Scholarly communication, scholarly publication and the status of emerging formats

Leah Halliday
Department of Information Science
Loughborough University
UK

Abstract

A study was carried out to clarify the term 'scholarly publication' and to explore the role of this activity in scholarly communication. Desk research was supplemented by responses to a questionnaire from key figures in the development of emerging scholarly communication behaviour. This contributed to a working definition of scholarly publication consisting of a list of criteria, which may be used to analyse the degree to which emerging formats can be categorised as scholarly publications and to identify the means by which they may be supplemented so that their status may be promoted to that of 'scholarly publication', i.e., documents that meet all of the publication needs of scholarly communities.

Introduction

This paper is based on a study commissioned to scope the work of the scholarly communications group (SCG) of the Joint Information Systems Committee (JISC) of the UK Higher Education Funding Councils. The mission of the SCG is 'To make a leading contribution to the investigation and implementation of sustainable and cost-effective emerging behaviours across the various aspects of the scholarly communication process, on behalf of the UK educational and research communities, and in collaboration with relevant national and international partners'. The SCG supplied the terms of reference for the study. In essence, the purpose of the study was to outline the criteria for defining what is a 'scholarly publication' (see Appendix 1).

Methodology

The study consisted of desk research supplemented by the responses to a questionnaire (Appendix 2; hereafter referred to as the questionnaire) of library and publisher representatives and of researchers of scholarly communication. The questionnaire was distributed to all sixteen members of the SCG and to nine non-members on 31 January 2001. All recipients are actively involved in reviewing and developing scholarly communication practices and thus are well informed about the topic of the study. Those who are not members of the SCG were selected to contribute expert opinion from outside the library community and from the USA. The questionnaire was designed to be provocative, with a view to exploring respondents' opinions as well as gathering information. The schedule for the study was very tight and responses were requested within seven days of receipt of the questionnaire. Eight responses were received; six from members of the SCG and two from non-members. All respondents were UK-based. They included three librarians, one publisher representative, two academic researchers, and two higher education administrators. Where appropriate, their responses are identified as L1, L2, L3, P1, R1, R2, A1 and A2 respectively.

While the response rate to the questionnaire was very low, at 32%, it was felt that the persons involved were influential in the field of scholarly communication generally, played a significant role in the representation of this

subject to the Joint Information Systems Committee, and, therefore, presented views that could usefully illuminate the desk research. The data have been used in this way, rather than being presented as the results of a stand-alone survey."

Scholarly communication

The term 'scholarship' is defined, for the purposes of the UK research assessment exercise (RAE) as:

> the creation, development and maintenance of the intellectual infrastructure of subjects and disciplines, in forms such as dictionaries, scholarly editions, catalogues and contributions to major research databases (Higher Education Funding Councils, 2001).

Thus, the end products of scholarship are publicly available information objects. The term 'intellectual infrastructure' also implies formality and permanence.

Scholarship is not an activity undertaken in isolation. Borgman explained that 'Scholarship is inherently a social process… research in all fields is incomplete until it is validated through review processes and shared with others'. Thus, scholarship necessarily involves communication. The term scholarly communication refers not to an output but to an iterative process in which scholarship is communicated, used and developed within a community (Kling & McKim, 1999). Borgman defined and described scholarly communication as 'how scholars in any field… use and disseminate information through formal and informal channels' (Borgman, 1990: 13-14) but other writers commonly equate scholarly communication with a narrow segment of scholarly output, namely scholarly articles intended ultimately for publication in journals or similar formats (e.g., Alexander & Goodyear, 2000; Bowen, 1996; Buck, *et al.*, 1999; Graham, 2000; Odlyzko, 1999; Okerson, 1998; Rohe, 1998; Scarry, 1999; and Sosteric, 1998). This is explicable in that it generally occurs within the context of a discussion of the 'crisis' in scholarly communication which refers to library costs associated largely with increased subscription charges of scholarly journals.

For the purposes of this paper, the term 'scholarship' is used in a wider sense so the end products of scholarship would include but would not be limited to journal articles that report original empirical research, review papers and opinion papers, monographs, multi-contributor editions, software and databases. It is not intended that this definition be restricted to specific media. The rate and way in which emerging media are used for scholarly communication will vary across scholarly communities. Whether a particular application is appropriate is best judged by those communities (see, Covi, 2000; Hurd, 2000; Kling & McKim, 2000; and McCain, 2000).

Graham divided the scholarly communication process into three stages: communication within informal networks, which now occurs largely through the use of electronic media (see, Gabbay, 2000); the initial public dissemination of research through conferences and preprints; and finally, formal publication in 'prestigious journals' (Graham, 2000). Graham limited the potential output of scholarly communication to publication in prestigious journals. This paper is concerned with a wider range of outputs. Nevertheless, Graham's definition is useful for exploring the process of scholarly communication and considering the value of different types of information objects at different points on the continuum.

Scholarly publishing

Scholarly publishing is a means of communicating scholarship within a community. In the context of electronic media, the term 'scholarly publication' commonly refers to all forms of online distribution of documents (Kling & McKim, 1999; Borgman, 2000). This reflects the conceptual blurring of informal and formal modes of communication in an electronic environment referred to by Meadows in 1998. He defined informal modes of communication as, 'often ephemeral and… made available to a restricted audience only' and formal modes as, 'typically available over long periods of time to an extended audience'. He identified journals and books as 'archetypical examples of formal communications' (Meadows, 1998: ix). A contribution to an electronic discussion list with a substantial, international membership illustrates how electronic media blur this distinction. The message may be forwarded to numerous other forums and may be available in an archive of the discussion for a 'long period of time' yet few analysts would currently define it as a 'formal communication'. Nevertheless, a record of the informal discussion of an 'invisible college' may be deemed valuable to scholarship. The facility to capture that discussion raises the question of its status within the scholarly communication continuum and whether or not it is

sufficiently valuable to warrant investment that ensures its preservation and long-term accessibility.

The distinction between formal and informal communication formats is more difficult when a document is designed as a permanent record of research. Examples include conference proceedings, preprints, theses, reports, journal articles, and monographs. All but the latter two are categorised as 'grey' in a print environment because they can be difficult for users to identify and to access. The electronic environment makes these more accessible, thus moving them further along a scholarly publication continuum. Formats that are 'grey' in a print environment may, in an electronic environment, fulfil some of the functions of a scholarly publication more effectively than the archetypical formats to which Meadows referred. For example, it may be argued that an electronic preprint archive is a more effective vehicle for timely communication and for establishing chronological priority than a scholarly journal. Nevertheless, most preprints posted on **arXivX** (Los Alamos National Laboratory, 2001), the biggest and best-established preprint archive are also submitted for publication in peer-reviewed journals (Smith, 2001) indicating that an electronic preprint does not fulfil all of the publication needs of the author. The fact that readers in subjects well served by preprint archives are reluctant to forgo access to journals suggests also that preprints do not fulfil all needs of the reader (Mahé *et al.*, 2000; Smith, 2001).

Electronic preprints exemplify the use of ICT to distribute a format that existed previously in print as a draft of a scholarly journal article. New technologies however, are facilitating the development of modes of communicating scholarship and recording that communication that have no print analogue and cannot readily be accommodated by existing scholarly publishing systems (Lynch, 1999). Examples that already have a significant impact on scholarly communities are scientific databases of the size and scope of GenBank and the Protein Information Resource (McCain, 2000) and collaborative, evolving resources such as subject-specific web sites that combine access to primary source materials, analysis and commentary (Lynch, 1999). More complex examples are collaboratories described by Lynch as:

> Environments [that] variously combine videoconferencing, synchronous and asynchronous text-based messaging, shared control of scientific instrumentation, access to data (from observational sensors, simulations, or data archives), analysis and visualisation tools, shared whiteboards, literature databases, and authoring tools (Lynch, 1999).

Lynch explained that:

> These systems also document the research and knowledge creation process … it's important to note that events and activities that take place in these collaboratories can be stored, reviewed, replayed and annotated (Lynch, 1999).

These are examples of new formats for recording scholarly communication that present problems of description and organisation, certification and validation, acquisition, and preservation. For example, Lynch stated that:

> the materials involved are not only dynamic but also represent complex, interlinked and autonomously evolving works, which present major conceptual challenges to traditional descriptive practices (Lynch 1999).

Clearly, scholarly communication and scholarly publication in the 21st century cannot be understood exclusively with reference to or through analogy to print practices but all stakeholders require to distinguish between informal communication and formal communication or publication. This requires a flexible definition that can be applied to emerging formats.

Characteristics of a scholarly publication

The questionnaire distributed for this study asked respondents to select the necessary and sufficient criteria for a definition of scholarly publication. None of the suggested criteria was selected by all respondents. P1 refused to answer Question 1 and only R2 and L2 selected a single necessary criterion. These were that 'the relevant community is notified of [the document's] existence' (R2) and that 'it be refereed to ensure quality' (L2). Of the remaining five respondents, four or more considered it essential that the document be: durably recorded on some medium; reliably accessible and retrievable over time; and publicly available. Three or more considered it essential that there be a commitment not to withdraw it; and that the author intends that the work be made publicly available over the long term. Four or more of the remaining five respondents also considered it unnecessary that a scholarly

publication be immutable; accessible by readers at any reasonable time of day; be covered by at least one abstracting and indexing service; be equal to or longer than a minimum length; have publicly available metadata describing the work; and have guaranteed, long-term archival preservation. Three or more considered it unnecessary that the information object reports the result of original research. Although there was no consensus on the specific characteristics of a scholarly publication, the criteria that respondents to the questionnaire considered most important generally relate to the accessibility of a publication (see conclusion).

Functions of a scholarly publication

Two recent journal articles are useful for exploring the definition of 'scholarly publication'. The first is a paper produced by an international working group (IWG) of publishers and representatives from the higher education community that convened in 1999 on behalf of the International Association of STM Publishers to 'develop a position paper on how to define a scientific paper in the electronic era' (International Working Group, 2000). The second is an analytical approach to scholarly electronic publishing that explores that concept with reference to the three 'dimensions' of scholarly publication as a 'communicative practice within a community' (Kling & McKim, 1999). The first of these papers refers exclusively to scholarly articles or research papers but is useful also for exploring the definition of scholarly publishing in a wider sense. For the purposes of this paper, no assumption is made about the medium or format of scholarly publications.

The IWG definition of 'publication' was designed to 'assist the orderly development of scientific knowledge'. It acknowledged a continuum of scholarly communication formats in a networked environment and attempted to identify the most important 'fixed points' so that the status of information can be clarified. The IWG aimed to distinguish between informal communication and publication and outlined the following as criteria for qualification as the latter:

- It must be publicly available.
- The relevant community must be made aware of its existence.
- A system for long-term access and retrieval must be in place (e.g. Handle (Corporation for National Research Initiatives, 2001)).
- It must not be changed (technical protection and/or certification are desirable).
- It must not be removed (unless legally unavoidable).
- It must be unambiguously identified (e.g. by a Serial Item and Contribution Identifier (SICI) (University of California, Berkeley. *Sun SITE*, 2001) or Digital Object Identifier (International DOI Foundation, 2001).
- It must have a bibliographic record (metadata) containing certain minimal information.
- Archiving and long-term preservation must be provided for.

The IWG considered documents that fail to fulfil these criteria as 'not published' and thus not suitable for citation or as the basis of claims of priority or professional evaluation. It outlined in detail the criteria for identification as a 'first publication' and those for a 'definitive publication'. Essentially, the former becomes the latter when it has been evaluated and certified by the author's peers.

The IWG definition arises from what the group perceived to be the function of scientific publication.

> Publication is the hard currency of science. It is the primary yardstick for establishing priority of discovery, making the status of a publication a critical factor in resolving priority disputes or intellectual property claims. Academic tenure and promotion decisions are based in large part on publication in peer reviewed journals or scholarly books. To make these decisions fairly and with confidence, scientists and their institutions need assurances of what counts as a legitimate publication. The status of a published electronic document is critical in determining the trust that fellow scientists will have in it … scientists need to know the status of the information they encounter, whether they need to refer to it, critique it, or build on it to advance their own work. The document also needs to persist since in science identifying a clear context for later responses is essential to maintain the quality and integrity of subsequent scientific discourse. (International Working Group, 2000)

This definition emphasises the certified record and its role in career advancement at least as much as the communicative function of a document. Whilst acknowledging the other functions, Kling and McKim explored the concept of scholarly publishing as 'a communicative practice within a community' i.e. 'an activity engaged in by scholars who primarily want their reports to be widely read and credited by their target audience' (Kling & McKim,

1999). Kling and McKim's focus on community, authors and target audience emphasised use as much as production. They framed scholarly communication as a social process rather than a production process. They conceptualised 'effective publication' with reference to three 'dimensions' which serve both authors and readers: publicity, access, and trustworthiness.

Trustworthiness is determined primarily by peer review but also by the reputation of the publisher, by the reader's knowledge of the researcher (Kling & McKim. 1999) and the reputation of the researcher's institution, and, in the case of journal publication, the reputation of the journal title. Sometimes, the source of research funding is also an indication of trustworthiness. Research funded by, for example, a pharmaceutical company may be less credible than publicly-funded research (Borgman, 2000). In non-reviewed or weakly-reviewed outlets, the reputation of the author may be relatively more important (Kling & McKim, 1999). For example, it has been suggested that people using e-print archives tend to read only papers by authors whom they know by reputation (Lesk, 1997). This list of characteristics indicates that trustworthiness arises from a combination of institutionalised practices and the reader's personal knowledge (Kling & McKim, 1999). In the absence of the former, researchers must rely on the latter.

Publicity refers to the degree to which 'primary and secondary audiences are made aware of [the information object's] availability'. Access means that an information object 'can be readily located and obtained by interested scholars' in perpetuity. The minimum requirement for access is that it has a stable identifier (Kling & McKim, 1999). Also 'long-term, stable accessibility requires active stewardship, and is more reliable to the extent that stewardship is institutionalised' (Task Force on Archiving of Digital Information 1996 cited in Kling & McKim, 1999). Publication that may depend on goodwill and the voluntary contributions of individuals does not instil confidence in authors (see Gomes & Meadows, 1998).

Using this analysis as a framework, it is clear that some formats from the scholarly communication process are more effectively published in an electronic environment than in print. For example, most of the respondents to the questionnaire agreed that the term 'grey literature' refers to documents that are difficult to identify and access. In recent years, a variety of formats that, in print, were 'grey', for example, theses, preprints, research reports, and conference proceedings, have been made more accessible by being posted on the Internet. This reduces geographical barriers for those with access to the Internet but stable access over time is not necessarily guaranteed. This requires institutionalised stewardship. The trustworthiness of formerly 'grey' resources tends not to be improved by electronic availability. While theses and some research reports are subject to rigorous peer review, preprints, by their nature, have yet to be reviewed; the delivery medium does not change this. The degree of publicity for these materials varies. Preprint archives such as **arXivX** (Los Alamos National Laboratory, 2001) provide services for alerting users of all new submissions on a weekly basis. Databases of electronic theses and research reports, for example, the Networked Digital Library of Theses and Dissertations (2001; Fox, 1999), often do not provide current awareness services. Publicity may also be improved by intermediary services such as subject-based gateways that index the Internet. Thus, while access to these 'grey' documents is improved in the electronic environment, none fulfils Kling and McKim's three criteria for 'effective publication'.

Kling and McKim's framework suggests that, in a print environment, to communicate research results most effectively to readers within a scholarly community, an author should publish in one of Meadows's archetypical formal communication formats. While other formats on the continuum fulfil one or two of Kling and McKim's 'dimensions', printed scholarly journal articles and monographs fulfil all three. In an electronic environment, the status of scholarly journals and monographs need not change as long as they reach a substantial proportion of the targeted audience and there is an institutional commitment to preserve and archive these formats over the long term (Kling & McKim, 1999). Clearly, the electronic environment facilitates publication in a wider range of media. This is likely to result in greater variety of formal formats. This paper cannot predict what these formats will be but Kling and McKim's framework is useful for determining the degree to which a new or adapted format is a scholarly publication or formal communication. The remit of the IWG was significantly different from that of Kling and McKim but its definition of a publication observes Kling and McKim's dimensions.

Together these documents explore the functions of scholarly publications and identify the mechanisms by which these functions are fulfilled. In a print environment, these functions tend to have been fulfilled best by documents that are published and distributed by established publishers. In an electronic environment, this need not necessarily be the case. The framework facilitates analysis of the degree to which 'alternative' formats are 'scholarly publications' and thus is useful for identifying the means by which these formats may be supplemented to promote their status to that of 'scholarly publications', that is, documents that meet all of the publication needs of scholarly communities.

A continuum of publishing formats

The characteristics of formats that, in a print environment, were grey, and the degree and manner to which these conform to a definition of scholarly publication, vary. What they still have in common is the fact that, regardless of their relative status on a continuum of scholarly publications, these formats are all frequently the basis of formal publications; journal articles and monographs. Thus, initiatives that make grey literature more accessible duplicate the efforts of scholarly publishers to varying degrees. The two formats overlap on the continuum of scholarly publications.

Publication in a second format is useful only to the degree that this fulfils functions not fulfilled by the first. For example, electronic preprint publication may be most effective for timely communication and for establishing chronological priority while publication in a peer reviewed journal is required for certification (Harnad, 1999a, 1999b; Smith, 2001). Where the extent and structure of formats is similar, for example, as for preprints and journal articles, or for theses and monographs, it may be more efficient to supplement the functions of the service delivering the informal format so that it fulfils Kling and McKim's criteria for 'effective publication' than to duplicate publication formats. A1, A2, and R2 (above) considered this possible with reference to eprint archives.

> A1: The absence of scholarly journals need not equate with the absence of peer reviewing processes. Similarly, the organisation of e-print archives could readily replace the function of journal titles in guiding academics to relevant bodies of work. Indeed, they may offer distinct advantages in assisting academic inquiry though the opportunity to develop more effective search tools.

> A2: A well-designed, authoritatively run, e-print archive with peer review and indexed by appropriate secondary sources could probably replace and expand the communication function of scholarly journals by widening the range of potential readers (third world, small institutions, researchers for whose institutions the journal would not be mainstream etc.).

A2's response implies that an e-print archive would necessarily be available to institutions that cannot currently afford to subscribe to related journals. Clearly, this depends on the economic model on which the development, maintenance and delivery of the e-print archive is based.

While not all functions related to the publication of a scholarly format need be the responsibility of a single agency, it would be inefficient to disaggregate and distribute functions in a manner that results in unnecessary duplication of effort. In instances where the format of an information object is significantly altered for subsequent publication, for example, where the content of a thesis is the basis of a journal article, both documents may be useful. A thesis is much longer and more detailed than the journal article. While most readers prefer the content distilled to article length, a subset of them is often interested in the detail.

The IWG considered it essential that a reader of one format be aware of the existence of another. Questionnaire respondents agreed that this is important but did not consider it essential. It can be achieved most effectively through creation, maintenance and dissemination of accurate and comprehensive metadata (International Working Group, 2000). When versions are created and published by different bodies, for example, universities and publishers, responsibility for maintaining accurate metadata about all versions is not clear. The IWG considered it an institutional responsibility rather than the responsibility of individual authors. Generally, questionnaire respondents considered this the responsibility of the publisher but in the case of pre-prints, two of them (L1 and L2) said that the author should be responsible. R2 considers the responsibility to lie with 'The owner or maintainer of the origin or source page of any version of the data'. These responses imply distributed metadata about different versions rather than a single file detailing all versions. R1 recognised this and suggested that accurate and useful metadata will be rare in future and L3 suggested that this may be a role for the information professions.

> R1: This is where the publisher can really add value. The publisher could have as part of their value-adding process, the addition of consistent metadata. If publishers no longer are active or cannot invest in this sort of thing, it ought to be the author, but whether authors will do it reliably is a very moot point. I think we're going to have some pretty crude searching going on. I think good metadata will exist for a minority of items. For the majority, you'll just have to do some pretty crude, free-text searching.

> L3: I cannot see authors producing standardised metadata, except in a closed community. This may be a

task for 'library-type' staff or for indexers such as usually work for abstract and indexing services (though this would have cost implications).

Networked communication and publishing formats

The networked environment has the potential to change the structure and format of scholarly publication, for example, by facilitating development of documents through more dynamic interaction between researchers and through interlinking. Several respondents referred to interaction with reference to documents that are made public initially and revised in response to feedback.

> R2: If done in a timely manner, that is, as soon as possible after the research has been done, [the facility to append comments to a scholarly work] has very profound implications to improve the quality of dialogue and thus of research. It is important to note that it is this possibility of timeliness that's new, and by association the degree of informality in feedback. The environment to stimulate immediate feedback and debate on formal papers needs to be crafted very carefully. It is not enough to be electronic.

> L1: This potentially makes publications more dynamic… the whole concept of a stable publication in the way that we've known it in a print environment could, potentially, be undermined… I think the number of different versions of a publication is the biggest implication here.

> A1: Intention by the author to disseminate the work in a durable form is certainly highly desirable and should be regarded as best practice in scholarly communication. However, it would be unfortunate if this were to constrain more interactive or developmental forms of publication where work is published in order to attract comment and discussion with a view to a future definitive version. The two are not necessarily incompatible provided that the purpose of the publication is clearly articulated and that that purpose is retained along with the item itself.

Networked communication also facilitates collaboration and discussion on a much broader scale than was possible in a print environment This presents the potential for 'community-authored' information objects whose status and function must be accommodated within scholarly communication systems.

The effect of networked communication on the definition of a scholarly journal

It has been suggested that the journal format is a consequence of the technology used, that is, printing (Barry & Richardson, 1997). This suggestion would be contested by publishers who consider the format to be functionally important, that is,. they consider a journal title to be a 'brand' that carries a number of messages about the subject and value of the content published therein (Bide, 1997; Morris, 1998). Question 3 in the questionnaire proposed the following definition of a scholarly journal and invited respondents to comment:

> A scholarly journal functions to communicate, validate and certify, to establish priority, to contextualise research within a specific field through citation of earlier work, and as a formal record of research results. It also filters the best and most relevant content in a specific field.

To this definition, P1 added:

> (a) branding – collecting together articles in a recognisable 'envelope' which conveys signals of both relevance and quality, and (b) quantity control – obvious in print but also important electronically. The journal (in most cases – not quite all!) selects a quantity of material which the reader can actually cope with.

The suggestion that journal titles are useful for controlling quantity makes a virtue of the need to control the number of pages in print. A print journal receiving a volume of submissions that it cannot accommodate increases the rejection rate by revising editorial policy. Thus, high quality content that may otherwise be accepted is rejected. This is not necessarily a positive contribution to scholarly publication. While readers need effective tools for navigating and filtering published content, all high-quality research deserves effective publication. R2's response to question 3 in the questionnaire is apposite here. S/he said:

The primary function of a journal is to package and market a body of work. You can imbue it with the above characteristics to varying degrees depending on the nature of the journal you want to market.

Scholarly journals as brands

The value to authors and readers of specific titles as brands is contentious and may be greater in some fields than in others. One of the SPARC-supported journals has demonstrated that the principle value of a specific title is related to the editor and editorial board rather than the brand *per se.* In 1998, the editor of a Kluwer title, *Evolutionary Ecology,* Michael Rosenzweig, left that journal due to 'indefensible price rises' and launched a competitor title *Evolutionary Ecology Research* with SPARC support. His editorial board joined him and 90% of authors whose papers awaited publication in the Kluwer journal withdrew them and submitted them instead to *Evolutionary Ecology Research* which broke even within the first year (Rosenzweig, 1999).

While many publishers insist that journals as brands are important, the electronic publishing strategies of the largest commercial scholarly publishers appear to undermine individual journal titles by bundling them together and offering libraries a service rather than a collection of discrete products. This is portrayed as added value and apparently offsets librarians' demands for price reductions. The value of the additional content received in a large bundle is unclear and bundling may even support weak titles that otherwise are not viable. Clearly, publisher-added value is worth paying for only if it valuable to the research community. R1 expressed this in her/his response to question 4:

> I think publishers should be focusing on adding value and thus being able to justify a price that is acceptable. It is added value in terms of the perceptions of the readers. Unless they can do that the implications are that publishers may get squeezed out.

The basic functionality of the web also threatens brand awareness. One way in which authors of scholarly publications contextualise their work within a field is through citation. In an electronic environment, citations can be hyperlinked to the cited papers. The culmination of this linking facility was described recently as 'the holy grail for content providers and users alike… a seamless, integrated, transparent network that allows searchers to link quickly and painlessly to any document they seek' (Grogg & Tenopir, 2000). In a seamlessly integrated, hyperlinked environment, the visibility and, thus, value attached to specific titles is likely to be reduced as readers link from paper to paper without having to first identify the journal in which a cited paper is published.

Learning scholarly communication

Scholarly communication practices are learned within a community (Covi, 2000; Kling & McKim, 2000). This is important because communities of scholars have developed different ways of resolving issues of legitimacy and trust in traditional media. These concerns are protected when the media are computerised. Value-adding features may be appended but the basic legitimising features remain intact (Kling & McKim, 2000). For example, the development of a thriving pre-print culture in high-energy physics is related to a variety of characteristics of that community, characteristics that are absent in other disciplines such as chemistry (Till, 2001). Thus, attempts to generalise communication practices across all disciplines in an electronic environment may be at best unhelpful and may waste resources spent on attempting to establish new practices in inappropriate disciplines. Questionnaire respondents were divided on this issue. Half of them agreed that scholarly communication should not be treated as a homogeneous practice as behaviour varies between communities or disciplines. In contrast, R2's response accords roughly with the view that all disciplines can benefit by adopting free, open access to pre-prints. A1, A2 and L3 also identified benefits of considering issues of 'scholarly communication' as a whole.

> A1: Provided variations are remembered, there may, indeed, be benefits in identifying and sharing good practice…Probably, the old 'just because you can do something doesn't mean that you should' applies here. Nevertheless, heterogeneity is sometimes used by academics (by no means exclusively) as a cloak for unwillingness or unfamiliarity. Remember the Luddites!

> A2: It must be possible and should be the goal to develop systems which have inbuilt flexibility to allow all disciplines to archive their material in a way suitable to their traditions and way of working, but compatible and interoperable with those f other disciplines – diversity within a common framework.

L3: I think it is true to a degree that, at present, there is a tendency to see these as a single kind of product, even if some will be discipline-based and others will be institutionally-based. At the same time (a) there is some similarity across many disciplines and a worry about technological determinism should not hold back development, and (b) is the alternative worse? There may be a danger of the paralysis of analysis in this question. As someone said in another context: 'follow the light you have, and pray for more light'. I think we can move ahead in this way and at the same time look for other solutions in areas where this approach may not be appropriate.

The suitability of any format or practice to a particular community requires careful consideration.

The primary stakeholders in scholarly publication are authors and readers. To a significant degree, these are drawn from the same community. Neither group is routinely aware of the system-wide economic structure of scholarly publishing. Many researchers as authors support an economic structure that effectively deprives them, as readers, of access to all of the research papers they need. Furthermore, as authors, they fuel publication of a volume of content that, as readers, they have difficulty managing (Meadows, 1998). Authors represent the most significant source of demand for publications. Attempts to address the system-wide problem must include authors.

The point at which scholars learn scholarly communication practices is the point at which alternative behaviours might best be inculcated. For example, when Virginia Tech imposed mandatory submission of theses and dissertations in electronic form, they used that opportunity to teach postgraduate students about the scholarly communication system and to equip them with the skills to create scholarly documents in standardised, electronic form (Fox, 1999). Initiatives directed at postgraduate students have the potential routinely to inform a new generation of scholars about the scholarly communication system, the assignment of copyright and related issues and to train them in the production and self-publication of electronic scholarly documents as they learn scholarly communication practices.

Conclusion

This paper aimed to provide a working definition of 'scholarly publication'. The framework provided by Kling and McKim supplemented by the IWG definition and the criteria of questionnaire respondents was used to generate the following definition which divides criteria into those that are: essential (E); Highly desirable (HD); and preferable (P). Thus, a scholarly publication requires:

- Trustworthiness

    - Publications should not be changed (**HD**).
    - Different versions should be clearly identified (**HD**).
    - To satisfy all potential interest, trustworthiness should be based on 'institutionalised' measures such as peer review rather than on personal knowledge (**HD**).
    - Each publication should have at least one identifiable author (**P**).

- Publicity

    - The potential audience must be made aware that the publication exists (**HD**).
    - The publication should have metadata containing a minimum set of information, preferably including information about all versions (**P**).

- Accessibility: the document must be readily obtained by those who wish to use it.

    - The author must intend that the publication be made publicly available in a durable form over the long term (**E**).
    - The publication must be durably recorded on some medium (**E**).
    - The publication must be reliably accessible and retrievable over time (**E**)
    - There should be a commitment not to withdraw the publication (**E**).
    - The publication must be publicly available, i.e. available to any member of the public on demand as of right, whether for payment of a fee or not. (**E**).
    - The publication should have stable identifiers (**HD**).

The establishment of emerging behaviour as practices that fulfil the functions of scholarly publication requires that scholars understand how each element in this definition contributes to effective publication. This, in turn, requires understanding of the whole publication chain. If research communities and those that support them are to maximise investment in the chain, education on this issue should be routinely integrated with education on research methods and practices.

References

- Alexander, A. and Goodyear, M. (2000) The development of BioOne: changing the role of research libraries in scholarly communication. *Journal of Electronic Publishing,* 5(3) [www.press.umich.edu/jep/05-03/alexander.html Accessed 23 February 2001].
- Barry, T. and Richardson, J. (1997) Death of the journal – will it be replaced by document delivery? Paper presented to the *CAUSE in Australasia Conference '97*, Melbourne, April 1997 [www.bond.edu.au/Library/jpr/cause/cause97.html Accessed 14th February 2001].
- Bide, M. (1997) The primary scientific journal. In: *From N to X: the impact of online networks on the publishing value chain.*. Northwood, Middlesex: VISTA International.
- Borgman, C. L. (ed.) (1990) *Scholarly communication and bibliometrics,* Newbury Park, CA: Sage Publications, pp.13–14.
- Borgman, C. L. (2000) Digital libraries and the continuum of scholarly communication, *Journal of Documentation,* **56** (4), 412–430.
- Bowen, W. G. (1996) JSTOR and the economics of scholarly communication. In: *The economics of information in the networked environment. Proceedings of the Conference challenging the marketplace solutions to problems in the economics of information,* edited by M. A. Butler and B. R. Kingman), Washington DC, 18–19 September 1995. Washington DC: Association of Research Libraries, pp. 23–34.
- Buck, A. M., Flagan, R. C., and Coles, B. (1999) Scholars' forum: a new model for scholarly communication, [http://library.caltech.edu/publications/scholarsForum/ Accessed 23 February 2001]
- Corporation for National Research Initiatives (2001). *Handle System.* Reston, VA: Corporation for National Research. [http://www.handle.net/ Accessed 13 July 2001]
- Covi, L. (2000) Debunking the myth of the Nintendo generation: how doctoral students introduce new electronic communication practices into university research, *Journal of the American Society for Information Science,* **51** (14), 1284–1294.
- Fox, E. A. (1999) Networked library of theses and dissertations. *Nature Web Matters,* [http://www.nature.com/nature/webmatters/library/library.html Accessed 9 February 2001]
- Gabbay, S. (2000) Connecting minds: computer-mediated communication in scientific work, *Journal of the American Society for Information Science,* **51** (14), 1295–1305.
- Gomes, S. and Meadows, J. (1998) Perceptions of electronic journals in British Universities, *Scholarly Publishing,* **29** (3), 174–181.
- Graham, T. (2000) Scholarly communication, *Serials,* **13** (1), 3–11.
- Grogg, J. E. and Tenopir, C. (2000) Linking to full text in scholarly journals: here a link, there a link, everywhere a link. *Searcher,* **8** (10) [www.infotoday.com/searcher/nov00/grogg&tenopir.htm Accessed 14 February 2001]
- Harnad, S. (1999a) Online journals and financial firewalls. *Nature,* **395**, 127–128.
- Harnad, S. (1999b) Free at last: the future of peer-reviewed journals. *D-Lib Magazine*, **5** (12) December [http://www.dlib.org/dlib/december99/12harnad.html Accessed 12 July 2001]
- Higher Education Funding Councils (2000). *RAE 2001, Annex A: Definition of Research* Bristol: Higher Education Funding Councils. [www.rae.ac.uk/PanGuide/guideann.htm Accessed 22 February 2001]
- Hurd, J. M. (2000) The transformation of scientific communication: a model for 2020. *Journal of the American Society for Information Science,* **51** (14), 1279–1283.
- International DOI Foundation (2001). *Digital Object Identifier System* Reston, VA: International DOI Foundation. [http://www.doi.org/ Accessed 13 July 2001]
- International Working Group. (2000) Defining and certifying electronic publication in Science, *Learned Publishing,* **13** (4), 251–257.
- Kling, R. and McKim, G. (1999) Scholarly communication and the continuum of electronic publishing, *Journal of the American Society for Information Science,* **50** (10), 890–906.
- 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**

(4), 1306–1320.

- Los Alamos National Laboratory (2001) arXiv.org e-Print archive. Los Alamos, NM: Los Alamos National Laboratory. [http://www.arXiv.org Accessed 20 February 2001]
- Lynch, C. (1999) On the threshold of discontinuity: the new genres of scholarly communication and the role of the research library, Paper given at ACRL National Conference, Detroit, 9 April 1999. Chicago, IL: Association of College and Research Libraries. [http://www.ala.org/acrl/clynch.html Accessed 9 March 2001]
- McCain, K. W. (2000) Sharing digitised research-related infomration on the world wide web, *Journal of the American Society for Information Science,* **51** (14), 1321–1327.
- Mahé, A., Andrys, C. and Chartron, G. (2000) How French research scientists are making use of electronic journals: a case study conducted at Pierre et Marie Curie University and Denis Diderot University, *Journal of Information Science,* **26** (5), 291–302.
- Meadows, A. J. (1998) *Communicating research,* San Diego, CA: Academic Press.
- Morris, S. (1998) Learned journals the communication of research, *Learned Publishing,* 11 (4), 253–257. [Available from Catchword at http://www.catchword.com/alpsp/09531513/v11n4/contp1-1.htm].
- National Digital Library of Theses and Dissertations (2001) Information about the initiative. Blacksburg, VA: Virginia Polytechnic Institute and State University [http://www.ndltd.org/info/index.htm Accessed 10 February 2001]
- Odlyzko, A. (1998) The economics of electronic publishing. *Journal of Electronic Publishing,* **4** [http://www.press.umich.edu/jep/04-01/odlyzko.html Accessed 6 January 1999]
- Okerson, A. (1998) Scholarly communication and the licensing of electronic publications. In: *The impact of electronic publishing on the academic community,* edited by I. Butterworth. London: Portland Press. pp. 110–117.
- Rohe, T. A. (1998) How does electronic publishing affect the scholarly communication process, *Journal of Electronic Publishing,* **3** (3), [www.press.umich.edu/jep/03-03/rohe.html Accessed 22 February 2001]
- Rozenzweig, M. L. (1999) Reclaiming what we own: expanding competition in scholarly publishing. Plenary address to the ninth conference of the association of college and research libraries, Detroit, Michigan, 11 April 1999, Chicago, IL: Association of College and Research Libraries. www.arl.org/sparc/rosenzweig.html Accessed 15 February 2001]
- Scarry, P. (1991) Scholarly communication at work: a dialogue aimed at understanding how shcolars and studetns use information and related services and how their needs are evolving as technology advances. *Newsletter on Serial Pricing Issues,* NS12, [http://www.lib.unc.edu/prices/1991/PRICNS12.HTML#NS12.2 Accessed 26 October 1999]
- Smith, A. E. (2001) The journal as an overlay on pre-print databases, *Learned Publishing,* **14** (1), 43–48. [Available from Catchword at http://www.catchword.com/alpsp/09531513/v14n1/contp1-1.htm].
- Sosteric, M. (1998) At the speed of thought: pursuing non-commercial alternatives to scholarly communication. *ARL: a Bimonthly Newsletter for Research Library Issues and Actions*, October, 17–19.
- Till, J. E. (2001) Predecessors of pre-print servers, *Learned Publishing,* **14** (1), 7–14. [Available from Catchword at http://www.catchword.com/alpsp/09531513/v14n1/contp1-1.htm].
- University of California, Berkeley. *Sun SITE* (2001). *SICI: Serial Item and Contribution Identifier standard. ANSI/NISO Z39.56-1996 Version 2.* Berkeley, CA: University of California, Berkeley. Sun SITE. [http://sunsite.berkeley.edu/SICI/ Accessed 13 July 2001]

Appendix 1: Terms of reference

The JISC is seeking to commission a scoping study to clarify what should be encompassed in the term "scholarly communications" for the purposes of its Scholarly Communications Group. The remit of this group is to take forward relevant issues relating to electronic journals, e-prints, archiving, IPR, assessment and new publishing paradigms. The Group hopes to consider ways to support emerging behaviours, for example by ensuring greater accessibility to information and by making information available in different and new ways.

The discussion paper produced will define the meaning of scholarly communication, by looking at various criteria and definitions of how to describe a scholarly publication including:

- the existence of formal records
- availability in the public domain

- peer recognition
- intention of the scholar to publish
- costs/potential costs.

The study will consider the distinction between scholarly and publication and consider some of the issues. It is hoped that the study will map the key areas for the initial focus of the Scholarly Communications Group where the work of the Group could help to make an impact on the future process.

Appendix 2: Questions on scholarly communication and scholarly publication for selected stakeholders

Defining scholarly publication

1a. Which of the following are **NECESSARY** to define something as a scholarly publication? (Please indicate priority by numbering those that you include in your definition)

- That it reliorts the result of original research
- That it has at least one identifiable author.
- That it be durably recorded on some medium.
- That it be reliably accessible and retrievable over time.
- That it is immutable.
- That it is liublicly available.
- That it can be accessed by readers at any reasonable time on any day.
- That it is certified as authentic and cannot be changed after liublication.
- That it has a liersistent and unambiguous identifier.
- That metadata describing the work are liublicly available.
- That it is covered by at least one abstracting and indexing service.
- That there is a commitment to liroviding continuing access and retrievability.
- That the relevant community is notified of its existence and availability.
- That there is a commitment not to withdraw the document.
- That the metadata refer to all versions of the document and identify the definitive version.
- That it be refereed to ensure quality.
- That the certified version cannot be altered.
- That long-term archival lireservation is guaranteed.
- That it be internally consistent and in a standard style and format (e.g. that references are standard and correct to facilitate hylierlinking).
- That the audience is not necessarily known or identifiable to the author and/or liublisher.
- That the author intends that the work be made liublicly available in a durable form over the long term.
- That it is equal to or longer than a minimum length (if so, how long should it be?).

1b. Which of the following are **SUFFICIENT** to define something as a scholarly publication? (Please indicate priority by numbering those that you include in your definition)

- That it reliorts the result of original research
- That it has at least one identifiable author.
- That it be durably recorded on some medium.
- That it be reliably accessible and retrievable over time.
- That it is immutable.
- That it is liublicly available.
- That it can be accessed by readers at any reasonable time on any day.
- That it is certified as authentic and cannot be changed after liublication.
- That it has a liersistent and unambiguous identifier.
- That metadata describing the work are liublicly available.
- That it is covered by at least one abstracting and indexing service.

- That there is a commitment to liroviding continuing access and retrievability.
- That the relevant community is notified of its existence and availability.
- That there is a commitment not to withdraw the document.
- That the metadata refer to all versions of the document and identify the definitive version.
- That it be refereed to ensure quality.
- That the certified version cannot be altered.
- That long-term archival lireservation is guaranteed.
- That it be internally consistent and in a standard style and format (e.g. that references are standard and correct to facilitate hylierlinking).
- That the audience is not necessarily known or identifiable to the author and/or liublisher.
- That the author intends that the work be made liublicly available in a durable form over the long term.
- That it is equal to or longer than a minimum length (if so, how long should it be?).

1c. Please comment on the last two questions.

2. Is it possible to preserve and archive the web of hyperlinks in scholarly publications? What are the implications of this for your definition of scholarly publication?

Functions of a scholarly journal

3. A scholarly journal functions to communicate, validate and certify, to establish priority, to contextualise research within a specific field through citation of earlier work, and as a formal record of research results. It also filters the best and most relevant content in a specific field. Please comment on this description.

Peer review

4. A combination of initiatives such as eprint archives and NEAR (National Electronic Archive Repository) threaten to reduce the publisher role to that of co-ordinating peer review. Given that price of journals is based on perceived value rather than cost, this may not reduce the price to the scholarly community. If the price that the scholarly community is prepared to pay for peer review is significantly lower than the price currently paid for a journal, co-ordination of peer review may not be commercially viable for publishers. What are the implications for scholarly publishing?

5. Peer review is based on quality rather than popularity at the time of publication. Papers often remain in the archive unused for years before they are discovered and contribute considerably to scholarship. Publication based on this selection process is incompatible with libraries' desire to purchase only those articles that their users need. Please comment.

A continuum of scholarly publishing formats

6. It has been suggested that in the absence of a filtering mechanisms such as peer review and a journal title, academics using eprint archives read only papers of authors that they know. Can eprint archives replace the communication function of scholarly journals in a research community?

7. Is it appropriate that peer reviewed journal articles cite preprints? Why?

8. Is it necessary, when accessing the certified, definitive version of a scholarly publication, to know that other versions exist and/or to know the form in which it was first made available to the public?

9. In a print environment grey literature is difficult to identify and difficult to access. In an electronic environment research reports, theses and preprints are freely available on the internet and are often indexed by subject-specific portals or by libraries. How grey are these electronic publications?

10a. In an electronic environment readers may append comments and/or criticisms to scholarly works. What are the implications of this for scholarly publication?

10b. Are these comments and/or criticisms scholarly works, are they elements of the work to which they are appended or neither?

11. Is it possible/desirable to retract work that has been published?

If so, what type of publications should be retractable (e.g. pre-print, conference proceedings, peer-reviewed journal article).

12. What are the implications for scholarly communication of making work available free of charge to a global audience (and potentially to mass media) before it has been validated and certified (e.g. as preprints, conference proceedings)?

13. Browsing in a print environment facilitates serendipitous discovery which may facilitate creativity. Is this facility lost in an electronic environment and, if so, what are the implications for scholarly communication?

Metadata

14. Who should be responsible for ensuring that accurate and standardised metadata about a publication are publicly available and are altered e.g. when errata are published? (Please refer to different types of publication, e.g. preprints, conference proceedings, grey literature, journal articles).

Heterogeneous practices in scholarly communication

15. Scholarly communication practices are heterogeneous across disciplines, e.g. some disciplines have a well-developed pre-print culture while others do not, some disciplines rely more heavily than others on conference proceedings to communicate results. What are the implications for scholarly publishing? Is it useful to discuss scholarly communication and/or scholarly publication as if these are homogenous activities?

16. Initiatives that aim to encourage the development of eprint archives across all STM disciplines reflect technological determinism and fail to recognise heterogeneity across scholarly communities. Please comment.

Publishing context

17. In a hyperlinked environment, it is easy to repackage and reposition content by collecting bundles of links. What are the implications of this for authors?

18. The electronic environment facilitates analysis of use of individual documents so that most frequently used documents can be readily identified. It also facilitates collaborative filtering. What are the implications of this for scholarly communication?

19. Please comment further on any issue raised or on any issue not raised that you consider relevant.