more in-depth analysis of background literature, will take into account these limitations and help to mitigate some of the issues encountered with the survey and focus groups methodologies.

2 Survey analysis

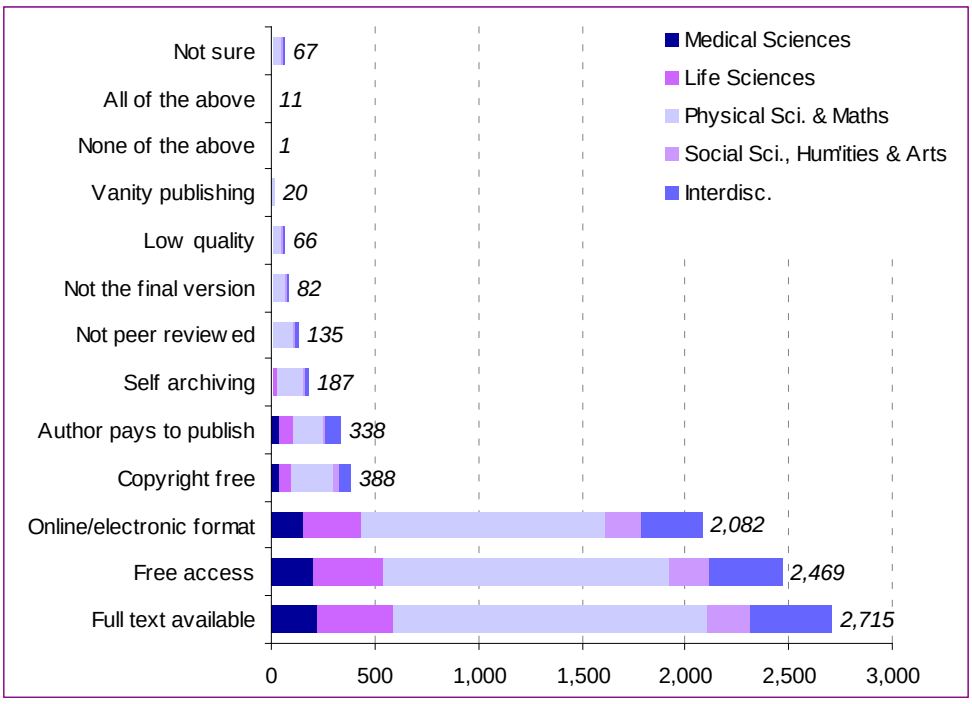
This section outlines the preliminary results from the survey, based on all respondents, with selected differences by broad subject and research experience noted. Further analysis remains to be done, and this will be incorporated in the PEER joint workshop presentation (March 2010) and final report due at the end of the project (June 2011). Summary tables detailing the quantitative responses to all questions are included at Appendix 1.

2.1 Awareness of open access repositories

Awareness of open access was high amongst survey respondents, with more than two-thirds indicating that they understood open access to mean free electronic access to the full text of articles. Less than five percent considered open access to mean low quality, not peer-reviewed, not the final version of an article, or vanity publishing. Six percent equated open access with self archiving, and 11% with author-side payment *(Figure 2.1).* Some differences by subject area were noted:

- Respondents from Physical sciences & mathematics and Social sciences, humanities & arts were less likely to equate open access with author payments to publish than those from Medical sciences, Life sciences and Interdisciplinary researchers.
- Respondents from Medical sciences and Life sciences were less likely to think open access equated to articles not being peer reviewed, and not the final version, than those from Physical sciences & mathematics and Social sciences, humanities & arts; this was also the case for Interdisciplinary researchers, but to a lesser extent.
- Respondents from the Social sciences, humanities & arts were most likely to be unsure about the meaning of open access.

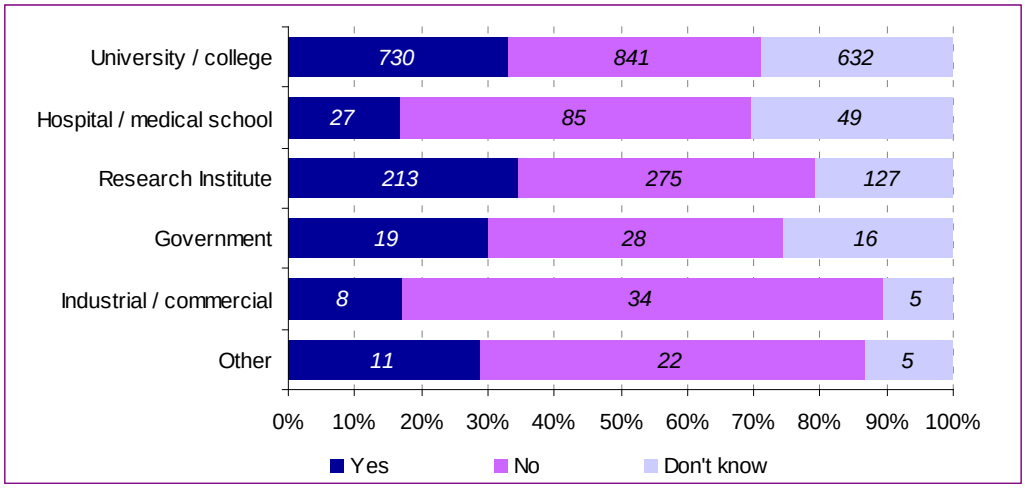
Figure 2.1 Awareness of open access, by subject



Total respondents = 3,136

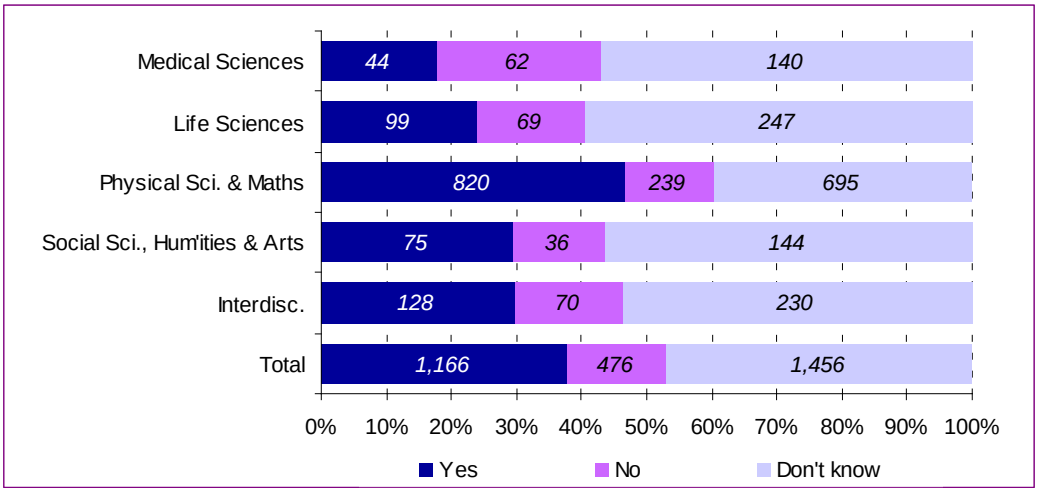
Just under one third of respondents overall indicated that their institution had its own open access repository for research outputs. There were statistically significant differences between the types of institution, illustrated in Figure 2.2. Although there were few respondents from industrial and commercial organisations, these respondents were most likely to know about any repository in their institution, and least likely to report that one existed. Researchers with less than five years' experience were less likely to know whether their institution had a repository than those with more experience; Creaser (in press) found similar results for the UK in 2008, with junior researchers less likely to know whether their institution had a repository.

Figure 2.2 Awareness of institutional repositories



Respondents were less likely to be aware of subject-based repositories. Almost half – 47% - did not know whether there was a suitable subject repository for their research, with 37% indicating that there was one available. Here, there were clear differences between subjects *(Figure 2.3)*, and again, relatively inexperienced researchers were least likely to know about such repositories.

Figure 2.3 Awareness of subject repositories



Respondents who indicated there were suitable subject repositories were asked to name these, and 1,160 gave details. In some cases the 'repositories' named were, in fact, open access journals and/or publishers (e.g. Biomed central, Public Library of Science). The archive most often mentioned was arXiv (by 602 respondents, mostly from Physical sciences &

mathematics). Other major archives mentioned included PubMed Central (52 respondents, from Medical sciences, Life sciences and Interdisciplinary); Citeseer (43 respondents, from Medical sciences, Physical sciences & mathematics and Interdisciplinary); HAL (43 respondents, from Physical sciences & maths); and RePec (25 respondents, from Social sciences, humanities & arts).

Respondents were also asked to name any other repositories which might be suitable for their research. In total, 401 respondents indicated that they knew of such a repository, while 1,205 said there was not one available to them. Of the 399 who provided details, the majority mentioned institutional repositories, open access journals, or alternative subject repositories.

The Registry of Open Access Repositories (ROAR) lists 135 'cross-institutional research repositories' on its web site[8]. Three repositories listed as having more than 100,000 records were not mentioned in the survey – one based in Australia (ARROW), and two in the US (Mountain West Digital Library and Ohio Digital Resource Commons). One of this size based in Europe (Max Planck Society Edoc Server) was mentioned by only one survey respondent.

2.2 Dissemination behaviours

Almost all respondents (99%) had published or disseminated their research in the last five years. Of these, 83% had 'frequently' published in peer reviewed journals over that time, and a further 17% had 'occasionally' done so. Less than 1% had not published in peer reviewed journals in the last five years. Since the survey was distributed via publishers, a high rate of journal publication was expected.

Figure 2.4 Preferred dissemination formats

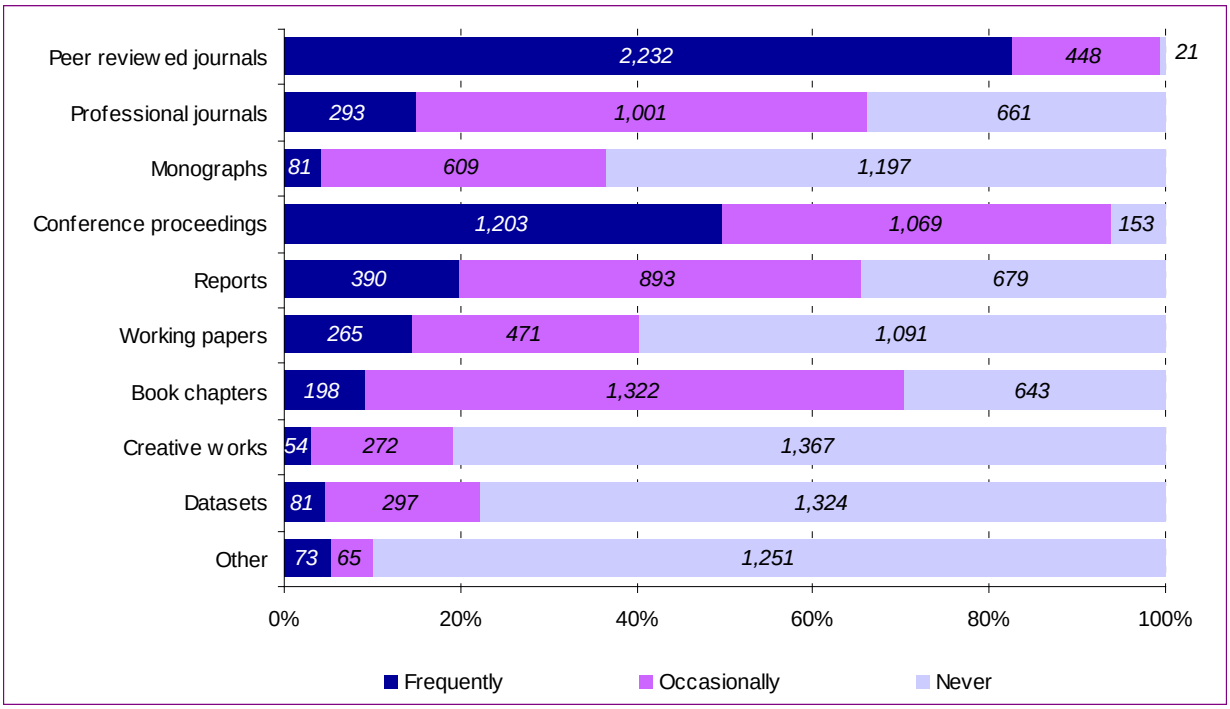


Figure 2.4 illustrates the frequency with which a variety of publication/dissemination formats were used. The formats are listed in order of importance, according to respondents' selection of

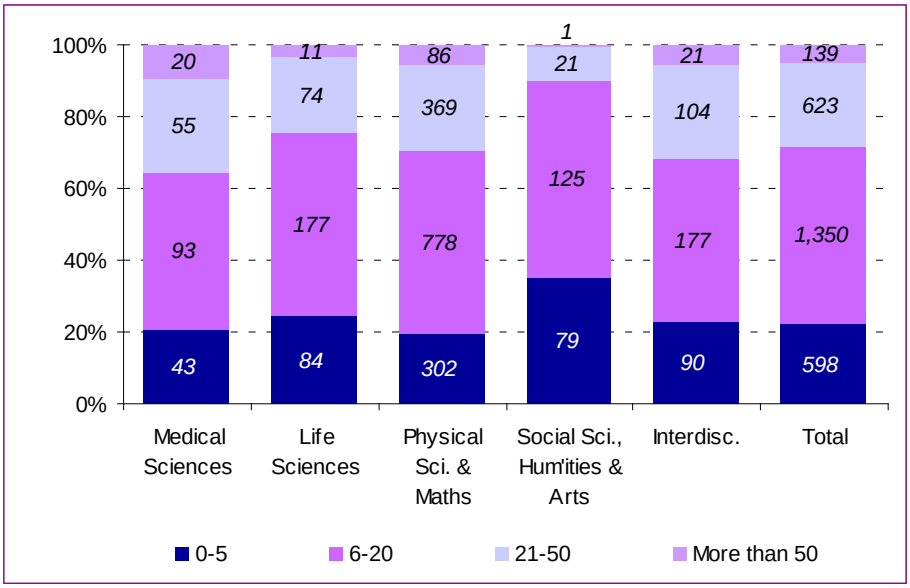
[8] http://roar.eprints.org/index.php, accessed 26-8-09

the most important type of output in terms of their career *(see Appendix 1, page 84)*. Peer-reviewed journals were clearly the most important, rated so by 89% of respondents, with professional journals seen as most important by 4%, monographs by 3% and conference proceedings by 2%. The other forms listed were rated as most important by less than 1% of respondents each. It is clear from these data that popularity does not necessarily equate to importance.

There are clear subject differences both in what is published, and the relative importance of different formats. Such differences have been documented in other studies (e.g. Hicks 2004, Research Information Network 2009), and will not be covered in depth here. Statistically significant differences between subjects were found in the frequency of publication of monographs, book chapters, professional journals, conference proceedings, working papers and datasets. There were insufficient data to formally test differences between subjects in the relative importance of the different types of output; however, Medical sciences and Life sciences appeared most likely to name peer-reviewed journals, with more then 90% of respondents in these subjects giving this as their most important output. The second most important output in all disciplines except Social sciences, humanities & arts was professional journals; in Social sciences, humanities & arts the monograph was the second most important output, with 14% of respondents, compared to 78% who named peer reviewed journals.

Half of all respondents had published between six and 20 peer-reviewed journal articles over the last five years, with approximately equal numbers publishing more or less frequently. There were statistically significant differences by subject, with respondents from Medical sciences more likely to have published over 50 articles than those in other subjects, and respondents from Physical sciences & mathematics less likely to have published fewer than six. Social sciences, humanities & arts respondents publish fewest articles on average, being more likely to have published fewer than six, and less likely to have published more than 20 over the last five years than respondents from the other subjects *(Figure 2.5).*

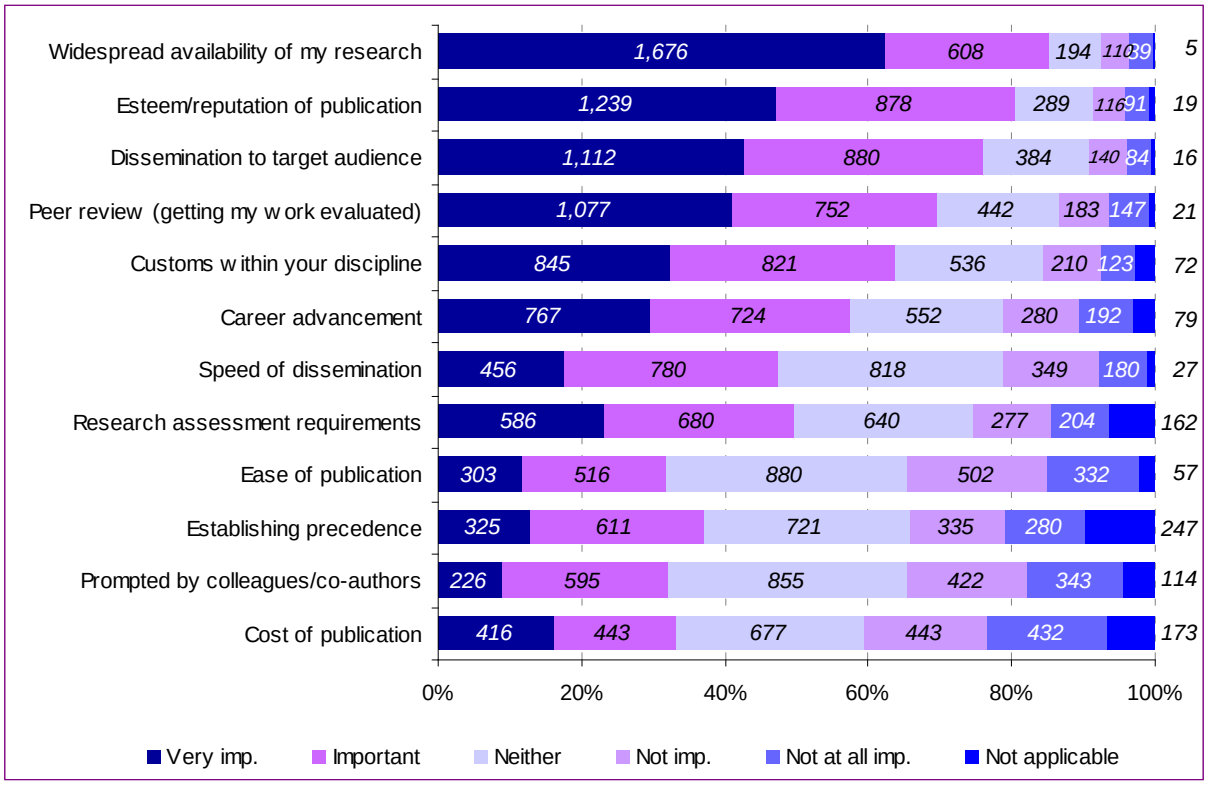
Figure 2.5 Number of journal articles published in the last five years



The most important factors in choosing a peer-reviewed journal (rather than, or in addition to, other forms of dissemination) on the most recent occasion were the widespread visibility of the research, the reputation of the journal, and dissemination of the research to the target audience.

Figure 2.6 illustrates the relative importance of a range of factors, ranked by average importance. The least important factors were the cost of publication, colleagues and co-authors, establishing precedence and the ease of publication. These results are broadly in line with the findings of previous studies (Research Information Network 2009). Not surprisingly, career advancement was significantly more important to the least experienced researchers (those with fewer than five years experience), as was the suggestion of colleagues and co-authors.

Figure 2.6 Factors in choosing peer-reviewed journals



When considering placing material in repositories, 20% of respondents had no preference for the type of repository they used, 46% preferred a subject repository, and 22% an institutional repository. Eight percent of respondents preferred to use their own or their departmental website, and 3% preferred not to do this at all.

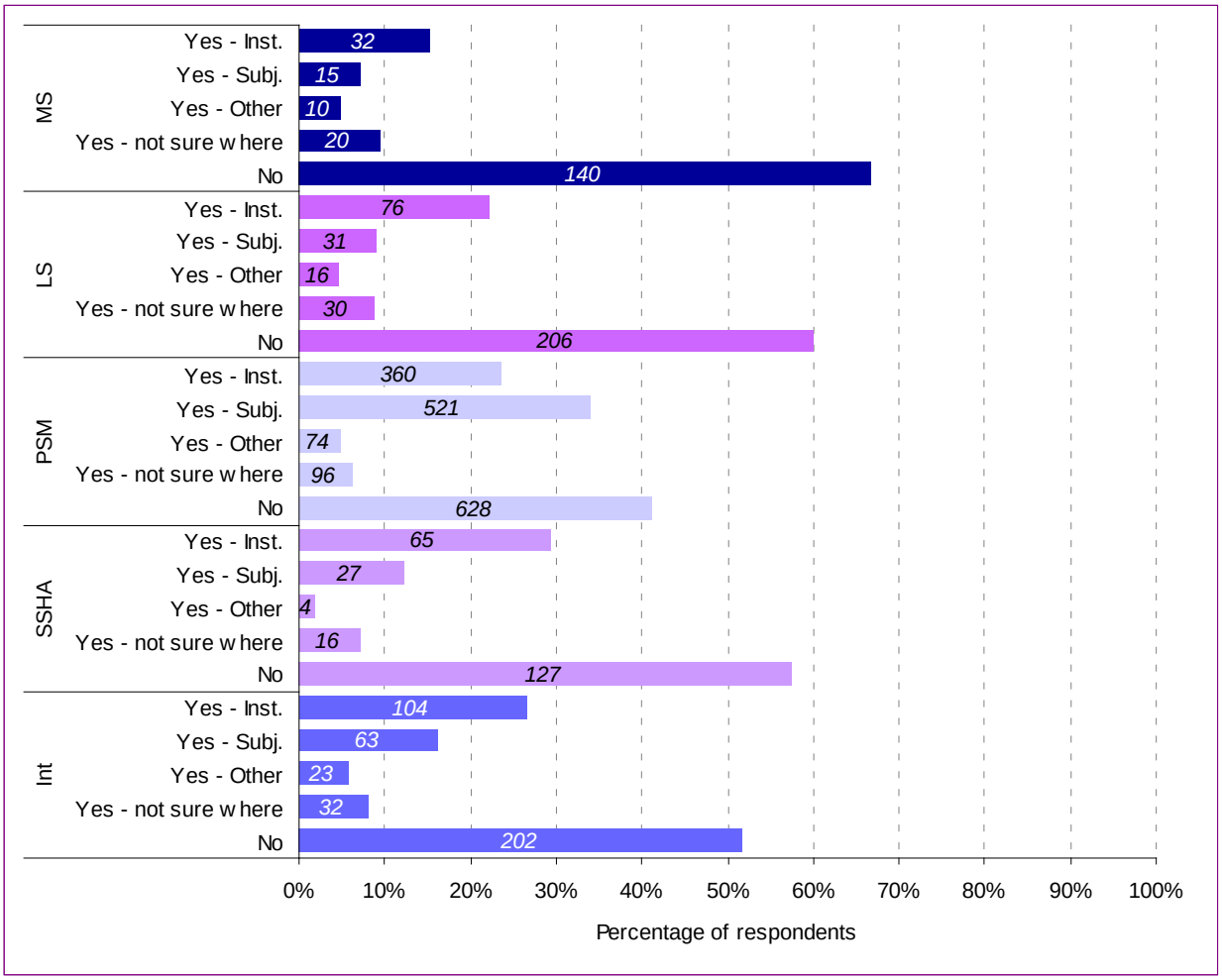
Almost half of respondents overall had not placed any peer-reviewed journal articles in a publicly available repository in the last five years, with approximately equal numbers using institutional and subject-based repositories. This raises some interesting questions in relation to stated preferences for the different types of repository described above. There were also some notable differences by subject, illustrated in Figure 2.7. Respondents from the Medical sciences were least likely to have deposited any articles in the last five years, with respondents from the Physical sciences & mathematics most likely to have done so. This relative lack of deposit from the Medical sciences is perhaps surprising, given that an increasing number of medical research funders internationally are now mandating open access to the results of the research which they fund (SQW & LISU, 2009). However this may be due to the fact that many medical publishers deposit in subject-based repositories such as PubMed on behalf of their authors.

Respondents from the Physical sciences & mathematics were more likely to have deposited an article in a subject-based repository than respondents in other subject areas. Respondents from

the Social sciences, humanities & arts were more likely to have deposited an article in an institutional repository than respondents in other subject areas. These results tie in with the high level of awareness of repositories and the use of arXiv in the Physical sciences & mathematics areas noted above.

Forty-four percent of authors had deposited the author's final version of the article, i.e. the stage-two manuscript. Forty percent had deposited a pre-print, while 32% had deposited the publisher's PDF, and 20% were not sure which version had been deposited (these percentages total more than 100 as respondents could tick more than one response to this question).

Figure 2.7 Repository deposit by subject



Of those authors who had deposited the stage-two manuscript, 70% reported that they did so voluntarily. Colleagues had suggested the deposit for 12% of respondents, while 11% were invited to do so by the repository, and 10% responded to requests from co-authors. Publishers had invited the deposit in 8% of cases, with a similar proportion being mandated by their employer. Only 3% had responded to a funder mandate (these percentages total more than 100 as respondents could tick more than one response to this question). These patterns raise some interesting questions, not least about funder and institutional mandates, which will be explored further in subsequent research and reporting. Interestingly, the most experienced researchers (those with 25 or more years experience) were most likely to report being invited by a publisher to deposit their article.

The most important factors driving authors towards depositing material in open access repositories are concerned with making their work as widely accessible as possible *(Figure 2.8).* Mandates are again seen to be relatively unimportant in this context. There were also some notable differences in the importance of the various factors by subject group. Respondents in the Life sciences were more likely to rate the principle of free access to all as 'very important'. Speed of dissemination was less important for respondents in the Medical sciences than those in other subjects. Those in Social sciences, humanities & arts were most likely to rate the possibility of increased citations as 'very important'. Career advancement was less important for respondents in Medical sciences and Physical sciences & mathematics than those in other subjects. Mandates, by the institution or the funder, were more likely to be rated as 'not at all important' by respondents from the Physical sciences & mathematics than those in other subjects.

Figure 2.8 Drivers for repository deposit

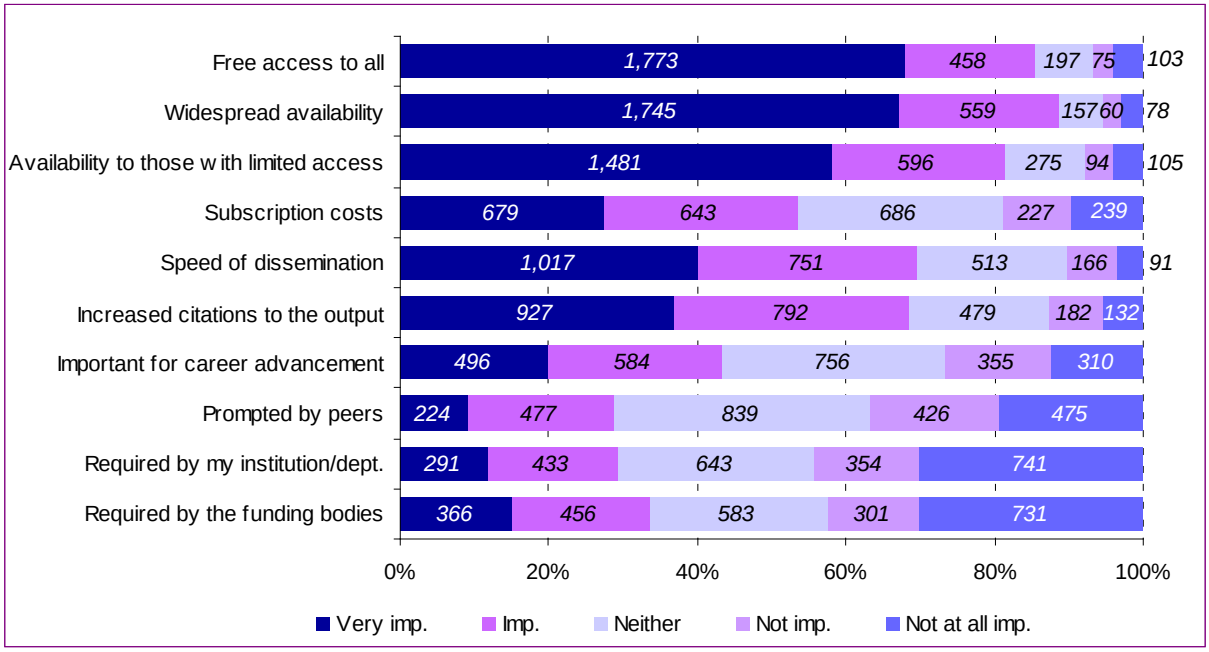


Figure 2.9 Barriers to repository deposit

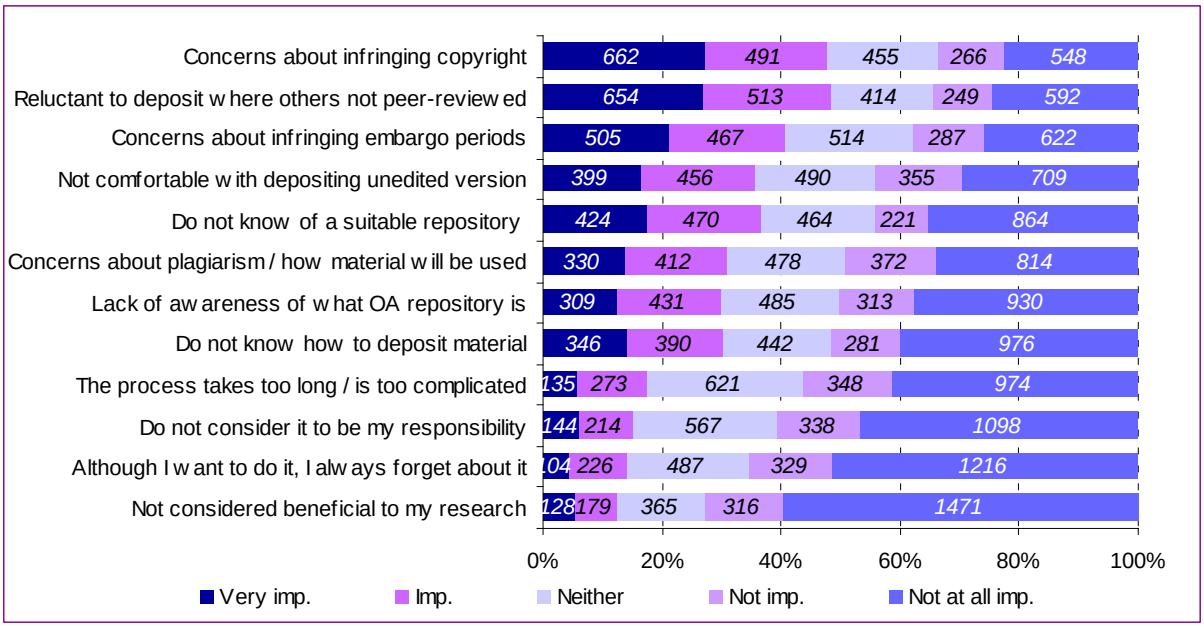
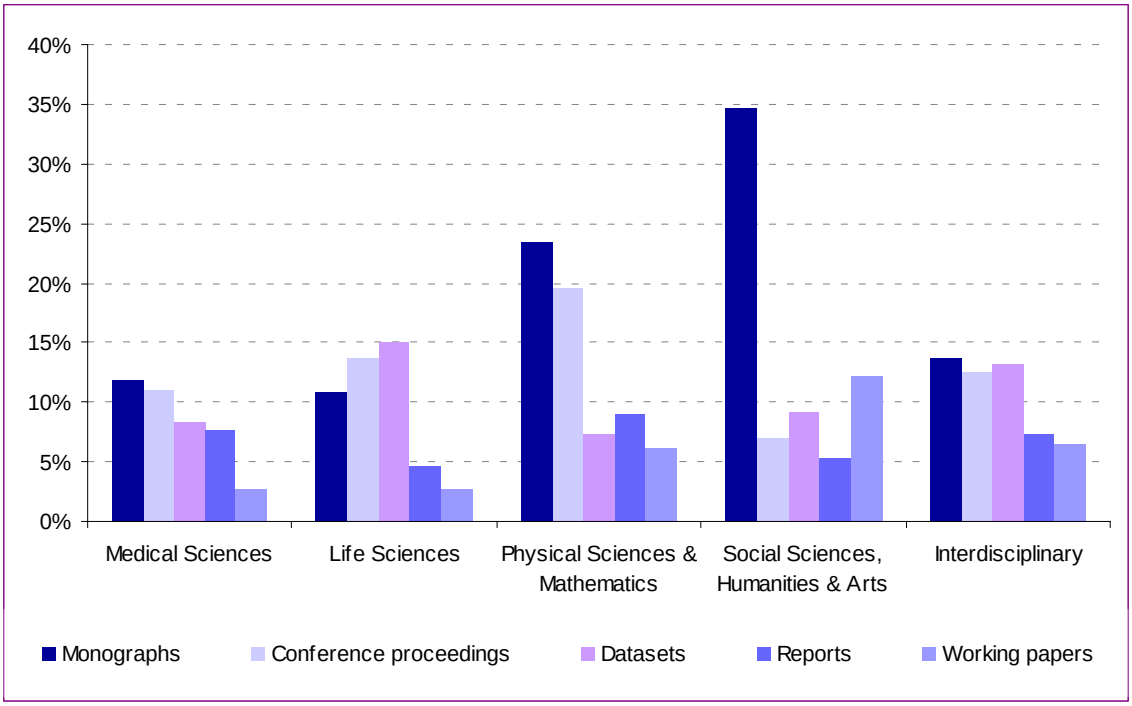


Figure 2.9 summarises the importance of various barriers to depositing material in open access repositories. The most important overall are concerns about infringing copyright, a reluctance to place research publications in a repository where other material has not been peer-reviewed, and concerns about infringing embargo periods. These were the top three concerns in all subject areas except Medical sciences, where a lack of knowledge of how to deposit material was the second most important reservation, and concern about infringing embargo periods came much further down the list. The three least important barriers were the same for all subject areas.

2.3 Information behaviours

The next section of the survey focussed on respondents' behaviours, attitudes and opinions as consumers of research publications, and in particular on their information searching behaviours. Over 90% of respondents rated peer-reviewed journals as 'very important' to their research, although this varied by subject from 95% in the Life sciences to 86% in Social sciences humanities & art. There were statistically significant differences between subject areas in the relative importance of various other types of output when considered as resources for research. Social sciences, humanities & arts and Physical sciences & mathematics respondents considered monographs to have greater importance than did Life scientists and Medical scientists. Physical sciences & mathematics respondents considered conferences to be more important than did those from the Social sciences, humanities & arts. Life scientists and Interdisciplinary researchers gave the highest ratings to datasets. Life scientists gave the lowest importance rating to reports, while the Medical scientists rated these most importantly. Social sciences, humanities & arts respondents gave the highest importance ratings to working papers. Figure 2.10 summarises these differences, showing the percentages of each discipline group rating each of these types of output as 'very important'. The importance of professional journals did not differ by discipline; other types of output had insufficient data for formal comparisons to be made.

Figure 2.10 Proportions rating outputs 'very important' as resources for research (excluding scholarly journals)



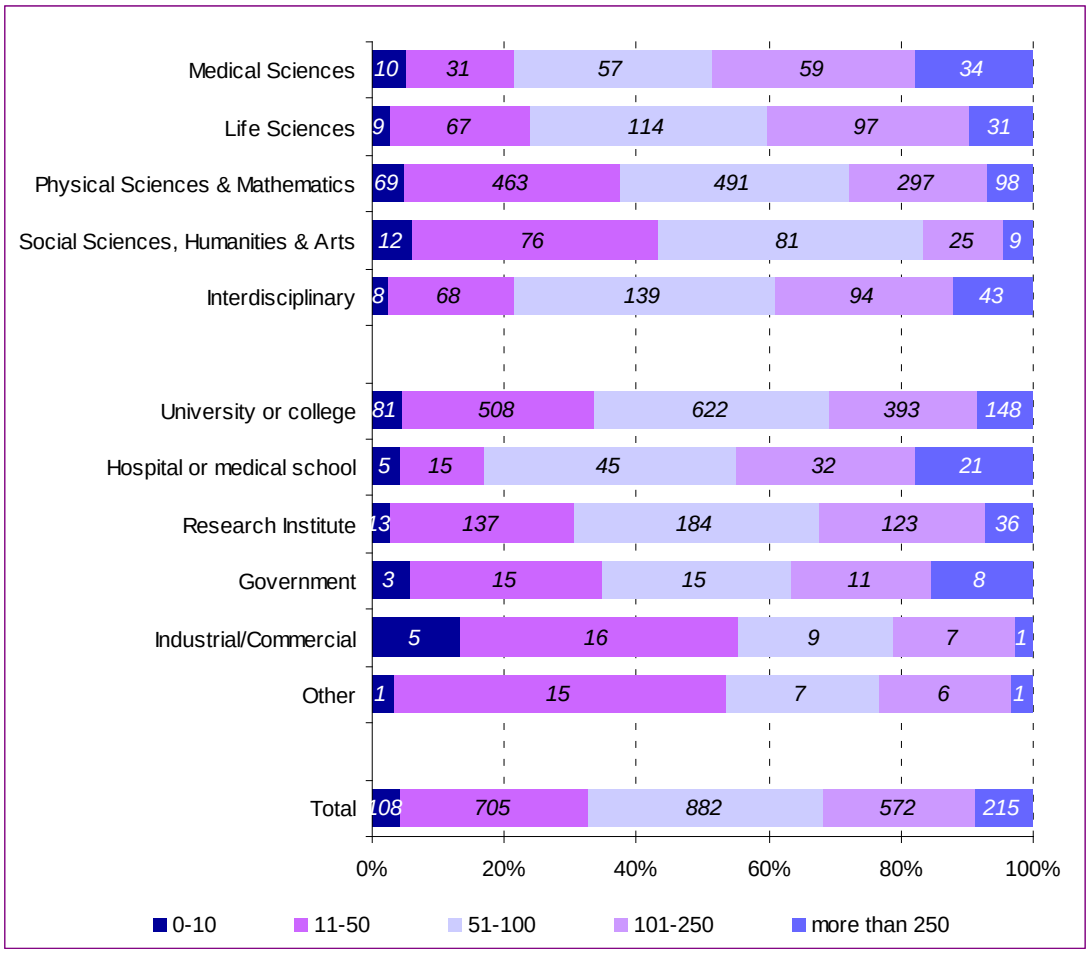
The total number of respondents on which the percentages are based varies by type of output, within the following ranges: Medical sciences 189 - 198; Life sciences 298 - 313; Physical sciences & mathematics 1,291 - 1,493; Social sciences, humanities & arts 197 - 208; Interdisciplinary 348 - 360.

Given the importance of journal articles as research resources, it is not surprising that over 90% of respondents read more than 10 peer reviewed articles per year, on average *(Figure 2.11)*. Just over one third of the survey respondents based in universities indicated that they read between 51 and 100 peer-reviewed articles per year on average, and less than one quarter of respondents read between 101 and 250 peer-reviewed articles per year. This implies a lower frequency of reading journal articles than found by Tenopir and King in a series of studies conducted in the US and Australia (Tenopir and King, 2000; Tenopir *et al.*, 2009a). There are number of differences between the studies, not least in the way in which the questions were asked, as well as the different geographical base, which may go some way towards explaining these differences, which will be investigated further in our analysis of the survey data. There were interesting differences in reading frequency by institution type, for example, illustrated in Figure 2.11, although there was insufficient data from some of the institution types to formally test the statistical significance of these apparent differences.

There are also clear differences between respondents from the five broad subject groups, with those from Medical sciences and Life sciences reading more, and the other groups reading less peer reviewed journal articles than average. Almost half of Medical scientists (49%), and two in five Life scientists (40%), read more than 100 such articles per year compared to just over one quarter (27%) of Physical scientists & mathematicians, and 16% of respondents from Social sciences, humanities & arts. One explanation for the finding that Physical sciences & mathematics respondents tend to read fewer peer-reviewed journal articles than other science-based disciplines may be the importance attached to pre-prints in some subjects within this broad discipline. It is possible that, having read the pre-print, researchers do not then go on

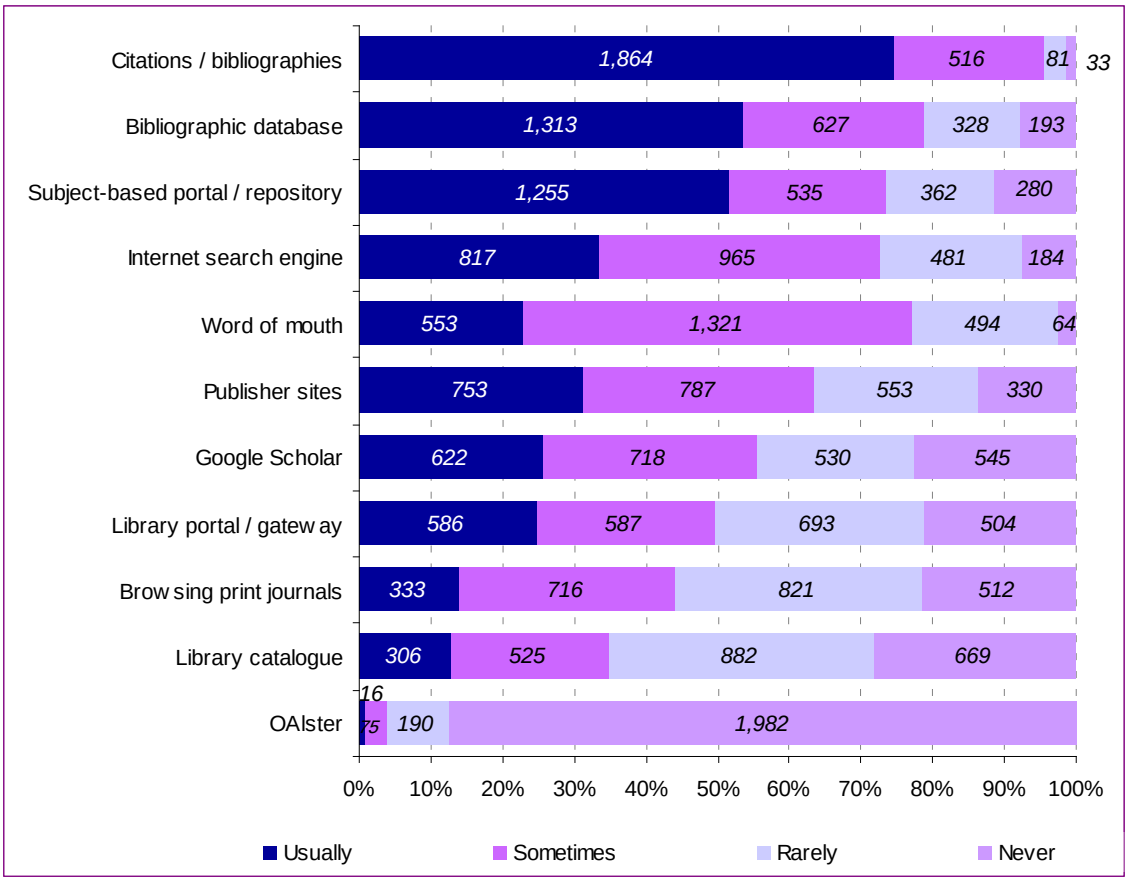
to read the peer-reviewed version of the article when it becomes available. This is only an assumption at this stage, and remains to be followed up with further research.

Figure 2.11 Numbers of journal articles read



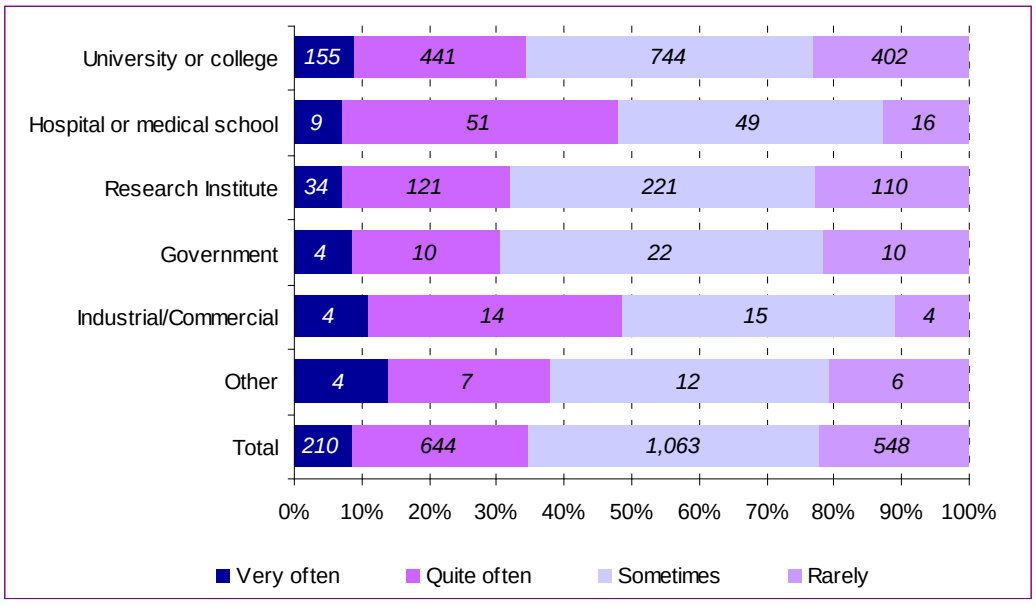
The most frequently used methods for identifying relevant journal literature are bibliographies and citation lists, 'usually' used by 75% of respondents, bibliographic databases (53%) and subject-based portals or repositories, by 52% *(Figure 2.12)*. General internet search engines are more popular than Google Scholar, and the specialist OAIster is little used. There are again differences between subjects, with Medical scientists more likely to use subject portals/repositories, Life scientists more likely to use bibliographic databases, and respondents in Social sciences, humanities & arts more likely to use library resources and search engines, than respondents from other subjects. Some interesting differences by length of career have also emerged; in particular that less experienced researchers (less than five years) are more likely to use Google Scholar, while the most experienced (25 or more years) are more likely to browse print journals, than the other groups.

Figure 2.12 Means of identifying relevant journal literature



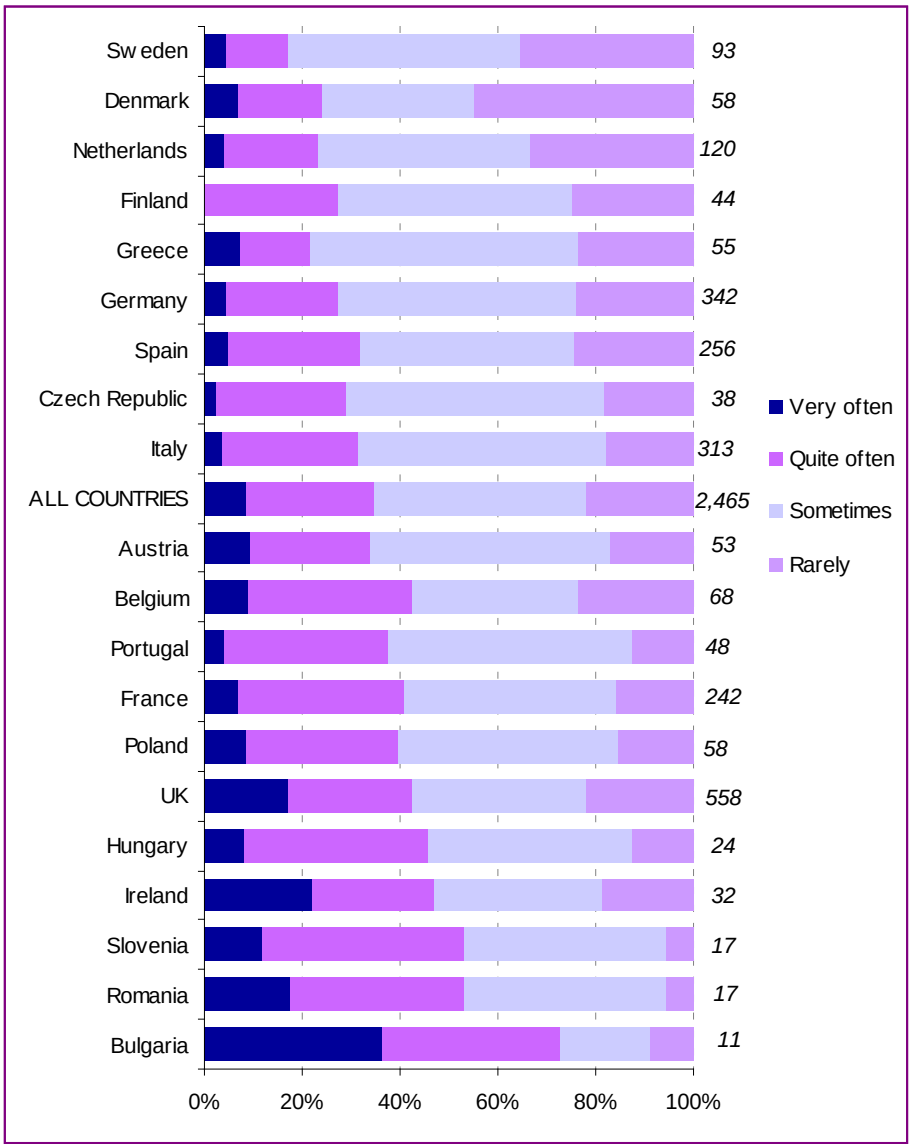
More than three-quarters of all respondents reported being unable to obtain access at least 'sometimes' to a peer-reviewed journal article which they had identified, with 9% reporting that this happened 'very often', and 26% 'quite often'.

Figure 2.13 Frequency of lack of quick and easy access to peer-reviewed articles, by institution type



There were some interesting differences between respondents from different types of institution, although there were insufficient responses from government, industrial/commercial and 'other' institutions to formally test the statistical significance of the apparent differences illustrated in Figure 2.13. This is also an area where differences by country may be expected. Again, there were insufficient respondents from a number of countries to formally test the statistical significance of the reported differences, which are illustrated in Figure 2.14 for those countries with more than 10 respondents. Countries are ordered in this graph according to the relative ease with which researchers report being able to access peer reviewed articles. The research team plan to carry out further investigations by region, following consultation with the PEER Research Oversight Group to select meaningful regional definitions.

Figure 2.14 Frequency of lack of quick and easy access to peer-reviewed articles, by country



Figures in italics are the total respondents from each country;
'All countries' includes those not shown separately.

The most usual action reported when respondents were unable to access an article was to seek an open access source for the material, with half of respondents overall 'usually' doing this.

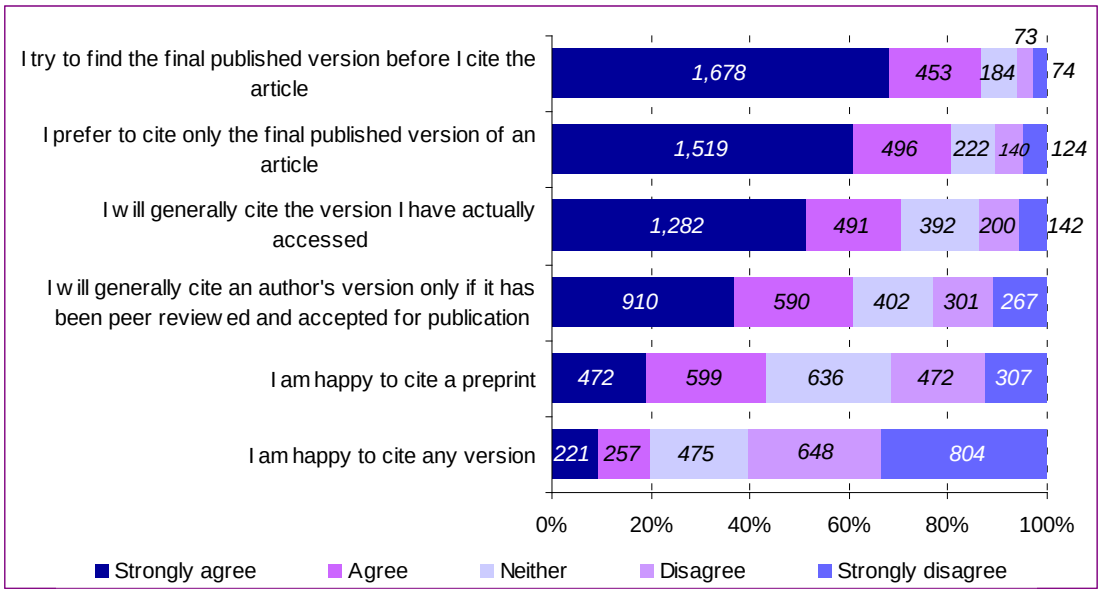
Alternative actions include asking colleagues, contacting the author, asking the library or seeking inter-library loan. Respondents were reluctant to pay to use or download the material, with 64% reporting that they 'never' did this, although the most experienced group (over 25 years experience) were slightly more prepared to do this than respondents with less experience. An interesting difference between subject groups was noted here, with 72% of Life scientists reporting 'never', compared to 49% of Medical scientists. Again, there were differences noted according to the length of research career, with the least experienced group (up to five years) less likely to contact the author, and more likely to ask their colleagues, than more experienced researchers.

When respondents were led to an open access source for a full text article, it was not always clear that this was the case. Only one-third of respondents reported that it was 'usually' clear that that articles were held in a repository, and this was most likely to be so in the Physical sciences & mathematics. Seventy-four percent of respondents overall felt that it was 'very important' or 'quite important' to know which version of the article they had found (i.e. a pre-print, stage two manuscript or publisher's pdf), but only 39% reported that they could 'usually' identify the version from the information provided, while just 37% reported that they 'always' trust the integrity of documents they find in repositories.

There are some interesting patterns in the responses by subject area concerning the use of articles in repositories. Respondents from Medical sciences and Social sciences, humanities & arts are most likely to consider the version of an article important, and least likely to feel that they can identify this from the information provided. Consequently, they are least likely to trust the integrity of documents in repositories. These issues require further investigation, as they are particularly pertinent to the PEER Observatory.

More than two-thirds of respondents in all subject areas will try to find the final published version of an article they wish to cite in their own articles, while less than half are happy to cite pre-prints *(Figure 2.15).* Respondents in Physical sciences & mathematics were more likely to agree that they were happy to cite pre-prints, or any version, of an article than those in the other subject groups.

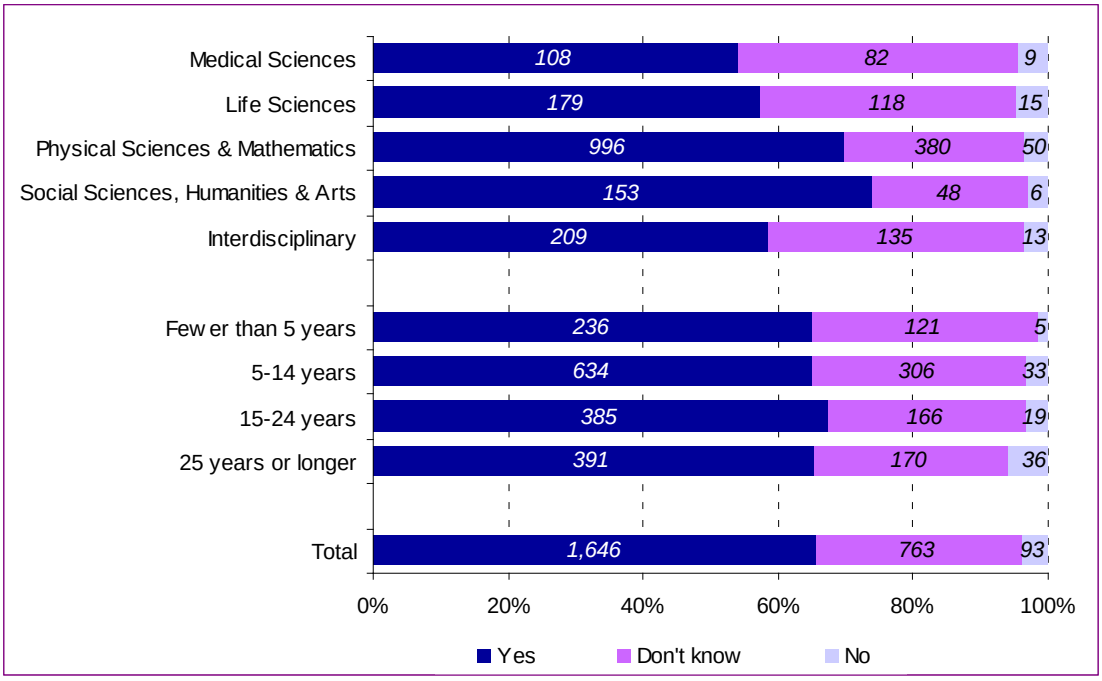
Figure 2.15 Citation preferences



2.4 General attitudes

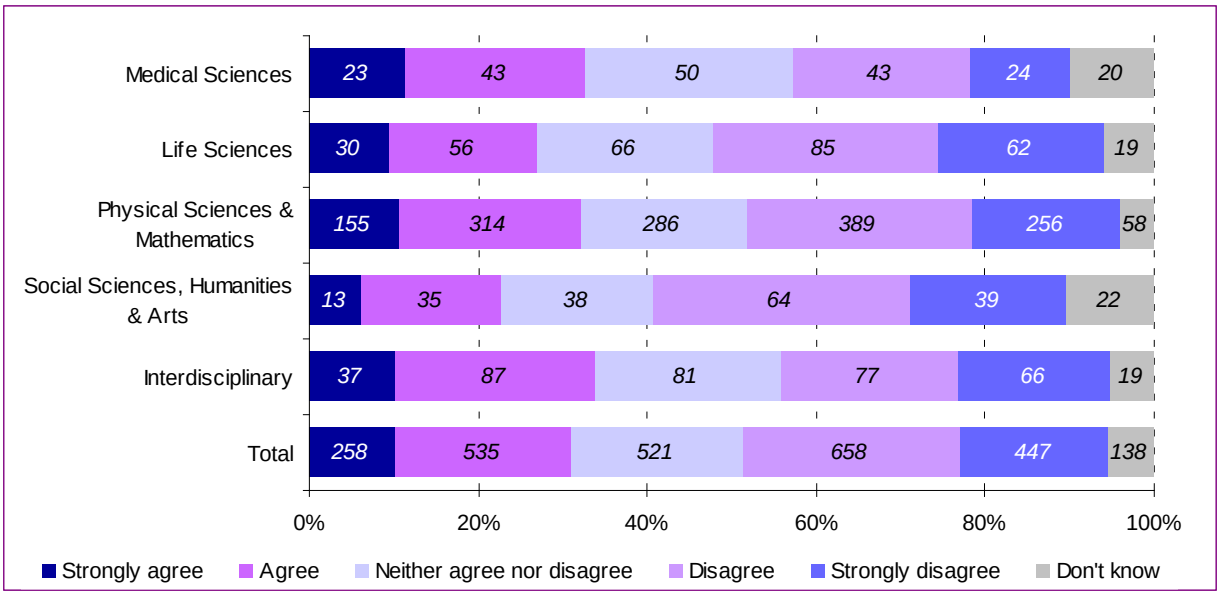
The final section of the survey sought more general opinions concerning the deposit of research material in repositories. Two-thirds of respondents were of the view that there is a role for repositories in scholarly communications, while 30% did not know. Just 4% thought that there was no role for repositories. The differences between subject groups shown in Figure 2.16 are statistically significant ($p<0.01$). The main differences are in the proportions who think that repositories do have a role in scholarly communications compared to those that have no opinion on the matter. The differences between research experience groups shown in Figure 2.16 appear small, but are also statistically significant ($p<0.01$)The more experienced respondents were, the more likely they were to have a view on this, and the more likely they were to consider that there is no role for repositories *(Figure 2.16)*.

Figure 2.16 Repository role in scholarly communications



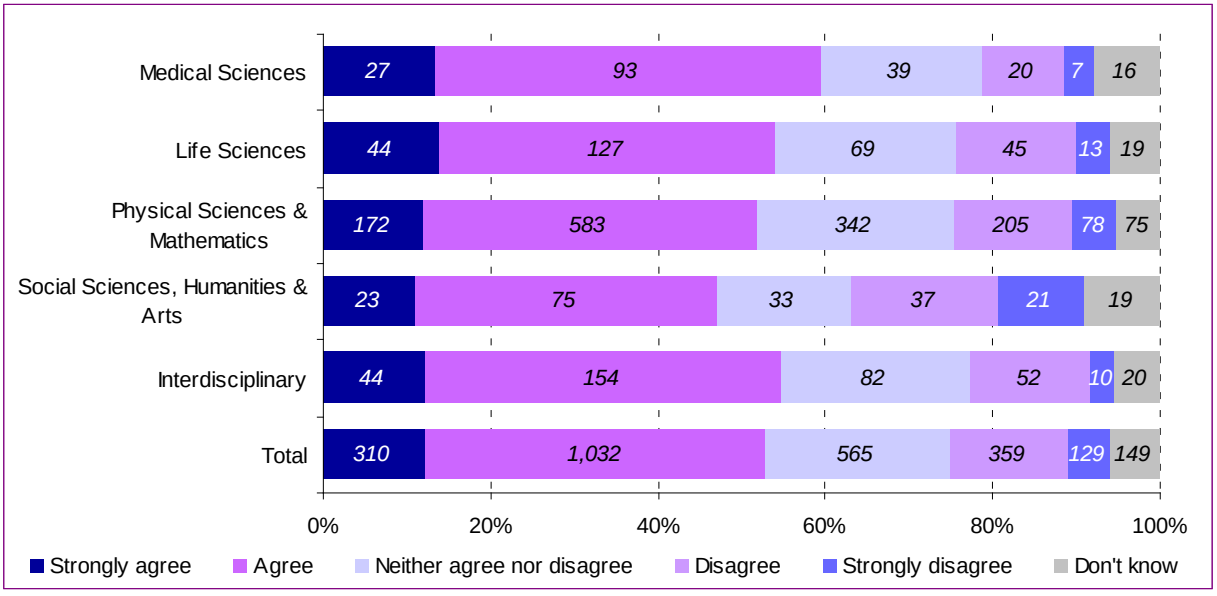
Opinion concerning the perceived challenge posed to the peer review system by depositing research material in repositories was divided. Although 43% of respondents disagreed with the perception that depositing material may challenge the existence of the peer-review system, 31% agreed, with 20% having no opinion. There were statistically significant differences noted ($p<0.01$) between the subject groups *(Figure 2.17)*. Respondents from Social sciences humanities & arts were least likely to agree, and most likely to disagree with the perception.

Figure 2.17 Challenge to the peer-review system



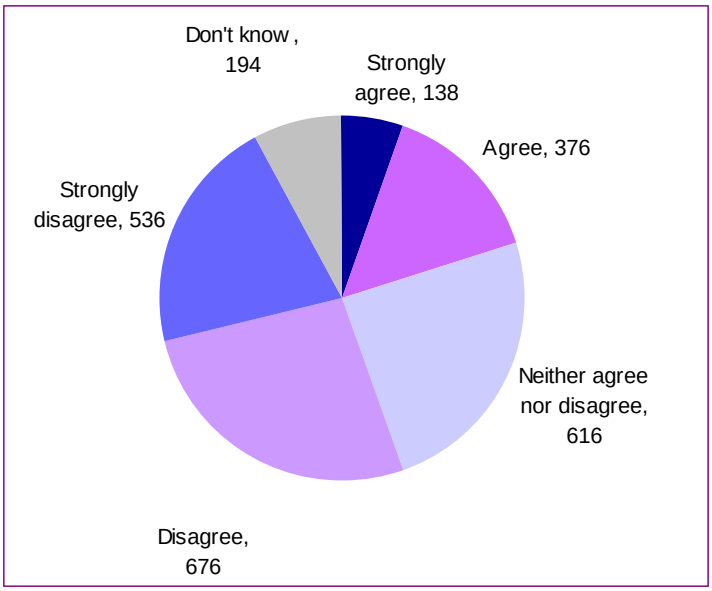
More than half of all respondents overall (53%) agreed with the perception that depositing research material in repositories might challenge the existence of subscription-based journals. Almost one in five (19%) disagreed, however, with 22% having no opinion. Again, there were statistically significant differences between the subjects ($p<0.05$), illustrated in Figure 2.18. Respondents from the Social sciences, humanities & arts were least likely to agree with this perception.

Figure 2.18 Challenge to subscription journals



With regard to the perceived challenge posed to the scientific societies by depositing research material in repositories, opinion was again divided *(Figure 2.19)*, although only 20% of respondents agreed that this might be the case. Almost half, 48%, disagreed, with 24% having no opinion. In this case there were no statistically significant differences between the subject areas.

Figure 2.19 Challenge to scientific societies



3 Focus group analysis

A number of broad themes emerged from the initial analysis of the focus group discussions, in terms of issues relating to:

- Dissemination and visibility – as motivations for researchers to publish in open access sources.
- Quality and reputation – both in terms of the prestige of journals and open access repositories, and the confidence researchers place in these resources, both as authors and readers
- Policy – including copyright, mandates and embargo periods, and tensions between these policy elements, and dissemination behaviour and attitudes to open access repositories.
- Locating and citing resources – factors influencing researchers' information seeking behaviour and their attitudes to using open access sources
- Repository services – including issues related to the management of versions, and the functionality of open access repository

Presentation of the analysis is arranged according to these themes. The demographic characteristics of participants were detailed in section 1.7.2.2. A description of the focus group protocol is in Appendix 2.

3.1 Awareness of open access repositories

In order to explore their perceptions and awareness of open access repositories, participants were asked to place themselves in a situation where they have just finished writing a paper and were asked whether uploading this paper onto a repository comes to their mind at some point. From this scenario, it was possible to gauge the extent of researchers' awareness of open access repositories, as well as motivations, views and immediate issues arising from uploading material onto an open access repository.

The extent of awareness of open access repositories across the four focus groups was generally low, ranging from most participants having a relatively confused idea of what open access repositories are and their purpose, to a minority of participants using open access repositories on a regular basis. Although the focus groups were arranged in broad discipline categories, some fine-grained disciplinary differences emerged and will be detailed later in this section. Although only a small minority of participants, mainly within the sub-group of physicists, had already uploaded material onto a publicly accessible repository, the participants of all four focus groups generally showed a great interest towards open access repositories, as might be expected given the self-selecting nature of the participants.

Participants with a low level of awareness of open access repositories seemed to have, however, a general understanding of what the 'open access' concept is and generally had a good understanding of open access journals. Although two Life sciences participants tended to confuse open access with online access on several occasions, other participants showed a good understanding of the 'open access' concept. The term 'repository' was, however, generally less obvious to them. They associated it with some sort of database, digital location or digital collection, but they were relatively unsure about what was held in these repositories. Several participants across all four focus groups commented on their own institution having set up an

open access repository in the last few years and encouraging researchers to make their research available through the repository, but they were generally still unsure about what it was exactly and what it meant directly for them. A recurrent question across the four focus groups, was to ask for clarification as to whether the discussion was about repositories of peer-reviewed material or any sort of material, including pre-prints, and some participants seemed to assume that repositories are only for pre-prints. The physicists proved to be the exception to this, since they routinely use arXiv and use both pre-prints and author's final versions.

Interestingly, participants seemed more aware of open access repositories as readers rather than as authors. A connection between open access repositories as a tool for authors and open access repositories as a tool for readers was not very apparent at any of the focus groups. This was particularly evident for the Medical sciences field where all participants were very familiar with PubMed Central, without really associating it with being an open access repository. Similarly, some of the participants in the Life sciences focus group were familiar with Medline as readers but had never placed any material into an open access repository.

It emerged from the focus groups that only the sub-group of physicists had a long tradition of using open access repositories (mainly arXiv, and some were familiar with the NASA-funded and Harvard-operated ADS[9] repository), both as authors and readers. A lawyer in the Social sciences, humanities and arts group clearly showed a good knowledge of open access repositories such as SSRN (Social Sciences Research Network) as a reader but was more reluctant to put full-text articles up into this very subject-based open access repository. He said *"I've uploaded four abstracts but I haven't uploaded any manuscripts, so I belong to the conservative area of lawyers who haven't yet dared to upload anything so far."*

With the exception of the physicists and one or two life scientists, self-archiving was not part of the current practice of the participants. Hence, in the discussion that follows the self-archiving motivations put forward by the group of physicists are grouped with potential motivations indicated by the rest of the participants who do not currently self-archive at all.

3.2 Dissemination and author visibility

Reaching target audiences was seen by the participants as the primary motivation to upload material onto an open access repository platform. They thought dissemination was central to scholarly communication and anything providing greater dissemination, and visibility for their research would therefore enhance scholarly communication for the benefit of all researchers.

Participants familiar with open access repositories identified 'greater visibility' for their research as a motivation to upload material. Visibility was also linked to the need for researchers to build up a reputation in their research area which comes also through citation. Participants thought that the more they had articles on open access, the more they were likely to be cited, *"since our reputation is built, unfortunately, on the number of times we are cited, it would be good to be totally available, because more people cite us."* A few participants also indicated that open access repositories could help disseminate research in disciplines for which there were too few journals with high impact factors, or for inter-disciplinary fields for which there were no specific journals in which they could publish.

[9] ADS: Astrophysics Data System

Although the discussion was focused on peer-reviewed journal articles, some participants in all focus groups were happy to comment on other research formats they use and create for which dissemination is not as good as for journal literature. For these participants, potential difficulties in accessing journal literature were acknowledged but were not seen as a major issue. They preferred to focus the discussion on dissemination of other research formats, such as conference proceedings or book chapters. They thought open access repositories could have a major role in improving dissemination of these types of research format,

> *"because the journal is important in itself and allows generally an international access by the researcher. For other kinds of papers, in my opinion, there are some conference proceedings that are not so well distributed, well known, in the community. It is difficult to have access to some good-quality conference proceedings."* (Physical sciences & mathematics)

or

> *"books and chapters in books in principle have a lower/smaller circulation because maybe people are only interested in one chapter rather than the whole book, so I think that from my perspective, looking around, it would be more interesting to have a repository of chapters of books than [a repository] of articles because journals are available, you can find them in libraries, […] and after all it does not cost very much just to buy one single article whereas to buy a whole book just for one chapter, that could be a bit costly."* (Social sciences, humanities & arts)

Speed of dissemination was also a key motivation. Those participants using the repository arXiv indicated that speed of dissemination was a crucial element in Physical sciences & mathematics, where competition is very high between research groups. Such intense competition was identified as a reason why many physicists tended to deposit pre-prints in arXiv rather than authors' final versions. In some cases, research groups just could not afford to wait for the paper to be reviewed and accepted by the journal, and therefore upload onto the arXiv repository was taking place while the paper was going through the peer review process. Participants explained that such practice was aimed at putting a 'date stamp' on ideas and making papers available on open access as soon as possible both via the arXiv platform and via their Web pages.

Linked to timeliness, high competition and peer pressure were identified as two strong motivations for physicists to use arXiv, despite the reservations some of them had about this practice:

> *"From my point of view, I agree that the existence of the arXiv is a bit of a pain in the neck, it's something that I see the community uses, if I want to stay in competition, I have to use it too, but really very seldom I catch an arXiv paper that I would sort of learn the same by waiting when it comes out as a paper, so it is something that I have to do because if you want to stay in the game, you have to do this."*

Some participants also mentioned that they were feeling a constant pressure, not only to publish in high impact factor journals, but also to be cited, *"the point is, as you know, we are pushed for citation, and we are pushed for impact."* In such context, the general opinion of participants in the Life sciences group was that open access articles were more likely to be cited, *"if it appears as early as possible, it will be cited as early as possible, more citations per page […]. And then therefore the open access for us […] is a very good thing, because then everybody can easily get it and can cite."*

As a corollary of speed of dissemination, publication delays were put forward on several occasions as a further motivation to put peer-reviewed articles onto publicly accessible repositories to mitigate those delays, for example:

> *"Once a paper is accepted, and before it goes through the actual publication in our journals, typically takes another month or so, we put it on arXiv and then it can be uploaded from our web page. So the arXiv number is there, people click that and get the manuscript before the final thing is published."* (Physical sciences & mathematics)

and

> *"What I use and what probably people who work in economic law, which is very international as you can imagine, is the SSRN, Social Science Research Network which I hadn't heard of before until I went to the USA in 2004 and everybody was talking about SSRN because you upload the working papers before they get published, because the publishing procedure takes time, you can gather information about what your colleagues are doing and so forth."* (Social sciences, humanities & arts)

Participants indicated that reasons for preferring one type of repository to another included the targeted audience, the reputation of the repository and what content and formats were already held in the repository. When asked whether they would have any preference between institutional and discipline-based repositories, participants' views were varied across disciplines, but the majority indicated a clear preference for discipline-based repositories. Participants generally felt that the *"relevance to the audience"* was an essential element in their choice of repository, which might explain why they were more inclined to choose subject-based repositories. A Medical sciences participant commented that *"if I'm publishing something virological, I'd put it in a virology repository, or something like that."* However, some participants also found the idea of institutional repositories attractive. In relation to identifying articles by searching for the work of key authors and research groups in the field, a physicist commented:

> *"… so in a sense the repositories of local universities and local departments could be more useful to the way people operate than the unfiltered repositories that are there, and people will keep using them, and I'm afraid that they will move into all other sciences, it's not true to say that what physics has done is very special for physics, it has been really the writing on the wall, it's going to happen in every field".*

For those participants who accepted the idea of uploading material into institutional repositories, the reasons put forward were varied. It was noted that upload only into institutional repositories was rarely mentioned. Institutional repositories were always considered in combination with subject-based repositories. A few Social sciences, humanities & arts participants thought that researchers may have a close connection with their institution (financial dependence and sense of belonging to the organisation, for example), and therefore may want to place their research articles in the institutional repository:

> *"Probably the institution is a bit closer because they pay you, so probably that's what we would do if we had such a thing, and then second SSRN because my boss might see it and think 'He's published something nice'. Things like this are important."*

and

> *"I mean both have this positive side, though actually I don't see the negative sides for the institutional repositories, because it's accessible, people can see it, and if someone's looking for it they would probably look at the institutional home page first. It is more easy*

to find that way. So it's more visib[le]. Visibility is also important if you put it in a public repository or some specific repository."

Inertia of dissemination practices was identified, amid others, as a main barrier to the take-up of depositing copies of published articles in repositories. Although participants acknowledged that access to journal content could sometimes be an issue for them, some admitted that they had little incentive to deposit material in open access repositories, and therefore making their articles publicly available via an open access repository platform was not at the top of their priorities. Participants were unanimous in their view that their top priority was to get published in high impact factor journals in order to gain the recognition of their peers for their contribution to the development of their discipline. It was felt that everything open access repositories could bring to them was perceived as a valuable, but not essential, aid to reaching this objective.

A participant in the Physical sciences & mathematics focus group observed that the dissemination itself is more important than the mode of dissemination: *"The other thing that people use, and it's more seriously of advantage in disseminating communication, is in fact the existence of individual web pages."*

3.3 Quality, reputation and confidence

Although their understanding of 'open access' was generally sound, most participants did not publish in open access journals because of concerns about prestige and reputation; they thought open access journals usually have a low impact factor as it takes time for journals to build up a reputation and achieve a high impact factor. As one participant in the Medical sciences focus group stated:

"Open access journals that are in my field […] have a low impact factor, so I submit my papers to the one[s] that are not open, because basically […] they are better".

The quality standards of the repository were also mentioned as a decision criterion that would apply when making articles open access via a repository platform: *"if it is a repository, which is known to have a very low standard [of] quality, maybe I wouldn't put it there, just for the sake of making it available."* In terms of quality standards, participants often mentioned that the content held in the repository would influence their decision. Although they did not explicitly talk about material other than research material (i.e. teaching material, student's works etc.), it was felt that participants were clearly only considering repositories of research material as suitable locations for their research articles, probably even in the case of institutional repositories. Opinions were, however divided about what sort of research material would be appropriate in a repository. While some participants, mainly in physics, found pre-prints acceptable for scholarly communication, other participants expressed concerns over this format and would prefer to have only articles accepted for publication, whether the publishers' PDF file or the authors' final version. A Social sciences, humanities & arts participant raised doubt over the quality of repository material from his perspective as a reader: *"Because these things [peer-reviewed journal articles] were already published so they are reliable, whereas with the SSRN it might be that the working papers are not finalised, so I wouldn't trust them that much".*

Reputation of repositories was an issue which came up regularly in the focus groups. Most participants unfamiliar with open access repositories questioned whether there were impact factors attributed to repositories in the same way journals have, *"I think that the problem of repository that we are talking about is that it would have to build up a reputation […], and I*

wouldn't know any other means but some sort of peer review to build up reputation." Some participants indicated that if open access repositories were to become an additional way to disseminate their journal publications, then they thought it would be interesting to have impact factors assigned to repositories too. Regarding esteem and reputation, a Life sciences participant made a point in saying *"in that case there might be one thing to consider, that you wouldn't put your paper in a repository or not you yourself use it actively, or at least I think so. So I mean what's the point if you don't have a high enough esteem for that repository to put your paper in there."*

3.3.1 Confidence in open access dissemination

It was also apparent that research workflows and confidence in one's own research vary considerably between disciplines. The discussion showed that some disciplines, such as Physics, have a research culture that prompts authors to disseminate ideas as soon as possible.

> *"The question here is are people using at the same time open access archives to speed up the distribution of what they have done, and at which level do they do it? That is why I was talking about self-confidence. If a research group is fairly successful, and hence the perception of what they are doing is very important to that community, let us say they are running for the Nobel prize, or something like that, then as a matter of course, I know this for a fact, people send their papers to whatever journal they choose […], and at the same time, they deposit the PDF so it is unchangeable on the arXiv. So they immediately have an arXiv number which appears on their web page … . But that means whatever you have prepared and you have sent to a journal, you are actually distributing it before you get it all evaluated. So this is the level of maximum self-confidence."* (Physical sciences & mathematics)

The Italian sub-group of physicists was divided regarding this practice. Some did not share such high levels of confidence and therefore did not deposit pre-prints in arXiv. They preferred to wait for the paper to be peer-reviewed and accepted before any upload onto arXiv, for example:

> *"So, the answer to your questions is that we all want to put it on arXiv as soon as possible, and in my view, we do it only after we get the first round of referees' reports. In some cases, if that takes too long, if you send it to a journal and the journal takes, say, two months to let you know what happens, then you may not want to wait two months, but you think that the particular community that refers to that paper would like to know about it before, and this is a self-confidence level. I mean you have to have a perception of what you have done, it is really so interesting that people can't wait to read it and if you have that perception, good for you then. You don't wait for the journal's answer and you put it on arXiv. I see, for instance, that the American community which is well-known for a high level of self-confidence uses arXiv as a matter of routine. So everybody puts its papers as they send it to the journals on arXiv. We are sort of in-between, we wait for the first round of referees, before you put it on arXiv."* (Physical sciences & mathematics)

Other participants felt comfortable depositing papers in arXiv before they had been through the peer review process. They however indicated that such practice had consequences on where they could publish those papers as not all publishers accepted papers that have been already made available publicly, for example:

> *"Basically, I put things on the arXiv at the same time that I send to the journal but this attitude varies among colleagues but the point is also the attitude of the publisher is not*

the same. So some publishers allow you to do that, like the Journal of American Physical Society […] I mean they accept that." (Physical sciences & mathematics)

Other disciplines, such as the Medical sciences, appeared to be more cautious with their results and wait for the validation of their research by their peers through the journals' peer review system, *"I would be very reluctant to put something out in the public domain until I had a publisher's imprimatur on it".* Regarding the deposit of pre-prints, a Medical sciences participant also indicated that she could see some risks associated with such practice in terms of competition in that specific discipline, for example:

"I think that's particularly important if you're working in perhaps a small group, because if you come up with a novel idea, and it's out in the public domain, a large group can beat you to the publication post, because they can actually work faster and more efficiently than you can perhaps, and I think that's something we're going to have to bear in mind. […] I think that's very important, because you can lose your footing in a field by allowing your information out too soon"

A few participants were very concerned about the protection of research material held in repositories, not so much for articles but rather for datasets, for example:

"we already circulate among our colleagues or peers, but not too open because there is also a problem of people just stealing your work […] I'm completing a paper which I hope will be published, using a lot of original data which I took a lot of trouble and time in collecting. If I put this in open access, this paper with nice tables which I worked a lot doing, and then somebody then just copies and pastes this, then even if they are quoting me, when I am going to publish it, I am no longer the original, I would have thrown away all my work. There is a lot of stealing around." (Social sciences, humanities & arts)

and

"I don't know really if such repository, open access repository, may have […] some protection on the data you put, because of course anyone can steal your data, and to publish other papers, and so on, so this could indeed create tension with young researchers, which of course they are very interested in protecting their data, because their career is anyway linked to publication" (Physical sciences & mathematics).

3.3.2 Confidence in open access information seeking

When asked to indicate whether they would be comfortable using an article available in an open access repository and whether they would trust the source, participants voiced quite strong, and sometimes contrasting, opinions about using material held in open access repositories. Key findings were that:

- The notions of trust, quality and peer review are intimately intertwined. Peer review is generally seen as the primary indicator of quality and researchers are comfortable using materials they trust are quality materials.
- A minority of research fields has already embraced the use of material held in open access repositories, including pre-prints. This is mainly the physics and mathematics community. Other research communities still harbour suspicion towards open access material.

Although differences might be occasionally found within a discipline, disciplinary norms seemed to apply and guide participants' behaviour and, therefore, results are presented for each disciplinary grouping.

Medical sciences

Medical sciences participants indicated that, in this specific case, the name of the authors of the article was the primary criterion determining their decision to access, read and maybe use the resource. There was a consensus within the group on this. A participant clearly answered *"it depends on the authors"* and added *"even more so!"* when presented with the case of a pre-print article held in a repository. Therefore, pre-prints were not totally ruled out for some participants as long as they knew the authors and because, one participant argued, *"in the end […], you've actually got to read the paper and actually make you own mind up, the lead and the content."*

In the case of an article written by authors they have never heard of but affiliated to a prestigious institution, participants held mixed feelings. Whereas a participant clearly said *"I wouldn't go for that, I wouldn't trust it […] because it could be any student's work. It could be just a summer project, it could be anything,"* another indicated a more balanced view and said *"I don't think I would distrust it"* - though this very participant insisted again that s/he would only trust pre-prints from authors s/he already knew. Participants also indicated that their decision to use pre-print material was also guided by what they intended to use the resource for, and linked up to publishers' rules regarding citation of unpublished material. This point is developed later in the section on citation practice *(Section 3.6.2)*

When presented with the case of a peer-reviewed article from an unknown person from an unknown institution, participants indicated they would generally consider it *"because there are always new people around […] and there are new investigators coming in, and if it's been peer-reviewed, it means it's been read three times, three or four times."* Because they generally had a good idea of the peer review standards held by individual journals, participants felt quite comfortable in deciding whether they should use the designated article or not only on the basis of the reputation of the journal. Although peer-review and journal reputation were not clearly cited as primary decision criteria, the group indicated that these were however important factors, sometimes as important as authors' names, in influencing their decision to use a resource or not.

Social sciences, humanities & arts

The Social sciences, humanities & arts group indicated that peer review was an important element factoring in their decision to use and therefore trust an article available in an open access repository.

Building upon this, most participants insisted on the fact that metadata were a crucial element of any repository infrastructure. They thought it was essential to have appropriate metadata to inform the reader about the status of the article they were about to access and use.

Commenting on their experience with SSRN, a researcher in law said *"I would probably only take note of articles which were accepted for a serious journal, with others I might be more sceptic and not even spend enough time on it."*

Life sciences

There was a consensus within the group to say that they would feel comfortable using a research article held in an open access repository if they could ascertain that this article has been through peer review. So, peer review appeared as the main decision criterion in regard to open access material. They stressed the fact that this information needed to be clearly stated for readers to make informed decisions about using such open access sources. This links to the essential role metadata play in an open access context.

The Life sciences group showed no preference between accessing an author's final version and a publisher's PDF file. When asked whether it mattered to them, a participant commented *"not really, no. Well if it looks better […] in the final version, then of course I would choose that. But the most important information that a scientific paper can show to us is usually not the colours and not the pictures. So usually it's okay if it's a manuscript."*

Physical sciences & mathematics

Regarding the question of trusting material in open access repositories, views and opinions were divided between those who do not currently use any repository and those who currently use arXiv or similar repositories. Non-users of repositories tended to show suspicion towards any material which has not gone through peer review and thus has not 'earned' the quality stamp affixed by journal publishers. A participant in the engineering field said *"I would be happy and I would trust the source, if I knew that it would be a sort of reviewed version of paper […]. Probably with a paper on an open access repository, I would trust it, as I say, if I knew it is an accepted version […]."* By contrast, users of repositories showed more openness towards the diversity of material held in repositories, and so were not deterred by the fact that some articles may have not gone yet through any peer review process. Critical judgement and critical thinking were put forward as a better system for checking quality than a publisher's peer review stamp. A physicist participant questioned a possible over-reliance on the peer review process as a safeguard for quality:

> *"You ask whether you trust – of course, I think you will give the same [trust], sometimes you find wrong results, even in peer review article[s], because then maybe this will open up another issue which is not debated today, how accurate is the peer review process. […] So, of course you are naturally suspicious, but I would say you put the same attention that you would dedicate to an article, so I would read what is written, but I can say that I will take that seriously, but of course [...] if I get convinced that it's wrong, OK it's wrong, but at the first stage I will take seriously, and then if it's right, it's right."*

3.4 Policy

Various issues arose during the focus group discussions relating to policy, notably tensions between perceptions of publisher policy and researchers' preferred methods of dissemination. Key barriers to publishing in open access repositories emerged in the form of participants' understanding of publishers' copyright transfer agreements (CTA), and general uncertainties about the deposit process. Discussions related to mandates were limited because participants showed limited interest in this area, and most comments focused only on institutional mandates to deposit research outputs in institutional repositories. In the Physical sciences & mathematics focus group there was some discussion of the motivations of institutions which mandate their researchers to deposit in the institutional repository, in so far as questioning whether the purpose is more for research evaluation than for research dissemination. Researchers' attitudes to embargoes were also explored during the focus group sessions. Although awareness of

embargoes was generally low, the issues of the impact of embargoes on the timeliness of research outputs was raised.

Participants across the four focus groups identified publishers' copyright transfer agreements as a major barrier to deposit material into publicly accessible repositories, for example:

> *"that lots of publishing houses don't want pre-publication in the form of an on-line working paper."* (Social sciences, humanities & arts)

> *"Some of the journals require you that you sign off all your rights, even to putting it on your own home page."* (Life sciences)

The perception that publishers' CTA require authors to sign away most, if not all, of their rights over their material was a recurring issue, with most participants convinced that they could not make their papers available on open access before (as pre-prints may jeopardize publication), or, indeed, after, the CTA was signed off.

Participants did not seem to be aware that some publishers actually had OA-friendly self-archiving policies and so did not always require authors to sign away all their rights. Participants, with the exception of the physicists who demonstrated a greater knowledge of self-archiving policies, were very confused about the use of pre-prints and author's final versions. They were, however, almost certain they could not use the publisher's PDF file for open access purposes. This contrasts sharply with the findings of a SHERPA Romeo survey[10] which highlighted that 51% of publishers surveyed allowed self-archiving of post-prints, compared to 61% allowing self-archiving of pre-prints

Changing the terms of the CTA was not current practice amongst the respondents. Only a senior participant in the Social sciences, humanities & arts group indicated that s/he always changes the terms of the copyright agreement in order to retain as many rights on her/his work as possible, *"I amend them, I never sign the original contract*", and makes much use of the licensing option wherever possible. This participant called for a directive from the European Union to support the concept of licence to publish which guarantees authors rights over their work, and so enables them to distribute their research as they think is most appropriate in parallel to traditional journal publication.

It was acknowledged by the Social sciences, humanities & arts group that changing the terms of the contract could prove difficult in most cases, especially for young researchers who cannot take the risk of seeing their publication rejected by the publisher because they wish to amend the terms of the contract. It was clear from the discussion that getting published was the utmost priority for young researchers:

> *"[T]hat this is based on changing the contract, and this would be I think very difficult in the area of law, especially as a young author because it depends on the major journals of course and you don't want to create problems about this."* (Social sciences, humanities & arts).

A few respondents in Medical sciences also indicated that the lack of guidance for PhD researchers was also a barrier to take action, for example:

[10] http://www.sherpa.ac.uk/romeo.php?stats=yes

"in my case, it's just lack of knowledge of how is the regulatory system for putting up papers, as I said before, so otherwise I would have done [it] probably"

and

"I think also, because I haven't done it before, we don't know how to do it as well, because there's so much [about] doing your PhD stuff, you know how to put a paper in a journal, but we have never been told how to put it in a repository"

3.4.1 Mandates

Several aspects pertaining to mandating deposit of research outputs in open access repositories were discussed by the participants. Overall, participants seemed to have limited interest in the question of mandates and therefore did not elaborate much on this topic. Perceptions of mandates were varied and ranged from the extra administrative duty keeping researchers away from their research, to the feeling that it was a fair requirement considering that research was often paid for by public money. Although the question about mandates was very general and participants were encouraged to discuss funders' mandates, publishers' mandates or institutional mandates, participants mainly discussed institutional mandates to deposit research outputs in institutional repositories. On one occasion, a participant in the Social sciences, humanities & arts group referred to an EU-funded project s/he was working on and said *"I know for sure that […] after the expiration of our funding we will have to make it public which I think is fair because it was with public money."*

The Physical sciences & mathematics focus group and the Social sciences, humanities & arts focus group were the two groups showing relatively strong, and occasionally divergent, feelings about institutional mandates. Some participants in the Physical sciences & mathematics group had strong feelings about institutional mandates. They felt that research institutions now tended to misappropriate the concept of open access repository, which originally was for dissemination purposes, for quantitative research evaluation and economic reasons. The following quote is an illustration of the tension between evaluation and dissemination and provides a good example of what scholars more generally may think.

"I think that we should try and distinguish between two important aspects here, because the discussions seem to be moving into the effect that using this particular repository has on people's work evaluation, and originally all of these things were not formed to evaluate people, they were actually organised to disseminate knowledge, to allow people to be informed of something else, it's really a facility […]. Now, of course then this as a result can also be used as an instrument of evaluation, but the primary vehicle is really to circulate information, so we had to really understand these things, how we view them, how we consider them important, to circulate what we are doing, not so much what do we gain from it by being better evaluated, that's something else. […] I think one has to first ask one's self […] do we think it's advantageous to circulate that information for making other people aware of what we're doing? And I think the answer to this is very strong in your second scenario, most British universities, because of this academic evaluation exercise, require that the people in the various departments deposit in some sort of archive at the university all the things that they do, so the university can say what each department has produced, and so this becomes an instrument of evaluation more than an instrument of circulation of your effort, so this is another aspect, this is an evaluatory pattern, the original pattern of the arXiv was information improvement, so we have to decide which of the two we are following."

Participants generally felt that institutional motivations to implement mandatory self-archiving in institutional repositories did not meet authors' requirements for greater dissemination and visibility. One reason for that was that it was felt that institutional repositories were more an archival platform than a circulation platform. A regular user of the arXiv repository commented about institutional repositories *"when I put my things there [in the institutional repository], nobody will find out from it anything new about my work, it is just used for evaluation."* Participants did not show any sense of ownership towards institutional repositories. Their discussion suggested they did not think institutional repositories could take their research to their research community and readers in the same way subject-based repositories such as arXiv do.

Not all participants, however, shared the view that institutional mandates are solely an instrument of evaluation. A social scientist acknowledged that mandates could possibly enhance scholarly communication through a better dissemination of knowledge, *"granted that the journal does not resent it, we go back to the idea that it helps diffusion of knowledge. Granted that everything is OK from a legal point of view, we want to be required!"*

The question of mandating researchers to place their research in some open access repositories raised the issue of the actual management of deposits of research articles under mandate. Participants generally thought *"it's just more work"* and they did not have the capacity and time to do the extra work. In some disciplines, such as physics, research groups tend to produce a substantial number of research articles. Some participants indicated that complying with such mandates would require them to spend more time and energy depositing their articles in an open access repository.

Another physicist, who also deposits in the subject-based repository arXiv, thought that mandating bodies could set up an automated harvesting protocol to import research articles into their designated repository, and thus save time and duplication of effort for researchers*: "[it] can also, I think, can be automated quite a lot. I mean once something is in the arXiv, this is exactly the problem that some technical people have to solve, once something is in arXiv can be transferred to another archive. You don't have to do [it] manually yourself."* However, articles would, of course, need to be manually deposited somewhere first.

Aside from the workload associated with mandated deposit, many participants also raised the issue of publishers' copyright policies not allowing researchers to make their research publicly accessible, and felt that mandates were not compatible with publishers' policies. A few participants, especially in the Social sciences, humanities & arts, mentioned that journal publishers allowing open access were not always the journals with the highest impact factor. Looking at open access journals more than at open access repositories, participants in the Medical sciences group said that open access journals often have a lower impact factor as it takes time to build a reputation.

A Medical sciences participant thought that funder mandates could create tensions with publishers' current practice, *"because there [are] the journals who cannot allow you to put things in a repository, and you want to publish in that journal and […] the policy of the journals should also change then."*

In the discipline of law, researchers may be paid for publishing articles in journals. Although payments made to researchers are not extremely high (ranging from €400 to €1,000), one

participant indicated that this remuneration is important - if depositing in any open access repository would hamper him/her being paid, s/he would opt for the payment rather than for the open access principle, despite being in favour of open access in principle:

> *"I just wanted to say that, as I have indicated before, I would be more sceptical because if I have such an obligation by whoever I might have problems selling my article to a prestigious journal, they may say that it is already published and sorry we don't take it, and there is a second issue connected with books, also in the law as you were just speaking about, you need to pay if you want to publish a book, so quite often you get funding from the state so that this book gets published and then they will say, 'Sorry, we won't publish a printed version, you can do it online' and so there is a shift and this would be bad for your list of publications because of course you want a prestigious publisher to publish it as a journal or as a book."* (Social sciences, humanities & arts)

A few researchers indicated their disagreement with a 'mandate' approach. They thought deposit of institutional research outputs could be dealt with without resorting to mandates, *"I would avoid mandating."* A participant suggested that institutional mandates were perhaps not the ideal solution and one way to go would maybe to have publishers intervening in the deposit of articles in open access repositories - *"the best way for you to go about would not be by a mandate from the university but by direct remit with the publishers, and then of course we would take it as an honour that our journal has even managed to get more readers."*

Certainly, participants indicated that mandates could create tensions with their scholarly communication practice. Journal impact factors were cited repeatedly across all focus groups as a major element in research evaluation, both at researcher level with peer recognition and career advancement and institutional level with funding. Participants felt there was a real tension between the impetus to publish in high ranked journals and the desire to disseminate and provide greater access to their research articles. The Social sciences, humanities & arts focus group, which discussed this topic more than any other groups, indicated that they would not compromise on the impact factor for making their research available on open access. When asked to choose between publishing in a journal with a high impact factor but not allowing researchers to make their research publicly accessible or in a journal with a lower impact factor allowing open access, a social scientist indicated *"of course if the higher impact factor journal accepts my article, I would go for the impact factor."* In the case of institutional mandates, it was indicated that publishing in high impact factor journals could conflict with mandates to deposit research outputs on open access, and this same participant indicated that it was institutions' responsibility to redefine their objectives and the ways to achieve them, *"well my institution cannot be contradictory because my institution wants me to publish in the highest impact factor possible because the institution itself is evaluated on the basis of the [impact factor], so my institution can not require me to fulfil [a mandate]"* at the expense of the impact factor.

A participant from Medical sciences also referred to the conflict arising from journals with high impact factors and restrictive publishers' policies, and commented that it was *"possibly contradictory, because it's actually the funder who usually require to have publication on high impact factors to get funded."* Another described the situation as a *"catch 22 situation",* where career advancement motivations and peer-recognition on the one hand, and advancement of knowledge and funding requirements on the other hand, can conflict, forcing researchers to make unavoidable compromises.

It was not clear whether senior researchers had different views from young researchers when funding was brought into play, although it appeared that young researchers were more inclined to compromise impact factors for funding. A young researcher in medical sciences commented *"well, if the choice is between not getting the funding and getting [it with an open access condition], of course you say yes, and then you look for […] a journal that can have this requirement, I think [it] would be more fair if you can have like both things.*

3.4.2 Embargoes

Questions about embargoes did not attract many comments from the participants and reactions were quite different from one group to another.

Participants across the four focus groups did not show a great level of awareness of embargo periods, and for those aware of such practice, they tended to accept it, but not without resentment as it hinders access to research material. This said, there was some indication that authors are aware of the commercial relationship between authors and publishers, and appreciative of the commercial needs of publishers:

> *"from the humanities side, if you publish in a professional journal the publisher will tell you that they have to earn money and you have to help them make money, so you would say that's OK and you accept the embargo period […]That's a fair deal."* (Social sciences, humanities & arts)

or

> *"you cannot tell them not to, otherwise we would not have any journals to begin with. Most of the time the journals are actually under-funded, they are always on the brink, maybe not yours, but ours are always having some trouble so actually it's just fair. It's not only this but given the fact that now journals are so important to build our own reputations, much more than books, it's a lot of work, it was always a lot of work to have a journal, not only for the publisher and now there is much more work."* (Social sciences, humanities & arts)

Participants in the Medical sciences focus group were generally not aware of publishers' embargo practice, though they could see that not all articles were immediately available in PubMed, *"I hadn't actually appreciated that, but I had kind of noticed in far left field that when I go to PubMed, it's free access, and then three months later, it's not free, and then three months later it is free […] so I was kind of vaguely aware that they were [not always accessible].."*

Participants showed different feelings about embargoes. Some commented about the fact that remembering to place one's articles on an open access repositories after six months and the expiration of the publisher's embargo may prove difficult in terms of management. A Medical sciences participant thought that *"they have to make the process easy for the author to do it, otherwise one is kind of put off from the beginning, it is something very bureaucratic.[…] I mean if I really want to put [it] public, maybe I go through the 2,000 forms, but usually the easier is the process […] more likely is the person to do it."*

Life sciences participants did not comment much on publishers' embargo periods. Nevertheless, one participant indicated that the distinction between applied research and conceptual research was central to one's perception of embargoes. Although this participant was generally not in favour of embargoes, and did try to publish in journals without embargo periods, he also

indicated that there were circumstances in which even authors could be in favour of embargo periods, for example:

> *"normally in all fields where I would say the technology improvement can result in some kind of new territory for patenting, then the embargo period should be longer. Meaning, one, one and a half year. […] But practically all the other ones which are trying to influence the community behaviour, then we never accept the embargo period. […] Practically we are trying to choose for concept papers journals which do not maintain an embargo period."*

As noted earlier, timeliness of research outcomes was noted as an important element in scholarly communication for the disciplines of Physics, Life sciences and Medical sciences. Therefore, participants in these disciplines expressed concerns over the fact that embargoes may create some tensions as results may become outdated very quickly.

3.5 Repository services

The effort involved in managing versions and depositing articles was identified by some participants as a barrier to place material in open access repositories. It is interesting to note that this comment came mainly from participants in the physics group who had experience in depositing articles in publicly accessible repositories, and therefore who were talking from their own experience, for example:

> *"arXiv is one thing that all of us think is essential to send a paper. The point is, it is hard work. Hard work in the sense that you have to remember once you have finished your manuscript, once you are working at the submission and everything, to also access arXiv and which version do you put on."*

Some participants in the physics sub-group also indicated that, in their field, the number of articles produced per year by research groups can be overwhelming in terms of management, *"and sometimes if the group is fairly large, in my case […] it is a fairly large group that seem[s] to be very productive, and we just don't get around to using arXiv".* It was also indicated that it is common practice for large productive research groups to delegate the management of their research articles to their doctoral researchers, and so mitigate the time constraints often raised as a barrier to placing articles in publicly accessible repositories - *"I know that in some places, it is a matter of course that the post-doc that sends the paper to the journal, the final version, also takes care of putting it on arXiv."*. One physicist was of the opinion that the deposit process could be automated, so that articles deposited in arXiv could be automatically transferred to another repository.

The Life sciences group was the only group thinking that the workload involved in placing material on open access was not an issue at all, and probably minimal compared to the benefits yielded by making their research openly accessible. It was felt that, from the four focus groups, that the Hungarian focus group bringing together researchers in the Life Sciences discipline was the group experiencing the most difficulties in terms of accessing journal content. Their openness towards the concept of open access repositories might partly be explained by their own experience in regard to access to research content.

Some participants indicated that if repositories could offer more features delivering tangible additional value, then they could become a very useful tool to researchers in their day-to-day research workflow. Greater readership, greater visibility or speed of dissemination were seen as

extremely important in principle, but since they are not readily quantifiable, they were less likely to be of immediate interest. It was noted that researchers want tangible, immediate benefits such as download statistics, email alerts for new content of interest or an integrated author's profile within the repository platform where the author's research and interests are displayed along with the publications. A participant commented *"I think there needs to be some added value as an author to upload your document there",* such as *"the possibility that interested people can subscribe to a certain type of journal or a certain kind of research area so that they get an e-mail saying that there is a new working paper."*

A few participants in the Life sciences group mentioned 'interactivity' as an important feature that could help open access repositories to attract more interest from researchers, and enhance communication between researchers. Comments included: *"also if it would have a discussion system [..] that would be useful", "more interactive is always useful"* and *"maybe if you see an article, you don't write to the author, getting in contact is not that easy but if you could comment on a web page, maybe that would facilitate communication."*

One Medical sciences participant observed:

> *"obviously the nature of the way we're funded and the nature of research is actually changing, and if repositories are going to become important, it probably would be quite important for them to have some sort of interlinking, so that one repository might have some type of data, for example, you were saying about, when you're between two disciplines, if you could actually find a way of actually linking them in some way, that could be actually quite important, but I would imagine that would be quite difficult to do".*

Some participants in the Social sciences, humanities & arts group also commented on the fact that, although they appreciated why English was the main language of articles placed in open access repositories, they also thought that having articles, or even only abstracts, in different languages would maybe help the diffusion of these articles in non-English speaking countries. They wanted also to be relevant to their national audience and in some disciplines, where translation may change the meaning of a concept, English may not be the language of choice for a non-native English speaker, for example:

> *"I think that these repositories would be much more attractive and probably useful if they were not only in English, so if they did not collect only English publications written in English, because particularly as we who belong to large countries and want to be relevant to our own country and not only for […] readers at [the] international level [and] publish in our own language."*

3.6 Information behaviours

3.6.1 Locating resources and sources

In order to explore readers' information seeking behaviour, participants were presented with a scenario where they had only three days to put together a research proposal and were asked to describe how they would proceed to seek and locate relevant information sources needed for their research proposal.

Key findings are:

- Researchers show similar information-seeking patterns across disciplines

- Physicists and medical scientists place a lot of emphasis on the notion of trust with regard to information sources.
- When searching for information to confirm prior knowledge, researchers tend to conduct their search first by authors they know in the field or by keywords if they do not know any specific authors. They also browse specific journals in which they know they will find relevant information.
- When searching for information in a field in which they are not familiar, researchers tend to do a search by keywords first in order to identify relevant authors in the field and then they use the 'pearl growing' method to continue their search.
- Search engines such as Google and Google Scholar are sometimes used, not as an information retrieval or discovery tool, but rather as a tool to access research content usually hidden behind a toll barrier.

The following sources were identified by participants:

Mathematics, Physics & Engineering:	
Authors' WebPages	Scholarpedia
Cross Ref	Scifinder
Google	Web of Science
Google Scholar	Wikipedia
Repositories	

Medical Sciences:	
Authors' WebPages	ScienceDirect
Google Scholar	Social Science database for interdisciplinary fields
Mimas Zetoc alerts for current awareness	Web of Science
PubMed	

Life Sciences:	
Cordis (Community Research and Development Information Service)	ScienceDirect
EISZ (Electronic Information Service available in Hungary giving access to Web Of Science, Science direct and most of Hungarian journals founded by the Hungarian Academy of Sciences and other journals)	Selected open access and high impact journals across medical sciences and life sciences, such as *Environmental Health Perspectives*
Elsevier journals	Web 2 for non-mainstream science or emerging fields
Google scholar	Web of Science

Social Sciences:	
Google	SSRN
ScienceDirect	Westlaw or similar databases

A majority of participants across disciplines confidently indicated using Google or Google Scholar when looking for research material. A few researchers were, however, less confident in mentioning their use of Google. This was well illustrated by a researcher in engineering who

admitted to using Google after his peers had confidently and openly acknowledged using Google and Google Scholar when looking for information sources. He said *"they confirm my idea. I was using sometime[s] Google in doing research, I didn't think it was a good way, but if they say it's a good way, I can trust on Google"* or *"[…] Google, and also Wikipedia. I didn't think it was really good, so I just use it behind closed doors"*.

In addition, it appeared that search engines were used as much to identify journal articles as to retrieve journal literature. For a few participants concerned with difficulties in accessing journal articles, especially amongst Hungarian participants, Google and Google Scholar were mentioned as a way to bypass the subscription barrier when their institution did not subscribe to a particular journal. A Life sciences participant indicated that he was *"using Google Scholar because it has lots and lots of full-text articles"* that are available on open access. In this specific case, Google Scholar was used more as an access tool than an information retrieval tool. Although difficulties in accessing information content were mentioned in several instances across the four focus groups, this was not however seen by most researchers as a major obstacle in their information search, with maybe a slight difference for the Hungarian focus group where lack of funding for research (and hence for journal subscriptions) was raised a number of times. Whereas literature on access to research content often stresses the problem of access by researchers to subscription journal content, participants did not show strong feelings or opinions about this during the focus groups. They acknowledged the problem but they did not elaborate on it. Instead, they put forward ways to circumvent the access problem. Alternative ways to access journal articles were through libraries' interlibrary loans services or by requesting a copy directly from the author of the article, both ways working well according to all participants, as illustrated by the comments below:

> *"I have two affiliations that I could use, but if I get stuck, then I'll email the author, the senior author: 'please, send me a copy'. […] sometimes it's been a little while coming, but I don't think anybody's actually completely ignored me, and usually they turn it round quite quickly actually. I think most of us I think are egocentric enough to be pleased that somebody wants to actually read our paper".* (Medical sciences)

A Medical sciences participant also mentioned the common practice operating in research circles whereby PDF files are passed between colleagues/friends from different institutions. However, this is generally kept quiet owing to potential copyright infringement implications. When asked to comment about how often s/he comes across a situation where s/he cannot access articles, this participant replied *"quite often, and I have two affiliations, so sometimes I have to try with one, sometimes with another one, and there are still papers that I can not use […] That depends from how much you want that paper, if I really want it, I figure out what to ask to a friend to have access".*

The heterogeneity of the disciplines represented in each focus group made it difficult to detect any distinctive disciplinary information-seeking pattern. However, the following observations attempt to draw an overview of our participants' information-seeking behaviours:

- PubMed emerged as the primary information source in the Medical sciences. Although other sources were considered by participants, PubMed was clearly identified by and large as the primary tool used by participants in seeking and locating journal articles. When asked whether PubMed could be qualified as an authoritative source in the Medical sciences, one participant answered *"yes, certainly for my field it would have access to all the journals that I would expect. I might use Google Scholar, but then most of that actually comes via PubMed anyway".* It is interesting to note the difference

in behaviour that medical scientists adopt as readers and as authors: open access repositories currently fit particularly well with medical scientists' information search behaviour through the extensive use of PubMed (and PubMed Central). However, although medical scientists tend to agree with the generic principle of open access and publicly accessible repositories, open access repositories have not yet found a place in authors' research workflows. Such difference in behaviour may be explained by the fact that deposit onto PubMed Central is done at the publisher level, not at the researcher level. It is publishers' responsibility to deposit research articles into PubMed, and researchers are not involved, at any time, in the management process of making their articles available on open access via the PubMed platform. A participant expressed satisfaction that this procedure lies in the hands of publishers and said *"I tend to rely on the fact that the publisher, or PubMed, will upload contents and therefore the PDF and so on".* (Medical sciences)

- The notion of trust, in regard to open access material and websites, was debated in all focus groups but much emphasis was put on this topic by the sub-group of physicists and the group of medical scientists. Interestingly, it appeared that there are different levels of trust when it comes to using open access material, be it an open access article held in a repository or available on a website. Whereas most participants were happy just knowing that the open access material they were about to access has been, at least, through the first round of the peer review process, medical scientists only trusted open access material published by renowned journals, i.e. reputed for the quality and high standards of their peer review, or articles written by authors they already knew and recognised for the quality of their research. Alternatively, physicists tended to place more trust on a cluster of research groups who, they knew, were active researchers in the field rather than on the reputation of a journal or on the peer review itself. It is worth noting that it is common practice in physics to make pre-prints available on open access, and thus whether articles have been through peer-review, or not, did not matter that much to them. Physicists' information seeking behaviour was very much based on the idea of following up researches undertaken by a selection of trusted research groups by regularly checking their publications made available on open access on their personal or institutional web pages, or through a link to the open access arXiv file. The notion of trust was also very much linked to citation behaviour *(see section 3.6.2)*.

A few individuals also indicated that specialist databases would deliver only mainstream information and so, they would happily search the Web for non-mainstream information if they had time to do so. For instance, a Life sciences participant referred to the Web 2.0 as an example of an information source where articles on the side of mainstream 'published' science are made available. Because mainstream science often does not recognise emerging research fields, this participant valued such non-conventional information sources, and commented *"especially when you want to invent something different from the mainstream science, you get more ideas and more information from that".* However, in the scenario presented to researchers, only three days were allocated to information search and therefore those individuals indicated that they would then go for the usual and mainstream information sources.

To confirm prior knowledge, researchers across disciplines tended to rely heavily on key authors they knew in the field and would search for articles published by these key people and cross-reference them. This pattern was well acknowledged across disciplines. To keep up-to-date with developments in their field, researchers indicated that they used various

strategies, from journal email alerts to systematic and regular keyword searches to regularly checking websites. This was particularly true for medical scientists and physicists for whom current awareness could be argued to be more important owing to the constantly changing nature of their discipline, as illustrated here:

> *"There are the three tools: the Web of Science, the cross references to papers you know and Google Scholar, and the fourth element we use are repositories [which] are really a repeat of the web pages of people I know of" (Physical sciences & mathematics)*

or

> *"None of us have got time to go through [the journals' table of contents]. I have Mimas Zetoc alerts for journals I am particularly interested in, so I go through those to try and keep up to date […] Just to make sure I am not missing anything. But also I have regular search terms that I'll do through PubMed, just to make sure that, particularly as research is going on, that somebody is not…, I'm not moving like this in front of another group, and make sure you're not missing out".* (Medical sciences)

or

> *"Personally the way I work, I mean in the old days we used to go to the library, and people used to go Saturday morning or whatever to the libraries and go through the journals […], now this doesn't happen any more, we have the electronic information, but searching for things is becoming more complicated, because you don't spend the time accessing the important journals. One practice that I know several people do is, if you make your own selection of 20 or 30 groups in the world that you trust, and you know they're very active and they're doing things in the field that you work in, and all you do with regularity, go to their web pages, which are usually well kept, and see what they have done, download the papers, or they get their arXiv papers from that, so it's a filter that you do by selecting about, in my case I have about 40 groups around the world that I try to follow, and this is of course, it's not complete, then you miss the young unknown newcomers, but that's life, you cannot really follow everything" (*Physical sciences & mathematics)

The use of websites as a source of information, along with traditional scholarly journal literature, is well acknowledged in the information seeking literature and participants' information behaviour confirmed this: many of them mentioned at some point checking regularly personal or institutional websites of specific individuals or research groups to keep up to date with developments in their field. This was particularly common practice for the group of physical scientists and mathematicians and the group of life scientists. For the two other groups, the Medical sciences group and the Social sciences, humanities & arts group, attitudes towards the use of websites as information source were quite different to those of life scientists and physicists. Medical Scientists indicated that collaboration was the prime motive for them to visit websites of specific individuals or research groups. When asked whether they were visiting websites of some people they knew in their field, a Medical sciences participant said *"only if I'm looking for some collaboration. [..] OK, he's from the same field – let's see what he has in his lab, what facility he has in his lab for a particular grant"*, whereas another Medical sciences participant added that, although collaboration was the prime motivation, they also wanted to keep abreast of what people do, especially people they know in the field or people with whom they had done some work:

> *"similarly, but also I would search for individuals whose work I know and I want to keep track of what they're up to, as it were – former collaborators to see what direction they've*

gone into, and maybe perhaps with the aim of actually re-collaborating with them, if it's appropriate. And certainly for looking for people, that maybe you want to set up a collaboration within the EU, because there's that desperation actually, if you want to get the EU grant, you've got some collaborators in place."

Although the objective of the focus groups was not to make any generalisation about information search strategies, it is interesting, however, to note that researchers within what would appear as a fairly homogenous community such as Medical sciences may have a different approach or practice regarding the use of information sources such as personal or institutional websites. The Social sciences, humanities & arts focus group did not explicitly refer to this practice. One participant did, however, refer to checking authors' personal web pages on SSRN in order to keep up to date with people's research profile, but not to seek and locate information. Commenting on using SSRN as an information search tool, this participant said:

"Because these things were already published so they are reliable, whereas with the SSRN it might be that the working papers are not finalised, so I wouldn't trust them that much. There's also a problem with researching in the SSRN, you can't really search within the papers, which is really bad, at least I cannot. I find it really difficult to search within the research papers uploaded on the Social Science Research Network, this is why I would search within Westlaw where I would have far better hits."

Alternatively, to acquire new knowledge in an area of their discipline they were not familiar with, researchers tended to use the keyword search options available in search engines and discipline-based databases in order to identify the key people in the field and then look at their publications. From there, they would use the pearl growing method to get to other authors. No clear preferences between discipline-based databases and search engines were observed. For some participants, Google and Google Scholar appeared to be the first place to conduct keyword searches for unfamiliar research areas or new knowledge. A medical scientist said *"for new knowledge, I often go to Google, and then it gives you a lot of options, and then it can redirect you to the right part as well".* But for most participants, it was unclear whether they were using search engines more than discipline-based databases as both methods seemed to be fully part of their general information-seeking pattern.

3.6.2 Citation practice

Participants were asked to describe and explain their current citation practice when they choose to use a resource that is held in an open access repository. Participants held very different views about whether or not it is acceptable to cite articles in repositories.

The sub-group of physicists indicated that it was common practice within the physics research community to cite articles held in arXiv, including pre-prints. A physics participant said *"I think […] in some sense the archive is so established that nobody really has a problem in citing an article from the arXiv, because there is a code, so it's well identified, and so sometimes you may cite an article in the arXiv with the arXiv number of that article."* Another participant, building upon previous discussion on researchers' information seeking behaviour and linking this to the notion of trust, said *"certainly most people will [be] looking at things that are often coming out of groups that they know"*, and so they would trust the source and therefore be comfortable citing it.

A more senior physics participant also indicated that consideration of the target audience could also influence decisions to cite an open access material. S/he explained:

> *"I have no hesitation using something that is in open access, I think that anything that you look at, you use your judgement to evaluate whether it's good or not. If I have to quote, I would prefer to quote the published version, but in some cases I also prefer to put the arXiv [version], thinking of people who may have a difficulty to reach the published version, if it's in a journal they have no access to. But still, I would emphasise that the existence of peer review with all its defects you know, nothing is perfect, but is really fundamental, and it makes the difference between science and not science."*

Such consideration about who will be the audience for their article also prompted a social scientist to amend his current citation practice on a specific occasion. S/he said *"I've done it a little differently when there was an article published in Macedonia because I knew that it is a country that doesn't have a lot of money, so I was trying to cite both versions, but normally I always go for the printed one."*

The discussion about citing an article made available on open access in a repository indicated that most researchers across disciplines were rather conservative in terms of citation practice, and would therefore try to track down the published version of the article they wish to reference. A participant thought that citing a repository would *"look[..] as though you may not have consulted a proper library which is sort of academic standard."* (Social sciences, humanities & arts)

Participants across the four focus groups often mentioned that decisions about citing an open access repository resource would depend very much on the journal they wish to publish in. Participants tended to hold strong beliefs that their papers could be rejected if they do not follow publishers' citation guidelines. It was also said that some journal publishers may sometimes preclude citation of unpublished material.

> *"I think that would depend on where I was trying to publish the piece of work that I was putting together, because there's a lot of journals who won't actually accept things in press…"* (Medical sciences)

> *"It depends on the journal where you want it published because there are several journals which do not accept [this] in citations"* (Life sciences)

> *"I have even heard of a US journal in social sciences which can turn down articles if these articles used references which are mostly books, or chapters in books and not actual published articles, so it would appear that if you do not correctly cite the journal and the pages, it would like that you are using some kind of online paper which doesn't have a clear status."* (Social sciences, humanities & arts)

A few participants mentioned the fact that authors' final versions not having proper pagination made it difficult to reference such articles made available on open access:

> *"it has been uploaded as a working paper and you don't find any pagination, so for this you would still need to go to another database and figure out where precisely it is, if you want to cite it you will still have to refer to the printed version and I think that is a clever technique"* (Social sciences, humanities & arts)

or

> *"there is a problem […] that without the page numbering you cannot cite."* (Social sciences, humanities & arts)

The issue of permanency was raised in the Life sciences focus group, which touched upon citing the URL of a resource made available in an open access repository. Drawing a comparison between repositories and WebPages, a participant said *"this can cause problems because you are citing a web page, and then the web page disappears, and the article is no longer available. It happens many times."* The Physical sciences & mathematics group also touched upon this issue and the sub-group of physicists led the discussion by taking the example of the repository arXiv which is so well established within the physics community that nobody questions its long-term future. They also explained that a unique identification number is attributed to each research article uploaded onto arXiv. Different versions of the same article are thus grouped together under this single identification number and arXiv allows readers to see the history of the article, i.e. any previous versions of the article which have been uploaded onto arXiv are made available, including pre-prints. The group concluded that with a version management like that of arXiv, permanency may be less of an issue.

4 Summary

Sections 2 and 3 of this report provide an initial analysis of the quantitative and qualitative data gathered in the first phase of the research. An in-depth analysis and synthesis of these data will be presented at the PEER joint workshop (March 2010) and, on completion of the second phase of the research (September 2010 – June 2011), in the project final report, which will fully address the project's aim to contribute to understanding the extent to which authors and users are aware of open access, the different ways of achieving it, and the (de) motivating factors that influence its uptake.

4.1 Discussion

Where appropriate we outline below ways in which the initial data gathering and analysis address the project's research questions (abridged below and given in full in the Introduction).

4.1.1 Common perceptions held by authors and readers in relation to repositories and ways in which those perceptions influence behaviours.

The survey data indicate that there is a general awareness of open access, with more than two-thirds of respondents understanding open access to mean free electronic access to the full-text of journal articles, although focus group participants expressed uncertainty about what open access entails. Our findings show that general awareness of open access is growing compared to results from earlier seminal studies conducted by Rowlands et al. (2004) and Swan and Brown (2004, 2005). A key feature of focus group participants' uncertainty was what to expect in terms of quality, be it the reputation of a repository itself in the case of authors, or the quality of the articles deposited in the case of readers. This was in contrast to the survey where respondents indicated that open access does not equate to diminished quality. There was a feeling amongst some focus group participants that, from an author perspective, unrefereed preprints of variable quality may not be appropriate alongside peer reviewed stage-two manuscripts. A reluctance to deposit articles where other material had not been peer-reviewed was a significant concern for survey respondents in all subjects. This was not a concern raised within the Physical sciences & mathematics focus group, whereas amongst some Life sciences participants concern was expressed over unrefereed papers appearing in the same repository as peer reviewed articles. Visibility of an author's work was raised in the focus group discussions and there was a perception that unless a repository itself had visibility and kudos, then the motivation to deposit papers or articles in it would be low. The main driver for this concern seemed to be that if repository content was fragmented in terms of quality this will have a negative influence on the overall prestige of a repository and thus reduce the visibility of authors contributing to that content. This resonates with Borgman's (2007) argument that the 'legitimatization' of sources and resources becomes increasingly important in the context of emergent publishing paradigms, such as self-archiving, and the fuzzy boundary between traditional and emergent publishing paradigms. Furthermore, the negative correlation between perceived low prestige of a particular repository and willingness to deposit to it signifies that studies (Swan and Brown, 2004; Watson, 2007) linking research evaluation mechanisms, such as the Research Assessment Exercise in the UK, with disincentives for academics to disseminate in OA journals with a low impact factor are also relevant for other types of information resource.

Amongst survey respondents the level of awareness of open access did not differ greatly across disciplines, although researchers in the Life sciences and Medical sciences had slightly differing views on what open access entails to researchers from other disciplines. Outside of the Life sciences and Medical sciences researchers expressed concern that open access equated to 'not peer reviewed', although the percentage of respondents who held this opinion was low across all disciplines. Within the Life sciences and Medical sciences, open access was more closely associated with the 'author pays' model of open access journals, rather than self-archiving.

The term 'repository', however, was not universally understood to mean the same thing amongst focus group participants and some did not have a clear idea of what a publicly available open access repository was. A lack of awareness of the existence of repositories was also indicated by survey respondents, with over 26% of respondents being unsure as to whether or not their institution had a repository and 47% being unsure as to whether a subject-based repository was available to them. Physical science & mathematics survey respondents appeared slightly more aware of the subject-based repository options available to them. This increased awareness was also reflected amongst participants in the Physical sciences & mathematics focus group where for a number of participants the use of repositories, such as arXiv, was described as part of the daily or weekly workflow, both in terms of depositing and locating papers. Community-based information services, such as SPIRES[11] or arXiv, play a major role in the information landscape of the High Energy Physics (HEP) community, with 87.9% of HEP scientists reporting using these two information sources in most of their information searches (Gentil-Beccot, 2009). It has been recognised that the uptake and use of non-market based information resources, such as institutional and subject-based repositories, has been problematic - with long lead times in terms of content population (Duranceau, 2008). Yet, amongst the High-Energy Physics community almost the opposite is true. Our findings do not particularly extend current understanding with regard to this phenomenon and it is likely to be explained by the fundamental cultural characteristics of the High Energy Physics community. Our findings do, however, emphasise the importance of connective structures in digital information environments (Palmer, 2005) and the 'location' of repositories in the broader context of competing resources, with search preferences likely to play an influential role in resource discovery and use.

Focus group participants appeared to show different perceptions and behaviours with regard to open access materials when considering their position as readers and users of research materials or as authors of journal articles. This will be investigated further in the secondary analysis of the survey data.

Although open access was not an unfamiliar concept to most focus group participants and survey respondents, influences and perceptions of the use of publicly available open access repositories varied greatly. In some cases these differences can be attributed to disciplinary practices and in others they are seemingly due to individual idiosyncrasies. Important influences appear to be perceptions of quality, peer review, confidence, trust and visibility, and in-depth analysis is required to untangle the relationship between influences, perceptions and use. What

[11] SPIRES is a metadata-only search database for the field of high energy physics. It collects metadata from repositories and journal literature and proposes additional services such as citation analysis, keywords, authors' affiliation, matching pre-prints with publications, etc. Search facilities in arXiv are extremely limited and, as pointed out by Gentil-Beccot (2009), SPIRES and arXiv may be seen as the two ends of a single information system, with arXiv focussing on data storage and SPIRES offering users search facilities and other services.

is clear is that there is still a need for the promotion and advocacy of repository-based deposit of research articles across disciplines.

4.1.2 The choices authors make in locating and selecting appropriate outlets for publication and the underlying influences.

Peer reviewed journals were the preferred mode of disseminating research, with over 99% of survey respondents having published at least one journal article in the last five years. Social science, humanities & arts researchers are arguably more likely to publish monographs and book chapters than researchers in other disciplines, but none-the-less over 98% of respondents in this category had published a peer reviewed article (this high percentage is also a direct consequence of the sampling method used for the survey, and should therefore be interpreted with caution). Focus group participants whose research areas demonstrated strong interdisciplinary characteristics highlighted that the limited availability of peer-reviewed journals in their area was problematic, particularly in terms of submitting papers for publication. Institutional or subject based repositories were seen as a potential solution to this problem; however, participants emphasized that, despite the limited opportunities, publishing papers in reputable peer-reviewed journals was a priority when considering how and where to publish their research.

In terms of dissemination, defined more broadly than publication alone, focus group participants and survey respondents indicated that the most important influence in their choice of journal was widespread dissemination of research, e.g. reaching their target audience(s). Having research evaluated by peers was also an important factor and this was raised by focus group participants in terms of having confidence in one's own research – an influence that seemed particularly pertinent in terms of the use of subject repositories such as arXiv which are not peer reviewed. Participants in the Physical sciences & mathematics focus group, in particular, felt strongly that there was an interrelationship between feeling confident in the quality of one's own research or reputation and willingness to deposit unrefereed papers in a repository. Journal reputation was also given as an important influence in choice of publication outlet amongst survey respondents. This may imply that if stage two deposit is to become the norm, it is important for publishers of high quality peer reviewed journals to support it. Focus group participants who were from a physics background, a discipline with an established record of deposit into the arXiv repository, seemed more aware of the benefits of depositing all versions of research papers – for instance it was mentioned that depositing pre-print material in a repository was a way to 'date stamp' research and that, in a competitive research environment, this may be an important by-product of making research available in a repository. Other focus group participants expressed this explicitly as an issue of 'trust' and there was a perception that one's research was less likely to be plagiarised if papers had gone through the peer review process and there was a clear and permanent way of citing those papers.

Despite the fact that widespread dissemination of research was identified by survey respondents and focus group participants as a factor influencing dissemination and publication behaviours, nearly half of survey respondents (48%) had not placed an article in a publicly accessible repository. This figure was lower in the Physical sciences & mathematics (41%) and higher in the Medical sciences (67%). Both the physical and medical sciences have a number of esteemed subject repositories that are arguably central to each respective information landscape. For those Physical sciences & mathematics focus group participants who routinely used repositories, both as authors and readers, such a behaviour was expressed as a norm within their research community, a smaller number of the Life sciences focus group participants

also routinely used repositories and again it was expressed as norm in their disciplinary community, partly driven by specific journal publishers.

An explanation for differences in the self-reported deposit rates across these two broad disciplinary groupings could be the way in which the deposit process is managed by the most popular repositories in these disciplines. For example, in the Medical sciences, journal publishers often deposit articles into PubMed Central on behalf of authors, this contrasts with arXiv where authors often deposit articles themselves. Although many medical scientists tend not to deposit their papers or articles themselves in publicly accessible repositories, many do use PubMed Central for locating sources and the focus group and survey findings indicate that researchers appreciate its role as an important repository of material. Whilst the survey findings did not indicate that the process of depositing was a major barrier to deposit, some focus group respondents did outline that deposit is an extra task that takes time and has to be managed.

Mandatory deposit might also be an explanatory factor, since in recent years funding agencies in most disciplines have developed open access policies or at least position statements on open access. Furthermore, where funding agencies do stipulate mandatory deposit or have position statements they also encourage grant applicants to apply for the necessary funds to make articles publicly available via open access e.g. financial support for author pays mechanisms. It is therefore of interest that funder, or indeed, institutional, mandates were considered relatively unimportant as drivers for repository deposit by survey respondents.

The physicists who participated in the Physical sciences & mathematics focus group had particularly strong feelings with regard to deposit workflows. In their experience, the phase of the workflow in which papers were deposited to a repository varied, with participants reporting colleagues who deposit papers at the same time that they submit a manuscript to a journal, compared with their own practice, which tended to be to deposit papers once the manuscript has been peer-reviewed and accepted by a journal. Experimental physics typically involves working in large collaborative teams where a high volume of papers are produced. A senior physicist explained that keeping track of different versions of papers, e.g. pre and post peer-review, and whether or not they had been deposited to arXiv, for example, was time consuming and not fool-proof. In short, self-deposit is perhaps feasible for individuals or small groups of authors producing a modest number of papers per year, but automatic deposit by publishers becomes more desirable for large-scale distributed groups of authors such as those exemplified by international particle physics experiments.

The ethos underlying open access repositories was appealing to most focus group participants, but lack of knowledge (both of appropriate repositories and of the deposit process) was articulated as a barrier for those participants who would consider depositing articles in a publicly accessible repository, but had not yet done so. For example, there was a general perception amongst those focus group participants who had not deposited in a repository that to do so would infringe any copyright agreement that they had signed with a publisher. The role and profile of services such as SHERPA/RoMEO[12] need to be considered here (Jenkins et al, 2007), particularly as a number of focus group participants were unaware of the policy of many publishers towards open access – most were of the opinion that publishers were much less open-access friendly than their stated policies would suggest.

[12] SHERPA/RoMEO - Publisher copyright policies & self archiving details available at: http://www.sherpa.ac.uk/romeo/

Based on some of the perceptions illuminated by the focus group data, it would appear that publisher-mediated deposit, or clear guidance from publishers as to their attitude towards deposit of published articles, would be of considerable help in overcoming author concerns. This certainly supports the need for further investigation into the deposit process and justifies one of the aims of PEER - to gauge the impact of publisher involvement in the deposit process.

There was a concern, however, amongst focus group participants that if there was a growth in mandatory deposit policies or that if existing policies were more strictly implemented institutions would have a tendency to use repositories as part of evaluation mechanisms e.g. for tenure and promotion – in much the same way that the *ISI Citation Indexes* have become a dataset for evaluating journal impact and a means to measure impact of research more broadly. With an increasing emphasis on measuring research impact, as opposed to inputs and outputs, amongst funding agencies and institutions this is perhaps a reasonable assumption for researchers to make.

Survey respondents expressed a preference for subject-based, rather than institutional repositories, mainly indicating that these would be more relevant to the discipline, whereas focus group participants had a mixed preference, with some strongly in favour of institutional repositories as they were thought to support the institutional research support strategy in the same way as staff profiles listing research interests and publications. The concept of service providers being able to harvest metadata from institutional repositories and provide subject-based services was not directly addressed in the survey or focus groups, and general levels of awareness implied from the focus groups indicated that this was not an area that had been considered by many participants. Focus group participants did, however, feel strongly that repositories should provide added-value in some way, such as download statistics. It should be noted that the preference for subject-based repositories over institutional ones does not necessarily mean that subject-based repositories are the only way forward; in fact focus group participants did appreciate the role of institutional repositories. None-the-less, it would seem that subject-based services are important and should be developed if institutional, rather than subject-based, repositories become the most common type of repository.

4.1.3 The role played by repositories in information seeking behaviours

In general scholars rely mainly on scholarly articles to inform their research work, with 90% of survey respondents rating scholarly journals as ‘very important’ for their research, although this result may be biased owing to the method used to contact authors. The importance of journal articles highlighted in the survey was well supported by the amount items read by scholars. These findings regarding the importance of scholarly articles in scholars’ information landscapes corroborate results recently published by Tenopir *et al.* (2009) and King *et al.* (2009) who found that, in both studies, scholarly articles counted for over 90% of scholars’ information sources.

From the perspective of readers, it was clear that an inability to access journal articles had been a problem for many researchers with only 22% of survey respondents stating that access to research was ‘rarely’ a problem. Researchers from all disciplines were likely to seek an open access source if needed, though this behaviour was more noticeable amongst researchers in physical sciences. Across all disciplines, a range of other approaches (from inter-library loan, to asking the author for a copy) were adopted by researchers seeking articles. Interestingly physical scientists seemed slightly more prepared to forego using a resource if it were not easily

accessible to them – this may be a consequence of having a well developed culture of open access repository deposit.

Reputable repositories were recognized as important sources of material (for instance PubMed Central was clearly of importance for medical scientists), however it was clear that when citing articles, it was preferable in all disciplines to cite the final published version rather than any other version (although over 60% of survey respondents indicated that they would cite an author's version as long as it had been refereed). The focus groups indicated that the citation policy of the journal publishing the article may also have an influence on authors' citation behaviour and choice of version. It is clear that users wish to be fully aware of the version and status of any papers that they use and cite, but that currently this is not always obvious from the metadata provided by repositories. It would seem that appropriate metadata describing the version of an article is important and that any links between related versions of articles would be beneficial to users.

4.1.4 Coarse grained characteristics of authors and readers that influence their behaviour

A number of factors were investigated as potential contributors to researcher behaviour when using repositories. Further analysis remains to be done on many of these factors, but some initial points have emerged.

It is clear that subject norms influence the behaviours of authors in particular, not least as some disciplines have a long history of using subject repositories to disseminate their research, both as pre-prints and as stage–two manuscripts. The use of arXiv in physics and RePec in economics are perhaps the best known of these. The focus groups found that only physicists were in the habit of depositing articles in repositories and the survey confirmed this, with more than half of respondents in Physical sciences & mathematics having deposited material in a repository in the last five years. Medical scientists were least likely to have deposited material.

As readers, disciplinary differences were less apparent, although there was a greater acceptance of pre-prints in Physical sciences & mathematics than in the other disciplines. In Medical sciences, knowledge of the author was key to acceptance of and trust in pre-prints, although the reputation of the journal was equally important when considering peer-reviewed articles in repositories. Social sciences, humanities & arts and Life sciences focus group participants placed great emphasis on peer review and journal reputation to assess the quality of research articles. In the Physical sciences & mathematics focus group, there was a dichotomy between those who used repositories and those who did not, and emphasis was placed on the users' own critical judgement rather than a peer-review 'stamp'. The survey data showed that researchers from the Medical sciences and Social sciences, humanities & arts were most likely to consider the article version important, and least likely to trust documents they found in repositories.

The type of institution in which researchers were based was found to influence their awareness of repositories. Survey respondents from the industrial and commercial sectors were most likely to know whether their institution had its own publicly available repository, and least likely to report that their institution did have one.

Career stage is an important influence for authors in choosing how and where they disseminate their research. This is evidenced in the prior literature. Early career researchers in the focus

groups reported being less likely to try to amend publishing contracts to retain their rights to self-archive. Differences were also found in information seeking behaviours, with early career researchers more likely to use Google Scholar and the most senior researchers more likely to browse print journals when seeking to identify relevant journal literature for their research.

Access to research information for readers in developing countries is often cited as one advantage of open access, whether through open access journals or self-archiving in repositories or elsewhere. Participants in the Budapest focus group commented that a lack of funding for research in their institutions also restricted journal subscriptions, so that open access sources were particularly helpful, although other methods of sourcing material were also used. Focus group participants also commented that more repository content in non-English languages would address a gap in dissemination technologies.

4.2 Elements for further research

Below is a selection of some interesting points that have arisen from the analysis so far and which should be considered for further research:

- An individual's attitude towards open access may change dependant on whether they are an author or a reader
- Researchers in certain disciplines may lack confidence in making preprints available and to some extent this is due to differences in work organisation across research cultures e.g. researchers who work in close-knit groups, such as high-energy physics, tend to have systems in place for internal peer review. Other factors are likely to include career stage and centrality of research to the parent discipline
- Value-added services such as download statistics, email alerts, etc would contribute to the perceived usefulness of repositories and would help them gain popularity in what is an increasingly competitive information landscape
- There is a perception amongst authors that if institutions mandate deposit in an institutional repository the dissemination function will be overshadowed by research assessment. This is perhaps reflected in the perception amongst authors that subject-based repositories are more appropriate for dissemination purposes
- Readers across many disciplines often need to go through a variety of processes to access all the articles that they require and that widespread open access may reduce the need for this time consuming practice.

One final point of note is that focus group participants were appreciative of the needs of publishers, as indicated in earlier studies such as Swan and Brown (2003), and a substantial number of respondents to the survey felt that open access repositories may affect subscription-based journals. Perhaps more importantly a smaller, though not insignificant, number (31%) felt that these repositories may challenge the peer review process. Given the importance placed on peer review by most participants in this research, it is important that the integrity of the peer review process is not compromised by any open access developments and that researchers are aware of this. Such findings highlighting scholars' commitment to scholarly journals and to the quality approval stamp they bring through the peer review system are not anecdotal, nor isolated. It is well in line with what is currently taking place in the High Energy Physics community where innovative ways to maintain both the various value-added services (one of which is the peer-review system) offered by the scholarly journal publishing industry and

sound business models in an open access environment are developed (Bianco *et al.*, 2007). The SCOAP[3] model[13] whereby a consortium of private and public institutions (funding agencies, libraries, research laboratories etc.) aims to provide open access to peer-reviewed literature while maintaining the integrity of journal publishers through the payment of a unique subscription is an example of scholars' willingness to see the publishing industry developing alongside repositories.

4.3 Planned future work

Looking ahead to the ongoing background research for the project and the second phase of data gathering and analysis we have identified the following specific areas where more in-depth analysis of our existing dataset and background literature are required in order to meet the project's objectives.

We anticipate that the second phase of data gathering and analysis will provide fuller answers to the questions directly addressed in the Baseline report since the longitudinal nature of the research will enable us to identify any changes in authors' and users' awareness, perceptions and behaviours relating to open access. There are two research questions that we have some data to support, but have not covered explicitly in this report. Due to the complex social issues to which they relate, additional input is needed from the ongoing background research and second phase of data gathering. These two questions are: "How do social/institutional factors influence author and reader behaviours (e.g. mandates, embargoes, research cultures)?'; and 'What tensions, if any, exist between institutional (e.g. employer/funder/publisher) policies and practice, and disciplinary norms and practices? In what ways do such tensions influence authors and readers?'

These questions will be partially addressed by analysing the free-text responses to the survey and conducting follow-up interviews. There is also evidence from the survey data which will contribute to the first question. With regard to the second, a major activity in exploring these tensions will be to obtain an up-to-date comparative overview of open access policies amongst those countries most prominently represented by the survey respondents and also the availability of institutional or subject-based repositories in those countries.

Additional analysis is also needed to provide a more robust investigation of the issues surrounding the coarse-grained characteristics of authors and readers that influence their behaviour. This will be addressed by an in-depth analysis of our existing survey dataset by exploring the relationship between institutional type (where there is sufficient data), discipline, region etc and those questions in the survey that provide a robust representation of behaviour. In respect of this analysis we seek advice from the PEER Research Oversight Group as to ways in which individual country data can be aggregated into meaningful regions.

In terms of further analysis of the survey data we propose also to analyse interrelationships between responses, for example, differences between where researchers disseminate material and where they go to locate material.

[13] SCOAP: Sponsoring Consortium for Open Access Publishing in Particle Physics

Bibliography

Anderson, R., 2007. Open access – clear benefits, hidden costs. Learned Publishing, 20(2), pp.83-84.

Antelman, K., 2004. Do open-access articles have a greater research impact? College and Research Libraries, 65(5), pp.372-382.

Antelman, K., 2006. Self-archiving practice and the influence of publishers' policies in the social sciences. Learned Publishing, 19(2), pp.85-95.

Aschenbrenner, A., Blanke, T, Flanders, D., Hedges, m. and O'Steen, Ben, 2008. The future of repositories? Patterns for (cross-) repository architectures. D-Lib Magazine, 14(11/12).

Bailey, C., 2009. Scholarly electronic publishing bibliography. Version 76. Available at <http://www.digital-scholarship.org/sepb/sepb.html>, [accessed 02.12.09].

Barjak, F., 2006. Research productivity in the Internet era. Scientometrics, 68(3), pp.343-360.

Becket, C. and Inger, S., 2007. Self-archiving and journal subscriptions: co-existence or competition? An international survey of librarian preferences. Report prepared for the Publishing Research Consortium by Scholarly Information Strategies Ltd.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003.<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>, [accessed 14.09.09].

Bethesda Statement on Open Access Publishing, 2003. <http://www.earlham.edu/~peters/fos/bethesda.htm>, [accessed 14.09.09].

Bianco, S., Ellestad, O.-H., Ferreira, P., Friend, F., Gargiulo, P., Hanania, R., Henrot-Versille, S., Holtkamp, A., Igo-Kemenes, P., Jarroux-Declais, D., Jordão, M., Kämper, B.-C., Krause, J., Lagrange, T., Le Diberder, F., Le Masurier, A., Lengenfelder, A., Lindqvist, C.M., Mele, S, Plaszczynski, S., Schimmer, R., Vigen, J., Voss, R, Wilbers, M., Yeomans, J. and Zioutas, K., 2007. Towards open access publishing in High Energy Physics. Report of the SCOAP^3 working party. Available at <http://cdsweb.cern.ch/record/1034368/files/open-2007-015.pdf?version=1>, [accessed 01.12.09].

Borgman, C. L., 2007. Scholarship in the digital age: information, infrastructure, and the Internet. Cambridge, MA: MIT Press.

Brown, S. and Swan, A., 2007. Researchers' Use of Academic Libraries and their Services. Published by RIN in association with CURL. Available at <http://www.rin.ac.uk/researchers-use-libraries>, [accessed 24.09.09].

Brown, B. Found, C. and McConnell, M., 2007.Federal science eLibrary pilot:seamless, equitable desktop access for Canadian government researchers, The electronic Library, 25(1), pp.8-17.

Budapest Open Access Initiative [BOAI], 2002. Available at <http://www.soros.org/openaccess>, [accessed 02.12.09].

Craig, I., Plume, A, McVeigh, M.E., Pringle, J. and Amin, M., 2007. Do open access articles have greater citation impact? A critical review of the literature. Journal of Infometrics, 1, pp.239-248.

Creaser, C., in press. Open access to research outputs – institutional policies and researchers' views: results from two complimentary surveys. New Review of Academic Librarianship

Davis, P.M. and Fromerth, M.J., 2007. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics, 71(2), pp.203-215.

Davis, P.M., Lewenstein, B.V, Simon, D.H., Booth, J.G. and Connolly, M.J.L., 2008. Open access publishing, article downloads, and citations: randomised controlled trial. BMJ 2008;337:a568, doi:10.1136/bmj.a568.

Duranceau, E.F., 2008. The "Wealth of Networks" and institutional repositories: MIT, DSpace, and the future of the scholarly commons. Library Trends, 57(2), pp.244-261.

Evans, J.A., 2008. Electronic publication and the narrowing of science and scholarship. Science, 321 (5887), pp.395-399.

Evans, J.A. and Reimer J., 2009. Open access and global participation in science. Science, 323 (5917), p.1025.

Foster, N.F. and Gibbons, S., 2005. Understanding faculty to improve content recruitment for institutional repositories. D-Lib Magazine, 11(1).

Fry, J. and Thelwall, M.,2008 Measuring the Impact of e-Research: How to Account for Disciplinary Differences. 4th International Conference on e-Social Science. Manchester, 18-20th June 2008. Available at: <http://www.ncess.ac.uk/events/conference/programme/fri/3djfry.pdf>, [accessed 16.09.09].

Fry, J., Virkar, S., and Schroeder, R.,2008. Search Engines and Expertise about Global Issues: Well-defined Landscape or Undomesticated Wilderness? *In:* Spink, A. and Zimmer, M. (eds) *Web Search: Interdisciplinary Perspectives*, pp. 255-276.

Gentil-Beccot, A., Mele, S, Holtkamp, A., O'Connell, H.O., and Brooks, T., 2009. Information resources in high-energy physics. Surveying the present landscape and charting the future course. Journal of the American Society for Information Science and Technology, 60(1), pp.150-160.

Hagstrom, W., 1965. The scientific community. London: Basic Books inc.

Hajjem, C, Harnad, S and Gringas, Y, 2005. Ten-year cross disciplinary comparison of the growth of open access and how it increases research citation impact. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 28(4), pp.39-46.

Harley, D., Earl-Novell, S., Krzys Acord, S.L. and King, C.J., 2008. Assessing the Future Landscape of Scholarly Communication: An In-depth Study of Faculty Needs and Ways of Meeting Them. Centre for Studies in Higher Education, UC Berkeley: Interim Report. Available at: <http://cshe.berkeley.edu/publications/publications.php?id=300>, [accessed 16.09.09].

Harnad, S., 2006. Opening access by overcoming Zeno's paralysis. *In*: Jacobs, N., ed. Open access: key strategic, technical and economic aspects, pp.73-85.

Harnad, S. and Brody, T., 2004. Comparing the impact of Open Access (OA) and non-OA articles in the same journals. D-Lib Magazine, 10(6).

Hesse, B. Sproull, L.S., Kiesler, S.B. and Walsh, J.P., 1993. Returns to science: computer networks in oceanography. Communications of the ACM, 36(8), pp.90-101.

Hicks, D., 2004. The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics, 44(2), pp.193-215.

Houghton, J. , Rasmussen, B., Sheehan, P., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, H. and Gourlay, A., 2009. Economic implications of alternative scholarly publishing models: exploring the costs and benefits. JISC EI-ASPM Project. A report to the Joint Information Systems Committee (JISC). London : JISC.

Jenkins, C., Probets, S.G., Oppenheim, C. and Hubbard, B., 2007. ROMEO Studies 8: Self-Archiving: The logic behind the colour coding used in the copyright knowledge bank, Program, 41(2), pp.124-133.

Joint, N., 2006. institutional repositories, self-archiving and the role of the library. Library Review, 55(2), pp.81-84.

Kim, J., 2007. Motivating and impeding factors affecting faculty contribution to institutional repositories. Journal of digital information, 8(2).

King, D., Tenopir, C, Choemprayong, S. and Wu, L., 2009. Scholarly journal information-seeking and reading patterns of faculty at five US universities. Learned Publishing, 22(2), pp.126-144.

Kling, R., Spector, L. and Mckim, G., 2002. Locally controlled scholarly publishing via the Internet: The Guild Model. The Journal of Electronic Publishing 8(1). Available at: http://www.press.umich.edu/jep/08-01/kling.html, [accessed 28.09.09].

Kurtz, M. J., Eichhorn, G. Accomazzi, A, Grant, C., Demleitner, M., Henneken, E. and Murray, S.S.., 2005. The effect of use and access on citations. Information Processing and Management, 41(6), 1395-1402.

Mabe, M., 2006. (Electronic) journal publishing *in:* UKSG, The E-resources Management Handbook, UKSG. Available at <http://uksg.metapress.com/media/4nc6a0bntk0unhddva6g/contributions/3/c/v/j/3cvjfqplkkemj3w0.pdf>, [accessed 14.09.09].

Mackie, M., 2004. Filling institutional repositories: practical strategies from the DAEDLUS project. Ariadne, 39.

McKay, D., 2007. Institutional repositories and their 'other' users: usability beyond authors. Ariadne, 52.

Moed, H., 2007. The effect of 'Open Access' upon citation impact: An analysis of ArXiv's Condensed Matter section. Journal of the American Society for Information Science and Technology, 58(13), pp.2047-2054.

Norris, M., Oppenheim, C. and Rowland, F., 2008. The citation advantage of open-access articles. Journal of the American Society for Information Science and Technology, 59(12), pp.1963-1972.

Oppenheim, C., 2008. Electronic scholarly publishing and open access. Journal of Information Science, 34(4), pp.577-590.

Palmer, C.L., 2005. Scholarly work and the shaping of the digital access. Journal of the American Society for Information Science and Technology, 58(11), pp.1140-1153.

Palmer, C. L. and Neumann L.J., 2002. The information work of interdisciplinary humanities scholars: Exploration and translation. Library Quarterly, 72(1), pp.85-117.

Pinfield, S., 2009. Journals and repositories: an evolving relationship? Learned Publishing, 22(3), pp.165-175.

Prosser, D., 2005. Fulfilling the promise of scholarly communication – a comparison between old and new access models. *In:* Nielsen, E.K., Saur, K.G. and Ceynowa, K. (eds), Die innovative Bibliothek: Elmar Mittler zum 65.Geburstag. Munich: K.G. Saur, pp.95-106. Available at <http://eprints.rclis.org/archive/00003918/>.

Research Information Network (RIN), 2008. Activities, costs and funding flows in the scholary communications system in the UK. Available at <http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu>, [accessed 12.11.09].

Research Information Network (RIN), 2009. Communicating knowledge: how and why UK researchers publish and disseminate their findings. Available at <http://www.jisc.ac.uk/publications/documents/communicatingknowledgereport.aspx>, [accessed 24.09.09].

ROARMAP (Registry of Open Access Repositories Material Archiving Policies), [2009.]. Available at <http://www.eprints.org/openaccess/policysignup/>, [accessed 15.09.09].

Rowland, J.F.B., Swan, A., Needham, P., Probets, S.G., Muir, A., Oppenheim, C., O'Brien, E.A. and Hardy, R.L., 2004. Delivery, Management and Access Model for E-Prints and Open Access Journals, Serials Review, 30(4), pp.298-303.

Rowlands, I., Nicholas, D. and Huntingdon, P., 2004. Scholarly communication in the digital environment: what do authors want?. London: CIBER. Available at <http://www.ucl.ac.uk/ciber/ciber-pa-report.pdf>, [accessed 29.09.09].

Rowlands, I., Nicholas, D. and Jubb, M., 2009. Does e-journal investment lead to greater academic productivity? Library & Information Update, issue july/August, pp.45-47.

Sherpa-Juliet, 2009. Available at <http://www.sherpa.ac.uk/juliet/financialstats.html>, [accessed 28.09.09].

Simpson, P and Hey, J., 2006. Repositories for research: Southampton's evolving role in the knowledge cycle. Program: electronic library and information systems, 40(3), pp.224-231.

SQW Consulting and LISU, 2009. Open access to Research Outputs, Report to RCUK, Swindon: RCUK. Available at <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/news/oareport.pdf>, [accessed 16.09.09].

Swan, A., 2007. Open access and the progress of science. American Scientist, 95, pp.198-200.

Swan, A., 2008. Key concerns within the scholarly communication process. Report to the JISC Scholarly Communications Group. Truro, Key Perspectives. Available at <http://www.jisc.ac.uk/media/documents/aboutus/workinggroups/topconcernsreport.doc>, [accessed 15.09.09].

Swan, A and Brown, S., 2003. Authors and electronic publishing: what authors want from the new technology. Learned Publishing, 16(1), pp.28-33.

Swan, A and Brown, S., 2004. Authors and open access publishing. Learned Publishing, 17(3), pp.219-224.

Swan, A. and Brown, S.,2005. Open access self-archiving: an author study. Truro, UK: Key Perspectives Limited. Available at

<http://www.jisc.ac.uk/media/documents/themes/infoenvironment/oaselfarchivingauthorstudy.pdf >, [accessed 15.09.09].

Swan, A. and Carr, L., 2008. Institutions, their repositories and the Web. Serials Review, 34(1), pp.31-35.

Tenopir, C., 2003. Use and users of electronic library resources: an overview and analysis of recent research studies. Washington, D.C.: Council on Library and Information Resources. Available at <http://www.clir.org/pubs/abstract/pub120abst.html>, [accessed 16.09.09].

Tenopir, C., 2007. Measuring impact and quality. Library Journal, issue 14 (Sept 1, 2007), p.30.

Tenopir, C. and King, D., 2000. Towards electronic journals: realities for scientists, librarians, and publishers. Washington, DC: Special Libraries Association.

Tenopir, C., King, D., Edwards, S. and Wu, L., 2009a. Electronic journals and changes in scholarly article seeking and reading patterns. Aslib Proceedings New Information, 61(1), pp.5-32.

Tenopir, C., King, D.,Spencer, J. and Wu, L., 2009b. Variations in article seeking and reading patterns of academics. What makes a difference? Library & Information Science Research, 31(3), pp.139-148.

Toledano, K., 2003. Scholarly communication and OAI – what are the issues for journal publishers? Serials, 16(3), pp.238-242.

Vakkari, P., 2008, Perceived influence of the use of electronic information resources on scholarly work and publication productivity, Journal of the American Society fro Information Science and Tecnology, 59(4): 602-612.

Vakkari, P. and Talja, S., 2005. The influence of the scatter of literature on the use of electronic resources across disciplines: A case study of FinElib. *In:* A. Rauber *et al.* (Eds.): ECDL 2005, LNCS 3652, pp.207-217.

Van Westrienen, G. and Lynch, C.A., 2005. Academic institutional repositories: deployment status in 13 nations as of mid 2005. D-Lib Magazine, 11(9).

Velterop, J., 2008, Open access and publishing *in:* UKSG, The E-resources Management Handbook. UKSG. Available at <http://uksg.metapress.com/media/788y661ntk0unhd0fl1y/contributions/d/p/l/a/dplay0kyn6nkvk7u.pdf>, [accessed 14.09.09].

Vernooy-Gerritsen, M., Pronk, G. and van der Graaf, M., 2009. Three perspectives on the evolving infrastructure of institutional research repositories in Europe. Ariadne, 59.

Wates, E. and Campbell, R., 2007. Author's version vs. publisher's version: an analysis of the copy-editing function. Learned Publishing, 20(2), pp.121-129.

Watson, S., 2007. Authors' attitudes to, and awareness and use of, a university institutional repository. Serials, 20(3), pp.225-230.

Wojciechowska, A., 2007. Analysis of the use of open archives in the fields of mathematics and computer science. OCLC Systems & Services: International digital library perspectives, 23(1), pp.54-69.

Xia, J., 2007. Assessment of self-archiving in institutional repositories: across disciplines. Journal of Academic Librarianship, 33(6), pp.647-654.

Xia, J., 2008. A comparison of subject and institutional repositories in self-archiving practices. Journal of Academic Librarianship, 34(6), pp.489-195.

Xia, J. and Sun, L., 2007. Assessment of self-archiving in institutional repositories: depositorship and full-text availability. Serials Review, 33(1), pp.14-21

Appendix 1 Summary of survey responses

Section 1: About you

1 Which of the following best describes your institution?

	Frequency	%
University or college	2,211	*70.4*
Hospital or medical school	161	*5.1*
Research institute	619	*19.7*
Government	63	*2*
Industrial/commercial	47	*1.5*
Other	38	*1.2*
Total	3,139	*100*

2 In which country is your institution based?

	Frequency	%
Austria	72	*2.3*
Belgium	81	*2.6*
Bulgaria	16	*0.5*
Cyprus	2	*0.1*
Czech Republic	52	*1.7*
Denmark	71	*2.3*
Estonia	6	*0.2*
Finland	52	*1.7*
France	310	*9.9*
Germany	437	*13.9*
Greece	78	*2.5*
Hungary	31	*1*
Ireland	37	*1.2*
Italy	400	*12.7*

	Frequency	%
Latvia	3	*0.1*
Lithuania	1	*0*
Luxembourg	4	*0.1*
Malta	3	*0.1*
Netherlands	156	*5*
Poland	82	*2.6*
Portugal	70	*2.2*
Romania	19	*0.6*
Slovakia	8	*0.3*
Slovenia	25	*0.8*
Spain	335	*10.7*
Sweden	123	*3.9*
UK	665	*21.2*
Total	3,139	*100*

3 How many years have you been employed in research?

	Frequency	%
Fewer than 5 years	494	*15.7*
5-14 years	1241	*39.5*
15-24 years	703	*22.4*
25 years or longer	701	*22.3*
Total	3139	*100*

4 Which field best describes your research area?

	Frequency	%
Medical sciences	248	*7.9*
Life sciences	416	*13.3*
Physical sciences & mathematics	1,773	*56.5*
Social sciences, humanities & arts	259	*8.3*
Interdisciplinary	440	*14*
Total	3,136	*100*

Percentages = % of those who responded to the question

Section 2: Awareness of open access and repositories

5 What does open access mean to you?

	Frequency	%
Full text available	2,717	*86.6*
Free access	2,469	*78.7*
Online/electronic format	2,082	*66.3*
Copyright free	389	*12.4*
Author pays to publish	338	*10.8*
Self archiving	187	*6*
Not peer reviewed	135	*4.3*
Not the final version	82	*2.6*
Not sure	67	*2.1*
Low quality	66	*2.1*
Vanity publishing	20	*0.6*
All of the above	11	*0.4*
None of the above	1	*0*

Percentages = % of total respondents that answered any part of the question

6 Does your institution have its own publicly available (i.e. open access) repository for research outputs?

	Frequency	%
Yes	1,008	*32.2*
No	1,285	*41.1*
Don't know	834	*26.7*
Total	3,127	*100*

Percentages = % of total respondents that answered the question

7 Are there any subject-based publicly available repository(ies) suitable for your research?

	Frequency	%
Yes	1,166	*37.6*
No	477	*15.4*
Don't know	1,457	*47*
Total	3,100	*100*

Percentages = % of total respondents that answered the question

7a. If 'Yes' in Q7, please name the repositories: 1,162 responses

8 Are you aware of any other open access repositories suitable for your research?

	Frequency	%
Yes	401	*13.1*
No	1,205	*39.4*
Don't know	1,451	*47.5*
Total	3,057	*100*

Percentages = % of total respondents that answered the question

8a. If 'Yes' in Q8, please name the repositories: 399 responses

Section 3: Authors – research dissemination behaviour

9. Have you published, or disseminated in any way, your research in the last five years?

	Frequency	%
Yes	2,725	*98.8*
No	32	*1.2*
Total	2,757	*100*

Percentages = % of total respondents that answered the question

10. How have you published/disseminated your research in the last five years?

	Frequently		Occasionally		Never		Total	
Monographs	81	*4.3%*	609	*32.3%*	1,197	*63.4%*	1,887	*100%*
Book chapters	198	*9.2%*	1,322	*61.1%*	643	*29.7%*	2,163	*100%*
Peer-reviewed journals	2,232	*82.6%*	448	*16.6%*	21	*0.8%*	2,701	*100%*
Other journals/magazines	293	*15%*	1,001	*51.2%*	661	*33.8%*	1,955	*100%*
Conference proceedings	1,203	*49.6%*	1,069	*44.1%*	153	*6.3%*	2,425	*100%*
Reports	390	*19.9%*	893	*45.5%*	679	*34.6%*	1,962	*100%*
Working papers	265	*14.5%*	471	*25.8%*	1,091	*59.7%*	1,827	*100%*
Datasets	81	*4.8%*	297	*17.5%*	1,324	*77.8%*	1,702	*100%*
Creative works (incl. exhibitions & performances)	54	*3.2%*	272	*16.1%*	1,367	*80.7%*	1,693	*100%*
Other	73	*5.3%*	65	*4.7%*	1,251	*90.1%*	1,389	*100%*

Percentages = % of total respondents that answered each part of the question

11. If 'Other' in Q10, please specify: 144 responses

12. Which is your most important type of research output in terms of your career?

	Frequency	%
Peer reviewed journals	2,411	*89.2*
Professional journals	108	*4.0*
Monographs	69	*2.6*
Conference proceedings	41	*1.5*
Reports	20	*0.7*
Working papers	18	*0.7*
Book chapters	16	*0.6*
Creative works	4	*0.1*
Datasets	2	*0.1*
Other	15	*0.6*
Total	2,704	*100*

Percentages = % of total respondents that answered the question

13. Approximately how many peer-reviewed journal articles have you published in the last five years?

	Frequency	%
0	11	*0.4*
1-5	587	*21.6*
6-20	1,351	*49.8*
21-50	623	*23*
More than 50	140	*5.2*
Total	2,712	*100*

Percentages = % of total respondents that answered the question

14 If you have published in a peer-reviewed journal, for your most recent article, please indicate the importance of the following factors in choosing a peer-reviewed journal rather than, or in addition to, any other form of dissemination *(please rate on a scale 1-5, 1 = very important and 5 = not important at all)*

	1		2		3		4		5		Not applicable		Total	
Widespread availability of my research	1,676	*62.5%*	608	*22.7%*	194	*7.2%*	110	*4.1%*	89	*3.3%*	5	*0.2%*	2,682	*100%*
Customs within your discipline	845	*32.4%*	821	*31.5%*	536	*20.6%*	210	*8.1%*	123	*4.7%*	72	*2.8%*	2,607	*100%*
Dissemination to target audience	1,112	*42.5%*	880	*33.6%*	384	*14.7%*	140	*5.4%*	84	*3.2%*	16	*0.6%*	2,616	*100%*
Speed of dissemination	456	*17.5%*	780	*29.9%*	818	*31.3%*	349	*13.4%*	180	*6.9%*	27	*1%*	2,610	*100%*
Ease of publication	303	*11.7%*	516	*19.9%*	880	*34%*	502	*19.4%*	332	*12.8%*	57	*2.2%*	2,590	*100%*
Cost of publication	416	*16.1%*	443	*17.1%*	677	*26.2%*	443	*17.1%*	432	*16.7%*	173	*6.7%*	2,584	*100%*
Career advancement	767	*29.6%*	724	*27.9%*	552	*21.3%*	280	*10.8%*	192	*7.4%*	79	*3%*	2,594	*100%*
Research assessment requirements	586	*23%*	680	*26.7%*	640	*25.1%*	277	*10.9%*	204	*8%*	162	*6.4%*	2,549	*100%*
Esteem/reputation of publication	1,239	*47.1%*	878	*33.4%*	289	*11%*	116	*4.4%*	91	*3.5%*	19	*0.7%*	2,632	*100%*
Prompted by colleagues/ co-authors	226	*8.8%*	595	*23.3%*	855	*33.5%*	422	*16.5%*	343	*13.4%*	114	*4.5%*	2,555	*100%*
Peer review (getting my work evaluated)	1,077	*41.1%*	752	*28.7%*	442	*16.9%*	183	*7%*	147	*5.6%*	21	*0.8%*	2,622	*100%*
Establishing precedence	325	*12.9%*	611	*24.3%*	721	*28.6%*	335	*13.3%*	280	*11.1%*	247	*9.8%*	2,519	*100%*
Other	42	*3.7%*	16	*1.4%*	14	*1.2%*	4	*0.4%*	81	*7.2%*	965	*86%*	1,122	*100%*

Percentages = % of total respondents that answered each part of the question

15 If 'Other' in Q14, please specify: 65 responses

16 Have you placed, or given permission to place, a version of any of your peer-reviewed journal articles in a publicly available repository in the last five years?

	Frequency	%
Yes – institutional	638	*23.6*
Yes – subject-based	657	*24.4*
Yes – other	127	*4.7*
Yes – but not sure where	194	*7.2*
No	1,304	*48.3*
Total respondents	2,920	

Percentages = % of total respondents that answered any part of the question

16a If 'Yes – other' in Q16, please specify: 200 responses

17 How easy was it to identify a suitable repository for these articles?

	Frequency	%
Very easy	630	*29.8*
Easy	429	*20.3*
Not very easy	172	*8.1*
Difficult	52	*2.5*
Don't know	834	*39.4*
Total	2,117	*100*

Percentages = % of total respondents that answered the question

18 Which version(s) of these peer-reviewed journal article(s) did you deposit?

	Frequency	%
Pre-print (electronic version prior to peer-review)	682	*40.2*
Author's final version (electronic version after peer-review, incl. amendments following referees' comments, also known as post-print or stage-2 manuscripts)	740	*43.6*
Published final version (publisher's PDF file)	541	*31.9*
Not sure	302	*17.8*
Total respondents	2,265	

Percentages = % of total respondents that answered any part of the question

19 If you have deposited the author's final version of any article, what prompted you to do this?

	Frequency	%
You did so voluntarily	723	*45*
Not applicable	465	*28.9*
You were invited by a publisher to do so	164	*10.2*
Colleague(s) suggested it	150	*9.3*
You were invited by the repository	131	*8.1*
Co-author(s) asked you to	129	*8*
You were required to do so by your employer	106	*6.6*
You were required to do so by your research funder	52	*3.2*
Other	39	*2.4*

Percentages = % of total respondents that answered any part of the question

20 How important are the following factors, or would the following factors be, in encouraging you to place peer-reviewed journal articles in a publicly available repository?

(please rate on a scale 1-5, 1 = very important and 5 = not important at all)

	1		2		3		4		5		Total	
Principle of free access to all	1,773	*68%*	458	*17.6%*	197	*7.6%*	75	*2.9%*	103	*4%*	2,606	*100%*
Widespread availability of my research	1,745	*67.1%*	559	*21.5%*	157	*6%*	60	*2.3%*	78	*3%*	2,599	*100%*
Availability to researchers with limited access to subscribed journals	1,481	*58.1%*	596	*23.4%*	275	*10.8%*	94	*3.7%*	105	*4.1%*	2,551	*100%*
Journal subscription costs charged by publishers	679	*27.4%*	643	*26%*	686	*27.7%*	227	*9.2%*	239	*9.7%*	2,474	*100%*
Speed of dissemination	1,017	*40.1%*	751	*29.6%*	513	*20.2%*	166	*6.5%*	91	*3.6%*	2,538	*100%*
Possibility of increased citations to the output	927	*36.9%*	792	*31.5%*	479	*19.1%*	182	*7.2%*	132	*5.3%*	2,512	*100%*
Important for career advancement	496	*19.8%*	584	*23.4%*	756	*30.2%*	355	*14.2%*	310	*12.4%*	2,501	*100%*
Prompted by peers to deposit in an open access repository	224	*9.2%*	477	*19.5%*	839	*34.4%*	426	*17.5%*	475	*19.5%*	2,441	*100%*
Requirement of my institution/department	291	*11.8%*	433	*17.6%*	643	*26.1%*	354	*14.4%*	741	*30.1%*	2,462	*100%*
Requirement of the finding bodies	366	*15%*	456	*18.7%*	583	*23.9%*	301	*12.4%*	731	*30%*	2,437	*100%*
Other	49	*9%*	6	*1.1%*	50	*9.2%*	8	*1.5%*	430	*79.2%*	543	*100%*

Percentages = % of total respondents that answered each part of the question

21 If 'Other' in Q20, please specify: 55 responses

22 What reservations do you have about placing your peer-reviewed journal articles in publicly available repositories?

(please rate on a scale 1-5, 1 = very important and 5 = not important at all)

	1		2		3		4		5		Total	
Do not consider open access to be beneficial to my research	128	*5.2%*	179	*7.3%*	365	*14.8%*	316	*12.9%*	1,471	*59.8%*	2,459	*100%*
Lack of awareness of what is an open access repository	309	*12.5%*	431	*17.5%*	485	*19.7%*	313	*12.7%*	930	*37.7%*	2,468	*100%*
Do not know of a repository suitable for my research material	424	*17.4%*	470	*19.2%*	464	*19%*	221	*9%*	864	*35.4%*	2,443	*100%*
Do not know how to deposit material in a repository	346	*14.2%*	390	*16%*	442	*18.2%*	281	*11.5%*	976	*40.1%*	2,435	*100%*
The process takes too long/is too complicated	135	*5.7%*	273	*11.6%*	621	*26.4%*	348	*14.8%*	974	*41.4%*	2,351	*100%*
Concerns about infringing embargo periods imposed by publishers	505	*21.1%*	467	*19.5%*	514	*21.5%*	287	*12%*	622	*26%*	2,395	*100%*
Concerns about infringing copyright	662	*27.3%*	491	*20.3%*	455	*18.8%*	266	*11%*	548	*22.6%*	2,422	*100%*
Concerns about plagiarism/how my material will be used by readers	330	*13.7%*	412	*17.1%*	478	*19.9%*	372	*15.5%*	814	*33.8%*	2,406	*100%*
Reluctant to put my research publications in a repository where other materials have not been peer-reviewed	654	*27%*	513	*21.2%*	414	*17.1%*	249	*10.3%*	592	*24.4%*	2,422	*100%*
Not comfortable with depositing a peer-reviewed version of my paper which has not been properly edited by the publisher	399	*16.6%*	456	*18.9%*	490	*20.3%*	355	*14.7%*	709	*29.4%*	2,409	*100%*
Although I want to do it, I always forget about it	104	*4.4%*	226	*9.6%*	487	*20.6%*	329	*13.9%*	1,216	*51.5%*	2,362	*100%*
Do not consider it to be my responsibility	144	*6.1%*	214	*9.1%*	567	*24%*	338	*14.3%*	1,098	*46.5%*	2,361	*100%*
Other	55	*12.1%*	13	*2.9%*	41	*9%*	6	*1.3%*	341	*74.8%*	456	*100%*

Percentages = % of total respondents that answered each part of the question

23 If 'Other' in Q22, please specify: 68 responses

24 When considering placing material in publicly available repositories, do you have a preference for the type of repository?

	Frequency	%
Institutional	574	*22.1*
Subject-based	1,197	*46.1*
No preference	508	*19.6*
I prefer to put material on my personal or departmental website	216	*8.3*
I prefer not to do this at all	73	*2.8*
Other	30	*1.2*
Total	2,598	*100*

Percentages = % of total respondents that answered the question

24a Please add any comments: 121 responses

25 Please add any other comments on repositories from your perspective as an author: 286 responses

Section 4: Readers – information searching

26 Please rate the importance of the following types of output as resources for your research *(please rate on a scale 1-5, 1 = very important and 5 = not important at all)*

	1		2		3		4		5		Total	
Monographs	503	*20.5%*	659	*26.9%*	671	*27.4%*	370	*15.1%*	246	*10%*	2,449	*100%*
Book chapters	466	*18.7%*	922	*36.9%*	745	*29.8%*	303	*12.1%*	62	*2.5%*	2,498	*100%*
Peer-reviewed journals	2,326	*90.1%*	149	*5.8%*	25	*1%*	19	*0.7%*	63	*2.4%*	2,582	*100%*
Other journals/magazines	185	*7.7%*	457	*19.1%*	807	*33.8%*	598	*25%*	341	*14.3%*	2,388	*100%*
Conference proceedings	402	*16.1%*	781	*31.3%*	725	*29%*	440	*17.6%*	148	*5.9%*	2,496	*100%*
Reports	189	*7.8%*	503	*20.8%*	714	*29.5%*	606	*25.1%*	407	*16.8%*	2,419	*100%*
Working papers	144	*6.1%*	291	*12.3%*	645	*27.2%*	633	*26.7%*	659	*27.8%*	2,372	*100%*
Datasets	219	*9.4%*	328	*14.1%*	524	*22.5%*	555	*23.9%*	700	*30.1%*	2,326	*100%*
Creative works (incl. exhibitions & performances)	57	*2.5%*	111	*4.9%*	310	*13.7%*	456	*20.2%*	1,324	*58.6%*	2,258	*100%*
Other	52	*8.2%*	20	*3.1%*	33	*5.2%*	21	*3.3%*	511	*80.2%*	637	*100%*

Percentages = % of total respondents that answered each part of the question

27 If 'Other' in Q26, please specify: 67 responses

28 Approximately how many peer-reviewed journal articles do you read on average each year?

	Frequency	%
0	3	*0.1*
1-10	105	*4.2*
11-50	706	*28.4*
51-100	882	*35.5*
101-250	572	*23.0*
more than 250	215	*8.7*
Total	2,483	*100*

Percentages = % of total respondents that answered the question

29 How do you identify relevant journal literature for your research?

	Usually		Sometimes		Rarely		Never		Total	
Library catalogue	306	*12.8%*	525	*22%*	882	*37%*	669	*28.1%*	2,382	*100%*
Library portal or gateway	586	*24.7%*	587	*24.8%*	693	*29.2%*	504	*21.3%*	2,370	*100%*
Subject-based portal, gateway or repository (e.g. ArXiv, PubMed)	1,255	*51.6%*	535	*22%*	362	*14.9%*	280	*11.5%*	2,432	*100%*
Bibliographic database (e.g. Web of Science, Scopus)	1,313	*53.4%*	627	*25.5%*	328	*13.3%*	193	*7.8%*	2,461	*100%*
Citations/references/ bibliographies	1,864	*74.7%*	516	*20.7%*	81	*3.2%*	33	*1.3%*	2,494	*100%*
Google Scholar	622	*25.8%*	718	*29.7%*	530	*21.9%*	545	*22.6%*	2,415	*100%*
OAlster	16	*0.7%*	75	*3.3%*	190	*8.4%*	1,982	*87.6%*	2,263	*100%*
Internet search engine (e.g. Google)	817	*33.4%*	965	*39.4%*	481	*19.7%*	184	*7.5%*	2,447	*100%*
Publisher sites (e.g. SpringerLink, ScienceDirect,Oxford Online)	753	*31.1%*	787	*32.5%*	553	*22.8%*	330	*13.6%*	2,423	*100%*
Browsing personal or library print journals	333	*14%*	716	*30.1%*	821	*34.5%*	512	*21.5%*	2,382	*100%*
Colleagues/word of mouth/personal contact	553	*22.7%*	1,321	*54.3%*	494	*20.3%*	64	*2.6%*	2,432	*100%*
Other	65	*13.3%*	27	*5.5%*	15	*3.1%*	382	*78.1%*	489	*100%*

Percentages = % of total respondents that answered each part of the question

30 If 'Other' in Q29, please specify: 90 responses

31 How frequently are you unable to get quick and easy access to the peer-reviewed journal article that you have identified?

	Frequency	%
Very often	210	*8.5*
Quite often	644	*25.9*
Sometimes	1,063	*42.8*
Rarely	548	*22.1*
Don't know	19	*0.8*
Total	2,484	*100*

Percentages = % of total respondents that answered the question

32 When you do not have quick and easy access to the peer-reviewed journal articles identified, which of the following do you do?

	Usually		Sometimes		Rarely		Never		Total	
Seek an open access source for the material	1,213	*50.2%*	600	*24.8%*	334	*13.8%*	271	*11.2%*	2,418	*100%*
Contact the author for a copy	529	*21.7%*	791	*32.5%*	754	*31%*	360	*14.8%*	2,434	*100%*
Seek a copy through interlibrary loan	638	*26.4%*	755	*31.3%*	609	*25.2%*	411	*17%*	2,413	*100%*
Ask a colleague	548	*22.9%*	998	*41.7%*	612	*25.5%*	238	*9.9%*	2,396	*100%*
Ask the library	466	*19.5%*	766	*32%*	682	*28.5%*	481	*20.1%*	2,395	*100%*
Pay to use/download an article	88	*3.7%*	224	*9.5%*	550	*23.3%*	1,498	*63.5%*	2,360	*100%*
Nothing – I do not use that resource	230	*11.1%*	656	*31.7%*	496	*24%*	688	*33.2%*	2,070	*100%*
Other	37	*8.9%*	17	*4.1%*	6	*1.4%*	355	*85.5%*	415	*100%*

Percentages = % of total respondents that answered each part of the question

33 If 'Other' in Q32, please specify: 56 responses

34 When your resource discovery path leads you to an open access source to obtain the full-text article:

	Frequency	%
Is it clear whether the article you are about to access is held in an open access repository?		
Usually	846	*33.5*
Sometimes	773	*30.6*
Rarely	335	*13.3*
Never	74	*2.9*
Don't know	497	*19.7*
Total	2,525	*100*
How important to you is it to know the version *(pre-print/author's final version/publisher's PDF file)* of the article?		
Very important	820	*32.3*
Quite important	1,059	*41.7*
Not very important	489	*19.3*
Not important at all	85	*3.3*
Don't know	86	*3.4*
Total	2,539	*100*
How often can you identify the version *(pre-print/author's final version/publisher's PDF file)* from the information provided?		
Usually	982	*38.9*
Sometimes	809	*32.1*
Rarely	293	*11.6*
Never	35	*1.4*
Don't know	403	*16*
Total	2,522	*100*

Percentages = % of total respondents that answered the question

35 Do you generally trust the integrity of the documents you find in repositories?

	Frequency	%
Always	913	*36.7*
Sometimes	1,096	*44*
Rarely	90	*3.6*
Never	17	*0.7*
Not sure	373	*15*
Total	2,489	*100*

Percentages = % of total respondents that answered the question

36 In terms of your citation preferences for journal articles, please rate the following statements on a scale of 1-5 *(1 = strongly agree and 5 = strongly disagree)*

	1		2		3		4		5		Total	
I will generally cite the version I have actually accessed	1,282	*51.1%*	491	*19.6%*	392	*15.6%*	200	*8.0%*	142	*5.7%*	2,507	*100%*
I am happy to cite a preprint	472	*19.0%*	599	*24.1%*	636	*25.6%*	472	*19%*	307	*12.3%*	2,486	*100%*
I will generally cite an author's version only if it has been peer reviewed and accepted for publication	910	*36.8%*	590	*23.9%*	402	*16.3%*	301	*12.2%*	267	*10.8%*	2,470	*100%*
I prefer to cite only the final published version of an article	1,519	*60.7%*	496	*19.8%*	222	*8.9%*	140	*5.6%*	124	*5.0%*	2,501	*100%*
I am happy to cite any version	221	*9.2%*	257	*10.7%*	475	*19.8%*	648	*26.9%*	804	*33.4%*	2,405	*100%*
I try to find the final published version before I cite the article	1,678	*68.2%*	453	*18.4%*	184	*7.5%*	73	*3.0%*	74	*3.0%*	2,462	*100%*

Percentages = % of total respondents that answered each part of the question

37 Please add any other comments on repositories from your perspective as a reader: 124 responses

Section 5: General opinions

38 To what extent do you agree or disagree with the perception that depositing research material in repositories may challenge the existence of:

	Strongly agree		Agree		Neither agree nor disagree		Disagree		Strongly disagree		Don't know		Total	
The peer-review system	258	*10.1%*	536	*21%*	521	*20.4%*	658	*25.7%*	447	*17.5%*	138	*5.4%*	2,558	*100%*
Subscription-based journals	310	*12.2%*	1,032	*40.6%*	566	*22.2%*	359	*14.1%*	129	*5.1%*	149	*5.9%*	2,545	*100%*
Scientific societies	138	*5.4%*	376	*14.8%*	616	*24.3%*	676	*26.7%*	536	*21.1%*	194	*7.6%*	2,536	*100%*

Percentages = % of total respondents that answered the question

39 Do you think there is a role for repositories in scholarly communications?

	Frequency	%
Yes	1,646	*65.8*
No	93	*3.7*
Don't know	763	*30.5*
Total	2,502	*100*

Percentages = % of total respondents that answered the question

39a Please give your reasons for this view: 914 responses