

Learned Publishing (2001)14, 265–272

Introduction

Authors, editors, publishers, librarians and, most importantly, readers – all are witnessing a change from paper to electronic media for the storage, dissemination and handling of scientific articles. As in early-modern times when print on paper was invented in Europe, completely new vistas of inter-human communication are opening up. It is worthwhile unplugging your computer for some days to reread Elizabeth Eisenstein's masterpiece *The Printing Press as an Agent of Change*,¹ or – if, due to heavy computer use, you have completely forgotten how to read plain text – her much shorter illustrated *The Printing Revolution in Early Modern Europe*.² As vividly described there, the change in technology deeply influenced our communication systems and hence our culture. The fascinating issue now is: to what extent will the electronic revolution play a similar role in changing our way of handling knowledge and information?

Although the pace of change seems enormous, we are still only at the beginning of developments, as illustrated by the rapid overtaking of one computer system or program by another. The situation has not yet stabilized and when we realize that in research environments people 'play' with devices with 64 gigabytes of RAM, it is clear that the future is not easy to predict. Despite the hype and the power of the technology, pressing us to respond to more and more so-called market demands, trumpeted by the computer industry, it is essential to take a more Olympian (in modern parlance 'bird's eye') view, using analysis from within the communication processes to identify the really novel developments and to understand how scientists and the scientific publishing community can prepare themselves for the future. It goes without saying that most of this is still a matter for research and that many an experiment has yet to be

New practices for electronic publishing

1: Will the scientific paper keep its form?

Joost G. Kircz

*KRA-Publishing Research and Van der
Waals-Zeeman Instituut, Universiteit van
Amsterdam*

© Joost G. Kircz 2001

*ABSTRACT: Discussion about the value of electronic documents is often hampered by starting from what is usual in the paper world and attempting to impose that on an electronic environment. In order to grasp the impact of the current electronic revolution, and formulate a policy for the future, we examine the aims and content of scientific communication. We then critically discuss the recommendations of an International Working Group [see *Learned Publishing* 2000:13(4) Oct. 251-8], and show the tension between these very reasonable recommendations and the reality of electronic publishing. We conclude that the scientific article will change considerably but that, in its new more composite form as an ensemble of various textual and non-textual components, it will retain many of the current cultural and scientific requirements with regard to editorial, quality and integrity.*

performed before we reach equilibrium again.³

At present, most electronic publications are simply paper products transposed onto electronic media. Neither the structure, nor the way language is used, is significantly different from earlier practice.

Nevertheless, we witness a sometimes heated debate on the value of such 'electronic documents', especially in the context of peer review. In my view, we have to differentiate between documents that look, smell and sound like paper documents but are stored and transmitted by electronic means, and documents that have been originally created for an electronic environment, and hence which are new animals in the zoo of scientific communications.

Discussion about the value of electronic documents is often hampered by the fact that it starts from what is usual in the paper world and attempts to impose that on an electronic environment. In the same way as early printed texts closely resembled the old manuscripts, so the scientific paper as we know it is a paper-based object that obviously can be cast into various technical forms, but intrinsically remains a paper object.

In order to grasp the impact of the current electronic revolution, and formulate a policy for the future, we need to examine the aims and content of scientific communication before we focus on a particular presentation medium. We need to step back and analyse what it means to write for an electronic medium and what it means to read material that is stored electronically. In a paper world, writing and reading are very close. Reading electronic articles, however, does not mean reading from a screen only. The presentation becomes flexible! In contrast to paper, electronic media allow distinct differences to exist between the author's favoured presentation and the consumer's reading practice.

An electronic document is not the electronic version of a traditional paper document with embellishments such as hyperlinks, colour pictures and illustrative animations. Rather, an electronic document is a document comprising a variety of different types of information presentations that are brought

together by an author in order to formulate a comprehensive scientific argument. Or to put it in other terms: in an electronic publication, images, animations and so on cease to be illuminating illustrations to the text, and become semi-independent knowledge representations that together with the text comprise the scientific argument communicated to peer scientists.

In order to develop new insights into an editorial policy that maintains the essential virtues of the paper document while incorporating all the new exciting features of the electronic document, I will firstly discuss the scientific paper as we know it. Subsequently, in the second part of this paper, I will examine new ways of expressing knowledge.

What is a scientific paper?

For the readership of this journal, it is not necessary to dwell at length on the evolution and practice of present-day scientific publishing. The reader is referred to Garvey's book *Communication: the Essence of Science*⁴ and the more recent book by Meadows *Communicating Research*⁵ and references therein. A good starting point for our discussion on 'what is new?' is the report of an International Working Group based on a workshop organized by the AAAS, ICSU Press and UNESCO and published in this journal under the title *Defining and Certifying Electronic Publication in Science. A Proposal to the International Association of STM Publishers*.⁶ We gain a clear understanding of what a scientific publication actually is from their well-formulated statement:

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 assurance of what counts as a legitimate electronic publication.

*We need to
step back and
analyse what it
means to write
for an
electronic
medium*

Thus, the challenge is to ensure that, independent of technology, the use and exchange value of this type of currency can be established universally for all participants in the world of science.

The Working Group proposes a list of minimum characteristics for a document to qualify as a 'publication'. It is worth comparing this list with, on the one hand, the expansion of the concept 'document' to include all coherent knowledge presentations whether textual, non-textual or a mixture, and, on the other hand, the list of communication needs presented by Kircz and Roosendaal,⁷ namely (i) awareness of knowledge, (ii) awareness of new research outcomes, (iii) specific information, (iv) scientific standards, (v) platform of communication and (vi) ownership protection. It immediately becomes clear that scientific communication encompasses a much wider range of interaction between scientists than mere formal publication. We have to incorporate, notes, drafts, preprints, excerpts from laboratory logs, etc. The fascinating issue is that in an electronic environment we are indeed able to integrate these various formal and non-formal means of communication – all the more reason to ensure that we maintain the integrity of our currency.

The Working Group makes a useful distinction between an informal notification, a First Publication and a Definitive Publication. They recommend that all publications should conform to the following characteristics, which we discuss in some detail:

1. Permanence (i.e. the document must be durably recorded on some medium)

This demand is self-evident: no communication or debate independent of time and location is possible without the object under discussion being fixed in some medium of communication. The move to an electronic environment implies that the notion of 'durably recorded' is under attack. In contrast to the paper world, where we can demand that the information is printed on acid-free paper according to an official standard, in an electronic environment we do not yet have any idea what an equivalent

form looks like. Almost every month we are confronted with a claim for an even more superior technology. On top of that, we have to expand the notion of a document to include non-textual objects such as images, simulations or other multimedia objects that might be the final output from a research programme. [We are not necessarily talking here about fashionable computer-game type presentations: the ability of electronic media to deal with civil engineering design drawings (not necessarily of a complicated CAD/CAM type) in the same way as with textual documents is a simple example and one we already find very difficult to deal with.] We have no idea what kind of optical, magnetic or other medium will be selected as the accepted standard in the coming years and we have no idea what the method of writing to that medium will be.

This means that the demand for permanence must be tailored towards a demand for the inalterability of the content of the said object. Thus we have to interpret it as a demand for a well-defined descriptive standard about the content of the document – a standard that enables the storage and maintenance of the integrity of the information independent of the carrier of that information, be it a clay tablet or a future DNA chip. It goes without saying that the current developments in descriptive languages such as Standard General Mark-up Language (SGML) and its successor eXtended Mark-up Language (XML) are of the utmost importance. If, finally, all the information in a document is properly coded according to such a language, we are dealing with simple ASCII or, better still, Unicode strings that can be handled in all conceivable material memory structures. File integrity can be guaranteed by an electronic watermark or signatures. Once the document has been stored, it must be capable of being retrieved and read by the future user using whatever popular medium is then current. An interesting initiative for the immediate future is the National Center for Supercomputing Applications (NCSA) Astronomical Digital Image Library (ADIL) – a repository providing astronomers with research quality images straight from the telescope to their desk over the web.⁸

the document must be capable of being retrieved and read by the future user using whatever popular medium is then current

2. Public availability (in principle not necessarily free of charge)

This demand is clearly medium independent and does not need any further consideration here, although we note that the discussion of how the *hard currency of science* is related to hard currency in the economic realm of publishing fills many pages, regardless of the medium.

3. Persistence (i.e. it should remain in the same form and at the same location, so that it is reliably accessible and retrievable over time)

This point overlaps with the first demand, and again we see a mix of old and new concerns. The persistence of the work has two aspects: the integrity of the appearance and the completeness of the content. Firstly, we have to deal with the problem of the integrity of appearance. This issue is also an important discussion point in the world of the archivist. In most cases, e.g. the figures of a town or departmental budget, only the content of the information is important. In others, e.g. an official certification or a signed treaty, the visual and textural aspects are an essential part of the archival object. Obviously, there are differences between text and non-textual material, where persistence of presentation form may be crucial. But even here we cannot be too conservative: while the pictorial presentation of a dataset may be essential for spotting a peculiar behaviour, in time more sophisticated presentations might reveal more details. This argument leads us to the notion that we have to differentiate in such cases between the basic non-figurative data and their presentation by the original author. Both need to be fixed and together they form part of the author's original publication. But the dataset must also be available separately from the presentation module so that future authors may use and/or integrate these data with new data or with new presentation techniques to develop new work for publication.

Secondly, we have the aspect of internal integrity and coherence. This is typically an XML issue since, with the aid of a full Document Type Definition (DTD) within the

XML paradigm, we can guarantee that articles have well-defined, and hence retrievable, types of information. This persistence aspect can be covered by introducing a complete list or map of contents as an integral part of every document. We have to maintain not only the bitstreams of every component of the document, but also the mutual relations between the various components. We also need a mechanism to check that all components are present. This last demand could become a serious problem in the future. More and more documents will be rendered from components residing in different databases. Take, for example, an astronomy article that calls for data extracted from a huge satellite measurement database. As an electronic publication is, in principle, a modular entity and not an essay,⁹ the persistence demand requires that a publication guarantee that all components remain available. This demand is closely linked to the problem of dead hyperlinks. All this converges to the discussion on the Digital Object Identifier (DOI) initiative. The International DOI foundation was '*created in 1998 and supports the need of the intellectual property community in the digital environment, by the development and promotion of the Digital Object Identifier system as a common infrastructure for content management*'.^{10,11} The DOI foundation is supported by almost all major (commercial) publishers and societies. The idea behind DOI is that every item that has an assigned copyright (hence also books) will get a unique identifier. In the course of development, this identifier will be endowed with metadata such as bibliographic information, genre, but also publishers' information and price. In the first round, as experimented with in CrossRef,¹² DOI is limited to a one-to-one link with the URL of scientific articles in a publisher's database. In the full implementation, it is envisioned that DOI will allow choices, e.g. to go to a copy of the identified entity or to a metadata record about the entity, or to an identical copy of the same entity at different location (mirror site). Adding metadata to DOIs will allow the reader to choose which type of realization of a particular document is required, e.g. a PDF file, an XML file or whatever other storage types are available. It is clear that the DOI approach is a strong attempt to ensure the

there are
differences
between text
and
non-textual
material

integrity of information entities viewed as intellectual property containers; it is also a step towards electronic commerce and trade in intellectual property rights.

A competitive scheme for reference linking, emerging from the scientists engaged in the world of preprint servers, is the Open Archives initiative (OAi). Its goal reads:

The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication.¹³

The aim of this initiative is to promote those active authors for whom self-archiving solutions are a preferred option. The interoperability between such archives becomes the prime research.¹⁴ Although DOI and OAi are approaching the problem from two very different philosophical backgrounds, both schemes, at the end of the day, must ensure the integrity and quality demands at the basis of proper scientific discourse.

The Working Group then suggests that the following features are required to ensure that a document may be securely referred to by other writers:

4. Version control (a bibliographic record must be attached to each version; a set of minimum details is suggested in the document)

As long as we talk about a document in its traditional form, version management can be straightforward. However, once we have a document composed of various modules, originating from different sources, new schemes have to be developed. The main point is that electronic documents are no longer the smallest exchangeable entities. Many electronic documents and most professional web pages are derived from a variety of dynamic databases. A feature of an electronic document can be that it changes with time (or outside temperature, or stock market index, or rocket launch date). This electronic document may be the result of some deep science or engineering

advance, and hence be a scientific publication. A bibliographic record (metadata) is essential for fulfilment of this recommendation. The issue of metadata that entail much more than the traditional bibliographic information will also be dealt with in the companion paper.¹⁵

5. Authenticity (i.e. versions should be certified as authentic and protected from change)

Although nobody will challenge the issue of the absolute need for authentication of every published document, we run into problems once we talk about the discussion of parts of a document. As described above, a necessary distinction is made between the document as the smallest unit of communication and documents that are built up from various components. We have to understand that there is a difference between reuse and multiple use. In the case of multiple use, the citing author integrates the full body of an 'information chunk' into the new work and uses it. A good example is the case of pattern recognition. The journals in this field are full of datasets (e.g. in the form of distorted pictures) on the one hand and methods on the other hand. Would it not be much more exciting if we could swap methods and datasets between authors and allow a true comparison between different methods unleashed on the same data set? This would be an extension of the practice, already current in some fields, especially astronomy, of tapping data from a common database: see, for instance, the French astronomical database Centre de Données astronomiques de Strasbourg.¹⁶ Rzepa and Murray-Rust¹⁷ recently discussed a chemistry application of this idea in more detail in this journal.

For a First Publication, next to version control, the Working Group recommends:

6. Notification (the community of one's peers must be informed as to the version associated with priority claims)

This is an obvious and essential demand for the awareness of current and new research outcomes and the free and democratic flow of information and knowledge. Notification

We have to understand that there is a difference between reuse and multiple use

can be enormously enhanced by electronic tools such as bulletin boards, newsgroups and current awareness services, in the open literature. In addition the original author can be automatically informed if (parts of) his/her work has been cited elsewhere in the literature.

7. Assignment and persistence of a web address/location for the record

In its document, the Working Group sometimes phrases this point as the need to identify the work unambiguously.

This demand is an obvious call for retrievability. We have already discussed above the DOI and OAI programmes that try to tackle this point. The DOI foundation especially is working on this issue, as its main goal is the identification and subsequent handling of intellectual property. The problem is that the persistence of a unique identification code can never be linked to a unique URL. It is much better to ensure a unique code per item, allowing that item to flow from database to database, provided that those databases have a searchable index capable of understanding the grammar of the unique identification code. Gkoutos and collaborators have presented a worked-out example of such a possible scheme.¹⁸

Another issue here is of a more archival nature, namely that at least one copy of every serious document is stored safely in an archive. This is an important and strong demand in a period where paper is declining and a plethora of digital media, each with its own way of data handling, is emerging. This point is closely related to the first point on permanence. It is also closely related to the metadata issue. It implies that a central organization such as the National Library of Congress in the USA, with its constitutionally assigned tasks, must install a legal depot of all items with unique storage codes in perpetuity.

8. Commitment not to withdraw (authors must agree, prior to commencing the selection process, that they will not delete the document from the electronic literature)

This recommendation is a clear statement

aimed at keeping lots of free-floating drafts and worse out of the mainstream of public scientific discourse. In practice, this will be a very difficult issue as, in some fields, people dump almost everything on their home websites and feel free to send all drafts to preprint servers. A problem arises when a second version of a draft has a slightly different title and a different number or order of authors. To impose strict adherence to version control and a commitment not to withdraw the final draft, i.e. the version open for peer review, will be very difficult, as the correction of a typo, a number, or the addition or deletion of a reference can be important as has already been proven in the paper world. A solution might be to link an erratum permanently to the original instead of storing it separately as has been the case in paper journals.

For the Definitive Publication, the Working Group recommends, alongside persistence and version control, assignment and persistence of a web address.

9. Quality control (vetted to ensure quality), in order to maximize usefulness for science and to establish a high level of trust among readers

With this issue, we enter the essence of quality control and the current heated debate on peer review. It is not the purpose of this contribution to discuss the various possible peer-review schemes in detail. The literature on this issue is abundant, ranging from full-scale books analysing a particular journal such as *Angewandte Chemie*¹⁹ to regular contributions on the pros and cons of double-blind refereeing, nepotism and sexism in peer review, and so on. An important new aspect is the self-publishing current in science. Here, new schemes for refereeing are regularly discussed in several internet lists and discussion forums such as the September forum,²⁰ and by individual protagonists, such as the cognitive psychologist and the editor of the e-journal *Psychology*, Stevan Harnad.²¹

Out of all this discussion, one thing becomes crystal clear, namely that the issue is very much domain dependent. Whilst in theoretical physics the pace of research is such that every new idea is immediately

*at least one
copy of every
serious
document is
stored safely in
an archive*

broadcast via preprint servers (although often after it has been internally peer reviewed by the researcher's institute), in more experimental fields the tempo is more relaxed. After all, it is easier to steal an idea than to redo an experiment. In medicine, the question is intrinsically more sensitive as new medical information often generates high levels of public fantasy (and fear). In this field, the discussion on ethics and misconduct is a permanent concern.²² For a recent review on the domain dependency of refereeing in e-journals, see Weller.²³

10. Commitment to archiving and long-term preservation

The same arguments hold here as for the persistence point. However, long-term preservation overshadows all other issues as a current concern. Within the archivist's world, an enormous effort is being made to design protocols for long-term storage. As mentioned above, the problem can be split into (i) storage of the digitized content and (ii) storage of the textural and visual appearance. One important ingredient in this discussion is the scheme of Jeff Rothenberg, in which he proposes to store, next to the information item itself, the software programs used, including the operating systems.²⁴ This very intriguing so-called emulation scheme is under severe attack from XML *aficionados*. A less fundamental but directly applicable scheme is discussed in the 'Draft recommendation for space data system standards: Reference Model for an Open Archival Information System (OAIS)'. This scheme allows the storage of heterogeneous information. Again, astronomy and space research are taking the lead here, as in these fields much information is already available only electronically.²⁵

Conclusion

In the above, we critically discussed the recommendations of an International Working Group. As we have seen, there is visible tension between these very reasonable recommendations and the reality of electronic publication, in that an electronic publication is not a paper publication stored in a different medium. Therefore, in the

hybrid period in which we now exist, we have to move forward by establishing two trails: (i) the introduction of full electronic equivalents to paper-based notions of integrity, authenticity and quality control; and (ii) the introduction of wide-ranging experiments with independent self-contained electronic manifestations of the traditional article. The first task is necessary to ensure the integrity of the continuation of the scientific discourse in the electronic era; the second task is the development of uniquely electronic methods of scientific publishing. Over time the experiments will lead to a new publication culture, although, just as with handwriting, the traditional article will not disappear. The conclusion of the above discussion is that the scientific article will change its form considerably but that, in its new more composite form as an ensemble of various textual and non-textual components, it will retain the cultural and scientific demands with regard to editorial, quality and integrity.

In the second article,¹⁵ attention will be given to the unique features of electronic publications and the need for clear publishing standards.

Note

This paper is an edited version of a presentation given at the 2nd ICSU-UNESCO International Conference Electronic Publishing in Science, Paris, 19-23 Feb. 2001 (<http://associnst.ox.ac.uk/~icsuinfo>).

References

1. Eisenstein, E.L. *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early-modern Europe* (paperback complete in one volume). Cambridge: Cambridge University Press, 1994.
2. Eisenstein, E.L. *The Printing Press in Early-modern Europe* (ed. Canto). Cambridge University Press, 1996.
3. Kircz, J.G. Scientific communication as an object of science. In I. Butterworth (ed.), *The Impact of Electronic Publishing on the Academic Community*. London: Portland Press, 1989. pp. 171-5.
4. Garvey, W.D. *Communication: The Essence of Science*. Oxford: Pergamon Press, 1979.
5. Meadows, A.J. *Communicating Research*. San Diego: Academic Press, 1998.
6. International Working Group. *Defining and Certifying Electronic Publication in Science. A Proposal to the International Association of STM Publishers*. Published as report. *Learned Publishing* 2000:13(4) Oct., 251-8.
7. Kircz, J.G. and Roosendaal, H.E. Understanding and shaping scientific information transfer. In D. Shaw and H. Moore (eds), *Electronic Publishing in Science: Proceedings of the Joint ICSU Press/UNESCO Expert*

an electronic publication is not a paper publication stored in a different medium

- Conference, Paris, Feb. 1996. Unesco Press, 1996, pp. 106–16.
8. Plante, R.L., Crutcher, R.M. and McGrath, R.E. The NCSA astronomy digital image library: from data archiving to data publishing. *Future Generation Computer Systems* 1999:16, 49–61 (<http://imagedlib.ncsa.edu/imagedlib>).
 9. Kircz, J.G. Modularity: the next form of scientific information presentation? *Journal of Documentation* 1998:54(2), Mar., 210–35. The final draft can be found on: www.science.uva.nl/projects/commphys/papers
 10. www.doi.org
 11. The DOI Handbook, Version 0.5.1, 11 Aug. 2000. www.doi.org/handbook_2000/index.html
 12. CrossRef. The central source for reference linking: www.crossref.org.
 13. Open Archive Initiative: www.openarchives.org.
 14. van de Sompel, H. and Lagoze, C. The Santa Fe Convention of the open archives initiative. *D-Lib Magazine* 2000:6(2) Feb. www.dlib.org/dlib/february2000/vandesompel-oai/02vandesompel-oia.html
 15. Kircz, J.G. New practices for electronic publishing 2: New forms of the scientific paper. *Learned Publishing*, in press.
 16. Centre de Données astronomiques de Strasbourg: <http://cdsweb.u-strasbg.fr/CDS.html>.
 17. Rzepa, H.S. and Murray-Rust, P. A new publishing paradigm: STM articles as part of the semantic web. *Learned Publishing* 2001:14(3) Jul., 177–82.
 18. Gkoukos, G.V., Murray-Rust, P., Rzepa, H.S. and Wright, M. Chemistry Markup, XML and the World-wide Web. Part III: Towards a signed semantic chemical web of trust. *Journal of Chemical Information and Computer Science*, in press.
 19. Daniel, H.-D. *Guardians of Science. Fairness and Reliability of Peer Review*, trans. W.E. Russe. Weinheim: VCK, 1993.
 20. September 1998 American Scientist Forum: september98-forum@listserv.sigmaxi.org.
 21. For publications on the internet by Stevan Harnad, see: <http://cogsci.soton.ac.uk/~harnad/intpub.html>.
 22. Hudson Jones, A. and McLellan, F. (eds) *Ethical Issues in Biomedical Publication*. Baltimore: Johns Hopkins University Press, 2000.
 23. Weller, A.C. Editorial peer review for electronic journals: current issues and emerging models. *Journal of the American Society for Information Science* 2000:51(14), 1328–33.
 24. Rothenberg, J. Avoiding technological quicksand: finding a viable technical foundation for digital preservation. Council on Library and Information Resources. Washington, DC, 1999 (www.clir.org/pubs/reports/rothenberg/contents).
 25. Consultative Committee for Space Data Systems. CCSDS 650.0-R-1: Reference Model for an Open Archival Information System (OAIS). Red Book, Issue 1, May 1999 (www.ccsds.org/review_books.html).

Joost G. Kircz

KRA-Publishing Research

Prins Hendrikkade 141

1011 AS Amsterdam,

The Netherlands

and

Van der Waals-Zeeman Instituut, Universiteit van Amsterdam

Email: kircz@kra.nl

Website:

www.science.uva.nl/projects/commphys

LEARNED PUBLISHING

Have you enjoyed this issue of our journal? Did you have to wait weeks for the office copy to get to you?

Extra 2002 subscriptions for members only cost £40/\$60 each, inclusive of postage.

All you have to do is contact the Member Services Administrator:

Lesley Ogg
47 Vicarage Road
Chelmsford

Essex CM2 9BS, UK

Tel: +44 (0)1245 260 571, Fax: +44 (0)1245 260 935

Email: members@alpsp.org.uk

Subscriptions for non-members cost £105/\$170 for institutions, £53/\$80 for individuals.