Knowledge Management Research & Practice (2009) 7, 218–233

www.palgrave-journals.com/kmrp/

The world wide web of research and access to knowledge

Eric T. Meyer[1] and
Ralph Schroeder[1]

[1]*Oxford Internet Institute, University of Oxford, Oxford, Oxfordshire, U.K.*

Correspondence: Eric T. Meyer, Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, Oxfordshire OX1 3JS, U.K.
Tel.: +44 (0)1865 287210;
E-mail: eric.meyer@oii.ox.ac.uk

Abstract
This paper examines the shift to online knowledge in research. In recent years there has been a major transformation in how formal and informal science communication is disseminated by electronic means. At the same time, researchers' practices in accessing knowledge and information have changed, particularly in the use of search engines and digitized resources apart from traditional journals. While we still know little about how this affects the nature of research, particularly in light of disciplinary differences, we reject here the idea that the simple growth of outputs and proliferation of outputs also leads straightforwardly to a richer and more diverse information and knowledge environment. Instead, we argue that gatekeepers such as search engines which shape online visibility, combined with competition for limited attention space at the leading edge of research, leads to a different model of how access to knowledge and information is being shaped.
Knowledge Management Research & Practice (2009) **7,** 218–233.
doi:10.1057/kmrp.2009.13

Keywords: e-Research; scholarly communication; disciplines; knowledge and information; knowledge creation; knowledge dissemination; socio-technical systems

Introduction

This essay examines how researchers gain access to knowledge at a time when scholarly communication and materials are increasingly moving online. This topic has so far mainly been discussed in terms of journal publication and readership. Here we take a broader view, partly because the focus on journals overlooks a number of trends favouring e-Research, where knowledge production and dissemination is broader than journal publications: e-Research also consists of efforts to develop distributed online tools, data and other resources. A second reason to take a broader view extends the horizon still further, since scientific communication and collaboration are not just undergoing change within the research community. Rather, these changes are also taking place, for example, in the light of how search engines affect what can be found online generally. New search behaviours are particularly evident among a new generation of scholars and potential scholars, and thus a wider picture is needed since search results are, in turn, affected by search behaviour. Hence we will look at changes in research as well as in the realm of online knowledge more broadly.

The essay will thus draw together recent research in a number of areas which, we will argue, are interrelated:

- how science communication is moving online;
- trends toward increased digitization and electronic availability of research materials;

Received: 3 December 2008
Revised: 16 March 2009
Accepted: 3 June 2009

- the growth of online tools and data;
- the broader context of how online information is used;
- how the use of search engines is shaping access.

It is the interrelatedness of these changes that is altering the research landscape, and the final part of the essay will show in detail how these areas are interrelated. To anticipate, with the growing importance of the web presence – and thus the online visibility of research – there is nevertheless only a small portion of this research which will be relevant within any one area of scholarship (Meyer & Schroeder, 2009). Thus despite the vast expansion of the online realm, there is still competition to dominate the attention space, which is shaped to a considerable extent by the gatekeeping function of search engines. There are implications not just for science communication, but also for the evaluation of research, which will increasingly rely on online measures of impact, a trend that is already influencing the ways that researchers disseminate their output. Hence there is a mutually reinforcing pattern to the ways in which research is shifting online and how its visibility is becoming ever important, in e-Research as well as in the larger domain of research generally.

The essay will argue that although any conclusions about the general impact of electronic resources on scholarly habits must be preliminary, digitized knowledge deserves close attention because its workings will have ever greater repercussions throughout the realm of research practices. The evidence upon which we have drawn to make these conclusions includes research done under the aegis of the Oxford e-Social Science (OeSS) project (www.oii.ox.ac.uk/microsites/oess/), which began in 2005. The numerous case studies of e-Research done by OeSS researchers, including those done by the authors, have focused on understanding the various social issues involved in implementing e-Research projects.

In this essay, we discuss how these case studies relate to scholarly communication and the wider issues in accessing and disseminating digitized knowledge. To do this, we will first sketch our overall view of the relations between the offline and online scholarly communication system, and this picture will be filled in with much more detail in the final section of the paper. In between, we will go back to how the shift towards an online system has taken place, including the various forms of informal communication. The key to understanding digitized materials, we argue, is to recognize that they are competing for visibility, and that this competition takes place within a whole system of online resources. There may be differences in how far this system applies to academic disciplines, but we go on to argue that despite these differences, mechanisms like search engines shape access to the system as a whole. So that when we return in the discussion to an elaborated diagram of our view of the scholarly communication system, and the place of e-Research within it, we conclude that there are self-reinforcing mechanisms which mean that certain types of research will become much more visible than others. Again, this requires a systemic view of online knowledge, which we shall now begin by sketching in simplified form.

This study of e-Research also potentially illuminates other domains in which activity is increasingly online, particularly those domains involving professional communication such as law, health and business. e-Research is a particularly pure version of professional communication online, however, because the very definition of e-Research discussed below is about digital tools and data used collaboratively over distributed online networks. This is why we focus on e-Research for understanding these trends that have clear parallels in other knowledge-work.

e-Research in the scholarly communication ecosystem

It is worthwhile to start the discussion with a schematic of the overall scholarly communication ecosystem that will be the focus of this paper. A more detailed diagram (Figure 2) will be presented below which adds complexity, but this simplified schematic can serve to introduce the outlines of our argument. Figure 1 illustrates several feedback loops that operate within the scholarly communication ecosystem. At the bottom of the diagram, we see a simplified model of the traditional path of knowledge creation and discovery in the pre-Internet offline era. Here, quite simply, individual researchers and teams of researchers draw upon the canon of literature that is transmitted via scholarly communication channels. The transmission lines are clear: printed journals and books are distributed either directly to academics or via research libraries. They then use that canon to inspire new research which they undertake, and finally feed the results of that new research back into the relevant scholarly communication channels. This is the primary feedback loop, traditionally. Some portion of this scholarly knowledge is translated for public consumption (shown in the bottom right-hand portion of the diagram) by popular scientific publications and educational media, but the communication to the public tends to be a one-way process, disseminating information for public consumption.

e-Research is only one part of the more recent internet-enabled scholarly communication system, but an important part. By e-Research we mean the use of digital tools and data (collectively *research materials*) for the distributed and collaborative production of knowledge. This definition sets e-Research apart from other uses of the internet for research, such as for scholarly communication or for the uses of digital materials by individual scholars. It also separates e-Research from the broader area of distributed work (Hinds & Kiesler, 2002) which encompasses a range of online collaborative activities which have been extensively studied by specialists in knowledge management for work organization settings. The two are nevertheless closely related in so far as

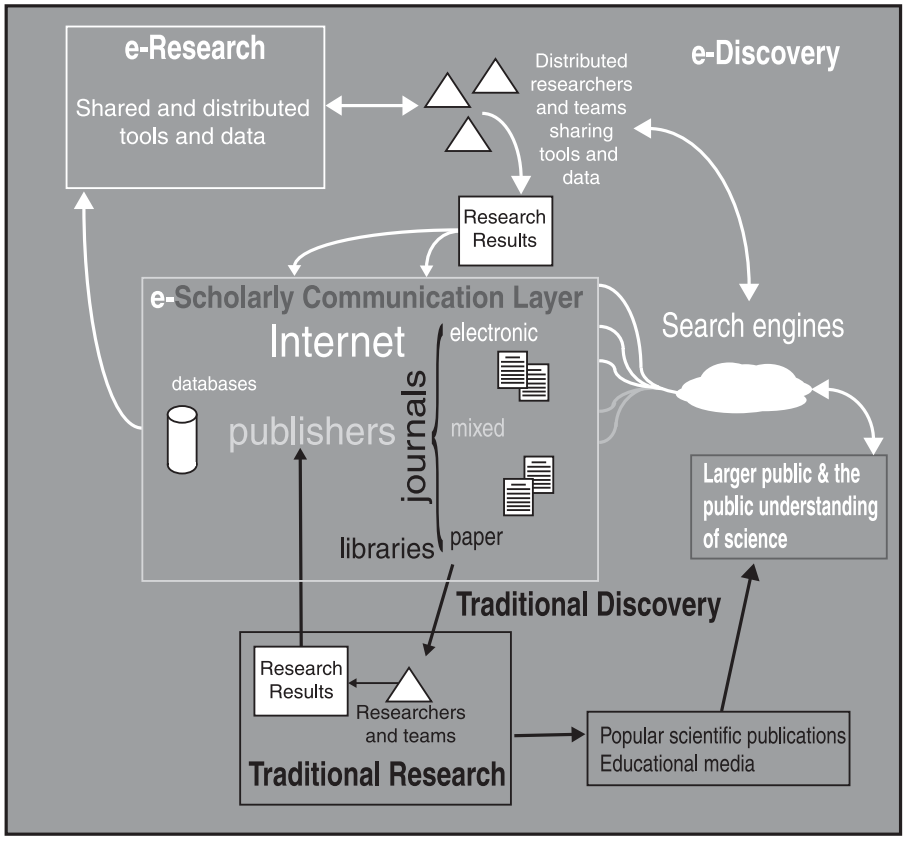

Figure 1 Schematic of the scholarly communication ecosystem.

research can be seen as a highly specialized type of work that is carried out using digital materials.

We can begin our sketch, at the top of the diagram, with the e-Research portion of the online ecosystem. As we can see (and will see in greater detail later), shifts towards online scholarship result in a more complex series of interacting feedback loops. On the top left, we see the activities that encompass the research portion of the ecosystem, where shared and distributed data and tools are accessed online. These data may be part of the cloud (a term derived from cloud computing), or live on the Grid (Berman *et al.*, 2003), or be part of federated data sets, and also include the tools used to manipulate and analyse data (as with visualizations of massive computations). Rather than being used by individual researchers or disconnected teams, in e-Research collaborative efforts performed by geographically distributed teams are the order of the day. In addition, multiple researchers and teams can access the same tools and data to generate new research. As with traditional research, e-Research also feeds back into the scholarly communication layer, but we will see later that the number of ways that this can happen has increased dramatically.

In addition to the loop in the top left-hand corner, there is a second, complementary feedback loop operating, as shown in the upper right-hand portion of the diagram. Here we see academics accessing the scholarly communication layer, but we also see that the paths are much more varied. No longer reliant simply on journal subscriptions and library access, academics can access more scholarly literature and more types of scholarly literature. The way that they find these resources is increasingly being mediated by search engines, and Google in particular. This reliance on a public search engine uncovers two related unintended developments related to the Web. First, Google was not designed primarily to allow academics to find research materials (although the later Google Scholar service was designed for this). Nevertheless, it has become a central tool in the researcher's toolbox for discovering knowledge. Second, academic results placed on the web were not primarily designed to be accessible to the public. Nevertheless, the fact that there is a single Internet, filled with everything from humorous cat photographs to peer-reviewed scientific papers on string theory, allows the public access to scientific results in a much less mediated fashion than traditional offline knowledge. This has implications for the relationship between science and the public.

It should be noted that the difference between the traditional research at the bottom of the diagram and the e-Research at the top is not strictly binary; there are shades of involvement with e-Research along the vertical

axis. However, we will argue that there are fundamental differences at the two poles of the vertical axis, and furthermore, that evidence supports our contention that scholars in general are migrating upwards in this ecosystem, away from offline modes towards online modes throughout the scholarly process. As we will see, this has implications for how the knowledge within the ecosystem is managed, including major shifts in the loci of organizational control. At this point, however, it will be useful to retrace the steps by which this online ecosystem emerged, and provide a fuller account of its components and mechanisms.

Background

Decades of hyperbole about the elusive promise of the paperless office notwithstanding (Sellen & Harper, 2001), the production, transmission, distribution, searching and consumption of academic knowledge is increasingly occurring paperlessly, and this trend has been enabled by Internet-based scholarly materials. While there remain disciplinary differences in the extent to which scholars rely on electronic resources instead of paper (Kling & McKim, 2000; Fry & Talja, 2004; Tenopir *et al*., 2004; Tenopir *et al*., 2005), the overall impact of electronic resources on the activities of scholars is undeniable.

Although there is still a limited amount of research on this topic and the world of scholarship is in a state of flux, there have recently been a number of studies which have examined different facets of how research is moving online. Caution is necessary: so far there is no comprehensive overview of how online scholarship is changing. The main notable exception to this is Borgman's (2007) work which covers a broad range of related topics, but her perspective is grounded in library and information science. In this essay there will be greater emphasis on the sociology of science and technology, though we shall come back later to the difficulty of addressing this topic from within any single disciplinary viewpoint. There is enough scholarship accumulating in this area, however, that it is possible to piece together a number of perspectives on the topic, and to ask whether they add up to being able to identify any larger overall patterns.

The shift to the use of online sources

There has been a marked shift to scholars accessing material online. Hallmark (2004), for example, reports on a study done at two time points (1998 and 2002) that examined how geologists and chemists were finding and retrieving research articles. The method in this study involved asking scientists how they found and retrieved a specific article that they had personally cited in a publication. In 1998, while 83% of their chemistry sample used Internet-based indices to find the article they cited, only 5% actually retrieved the article electronically. Most still relied on paper copies of articles retrieved from libraries, personal journal collections, or by contacting authors for re-prints. By 2002, however, the landscape had shifted entirely toward online search and retrieval. While 85% of chemists still used Internet sources to search for the article in question, 96% used the Internet to retrieve the article to either read it electronically or print it out for reading. The geologists in the study followed a similar pattern. It is remarkable that, for this sample at least, nearly all articles were being retrieved from electronic sources as early as 2002.

There appear also to be generational differences that are at play when trying to understand the shift to retrieving articles electronically. Sathe *et al*. (2002) conducted a study comparing the use of print and electronic journals. Using a small sample of journals in one library that were available both as print and as electronic editions, they surveyed users of each and found that younger scholars (Fellows, Residents and Students) were much more likely to have retrieved the journal electronically (57–70% across these categories) compared to clinical/research faculty (25%), who most often used the print sources at that time.

There are potential pitfalls, however, for younger scholars who rely on electronic resources. The highly publicized 'Google Generation' project and report (Centre for Information Behaviour and the Evaluation of Research, 2008) puts it this way:

> Most visitors to scholarly sites view only a few pages, many of which do not even contain real content, and in any case do not stop long enough to do any real reading. This is either a symptom of a really worrying malaise – failure at the library terminal – or maybe a sign that a whole new form of online reading behaviour is beginning to emerge, one based on skimming titles, contents pages and abstracts: we call this 'power browsing'. We urgently need to understand the root causes of this phenomenon. (p. 31)

The report also says that 'the ubiquitous use of highly branded search engines' entails, among other things, that 'many young people do not find library-sponsored resources intuitive and therefore prefer to use Google or Yahoo instead' (p. 12). 'Tools like Google Scholar', they say, 'will be increasingly a real and present threat to the library as an institution' (p. 13). 'Students usually prefer the global searching of Google to more sophisticated but more time-consuming searching provided by the library' (p. 31). This report in many ways puts lie to the meme that young so-called digital natives are sophisticated users and producers of information, seamlessly moving between online roles and connecting various electronic devices and resources through mashups and social networking. Instead, the CIBER report suggests that while young proto-academics may be quite comfortable with technology and tend to prefer electronic resources, the authors expressed serious concerns with the ability of children and college students to adequately search for and evaluate information, and they reject the notion that the Google Generation are 'expert searchers' (p. 20).

Another recent study (by the same research group) notes that 'undergraduate students tend to search the internet first, then go to library-based services, unless

they have been provided with and instructed on how to use a specific resource' (Nicholas *et al*., 2006, p. 1348). If this is the pathway for finding resources, however, then it is likely that different resources will be found and used than those that result from going to a library first, whether online or offline. Like the 'Google Generation' report's concern with the ability of younger generations to search and find relevant information, Nicholas *et al*. report that when trying to use digital libraries 'web users do not dwell, they examine just a few items/pages before they leave' (2006, p. 1363).

Much of what is known about access to online knowledge remains anecdotal: to what extent do students and scholars rely on electronic versions of papers rather than paper versions? As Borgman notes, whereas scholars may be able to make 'fine distinctions … in assessing the quality of a document … students, practitioners, scholars with minimal access to the published literature, and the general public usually are happy to read and cite any free version of a document they can find online' (2007, p. 84). However, even if college students do more than simply use Google searches for sources and make use of other online sources such as Wikipedia as well as offline sources, the very fact that they are uncertain about which sources to use (Head, 2007) suggests that visibility and access are important determinants of what they will find.

Informal scientific communication online

These concerns about the skills and judgement ability of younger generations of scholars are certainly a potential challenge for those responsible for educating them. These concerns may not apply to academics who have successfully negotiated the doctoral education system, and who will experience no particular difficulty in searching for information, even if they also face an increasingly varied set of online sources. However, the preference for digital resources, especially among younger generations of scholars, is being manifested in other ways as well. Channels of communication honed through use during student years are also appearing in modified forms in the academic arena. These communications, many of which fall under the now increasingly common title Web 2.0, include various novel forms of electronic informal scientific communication such as blogs, personal webpages, Podcasts, YouTube videos and Wikis. These are now being added to existing informal modes of academic communication, which include e-mail, e-mail lists, conferences and professional newsletters. While older electronic communications often represent technological replacements for existing modes of communication (such as e-mail replacing snail mail), the same is not true of many of the latest innovations. Blogging does not have a clear analogue in the paper-based world; journal writers may have kept track of their thoughts on a variety of topics, but they did not post them publicly unless they were published as memoirs at a much later date, often after their retirement or death.

The shift of research materials online thus involves a variety of informal means of scientific communication, and while it will not be possible to review these here, a few examples can suffice: Thelwall & Kousha (2008), for example, have examined PowerPoint presentations available on the Web looking for evidence of whether they could be used as non-traditional indicators of research impact. Using a combination of automated searching for PowerPoint files containing references to ISI journals and manual classification of an additional sample of presentations to look for other types of citations, they found that, in general, not enough presentations contained sufficient journal references to make the construction of an impact measure worthwhile. They did note, however, that online presentations often cite more popular resources such as Scientific American or Harvard Business Review, and thus may be an indicator that could be useful for tracking the popularization of research. Likewise, Wilkinson *et al*. (2003) also found that informal types of scholarly materials dominate the web-sphere. Wilkinson *et al*. extracted web-link information from 107 university websites, and found that almost 90% of the links were created for scholarly reasons, but that only a tiny minority were links to journal articles that could be considered the equivalent of a citation. The other links led to a variety of materials, including information for students (18%), but also material related to research resources (17%) and to libraries and e-journals (15%).

In general, if scientists are classic maximizers, they should only engage in alternative methods of informal communication if they are better in some measurable way to existing modes of informal communication. Matzat's (2004) study of scientific Usenet groups as channels for informal scientific communication concluded that while there was little support for the hypothesis that such groups had a democratizing effect on scientific communication, he did find evidence that participants in the groups reaped benefits in terms of research information and in maintaining weak ties with members of their extended networks.

Understanding that scientists communicate scientific information informally is not a new observation, but it is worth noting that the channels with which they do so are proliferating. Garvey & Griffith (1967) pointed out that 'scientists themselves create elements to fulfil the information needs that are not being satisfied by existing media' (p. 1012), and argued that these new elements would evolve over time and result in a shift in norms within a scientific field. Their example of certain fast-moving disciplines adopting increasingly speedy methods of exchanging pre-prints shows that tools such as arXiv.org are far from being completely new, technologically mediated innovations, but instead are the current incarnation of a trend noted in the literature 40 years ago (Garvey & Griffith, 1967, 1972). Specific scholarly behaviours may have changed, but the overall 'socioecological system' (Sandstrom, 2001) of scholarly communication has continued to evolve along

long-established lines (Hakken, 2003; Heimeriks & Vasileiadou, 2008).

Online visibility

For formal and informal academic materials to have any impact, they must be visible to their potential audiences. This is one area where the Internet offers much greater potential than the library-based paper publishing system ever did. Once academic material is on the web, particularly if located in open-access sources indexed by Google and other search engines, other scholars and members of the general public at least have a chance of finding the material. The apocryphal story of the doctoral dissertation on the library shelf still containing the $20 bill placed there decades earlier by its author reflects the understanding that few would bother to access something difficult to find and of such limited interest. Putting the same dissertation online doesn't make it more interesting to a wide audience, but it does make it much more likely that if someone is interested that they might take a look at it on the Web. The same is true of a variety of other academic outputs.

With regard to access by the general public, for much scholarly work there will always be a quite limited public audience. Nevertheless, the Internet is not compartmentalized and divided into separate physical spaces the same way that public libraries and academic libraries have been traditionally. By mixing one's academic work in with the other material in the cloud of information that everyone uses on the Internet, it becomes more likely that others may stumble on it than if it is locked away in dusty, little-visited academic libraries. Borgman puts it succinctly: 'content that is online gets more use than that which is not' (2007, p. 159). And, as Heimeriks & Vasileiadou (2008) point out, 'a scientist's visibility does not rely exclusively on the number of publications and their citations but can increasingly result from a well-designed and well-linked homepage providing scientific content' (p. 18).

Placing academic content online and allowing it to be freely used by others is referred to as open access publishing. Borgman (2007, p. 101) argues that enhanced visibility is one of the main motivations for open access. However, how the visibility of individual researchers is related to their impact as scholars is still unclear. Barjak *et al.*'s (2007) study of inlinks to scientists' personal webpages found conflicting results. While they found that full-text articles were the most linked content on the personal pages, they found confusing results when examining the collaboration networks of the scientists. Contrary to expectations, having a large number of collaborators had a negative impact on the number of inlinks, and productivity was similarly not reflected by the number of inlinks. The authors conclude that the main lesson of these confusing results is that our understanding of the role of visibility on the web is still incomplete (see also Houghton & Sheehan, 2006).

If one accepts the premise that online visibility is a growing trend, it becomes possible to speak of the online 'presence' or 'visibility' of research. However, presence and visibility have not yet been well-defined; a conception of the web as a whole, of the internet and web as a *system,* is currently missing from the debate, and thus there are insufficient models for understanding competition for attention within this space. The notion of the 'web sphere' (Schneider & Foot, 2005) has been used, for example, for analysing political phenomena, but it is unclear if this notion can be applied to research outputs, which are more varied. The reason for the lack of an understanding of the competition for visibility is that much of the focus has been on the producer side, on outputs or on those who put information and knowledge online. Equally important, however, is the consumer side; how are the use-patterns in information and communication practices of researchers changing in light of the other changes we have discussed in the scholarly communication system? Moreover, in relation to both production and consumption of online knowledge, there has been a focus on the question of whether the shift of materials online results in a winner-take-all system (the 'Matthew effect', see below) or in a more democratic system of scholarly attention. As we shall see, this question may be too limited. Yet in either case, the competition to dominate the attention space has moved onto the new terrain of the websphere, and, as a result, new players have entered the equation and have altered the role that different mechanisms such as search engines play in determining scholarly visibility.

Digitization of research materials

There is a larger ongoing digitization of research materials which has been discussed in terms of the materials being digitized (Nentwich, 2003; Borgman, 2007). Borgman points out that most journals in science, technology and medicine have been online for some time (2007, p. 181) and notes that 'scientific data are fastest-growing portion of the content layer' of scholarly communication infrastructures (2007, p. 182). Borgman's definition of data is 'a reinterpretable representation of information in a formalized manner suitable for communication' (pp. 119–120) and that 'the terms *data and facts* are treated interchangeably' (p. 120), although this still leaves a distinction between primary resources which consist of data – and secondary resources which are the means of accessing primary resources (p. 122).

Current efforts to put data and other research materials online and make them searchable and more easily manageable can be seen as attempts to cope with the 'data deluge' (Hey & Trefethen, 2003), especially in some of the natural sciences. Equally, however, a more long-term trend is simply the deluge of papers and research materials that researchers in all fields need to cope with. Collins summarized the trend before online material had become available: 'what we see around ourselves in recent decades has been an enormous expansion in

cultural production. There are over 1 million publications annually in the natural sciences, over 100,000 in the social sciences, and comparable numbers in the humanities' (Collins (1998, p. 521) citing de Solla Price (1986, p. 266)). Collins notes that scholars are increasingly 'buried in papers' (1998, p. 92), a phenomenon that today extends to electronic papers (although certainly figuratively rather than literally).

Heimeriks & Vasileiadou (2008) discuss this in terms of the science communication system, arguing that information and communication technologies (ICTs) contribute to 'the emergence of heterogeneity in methods of analysis, types of data, types of scientific output, modes of communication and coordination, modes of socialization and identity construction, types of career paths' (2008, p. 23). This increased heterogeneity leads to changing dynamics in the various layers of the scientific communication process, and also, they argue, leads to increased reflexivity about modes of communication as typified by the very public discussions of open-access publishing and copyright (see, for instance Ginsparg, 1996, 2006; Bachrach *et al.*, 1998; Harnad, 2001).

e-Research in the changing research landscape

e-Research has become an umbrella term for e-Science, e-Social Science and e-Humanities. Other terms that have been used are cyberinfrastructure (especially in the U.S.) and e-Infrastructure (in the EU). The various research programmes associated with this have been described in Schroeder & Fry (2007) and can be found on various agency websites detailing national agendas for scientific funding. Many of the projects can also be found in the various reports on cyberinfrastructure that have been released in the last few years (e.g., Atkins *et al.* (2003), Berman & Brady (2005), Arms & Larson (2007), and a variety of other reports). But apart from these funding programmes, as we have argued earlier, it is important to delimit e-Research analytically and to separate the distributed and collaborative production of knowledge using digital research materials from other uses of the internet in research.

This is not the place to describe the social aspects and implications of e-Research in detail. In this paper we can simply note, first, that e-Research materials constitute a subset of the overall digitization of research, namely, the subset that develops and uses these materials for online distributed collaborative research. Second, these materials have the further characteristic that they are intrinsically connected to the e-Scholarly communication layer (that we describe in more detail below) since e-Research materials inputs such as data and software and outputs such as data for analysis and re-use are online. Third, e-Research is also developing an infrastructure that is part of the communication layer, but also outside of it in terms of access to resources and tools on a longer-term basis.

As this material becomes increasingly prominent in the research landscape, it will be important to establish to what extent this becomes part of formal and informal scholarly communication. How, if these materials are part of an emerging infrastructure of research, will they be accessed and used by researchers? How will this vary across disciplines? More generally, how will e-Research shape and be shaped by the changing processes of scholarly communication that we have sketched here?

An online 'system' – and e-Research as part of it

e-Research contributes to the move of knowledge online, and by definition produces materials for online access. However, the visibility and dominance of online resources must also be seen in a context that is larger than search, fields, and formal and informal scholarly communication. Kling *et al*. have suggested that it is possible to see new electronic forms of scholarly communication such as 'electronic editions of paper journals, pure electronic journals, working article repositories, post-publication archives, pre-print servers, collaboratories, cross-linked Webs of resources, gene databases' and the like as part and parcel of a set of e-Scholarly Communication Forums (Kling *et al.*, 2003, p. 47). The authors go on to point out that this does not mean that these forums are therefore purely electronic since researchers also exchange information face-to-face.

These e-Scholarly Communication Forums could be regarded as overlapping with the emerging e-infrastructures or cyberinfrastructures, but the latter also constitute something larger. e-Infrastructures are systems of networked digital resources that will serve fields but also scholars across fields at a national or supranational level in the manner of a long-term support mechanism to support research (Borgman, 2007; Schroeder, 2007). Thus, the shift to online resources cannot be left on the level of scholarly communication practices, but must be raised to the level of transformation in the very systems of scholarly communication. Fry (2006) uses the term 'scholarly networked digital resources' to refer to the overall system beyond individual projects, digital libraries or discrete webpages. A broader conception such as this allows us to include both the infrastructure and its networked parts which make up the scholarly online ecosystem.

The online ecosystem thus consists of more than just scholarly communication. Within scholarly communication, a distinction is made between formal communication which is long-lasting and addressed to a wider audience and informal communication which is more ephemeral and between a more restricted audience, or between public and private communication. However, as Borgman (2007, pp. 48–49) points out, these lines are especially hard to draw with digital scholarship. These boundaries are blurring in various ways in e-Research and the digitization of scholarly materials. Hence Borgman also notes that, 'in digital environments, dissemination can be difficult to distinguish from access' (Borgman, 2007, p. 87). In other words, what is found online can be regarded as published for dissemination and what is

disseminated can be seen as being published, even if this interchangeability was not intended.

Disciplinary differences

Kling & McKim (2000) are sceptical of overly optimistic views of the power of ICTs to transform scholarly activity. The central argument they make is that while there are many examples of increasing digital scholarship, important field differences remain and shape the extent to which disciplinary actors are likely to actively engage in e-Research. (Field is used here when describing an emergent scholarly domain, such as a specialism or multi-disciplinary effort, which does not fit within the boundaries of existing disciplines.) For instance, in computer science and physics, articles are published online; computer scientists have been particularly aggressive in terms of placing their papers on their personal websites (Borgman, 2007, p. 102). Does this mean that these two disciplines will become more visible *vis-à-vis* other less aggressive fields? This would follow if we combine this trend with what Harnad (2001) noticed about how much more open access material was cited – but only if knowledge transcends disciplinary boundaries; if it does not, then the visibility should be 'contained' within the two disciplines.

Disciplinary differences can be overemphasized. Cummings & Kiesler (2005) found, for example, that cross-disciplinary collaboration is not so much of a problem for collaborative research as is research which spans across different institutions or that which bridges distances (in other words, with distributed teams). Similarly, Walsh & Maloney found that the 'structure of work', including size, distance, interdependence and scientific competition, are more problematic for collaboration in scientific teams than the mix of backgrounds which is often *seen* as problematic (Walsh & Maloney, 2007, pp. 1, 11).

Nevertheless, disciplinary differences are evident when it comes to the speed at which the processes described here are occurring. Fry, drawing on Whitley (2000), has argued that, in terms of 'the differential role of informal and formal communication across fields', Whitley's characteristics of fields have 'an influence on the production and use of scholarly networked digital resources' (Fry, 2006, p. 312), such that high-energy physics, with a high degree of mutual dependence and low degree of task uncertainty, are much more likely to produce and use these resources than fields like social/cultural geography, with low degree of mutual dependence and high degree of task uncertainty. There are, however, patterns which override differences like these; for example, researchers in all four fields that Fry *et al*. (2008) examined (terrorism, HIV/Aids, climate change and internet research) use scholarly networked digital resources in such a way – for example searching with Google – that they are even more reliant upon these resources. If, for example, you search for an individual scientist's personal webpage, visibility will be even more important in a field which makes low use of these resources. Other scholars have also looked at discipline-specific use of scholarly networked digital resources; see, for example Tenopir *et al.*'s (2005) paper on astronomers and Tenopir *et al*. (2004) examining medical faculty. It is clear therefore that field differences in terms of the extent to which electronic scientific communication is adopted will persist (Kling & McKim, 2000; see also Walsh *et al*., 2000). Yet even if humanities and social sciences 'lag', this must nevertheless be put in the context that all disciplines are moving in the direction of digitizing online resources. The key point is that *all* disciplines are doing this in different ways, and will thus be subject to the competition for visibility that we outlined earlier, even if this competition will take various forms.

Measuring and scoping online research

There has been a shift among some who measure impact away from traditional bibliometrics for measuring scientific output towards computing-based methods called webometrics. Thelwall describes webometrics as being closely related to bibliometrics except that webometrics focus on 'the quantitative analysis of web phenomena … typically addressing problems related to bibliometrics' (Thelwall, 2008, p. 7). Thelwall gives an account of how webometric methods are increasingly used to rank universities, the web visibility of research outputs, mapping research fields and the relations between research groups, and other metrics of knowledge production. He concludes that one of the drawbacks of webometric methods compared with more traditional bibliometric measures based on citation analysis is that the former are likely to reflect the social structure of the web, which 'is a very partial reflection of the activities of research' (Thelwall, 2008, p. 14). For example, only some papers are freely available online and researchers vary greatly in whether they maintain good, poor or no websites. What he fails to note is the self-reinforcing nature of this 'partial reflection': if researchers use search engines, and search engines reinforce high web visibility, then this partiality will increasingly reinforce itself in that searches will result in online materials being found and online materials being found by search because they are online. This sort of self-referential circle is arguably not a particularly good measure of scholarly impact. However, this is why a key question at this 'systemic' level concerns 'open access' as against an increasingly 'closed' regime whereby access is restricted because of cost and/or the national or other boundaries of infrastructures (Schroeder, 2007). As mentioned earlier, this openness at the various levels will influence visibility.

Some sceptics remain unconvinced that the trends such as those outlined in this paper represent anything approaching a qualitatively new set of scientific behaviours. Gläser (2003), for instance, argues that 'the social order of collective production in scientific communities is not affected by the internet' (p. 47). To reach this conclusion, Gläser focuses on the production

of knowledge (as opposed to the consumption of knowledge) and argues that while there have been some changes in the speed of communications and specific methods of communicating knowledge, none of the behaviours enabled by the Internet are truly new. One exception he notes is that of the sharing of raw data though online databases which has the potential, in his opinion, to open up an entirely new type of scientific activity. One reason for this, he argues, is that the publication of raw data is an activity quite unlike the publication of journal articles: 'The main difference between publications and submissions to databases is that publications contain knowledge claims, while databases contain raw data' (Gläser, 2003, p. 44).

Heimeriks & Vasileiadou (2008) argue, however, that Gläser's interest is too narrow, and is mainly looking at the scientizing level of e-Research (Heimeriks & Vasileiadou (2008) citing Rip (1990)). By taking a more systemic view, they argue, one sees more evidence of an impact: 'The main contribution of ICTs is the emergence of heterogeneity in methods of analysis, types of data, types of scientific output, modes of communication and coordination, modes of socialization and identity construction, types of career paths … The additional layer of communication that the ICTs provide in the sciences changes the dynamics of established layers of communication (such as the journal-based system), first, by providing a variety of new elements that can be recombined with more established elements' (Heimeriks & Vasileiadou, 2008, p. 23).

The world wide web of science and the Matthew effect

At this point, it may be useful to provide a brief concrete example of changing practices in the 'consumption' of research and its visibility. Here we can draw on previous work which has analysed this question both from a quantitative (webometric) and qualitative (interview based) perspective. In this project, we and our colleagues have examined the shift towards online knowledge in the context of a project ('The World Wide Web of Science') which specifically asked whether this shift concentrated or decentralized knowledge, sometimes referred to as the 'winner takes all' or 'Matthew effect'. The Matthew effect describes how scholars with high visibility are much more likely than others to have their new work noticed, read and cited, while less prominent scholars are less likely to have their work noticed (Merton, 1968). This is true even when prominent scholars co-author papers with their less prominent colleagues; when this happens, the prominent scholar is presumed to be responsible for the most important contributions to the work regardless of their placement in the list of authors as first, second, or even last author (although author order varies considerably by field). Merton argues that this phenomenon has both dysfunctional personal aspects (preventing ideas from entering the discourse because of an author's lack of prominence or minimizing the role played by a less prominent co-author) and functional systemic aspects (such as when prominent scientists are able to draw attention to the new discoveries of a less-prominent colleague through co-authoring).

One hypothesis about the impact of the wide availability of electronic resources is that search and notification systems enabled by electronic scholarship may lessen the Matthew effect. Merton argued that part of the social reason for the Matthew effect to exist at all was because of the problem of the 'Forty-First Chair'. The Forty-First chair refers to the French Academy, which allowed only 40 chairs at any time and excluded equally highly qualified scholars due to the lack of vacancies (Merton, 1968, p. 56). This implies that recognition of scholarship is a limited and constrained resource. In the age of digital scholarship, is this still true? When journals are no longer delivered in paper form to one's office but are instead accessed electronically, either openly on the web or through one's institutional library links, are the limits to attention the same? Can scholars set up search criteria that alert them to potentially interesting articles in a wider range of journals, and more importantly, do they?

In order to test whether electronic resources help alleviate the dysfunctional aspects of this system for the individual scholar, in previous work one of the authors and his colleagues examined how researchers in four fields gain access to knowledge. This was done in two ways: one using webometrics methods (see Caldas *et al*. (2008)) and the other using interviews (Fry *et al*., 2008). The results were mixed. While the project did not find evidence for a widespread Matthew effect, it did find evidence that search engines such as Google can function in some domains as a facilitator in accessing knowledge and in others as an influential gatekeeper (Fry *et al*., 2008). In a series of interviews with domain experts, Fry *et al*. found universal support for the notion that the Internet was a vital tool of scholarship in terms of accessing both published material and also for scoping new topics, accessing grey literature, and tracking the scholarly activity of colleagues. One key difference among different fields of study, however, is the degree of scatter in the field's topics and resources: in low-scatter fields, centralized resources and aggregators serve as prominent gatekeepers in a way that is less true in fields where information is highly scattered. In these high-scatter fields:

> Concepts are often contested, which leads to more open-ended undirected searches and increased uncertainty with regard to the appropriate keywords to search. Rather than search being for a particular specialized concept it is often for individual researchers, institutions or general concepts. In these domains access to online resources is more likely to depend on the indexing algorithms of Internet search engines and the online presence of particular institutions, organizations, people and resources. (Fry *et al*., 2008)

One related trend to note is the increasing dependence on teams to do science and to co-author papers. Wuchty

et al. (2007) studied data on publications and patents over five decades. With over 20 million items in their sample, they found that teams are more frequently cited than solo authors and produce the most high-impact research. This trend is true in varying degrees across a variety of domains including science and engineering, the social sciences, the arts and humanities, and patents. Even the domain with the least team-based publication, the arts and humanities, nevertheless showed 89% of fields in the domain experiencing increases in team size (2007, p. 1036). Comparable data collected by the authors of this paper suggest that for fields central to e-Research (computer science and engineering, for instance) team size is even larger than in the Wuchty *et al*. data when one looks at publications related to the topic of e-Research (Meyer & Schroeder, 2009). This is an indication that the Matthew effect may operate in the current landscape at the team level more than at the level of the individual: if a team of prominent scholars brings new members into their team, they are likely to increase the new members' impact as well. Whether this is more widely true would require additional research.

Shaping access with search engines

Search engines as gatekeepers

The very idea that commercial search engines, and one dominant one in particular, should be used to access scholarly knowledge, would have been unthinkable 10 years ago. As noted in Fry *et al*. (2008) above, Google and Google Scholar are increasingly playing a gatekeeping function in e-Research. Google has actively moved into an area formerly dominated by players such as Thomson/ISI. It would be hard to argue that the former gatekeepers were terribly democratic in their policies toward access to knowledge: most of these proprietary databases were carefully locked behind subscription-based walls guarded by university libraries. Google Scholar, however, is not a database but an index and search engine. The articles that it finds are still often locked away behind subscription-based interfaces unless one is accessing them from a prominent research university.

While the ISI Journal Impact Factor (JIF) is well documented as a strong influence on the behaviour of scholars (particularly in terms of tenure and promotion), the impact of Google is only starting to be discussed more widely (see, for instance, a recent editorial lamenting one new journal's lack of impact in Google Scholar (Spoelstra *et al*., 2007)). Hemminger *et al*., for instance, report that while 'researchers still primarily use library and bibliographic database searches … the use of Web search engines such as Google Scholar is almost as common' (Hemminger *et al*., 2007, p. 2215).

But how well does Google Scholar fare at finding relevant, high-quality scholarship? Meho & Yang (2007) conducted a study of the ability of Google Scholar to find citations to an author's work by comparing Google Scholar to Scopus and the Web of Science for the work of a single highly cited library and information science department. They determined that Google Scholar was superior to the other sources at finding citations to an author's work in conference proceedings (four times more), non-English sources (over six times more), and in works self-archived on a personal or institutional website (which are not covered by the other sources at all). This gives authors who choose to self-archive a 'dramatic advantage' in terms of their visibility in Google Scholar (p. 2118). In terms of accessing peer-reviewed journal literature, however, Scopus and Web of Science are better at filtering out low-impact sources of citations; most of the additional citations Google Scholar was able to find were from low-impact journals and conference proceedings, and their inclusion did not change the relative ranking of scholars' productivity. This wide coverage Meho & Yang report in Google Scholar is in marked contrast to an earlier study which found that Google Scholar suffered from 'massive content omissions' (Jasco, 2005, p. 208).

Kousha & Thelwall (2007) have also compared ISI citations to Google and Google Scholar using a different methodology. They also found variability in the effectiveness of Google Scholar, noting that fields with a bias toward valuing conference articles and placing them online such as computer science and some social science disciplines were better represented in Google Scholar. Overall, however, they found a strong correlation between Google Scholar citations and ISI citations.

Meho & Yang point out some structural problems with Google Scholar that limit its use as a bibliometric tool, but also have implications for scholars who rely on it as a tool for research. Google Scholar lacks full bibliographic information and metadata on the sources it finds. More importantly, Google Scholar 'ranks the items in a rather inconsistent way … [and] does not allow resorting of the retrieved sets in any way (such as by date, author name, or data source' (2007, pp. 2110–2111). However, if one wants to find more marginal literature (such as that located in what Anderson has called the 'long tail' (Anderson, 2006)), Google Scholar is a better source than Scopus or Web of Science. There are many reasons why a scholar might be interested in moving outside the mainstream journal articles: to examine an under-researched topic, to find references to a newly emerging topic that has not yet had time to appear in mainstream publications, or to find international perspectives on a topic of interest beyond the global north.

Discussion: social science approaches to understanding online research across the disciplines

It is interesting to consider how the social sciences are currently somewhat ill-equipped to understand the migration of research online, not only in the social sciences but also in the sciences, arts and humanities.

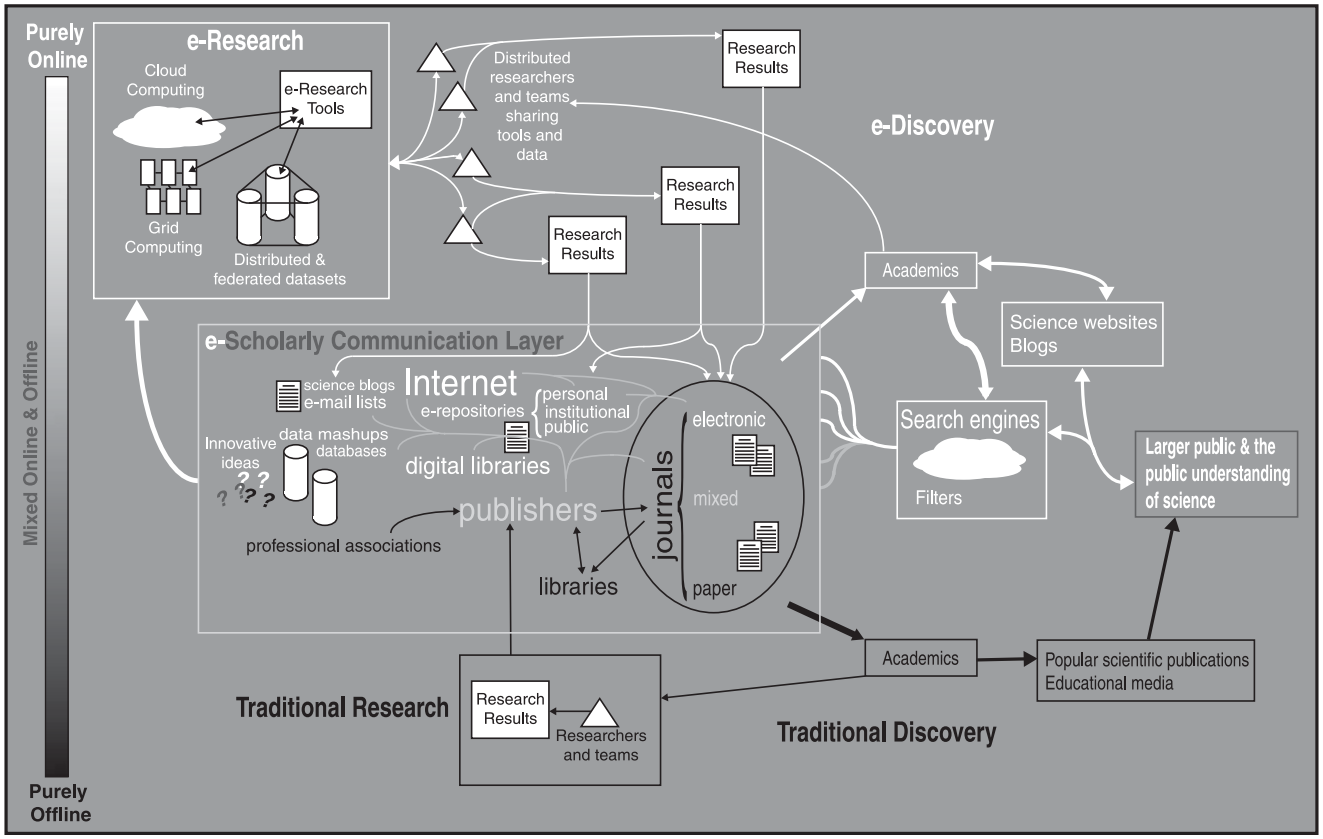

Figure 2 e-Research in the scholarly communication ecosystem.

This is due in part to disciplinary specialization, but also in part to the diffuseness of the object under consideration. As for the disciplines involved in advancing our understanding of how e-Research affects scholarship, these are foremost the sociology of science and technology and library and information sciences, though these do not deal, for example, with the broader changes outlined earlier in how young people access news about science (which might lie in the domain of educational or media studies), or how an attention economy is shaping the publishing world (which tends to fall under the scope of the economics of innovation). Orlikowski & Barley (2001) have argued that organizational studies and information science need to better learn from each other, by incorporating more institutional analysis in information science and a better understanding of the material properties of technology in organizational studies. We agree with this argument, but suggest also that to gain a fuller understanding one must also look at the broader sociology of the online realm. What would be needed is a wider understanding of how search is transforming the world of information-seeking generally beyond the boundaries of seeking scholarly information. This would require a broad sociology of knowledge and communication and of the online realm. It would also need to draw together very specialized areas, such as the public understanding of science, with much broader areas, like the shift from traditional to new media.

One way to get around the problem of various disciplinary approaches that have not integrated their insights is focus instead on the object under consideration. However, there is as yet no established conceptual apparatus for dealing with this object (web sphere, scholarly communication and e-Infrastructure, as we have seen, all have their limitations). For e-Research, arguably, there are three key elements: the material that has shifted into and become aggregated within this online realm; the gatekeepers and paths to this realm; and finally the users who seek and digest this material as part of their overall information and communication diet – or their information and communication ecology, if you prefer. The concepts of attention space, online visibility and gatekeepers cut across these and might allow us to get us a sense of how the leading edge of research is being shaped. One way to understand the relationships between these, then, might be as follows (see Figure 2).

The diagram in Figure 2 is a much more detailed version of the simple schematic presented in Figure 1. We can see here a number of additional elements of e-scholarly communication that constitute the online ecosystem that has been elaborated here. In this

illustration, again the bottom portion represents purely offline materials; moving upward passes through a range of mixed online and offline materials, to the purely online materials in the top portion of the illustration.

The cycle of traditional research, necessarily simplified for purposes of illustration, is as it was in Figure 1: academics discover existing knowledge through their subscriptions to paper journals, memberships in professional societies, book purchases, and the local and professional networks. Building on this knowledge, they do new research which generates research results. These are written up into new articles and books, which are submitted directly to publishers, and if all goes well with the peer-review process, the results are eventually published, allowing the cycle to continue. There are obviously variations that would need to be included if one were to build up a precise view of the old reality, but this simplified model is sufficient to serve as an illustration of the main elements of offline research. To the extent that the general public is allowed to see into this process, it is filtered through official channels such as popular science publications and educational films and television programmes. The system is relatively tidy, has quite clear lines demarcating various actors in the system and direct relationships between them, as indicated by the straight lines and arrows in the illustration.

At the top are the elements generally included in discussions of e-Science: the shared and distributed tools and data that are being digitized and networked in e-Science/cyberinfrastructure projects. This includes data sets, analysis tools, Grid-based and cloud-based resources, and a range of resources designed to enable e-Research. These resources are drawn from the e-Scholarly Communication Layer (e-SCL) in that they are themselves a form of scholarly communication. They do however, present new issues in research: From an organizational point of view, e-Research tools and data can be either more or less organizationally complex than traditional research.

Traditional research generally requires organizational support, in the form of universities and other research organizations, but the activities of individual researchers and their teams operate mainly independently of each other. e-Research, particularly that which relies on large infrastructure, often requires massive cross-institutional and even international cooperation. Interviews the authors have conducted with senior scientists in many different fields ranging from medicine to computer science and across the social sciences and humanities suggest that scientists involved in these distributed and collaborative efforts spend a lot of time on meetings and other organizational activities simply to support collaboration. For instance, one leading psychiatric scientist says: 'In my career, I've seen a lot more emphasis on collaboration and consortia and that has changed the way we think about doing our work … and I think that process is going to continue, where we recognize that work is being done by a big group and not by people in isolation' (Interview, 30 October 2008).

At the same time, there is a competing paradigm within e-Research that focuses on flexible, bottom-up approaches (De Roure & Goble, 2009). Sometimes called Research 2.0, referring to the flexibility of Web 2.0 in social networking software applications, these efforts often bypass organizational infrastructures entirely. Increasing numbers of resources that let researchers directly access data and tools without regard to their institutional affiliations and local infrastructure mean that in this case, researchers can lessen their dependence on organizations. This bottom-up approach taps into existing practices among many scientists and social scientists who already build their own tools to perform research (Meyer & Dutton, 2008). The difference between top-down approaches reliant on well-developed organizational structures and bottom-up approaches reliant on lightweight distributed tools and data are not mutually exclusive, however, and the availability of both allows researchers to choose the right tool for the job at hand. Supporting these tools and data, however, is more complex because of the multiple fluid paths to and through online resources.

The outputs of these distributed researchers and teams sharing tools and data, shown in the upper centre portion of the diagram, are research results, as with traditional research. However, the variety of forms that these research results can take has grown enormously. Rather than having a small range of options for submitting results for publication, the online e-SCL enables multiple fluid paths for different outputs and even for copies of the same output. Researchers can and do still publish their results via the traditional routes (articles in peer-reviewed journals and books), but they can now also take those same results and post about them on their blogs, send them to large groups of researchers via e-mail lists, and post pre-prints on their webpages and in institutional or public repositories. From the point of view of the researcher, these options complement the 'official' publication of the results in a journal of record, and increase the visibility of the results, as discussed above. Publishers, on the other hand, are struggling with maintaining their position as the freely available copies of the material for which they charge access proliferate.

In the middle of the diagram, at the e-Scholarly Communication Layer (e-SCL), we see a variety of these formal and informal modes of scholarly communication. The examples towards the left in the e-SCL box are generally newer arrivals on the SCL scene whereas those on the right represent more traditional forms of scholarly communication. Objects and actors in this layer are represented in white if they are primarily online, black if they are primarily offline, and gray if they are mixed. (This includes paper journals, which, even though they fall outside the 'e-' portion of the e-SCL, still have an influential role in e-SCL as a whole). The Internet has played a major role in enabling scientists to engage in

more widely disseminated forms of informal communication; some of these forms are enhancements of pre-Internet behaviours, others are novel forms of informal communication. Furthermore, the online elements have less direct lines connecting them. Instead of direct relationships between, for instance, publishers and libraries negotiating subscription terms, the online elements are tied together in a web of relationships that connects them to each other. These relationships can obviously be as simple as hyperlinks, but can also be as complex as the institutional demands of organizations like the American National Institutes of Health (NIH) to make funded research results available via freely accessible channels.

On the right-hand portion of the diagram is the process of discovery of research results. This is made up of the search engines (discussed above) which are one type of filter, along with elements like science websites and blogs geared to the general public, which filter access to scholarly material. Researchers may access materials directly, as indicated by the arrow directly from the e-SCL to academics, but are generally likely to access materials through one or more filtering mechanisms. Of course, the academics may also be those producing scholarship, as indicated by the arrow back to the distributed researchers, but they could equally be academics accessing knowledge that is outside the domain where they are the producers of knowledge. The barriers between fields and disciplines, while still strong in practice, are lowered for those who wish to engage with knowledge and data from outside their field. This is a serious challenge for those interested in knowledge management within organizations: the fluid paths to knowledge are difficult to manage, particularly as the Web itself is constantly evolving and shifting, as resources appear but also move and disappear over time.

Note that in the diagram, the organizational effects of being located to a greater extent in the online space (towards the top of the diagram) are not simple. As discussed earlier, the e-Research efforts in the top-left can either be much more organizationally structured than the traditional research at the bottom (as in the case of large e-Infrastructures) or they can be much less structured, in the case of Web 2.0 approaches to research. Similarly in the right-hand portion of the diagram, where the examples of traditional paths of discovery and communication with the public at the bottom are highly structured and dependent on organizations, whereas the paths to accessing information and knowledge at the top portion of the diagram are much more fluid, available through multiple paths, and generally more openly available. However, to ignore the fact that Google itself is a large organization, and has corporate policies that influence the information available to searchers, would be naïve. Indeed, while Google and other search engines in some ways make online information more transparently available, the actual workings of how it does this and how results are ordered are not at all transparent to the average academic or citizen. The search engine has been black-boxed, and the results that pop out of the box are taken as a given.

The wider public obtains their understanding of science both through researchers who engage in expanding the public understanding of science, but also increasingly through direct access to scientific information through filters (such as Google) which are widely available to those without university access to resources. This filtered access is a new phenomenon; other than the occasional enthusiast willing to go to an academic library and make photocopies of research articles, the general public in the pre-Internet age had very little access to scientific material outside of popular scientific publications such as science magazines and television programmes. This has also introduced a tension, particularly in some sensitive fields such as medical research where untrained readers may misinterpret scientific information and place their health at risk. Doctors have complained in anecdotal reports of increasing numbers of patients coming in to consultations armed with printouts of information from the Internet that the clinician considers to be of dubious quality or inaccurate. In this respect the role of scholars as the sole gatekeeper to the interpretation of scientific information has been weakened.

At this point we can put our model in a broader context: The processes described here have some affinity with the Socio-Technical Interaction Network (STIN) approach proposed by Kling *et al*. (2003) (see also Meyer, 2006), whereby there are pushes towards computerization by organizationally mobilized groups (in this case researchers). A key element here is that scholars and researchers will be increasingly aware of how their outputs will be affected by their visibility, which will in turn lead them to make these outputs more accessible. The online presence of these outputs, however, is not just shaped by their accessibility, but also by search engines and how these are used. Combining these two patterns produces a feedback loop, as well as highlighting a further element, which is where – in the e-Scholarly Communication Layer – these materials are located: as part of e-journal publication, or a repository (personal, institutional, or public), or some other forum or channel. The feedback loop thus operates via a new and expanded system with multiple parts – which makes the gatekeeping function of search engines and competition in the attention space more important, not less. Regardless of whether a winner-take-all *vs* democratizing (or concentration *vs* diversifying) effect takes place, there will therefore also be a novel type competition for online visibility which attaches to this new system, and shapes the kinds of knowledge that will gain prominence for researchers who are dependent on and use this system.

As we have argued in this essay, Google and other search engines will have an indirect effect on most areas of knowledge: firstly, a gatekeeper effect, whereby the

results found by the search engines like Google will achieve more prominence than others; secondly, it will play a role in the overall competition for visibility, as research increasingly needs to increase its visibility within a limited attention space; and thirdly, it will influence the extent to which manipulable digital materials are used in scholarship. All three trends are mutually reinforcing and therefore powerful, but the effect will be indirect since they largely shift existing offline processes into the online world, with the result that Google doesn't 'organize everything we know' (to borrow from the title of a journalist's book about Google), but is certainly a factor that shapes how knowledge is organized. To paraphrase Marx, researchers still make scientific knowledge, but they do not do so as they please. Instead, in the online world, they must instead increasingly pay attention to how knowledge is found.

Different disciplinary perspectives provide a limited handle on these processes (Meyer & Schroeder, 2009). The sociology of science and technology can tell us about the competition at the leading edge of research (Schroeder, 2008), and information science about workings of scholarly communication (Borgman, 2007). Yet we also need a larger context whereby the increasing use of search engines and the use of online materials among information seekers influences how the knowledge produced in society is becoming transformed both in terms of form and content. And although this is still a new system of online scholarly communication in-the-making, and it is therefore too early to talk about winners and losers or to quantify or assess the impact of these ongoing changes, it is possible to say that all disciplines, albeit in different ways, and the very conception of what is regarded as knowledge, will undergo a shift.

Anderson (2008) has gone so far as to claim that Google changes the very nature of science, such that it will be necessary to trawl through vast amounts of data with search engines and other techniques. This would call for a new theoretical approach to science and knowledge. What we have shown here instead is that, conceptualized properly, several mechanisms apart from Google operate, and while these do not make for a new world of science without theory, as Anderson claims, they do mean that we need to rethink science and knowledge in a frame that is broader than the sociology of science and technology and information science, and includes the whole way in which knowledge production is organized, and how this is changing in society.

Of course, examining the whole of this new interplay goes beyond the scope of this paper. Here we have focused on the part played by e-Research in an expanding e-Scholarly communication layer within the ecosystem of scholarly communities. Even in relation to e-Research, however, it is difficult to disentangle the shift toward distributed online tools and resources from larger ongoing shifts in the realm of search, accessing information on the Web, and the processes (competition for attention, online visibility and gatekeeping) that govern it. So that despite the proliferation of digital scholarly materials, in e-Research and beyond, we have argued that the online (or e-) scholarly communication layer will continue to be highly stratified because the search for materials within a limited attention space is likely to be dominated by a few resources which will, in turn, gain even more prominence because of the ways in which they have achieved this visibility.

As mentioned in the beginning of this paper, e-Research is a particularly pure version of professional communication as the definition is rooted in the online use of distributed and collaborative use of digital tools and data. Other domains which rely on professional communication, such as law, health and business, are also moving their communications increasingly online, either as part of a concerted effort or as part of an evolution of practices over time. While we have focused on e-Research in this paper, many of the general issues regarding attention and prominence also apply in these other domains. We encourage others to take up this task and contribute to an even broader understanding of online knowledge. In the realm of e-Research, we can conclude that academic research materials are moving online and thus partly now derive their visibility and prominence from a non-academic audience and non-academic tools such as commercial search engines. But this flow works both ways: how, for example, should we understand the use of references to Wikipedia in academic works, or the fact that many researchers have their own Wikipedia pages which highlight their publications and ideas? One possible reply here is that it has surely always been thus, that academics have gained prominence through the use of non-academic channels such as newspapers and television? This overlooks a critical difference, which is that, as we have shown, the electronic realm is to some extent a hermetic and feedback system: that is, online prominence reinforces itself, sometimes without reference to the offline world. This is just one example of how search and visibility can reinforce academic status in a self-referential loop, and the same goes for all online materials. What we have done in this essay is to point to a shifting ecology of online scholarly communication, a much expanded web of research in which, nevertheless, certain mechanisms operate to sift and determine what we know.

Acknowledgements

The work for this paper has been supported by the Economic and Social Research Council (ESRC) grant RES-149-25-1022 and is part of the Oxford e-Social Science project (http://www.oii.ox.ac.uk/microsites/oess/). The authors are grateful to the editors of this issue, Will Venters and Elaine Ferneley, and the anonymous reviewers for their helpful comments on this paper.

References

ANDERSON C (2006) *The Long Tail: How Endless Choice Is Creating Unlimited Demand*. Random House Business Books, London.

ANDERSON C. (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine,* 16 July. [WWW document] http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 27 June 2008).

ARMS WY and LARSON RL (2007) *The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship*. National Science Foundation and the Joint Information Systems Committee, Washington, DC.

ATKINS DE, DROEGEMEIER KK, FELDMAN SI, GARCIA-MOLINA H, KLEIN ML, MESSERSCHMITT DG, MESSINA P, OSTRIKER JP and WRIGHT MH (2003) *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation, Washington, DC.

BACHRACH S, BERRY RS, BLUME M, FOERSTER TV, FOWLER A, GINSPARG P, HELLER S, KESTNER N, ODLYZKO A, OKERSON A, WIGINGTON R and MOFFAT A (1998) Intellectual property: who should own scientific papers? *Science* **281(5382),** 1459–1460.

BARJAK F, LI X and THELWALL M (2007) Which factors explain the Web impact of scientists' personal homepages? *Journal of the American Society for Information Science and Technology* **58(2),** 200–211.

BERMAN FD and BRADY HE (2005) *Final Report NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences*. National Science Foundation, Washington, DC.

BERMAN FD, FOX G and HEY T (2003) *Grid Computing: Making the Global Infrastructure a Reality*. John Wiley & Sons, Ltd, Hoboken, NJ.

BORGMAN CL (2007) *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press, Cambridge, MA.

CALDAS A, SCHROEDER R, MESCH G and DUTTON WH (2008) Patterns of information search and access on the world wide web. *Journal of Computer-Mediated Communication* **13(4),** 769–793.

CENTRE FOR INFORMATION BEHAVIOUR AND THE EVALUATION OF RESEARCH (2008) Information behaviour of the researcher of the future. [WWW document] http://www.bl.uk/news/pdf/googlegen.pdf (accessed 16 January 2008).

COLLINS R (1998) *The Sociology of Philosophies: A Global Theory of Intellectual Change*. The Belknap Press of Harvard University Press, Cambridge, MA.

CUMMINGS JN and KIESLER S (2005) Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science* **35(5),** 703–722.

DE ROURE D and GOBLE C (2009) Software design for empowering scientists. *IEEE Software* **26(1),** 88–95.

DE SOLLA PRICE DJ (1986) *Little Science, Big Science … and Beyond*. Columbia University Press, New York.

FRY J (2006) Scholarly research and information practices: a domain analytic approach. *Information Processing and Management* **42(1),** 299–316.

FRY J and TALJA S (2004) The cultural shaping of scholarly communication: explaining e-journal use within and across academic fields. *Proceedings of the American Society for Information Science and Technology* **41(1),** 20–30.

FRY J, VIRKAR S and SCHROEDER R (2008) Search engines and expertise about global issues: well-defined territory or undomesticated wilderness? In *Web Search: Interdisciplinary Perspectives* (ZIMMER M and SPINK A, Eds), pp 255–276, Springer-Verlag, Berlin.

GARVEY WD and GRIFFITH BC (1967) Scientific communication as a social system. *Science* **157(3792),** 1011–1016.

GARVEY WD and GRIFFITH BC (1972) Communication and information processing within scientific disciplines: empirical findings for Psychology. *Information Storage and Retrieval* **8(3),** 123–136.

GINSPARG P. (1996) Winners and losers in the global research village. (Invited contribution for Conference held at UNESCO HQ, Paris, 19–23 February 1996, during session Scientist's View of Electronic Publishing and Issues Raised, Wednesday 21 February 1996). [WWW document] http://arxiv.org/blurb/pg96unesco.html (accessed 17 June 2001).

GINSPARG P (2006) As we may read. *The Journal of Neuroscience* **26(38),** 9606–9608.

GLÄSER J (2003) What Internet use does and does not change in scientific communities. *Science Studies* **16(1),** 38–51.

HAKKEN D (2003) *The Knowledge Landscapes of Cyberspace*. Routledge, New York.

HALLMARK J (2004) Access and retrieval of recent journal articles: a comparative study of chemists and geoscientists. *Issues in Science and Technology Librarianship* **40**, Article 1.

HARNAD S (2001) The self-archiving initiative. *Nature* **410(6832),** 1024–1025.

HEAD AJ. (2007) Beyond Google: How do students conduct academic research? *First Monday,* 12. [WWW document] http://firstmonday.org/issues/issue12_8/head/index.html (accessed 31 March 2008).

HEIMERIKS G and VASILEIADOU E (2008) Changes or transition? Analysing the use of ICTs in the sciences. *Social Science Information* **47(1),** 5–29.

HEMMINGER BM, LU D, VAUGHAN KTL and ADAMS SJ (2007) Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology* **58(14),** 2205–2225.

HEY T and TREFETHEN A (2003) The data deluge: an e-Science perspective. In *Grid Computing: Making the Global Infrastructure a Reality* (BERMAN F, FOX G and HEY T, Eds), pp 809–824, John Wiley & Sons, Ltd, Hoboken, NJ.

HINDS C and KIESLER SB (2002) *Distributed Work*. The MIT Press, Cambridge, MA.

HOUGHTON J and SHEEHAN P (2006) The economic impact of enhanced access to research findings. Victoria University Centre for Strategic Economic Studies Working Paper. [WWW document] http://www.cfses.com/documents/wp23.pdf (accessed 27 March 2008).

JASCO P (2005) Google scholar: the pros and the cons. *Online Information Review* **29(2),** 208–214.

KLING R and MCKIM G (2000) Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science* **51(14),** 1306–1320.

KLING R, MCKIM G and KING A (2003) A bit more to IT: scholarly communication forums as socio-technical interaction networks. *Journal of the American Society for Information Science and Technology* **54(1),** 46–67.

KOUSHA K and THELWALL M (2007) Google scholar citations and Google web/URL citations: a multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology* **58(7),** 1055–1065.

MATZAT U (2004) Academic communication and Internet Discussion Groups: transfer of information or creation of social contacts? *Social Networks* **26(3),** 221–255.

MEHO LI and YANG K (2007) Impact of data sources on citation counts and rankings of LIS faculty: web of science versus Scopus and Google scholar. *Journal of the American Society for Information Science and Technology* **58(13),** 2105–2125.

MERTON RK (1968) The matthew effect in science. *Science* **159(3810),** 56–63.

MEYER ET (2006) Socio-technical interaction networks: a discussion of the strengths, weaknesses and future of Kling's STIN model. In *IFIP International Federation for Information Processing, Volume 223, Social Informatics: An Information Society for All? In Remembrance of Rob Kling* (BERLEUR J, NUMINEM MI and IMPAGLIAZZO J, Eds), pp 37–48, Springer, Boston.

MEYER ET and DUTTON WH (2008) Top-down e-infrastructure meets bottom-up research innovation: fitting e-social science visions to the realities. [WWW Document] http://ssrn.com/abstract=1262211. In *Proceedings of the UK e-Science All Hands Meeting,* Edinburgh, UK.

MEYER ET and SCHROEDER R (2009) Untangling the web of e-research: towards a sociology of online knowledge. *Informetrics* **3(3),** 246–260.

NENTWICH M (2003) *Cyberscience: Research in the Age of the Internet*. Austrian Academy of Sciences Press, Vienna.

NICHOLAS D, HUNTINGTON P, JAMALI HR and WATKINSON A (2006) The information seeking behaviour of the users of digital scholarly journals. *Information Processing and Management* **42(5),** 1345–1365.

ORLIKOWSKI WJ and BARLEY SR (2001) Technology and institutions: what can research on information technology and research on organizations learn from each other? *MIS Quarterly* **25(2),** 145–165.

RIP A (1990) An exercise in foresight: the research system in transition. In *The Research System in Transition* (COZZENS SE, Ed), pp 387–401, Kluwer Academic Publishers, Dordrecht.
SANDSTROM P (2001) Scholarly communication as a socioecological system. *Scientometrics* **51(3),** 573–605.
SATHE NA, GRADY JL and GIUSE NB (2002) Print versus electronic journals: a preliminary investigation into the effect of journal format on research processes. *Journal of the Medical Library Association* **90(2),** 235–243.
SCHNEIDER SM and FOOT KA (2005) Web sphere analysis: an approach to studying online action. In *Virtual Methods: Issues in Social Research on the Internet* (HINE C, Ed), pp 157–170, Berg, Oxford.
SCHROEDER R (2007) e-Research infrastructures and open science: towards a new system of knowledge production? *Prometheus* **25(1),** 1–17.
SCHROEDER R (2008) e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. *Social Science Information* **47(2),** 131–157.
SCHROEDER R and FRY J (2007) Social science approaches to e-Science: framing an agenda. *Journal of Computer-Mediated Communication* **12(2)**, Article 11.
SELLEN A and HARPER RHR (2001) *The Myth of the Paperless Office*. The MIT Press, Cambridge, MA.
SPOELSTRA S, O'SHEA T and KAULINGFREKS R (2007) Marginal competencies. *Ephemera: Theory & Politics in Organization* **7(2),** 282–286.
TENOPIR C, KING DW, BOYCE P, GRAYSON M and PAULSON K-L (2005) Relying on electronic journals: reading patterns of astronomers. *Journal of the American Society for Information Science and Technology* **56(8),** 786–802.
TENOPIR C, KING DW and BUSH A (2004) Medical faculty's use of print and electronic journals: changes over time and in comparison with scientists. *Journal of the Medical Library Association* **92(2),** 233–241.
THELWALL M (2008) Bibliometrics to webmetrics. *Journal of Information Science* **34(4),** 1–18.
THELWALL M and KOUSHA K (2008) Online presentations as a source of scientific impact?: an analysis of powerpoint files citing academic journals. *Journal of the American Society for Information Science and Technology* **59(5),** 805–815.
WALSH JP, KUCKER S, MALONEY NG and GABBAY S (2000) Connecting minds: computer-mediated communication and scientific work. *Journal of the American Society for Information Science* **51(14),** 1295–1305.
WALSH JP and MALONEY NG (2007) Collaboration structure, communication media, and problems in scientific work teams. *Journal of Computer-Mediated Communication* **12(2),** Article 19.
WHITLEY R (2000) *The Intellectual and Social Organization of the Sciences* 2nd edn, Oxford University Press, Oxford.
WILKINSON D, HARRIES G, THELWALL M and PRICE L (2003) Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science* **29(1),** 49–56.
WUCHTY S, JONES BF and UZZI B (2007) The increasing dominance of teams in production of knowledge. *Science* **316(5827),** 1036–1039.

About the authors

Eric T. Meyer is a Research Fellow at the Oxford Internet Institute. Meyer studies the social implications of e-science and e-social science as part of the Oxford e-Social Science (OeSS) project. His publications include articles and chapters on social informatics, technology use by scientists, and online expressions of creativity. Meyer's Ph.D. in Information Science from Indiana University, which was named ProQuest Doctoral Dissertation of the Year, examined how marine biologists who rely on photographic evidence to identify individual marine mammals have seen significant changes in their everyday work practices as they switched from film photography to digital photography. His web page is http://people.oii.ox.ac.uk/meyer/.

Ralph Schroeder is senior research fellow at the Oxford Internet Institute at Oxford University. Before coming to the OII, he was Professor in the Department of Technology and Society at Chalmers University in Gothenburg. He is currently an investigator on the Oxford e-Social Science (OeSS) Project: Ethical, Legal and Institutional Dynamics of e-Sciences. His publications include books and essays on the sociology of science and technology, Max Weber, and virtual environments. His web page is http://people.oii.ox.ac.uk/schroeder/.