SUPERJOURNAL: THE PUBLISHER'S PERSPECTIVE

Michael Mabe

Paper presented at the UKSG 21st Annual Conference, Exeter, April 1998

In a unique collaboration betweeen publishers, universities and libraries, the SuperJournal project has developed a multiple electronic journal application for assessing user behaviour in both the humanities and sciences. In addition to showing how users actually use electronic journals (and the barriers that exist to use), the data also reveals the different patterns of scholarly activity in the humanities and the sciences.

Michael Mabe is Director of Material Science Publishing, Elsevier Science Ltd, Oxford OX5 1GB
E-mail: m.mabe@elsevier.co.uk

Introduction

It is a truth generally acknowledged that Oxford cocktail parties usually have at least one philosopher present. If pushed to define 'the wise man', the philosopher will generally fall back on Plato's description of Socrates:

"Socrates is a wise man because he knows what he does not know."

If this description is true, then journal publishers must be very wise indeed, for what we do not know about the way users use journals is very considerable. There can be few products with the venerability of the journal (333 years young last month[1]) where so little is actually known about how (and how often) it is used by its consumers.

A non-exhaustive list of questions that we might ask (and not know the answer) will contain at least some of the following:

-　how do readers actually use journals?

-　how often are journal articles consulted, and by how many (what proportion) of the readers?

-　what parts of the journals are used most, and by whom? The table of contents, the abstracts, the full text of the article? Do different information needs get satisfied by different user strategies?

-　what effects do format (especially the change from paper to electronic) have on authors, readers and librarians?

-　what elements or functionalities attract or put off authors, readers, etc. from using (electronic) journals?

The answers to these questions will have tremendous impact on the future of journal publishing. Establishing new economic models requires detailed research into usage. It is this concern that lies behind the great interest shown by 17 European journal publishers in being involved with the SuperJournal Project[2].

The SuperJournal Project

SuperJournal is a project in the electronic journals division of the Electronic Libraries Programme (eLib). The Joint Information Systems Committee (JISC) of the Higher Eduction Funding Councils and DENI (Department of Education, Northern Ireland) established eLib as a direct result of the Follett report on the future of libraries. SuperJournal is a project researching the factors that can make electronic journals successful and of value to the academic community. Many studies have explored *useability*: the barriers that need to be minimised so that readers will use an electronic journal. SuperJournal focuses on *usefulness*: the key features that make electronic journals valuable to researchers and *regularly* used. The SuperJournal project is a collaboration between four different parties:

- SuperJournal Consortium - 17 European commercial, academic and learned society publishers
- Manchester Computing at the University of Manchester
- HUSAT Research Institute at Loughborough University
- university Libraries at 13 UK universities[3]

Objectives

The SuperJournal Project has a number of objectives. Its principal purpose is to learn what features and functionalities of electronic journals are of real value to the academic community, and to explore this value from the perspective of all the stakeholders (authors, readers, publishers and libraries).

We can identify four objectives each aimed at one of the stakeholders:

- *Authors* - what are the benefits (or costs) of publishing in electronic journals, and what conditions will attract (or deter)?
- *Readers* - what features will make electronic journals useful and (more importantly) used regularly?
- *Publishers* - what changes need to be made in respect of editorial, production and distribution processes to ensure electronic journals do exist and have the features required by the academic community?
- *Librarians* - what organisational and technical changes are necessary to make electronic journals available to readers?

Research and design

The main objective of the research is to answer one simple question: what do authors and readers *really* want from electronic journals[4]?. The main problem with this question is that it cannot be asked directly, or if it is, readers usually cannot answer it in a meaningful way. Terence Conran described this problem quite neatly in an interview in the Independent last year: the trouble with market research was "[it] only tells you history...If you only show someone a red dress, they will say they want a red dress."[5] Research into electronic journals has been similarly plagued.

Hands on experience of using electronic journals is an essential pre-condition for answering the question of what readers really want. It is only when such a context has been provided for views and opinions to form that meaningful answers can be obtained. Bearing this consideration in mind, SuperJournal has, therefore, adopted the following methodology:

- create a multi-publisher, subject-cluster-based, electronic journal application on the world wide web;
- record the actual usage made of the application (and different elements of its functionalities) with a view to identifying the critical usage factors and barriers;
- explore the opinions and experiences of users before and after using the application;
- share their views with the stakeholders with a view to stimulating change.

The SuperJournal application now consists of four subject clusters of journals in the following areas: communication and cultural studies (CCS); molecular genetics and proteins (MGP); political science (PS); and materials chemistry (MC). Each cluster has been designed to provide as great a critical mass of quality material as possible, within the constraints of the coverage of the participating publishers and the time available. Each cluster contains roughly ten or more journals from four or more of the publishers.

SuperJournal application

The purpose of the SuperJournal application is to deliver features and functionality, so that readers can identify those that they value most. The application is, therefore, more a 'test-bed' rather than a true electronic journal. A list of value added features was developed for the application to test. These features were in the following areas:

- availability (e.g. 24 hour access);
- functionality (e.g. full text searching and alerting service);
- multi-media (e.g. graphics, 3-D etc.);
- content;
- timeliness (e.g. available sooner than print);
- convenience (e.g. customisation);
- performance (e.g. printing, download times);
- presentation (e.g. compared to the printed page).

A key consideration, in designing the application, was timing the resources. As the project is three years in length, the application needed to be up and running and available to all user sites within the year. A collection of off-the-shelf software was used by Manchester Computing to provide the functionality, assembled in such a way as to be seamless from the users' point of view. The first release of the application, available with the first journal cluster had the basic functionality. As each successive cluster was launched the application was upgraded with new features. By phasing these features with the launch of each new cluster, response to these features could be monitored as part of the evaluation study.

Baseline research and results

Thirteen UK universities are participating in the research and are partners in the SuperJournal project. Three or four universities participated in the evaluation research being conducted by the HUSAT Research Institute for each journal cluster, but all 13 have access to the application and all of its content.

The purpose of the baseline research is to generate a picture of how users currently use existing electronic journals and to assess the expectations and enthusiasm for them. At the time of writing, the baseline research has been completed on universities participating in the first three clusters. Potential users at the participating universities are asked to fill in a questionnaire available on the web. Representative users are invited to focus groups to explore their perceptions and expectations in greater depth. A typical focus group includes four to ten users, commonly researchers and academic staff but also some postgraduate students. While similar topics are covered, identical questions are not posed to each focus group. The agenda and the level of detail of the discussions are largely in the hands of each particular group. Each session might last about one to two hours and is held under the guidance of a researcher from HUSAT.

The tables 1, 2 and 3 summarise the results of the focus groups[6]. It should be noted that even if a box is not ticked, it may not mean it is unimportant; the topic simply did not occur in the particular discussion. The tables are divided by cluster and by target user site. There is general agreement about many issues facing readers and authors. Interestingly, there is some divergence between scientists and humanities users. The results can be discussed in three main areas:

- problems using printed journals;
- a 'wish list' for electronic journals;
- problems with electronic journals.

Problems with printed journals (Table 1)

Discussion in the focus groups centred on four main areas:

- availability issues;
- time;
- performance issues;
- photocopying issues.

Almost all the focus groups were concerned about the large numbers of missing items in their libraries. These items were missing through theft of the journal issue, vandalism or issues being bound up into volumes at the binders. A lower level concern, among the humanities readers, was the local availability of journal titles. All participants felt they had difficulty in keeping up-to-date with the literature and were concerned with time constraints generally. A leading concern, in respect of performance issues, was having to go elsewhere (i.e. other libraries or using

inter-library loans) to retrieve the information they needed. A large majority had difficulty in finding information in their library and with the slowness of the process. All participants complained about long queues for the photocopying of articles. A secondary concern from scientists was the difficulty of reproducing photos or colour images accurately on a photocopier.

'Wish list' for electronic journals (Table 2)

Discussions in the focus groups centred on the following key areas:

- content and coverage;
- availability and access;
- timeliness;
- searching;
- linking;
- multi-media;
- customisation;
- interactivity and education.

All participants wanted coverage to be as comprehensive as possible. The availability of back issues was especially important to humanities researchers. Almost all participants wanted desk-top access to information. Scientists, in particular, were concerned that access should be fast, instant and guaranteed. Where more rapid publications times were the norm for their subjects, participants expressed the hope that electronic journals were there to speed-up publication. The vast majority wanted to be able to research items over a wide range of journals. They also wanted better browsing and searching facilities in general. Many respondents wanted searching to be a seamless process, with a search linking them to abstract and article, in many cases the searches being conducted on full text. Linking was felt to be important by the majority, especially from references to articles and to supplementary data. Humanities users wanted direct links to source materials while scientists wanted direct access to deposited data such as structures, x-ray diffraction data etc.

Problems with electronic journals (Table 3)

Screen issues, technical or performance issues and author-related matters were foreseen as the most common problems for electronic journals. A majority of focus groups expressed the view that reading on screen was difficult and unattractive and that 'flipping' through articles is desirable but easier in print. The availability of good printers was a key technical issue. For scientists, long down-load times were also critical. There was a generalised, if low level, concern about multi-media matters for authors. Primary among these were the copyright position on multi-media elements and the dangers of other researchers taking and using your deposited data for their own work.

Preliminary usage results (Table 4)

So far, an enormous amount of log file usage data has accumulated. Its collation and analysis into meaningful datasets is proving to be more labour intensive than first expected. Although analysis of the key usage data is still on-going, some preliminary results can be reported. I am indebted to Ken Eason, Sue Pomfrett and colleagues at HUSAT for making this currently unpublished material available.

For all the clusters usage has built up slowly over time. The growth of usage has been uneven between the various clusters. There has been considerable growth of non-target-site cluster usage. That is, users at universities chosen as the selected sites for a specific cluster have also been using other clusters in the application. The number of users of the molecular genetics and proteins cluster (MGP) has grown faster than the communications and cultural studies cluster (CCS). In addition there are three times as many users in MGP as CCS. When considering repeat usage of the application a similar picture emerges. Nearly twice as many users of MGP are repeat users in comparison with CCS. The category of users who are repeat users divides into 34% researchers, 25% academics and 17% librarians. However, for MGP most of the repeat users fall into the research category.

Analysis of the actual use made of functionality and features is complex. It really needs long term repeat usage for a confident assessment of actual behaviour. However, some trends are already being discerned which may, or may not, be confirmed at the end of the project. Despite provision of a number of types of search engines in the SuperJournal Application, the actual usage

of any of these engines is low. In contrast there are high levels of browsing by all users. Browsing tends to lead to yet more browsing, a point emphasised by the low usage of abstracts compared to articles for both clusters.

Despite the differential in the numbers of users of the MGP and CCS clusters (145 repeat users MGP, 36 CCS) there is an unexpected *lack* of differential in terms of actual usage by the two groups. MGP users accessed 153 articles per month while CCS users used 98. One possible explanation for this behaviour may be current awareness browsing. MGP users are scanning the journals regularly to check that there are no articles relevant to their research area rather than to find some specifically to use.

In the focus groups humanities users made specific comments about the need for a large backfile to make electronic journal use more attractive to them. Despite this, the present usage data suggests that the reverse is true: humanities users are accessing current issues more often than back issues and using the currents issues more often than the science users. Whether this result is significant, or whether it is indicative of much greater back issue reading for all users than anticipated, must wait on further results.

With the exception of the capacity to browse from one article to another via hypertext links, users have made little use of other additional functionalities. It may well be that the value of these functionalities will be appreciated when repeat users become more familiar with the application, but there are already indications that the key functionalities for all users are the simple ones: content, browsing, printing.

Most recently[7], it has been possible to analyse the usage data for one of the clusters, in an attempt to begin to answer some of the questions posed at the beginning of this article. For the MGP cluster, Figure 1 shows the absolute number of items accessed per journal in the cluster for the period ending February 1998. Figure 2 recasts this information as a percentage of items accessed per title by normalising the usage levels of Figure 1 by the *actual number* of viewable items per journal which were made available on the SuperJournal Application in that period. While the final usage levels will naturally change as the journal articles remain available to users of the Application over

the coming months, the data already shows that three of the eleven titles (Journals Nos. 2, 12 and 15) have had over 30% of their articles accessed and (it is presumed) read, and in the case of one of these usage has exceeded 45%.

Figures 3 and 4 examine the nature of this access in more detail on three of the journals in the set. Figure 3 shows the absolute number of articles that were accessed once only, twice, thrice etc. Figure 4 normalises this result by the total number of items that it was possible to view for each title during the period the cluster has been available. The similarity between the curves for differing journals is intriguing. While, as expected, the highest proportion of the articles are accessed once only (18-26%), small but significant accesses are occuring between three and ten times. While it is not yet possible to say if all these accesses are by different users (authors admiring their own work cannot be ruled out!), there are glimmerings of reproducible behaviour patterns among readers which may have profound consequences for publishers and librarians, as they consider new models for the sale and distribution of the academic journal.

Preliminary conclusions

So what have we learned? On the basis of results achieved so far, it appears that the clustering of journals is of higher relevance to science users than to humanities researchers. In the focus groups, the lack of a large archive was said to be more of a problem by humanities scholars; at the time of writing the usage data does not support this. For all users, content, browsing and printing are the most important functionalities, although it must be emphasised that this conclusion may change as further data is analysed. It has become clear that obtaining articles containing true multi-media from authors for any of the journals in SuperJournal is going to be impossible within the project time scale. Most authors are reluctant to submit multi-media unless the creation of that multi-media element is directly part of their research work. All authors appreciate the advantage of including hypertext links and the use of colour. More complex multi-media elements such as video, audio, etc. are deemed less important. The benefits of inclusion are

outweighed by the costs in terms of difficulty and time available to a researcher.

From a technical perspective, building the application from existing software was relatively easy, and certainly easier than expected. However, getting consistent data formats from the 17 publishing houses involved (and by the extension the technical suppliers and typesetting companies engaged by them) was problematic. There are no universal *de facto* standards on-line. Even PDF, the closest approach to such a standard, during the duration of the project has evolved from ADOBE Acrobat 2.0 to ADOBE Acrobat 3.0. Files prepared by the latter are not readable by the former.

Many unanswered questions remain. Despite the first elements of detailed usage analysis reported in the last section, we still have to answer most of the academic usage questions posed at the beginning of this paper. We have had tantalising glimpses of the attractive and repulsive forces acting on readers and authors in their approach to electronic journals and a glimmering of the way readers actually use journals. One unexpected outcome has been the difficulty of making qualitative judgements from quantative data: usage levels are numbers, what exactly constitutes 'high' or 'low' usage? Benchmarks need to be established. Getting satisfactory answers to these questions was the starting point for the SuperJournal Project. Over its final year, answers are beginning to emerge that will illuminate some of the most critical issues facing the world of scholarly communication.

Acknowledgements

Further information about the SuperJournal Project is available at http://www.superjournal.ac.uk/sj/

Notes and references

1   *Philosophical Transactions of the Royal Society of London* commenced publication on 6 March 1665; an earlier but no longer extant journal was published on 5 January of the same year in Paris

2   The publishers forming the SuperJournal Consortium (SuperJournal Ltd) are: Blackwell Publishers, Blackwell Science Ltd, CAB International, Cambridge University Press, Carfax Publishing Ltd, Elsevier Science Ltd, Institute of Physics Publishing Ltd, Macmillan Publishing Ltd, Oxford University Press, Routledge, Royal Society of Chemistry, Sage Publications Ltd, Society for Endocrinology, Springer Verlag, Stockton Press,Taylor and Francis, Thomson Science

3   The universities involved are: Bradford University, University of Birmingham, Cambridge University, De Montfort University, Durham University, University College London, London School of Economics, National Institute of Medical Research (NIMR), University of Leeds, University of Oxford, University of Sussex, University of Ulster, University of Warwick

4   The work of Sue Pomfrett at HUSAT in analysing the raw MGP usage data and Ann Apps of Manchester Computing in providing item statistics, both at very short notice, is gratefully acknowleged

5.   Conran, T. in the *Independent*, 18 October 1997: "Market research only tells you history. It doesn't tell you what you want in the future. If you only show someone a red dress, they will say they want a red dress."

6.   Baldwin, C., private communication

7.   The work of Sue Pomfrett at HUSAT analysing the raw MGP usage data and Ann Apps of Manchester Computing providing item statistics, both at very short notice, is gratefully acknowledge

Tables 1, 2 and 3

C-Cambridge, O-Oxford, U-UCL, N- NIMR, B-Birmingham, Ul-Ulster, W-Warwick, S-Sussex, D-De Montfort, L-LSE

	MGP				CCS				PS	
	C	O	U	N	B	Ul	W	S	D	L
Availability Issues										
Local availability of journal titles						x	x	x	x	
Missing items: theft, vandalism, bindery		x	x		x			x	x	x
Someone else is using it, eg students		x			x		x			
Fragmentation of the collection	x								x	
Library opening hours								x		x
Time										
Time constraints generally		x	x		x	x				
Difficult keeping up to date		x	x			x	x		x	
Delays getting US journals		x	x							
Performance Issues										
Filtering, discovering what's relevant					x		x		x	
Difficult/slow finding it in the library		x	x				x		x	
Have to go elsewhere (other libraries, ILL)					x	x	x	x	x	
Tracing errata			x							
Photocopying Issues										
Time, inconvenience, eg long queues	x	x	x				x	x	x	
Can't reproduce photo or colour images	x	x	x							
Cost		x						x		

Table 1. Problems identified with using printed journals

Data grouped by journal clusters (MGP = molecular genetics and proteins, CCS = communication and cultural studies, PS = political science,MC = materials chemistry), and participating universities (C = Cambridge, O = -Oxford, U = UCL, N = NIMR, B = Birmingham, Ul = Ulster, W = Warwick, S = Sussex, D = De Montfort, L = LSE

	MGP				CCS				PS	
	C	O	U	N	B	Ul	W	S	D	L
Content/Coverage										
Comprehensive coverage		x	x		x	x	x	x		
Journals the library *doesn't* subscribe to							x	x		
Backfile	1		2		3		4		5	
Availability/Access										
Desktop access from anywhere	x	x	x		x	x	x	x	x	x
Fast, instant access	x		x		x	x			x	x
Guaranteed access	x	x	x		x					
Access any time			x					x		
Timeliness										
Timeliness		x	x				x		x	
Faster publication times	x								x	x
Immediate access to X-ray structure data (as now for protein sequence data)	x									
Searching										
Search over a wide range of journals		x	x		x	x	x	x	x	
Easier/better browsing and searching	x	x	x		x		x	x	x	
Seamless process: search, abstract, article		x	x				x	x	x	x
Traditional TI, AU, KW search	x	x	x		x				x	
Context searches, eg citations	x				x					x
Full text search	x				x	x	x	x		
Save and re-run searches		x					x			
Limit in time, eg last 3 months		x								
Easier to identify relevant articles						x	x			
Structure searching, eg protein structures	x									
Linking										
Expand footnotes			x		x		x			
From bibliographic references to articles	x	x	x			x	x		x	x
To (lists of) related material					x	x	x	x	x	
To supplementary data		x	x		x				x	x
To original source material					x	x	x			
Forward chaining to updates, errata, and comments on the article	x		x							
To deposited data, eg protein sequences	x	x	x							

Table 2. 'Wish list' for electronic journals

Data grouped by journal clusters (MGP = molecular genetics and proteins, CCS = communication and cultural studies, PS = political science, MC = materials chemistry), and participating universities (C = Cambridge, O = Oxford, U = UCL, N = NIMR, B = Birmingham, Ul = Ulster, W = Warwick, S = Sussex, D = De Montfort, L = LSE

* Backfile: 1 = 5 - 10 years; 2 = 2 - 3 years (ideally 5 - 10); 3 = to 1960s; 4 = 15 years; 5 = 10 - 15 years

	MGP				CCS				PS	
	C	O	U	N	B	Ul	W	S	D	L
Multimedia Content										
More graphics and photos	x				x	x	x			
Video and moving images		x	x		x	x				
3D graphics	x	x	x							
Colour	x									
Sound, music		x			x		x			
Overlay diagrams		x								
Animation (progression of a disease)		x								
Customisation										
Download references	x	x					x		x	x
Bookmarks to "my" articles	x		x		x	x				
Build personal collection of articles	x	x	x			x				
Alerting service, non-specific or based on TOCs					x		x		x	
Alerting service based on personal profile		x					x			
User specifies the set of journals for browsing	x		x							
Interactivity										
Serious comment on an article		x							x	x
Update your article		x							x	
Author email addresses		x								
Education										
Licence images to use in teaching	x		x							
Create/update student reading lists					x					x
Electronic course packets										x
Other										
Good quality printing	x	x	x		x					
Thumbnails of covers (where they vary)	x									
Look for journals to publish in									x	