research information network

E-journals: their use, value and impact final report

A Research Information Network report

January 2011

www.rin.ac.uk

This report is the second arising from a two-year project to describe and assess patterns of the use, value and impact of e-journals by researchers in universities and research institutes in the UK. It should be read in conjunction with the earlier report E-journals: their use, value and impact, which was published by the RIN in April 2009.

Acknowledgements

This research was conducted by the CIBER group at UCL: David Nicholas, Ian Rowlands, Peter Williams, David Brown and David Clark.

The RIN and CIBER wish to thank all those, in the case study institutions in particular, who gave freely of their time in enabling us to gather the data for the findings presented in this report.

Contents

About this report

Understanding the e-journal revolution and its implications: key phase one findings

Publishers began to provide online access to articles in scholarly journals just over a decade ago. Numerous studies have shown how much researchers have welcomed enhanced and easy access to unprecedented numbers of journals. But until recently there has been little detailed evidence about how researchers have changed their behaviours in response to this revolution in access, about how they make use of online journals, or about the benefits that flow from that use. This two-year-long study begins to fill that gap.

In the first phase of the project, we employed a technique called deep log analysis to investigate in detail how researchers are making use of e-journals. We analysed across ten institutions and for six subject areas – biological sciences, chemistry, earth and environmental sciences, economics, history and physics – the footprints that researchers leave behind when they visit the journal web sites of two major e-journal publishers: Elsevier's ScienceDirect and OUP's Oxford Journals. The key findings were set out in the RIN report *E-journals: their use, value and impact*, published in 2009:

Researchers at top-rated institutions behave differently

Users in the most research-intensive universities behave differently from those in less-research-intensive ones:

- they view and download more articles *per capita*
- they spend much less time on each visit
- they do not use many of the online facilities provided on the publishers' platform
- they are much more likely to enter via gateway sites

Researchers in different subjects and institutions behave differently

Patterns of use - in terms of concentration on a small number of journal titles, levels of usage, use of gateways, viewing of abstracts, and length of sessions - vary both by subject and by institution:

- researchers in the life sciences and physical sciences, for example, are much more likely than economists to enter publishers' sites via a gateway service, and much less likely to view abstracts
- the intensity of use per capita varies even across research-intensive institutions within the same subject.

Gateway services are the brokers of access

A large proportion of the traffic to e-journal sites comes via a small number of 'gateway' services, with Google predominant but also including services such as Web of Knowledge and subject-specific services such as PubMed. But the carefully-crafted search and discovery services provided on publisher sites are not much used:

- just four months after ScienceDirect content was opened to Google, a third of traffic to the physics journals on the site came via that route, even though physics was already richly endowed with information services such as SPIRES and ArXiv
- once users enter the publisher platform, they browse rather than search again using the internal search engine. Advanced search services are used rarely, and hardly at all by users in the most highly-rated research institutions.

Readers use e-journals well into the night and over the weekend

Nearly a quarter of ScienceDirect use occurs outside the traditional 9-5 working day. Weekends account for around 15 per cent of use.

In order to broaden and enhance the picture derived from the deep log analysis, and to test those findings, we also gathered and analysed UK-wide statistical data from a range of sources, including the Society of College, National and University Libraries (SCONUL), the Higher Education Statistics Agency (HESA), and proprietary publisher information; and we looked for any interesting patterns and associations.

Usage is rising and cost-per-use is falling

In the years from 2003-4 to 2006-7, the number of article downloads more than doubled, with growth at a compound annual rate of 21.7%. In 2006-7 users downloaded over 100 million articles, and each registered FTE library user downloaded on average 47 articles a year. As the number of downloads has risen, so the average direct cost of each download (excluding overheads, time and other indirect costs) has fallen, so that in 2006-7 it was £0.80.

High levels of expenditure are associated with high levels of use

There is a strong and positive correlation between universities' spend on electronic journals and the volume of downloads of articles *per capita*.

High levels of expenditure and use are associated with success in research outcomes

Per capita use of e-journals is nearly three times as high in Russell Group universities as it is in new universities, although there is no significant difference in cost per download.

Across the university sector as a whole, indeed, there are strong and positive correlations, irrespective of institutional size, between *per capita* expenditure and use of e-journals, and numbers of papers published, citation impact, numbers of PhD awards, and research grant and contract income.

Objectives and methods for Phase Two

The objectives for this second phase of the study were:

1. to establish a deeper understanding of what lies behind the patterns of use and information-seeking behaviour portrayed in the logs to answer questions such as:
    - why do users spend so little time on each visit?
    - why do researchers use gateway sites?
    - why do few researchers use advanced searching?
    - do high levels of use imply high levels of user satisfaction?
2. to investigate reasons for the diversity in information-seeking behaviour and usage shown in the logs, especially with regard to research status and seniority, institutional size and research strength, and subject or discipline.
3. to determine how online searching and use of e-journals relates to researchers' general behaviour in seeking and using information, and to scholarly and research workflows.
4. to investigate further the relationships between levels of expenditure on journals, levels of use, and research outcomes (e.g., does good e-journal provision drive research outcomes, or do libraries benefit from the additional revenue that research success creates?).
5. to analyse any trends in author referencing behaviour over a long period, and to investigate whether these have changed alongside the development of easier access to scholarly literature.

In this second phase, deep log analysis gave way to a more qualitative approach, engaging researchers and students through interviews (face-to-face and telephone), direct observation and surveys. The purpose was to interpret, understand and test the log findings.

We gathered data through:

- two surveys, one with researchers, from PhD candidates to senior academics (308 responses), and one with undergraduate and taught post-graduates (961)
- interviews with researchers (87) and with undergraduate students (15)
- observation of online searches (27 researchers, six undergraduates)

This work covered nine of the UK institutions (eight universities and a Government laboratory) for which we had previously analysed the log data, in the same six disciplines covered in the Phase 1 report: biological sciences, chemistry, earth and environmental sciences, economics, history, and physics. These are large and strategically important areas of research that account for around 41 per cent of the world's peer-reviewed article outputs; and the case study departments cover more than 20 per cent of all research-active staff in the UK in these subjects. Although researchers were the primary focus of the study, students' use of e-journals was also evaluated for comparative purposes.

Interview and survey questions, for both researchers and students, covered such issues as

- the importance of journal literature, printed as well as online, in each of users' various roles (research, teaching and learning, administration);
- the interplay between e-journals and other scholarly works, and the context in which e-journals are used
- routes to access to e-journals (e.g. gateway services, keyword searching, alert services, browsing, etc.)

- what training if any they had undertaken, and its effectiveness
- problems of access and how users seek to overcome them.

Questions for students focused in particular on the role of e-journals within the range of learning materials available to them, and the extent to which their tutors encouraged them to use e-journals.

Alongside the interviews and surveys, we also observed researchers and students at work, noting in particular:

- routes for access (use of gateway sites, alert services etc.)
- search terms and use of advanced facilities
- selection of articles to be read (e.g. by perusing only the title or author, abstract, tables, etc.)
- criteria used to download full text (i.e. how relevance judgments are made).

Participants were asked to 're-run' their last keyword search and to provide a commentary (or cognitive walkthrough) as they did so. This approach was appropriate because:

- the 'last' search, and the reasons behind it, may be well-remembered, and
- it minimises any bias that might occur when asking respondents to 'replay' a 'memorable' or 'interesting' search (which, by definition does not constitute a 'typical ' one).

In addition to this qualitative work at our nine case study institutions, we gathered and analysed UK-wide statistical data from SCONUL and HESA for four additional years, in order to extend our understanding of the correlations between journal use, spending, and research outcomes. This made it possible to introduce time lags into our earlier analyses and thus ask questions of the form, `is there a pay-off from spending on e-journal resources three years down the line?'.

Finally, we collected information regarding the length of reference lists at the end of journal articles over nearly thirty years using an online method. This enabled us to identify any trends in referencing behaviour and to respond to the question whether there is any evidence that reference lists have become longer as a result of the revolution in access that e-journals and associated discovery tools represent.

No other study has subjected a UK research community to such intense scrutiny: logs, questionnaires, interviews, observation and statistical datasets were used to enrich and triangulate the findings presented in this report. More detailed findings can be found in the working papers to be found on the RIN and CIBER websites, and in the journal articles that have arisen from the project. Details can be found at the end of this report.

Summary of key findings from Phase Two

Triangulating the evidence from different sources

Our aim in Phase Two was to test and examine the reasons underlying the behaviours we identified in Phase One, under each of our five objectives.

1. **to establish a deeper understanding of what lies behind the patterns of use and information-seeking behaviour portrayed in the logs to answer questions such as:**

- **why do users spend so little time on each visit?**

Our qualitative data confirms the findings from the log analyses, that many users spend only a few seconds on journal platforms, because they arrive there via gateway services and then move swiftly through links to the fast bag collection zone in order to access the article they want.

- **why do researchers use gateway sites?**

Gateway sites such as Google, Google Scholar, Web of Knowledge and PubMed are attractive to researchers for many reasons:

- they cover a vast number of journal titles and articles, and a lot more besides including conference papers, technical reports and dissertations. Even though Google Scholar does not index all the journal literature (no service does that), there is simply more in the shop window than on any publisher's platform;
- they facilitate serendipity, both through the variety of the literature they retrieve and through the links they provide;
- they are intuitive to use.

For all these reasons, the large gateway sites are massively powerful and influential. They are pivotal players in the scholarly information chain and their role deserves further research if we are to understand the behaviours we see at the publishers' sites.

- **why do very few researchers use advanced searching?**

Researchers use the advanced search facilities on publisher platforms very rarely. If they do use such facilities, it is on gateway sites. But generally they employ simple searches because

- often the search terms they use are so specialised that a simple search retrieves only a small number of hits;
- it is often relatively easy to narrow the search after an initial hit-list is generated;
- it is easy to scroll through even big hit-lists, especially since researchers (notably in the science-based disciplines) tend to make nearly all their relevance judgments on titles, journal names and authors alone;
- the top few entries in simple searches are generally found to include all that is needed to obtain the information required.

- **do high levels of use imply high levels of user satisfaction?**

Downloads of journal articles are rising faster than the worldwide growth in the number of articles published each year. This, together with the rise in the number of references researchers include and the range of unique sources from which they cite, suggests that they are reading a growing proportion of the worldwide scholarly literature.

In common with previous studies, our qualitative research shows that researchers emphasise the importance of scholarly journals (to teaching and learning as well as research), and that they value the enhancements to access over recent years. Scholarly journals are the lifeblood of research and are increasingly important in teaching too. The great majority of researchers use journals ' most days', if not 'every working day'.

2. **To investigate reasons for the diversity in information-seeking behaviour and usage shown in the logs, especially with regard to research status and seniority, institutional size and research strength, and subject or discipline.**

Our qualitative research confirms the findings from the log analyses that there are significant variations between researchers in different disciplines as to their patterns and levels of usage of e-journals, including the frequency and amounts of time they spend online. Some variations are associated with the differing characteristics of disciplines, and the place that journals play among other resources in the research process. Historians, for instance, spend more of their time reading than most scientists; but journals are not so dominant among the many different kinds of information resources they read. Further work is required, however, before we can fully understand all the variations between disciplines and such issues as the place that e-journals occupy alongside other information resources, concentrations of usage, and how e-journals fit into research workflows.

Our qualitative research also confirms the finding that researchers in top-rated research institutions and departments tend to behave differently from those in less highly-rated institutions. They use gateways more, and their greater knowledge of the literature means that they can scan long hit-lists and make relevance judgements very quickly. Any viewing of abstracts is done on gateway sites, from whence researchers go straight to the full text on the publisher platform. Again, however, there is much more work to be done before we fully understand the characteristic information-seeking behaviours of top-rated researchers as compared to their less highly-rated colleagues.

3. **to determine how online searching and use of e-journals relates to overall information seeking, use, reading and citing behaviour, and to overall scholarly and research workflows.**

As noted in earlier studies, researchers' information-seeking now goes on mostly outside the library, whether in the lab, the office, at home or on the move. Once they have found an article that looks interesting, they may read it in full, but often they will browse to get the main points, or read only the sections that are especially relevant to their work. Much of researchers' information-seeking and reading goes on outside normal office hours and at the weekend: researchers are aware that they can carry the library and its resources with them wherever they are. Only a small minority (14%, mostly in the humanities) visit the library building to browse or to read hard copy journals.

4. **to investigate any evidence of cause-and-effect and directionality in the relationships between levels of expenditure on journals, levels of use, and research outcomes (e.g., does good e-journal provision drive research outcomes, or do libraries benefit from the additional revenue that research success creates?)**

Our analysis of the relationships between levels of expenditure and use, and research outcomes, is now based on data over a five-year span. We have used a statistical modelling technique to test a number of hypotheses, and it shows that intensive use of e-journals is a very strong predictor of future research success. This is an important

finding for all those who are concerned in research and scholarly communication, and deserves further and more detailed investigation.

5. **to analyse any trends in author referencing behaviour over a long period, and to investigate whether these have changed alongside the development of easier access to the literature.**

Analysis of data from reference lists over the period 1990 to 2007 suggests that enhanced access to more titles in e-form, together with the use of generic rather than discipline-specific search tools has had a massive impact on the referencing behaviour of researchers both in the UK and worldwide. UK researchers are producing more articles, with more references, from a wider range of sources than they were two decades ago. The average number of references included, and the number of unique sources cited, in UK-authored articles is higher than the world average, and this lead is being sustained over time. The rise in the number of unique sources they cite suggests that UK researchers are keeping pace with increases in the volume and range of literature published worldwide, even if the average number of references in each article reflects a diminishing proportion of the global literature.

"Publishers's sites are just clunky ... they are just not easy to use."
Physics researcher, age 40-49.

"I go to PubMed - always. I ... don't really notice the publisher page at the end."
Life sciences researcher, aged 20-29.

Detailed findings from Phase Two

1. Finding articles

Search behaviour on publisher e-journal platforms

Our earlier work showed that researchers make little use of the search facilities on publishers' own platforms (see Table 1). The journals on the ScienceDirect and Oxford Journals platforms are heavily used, but the publisher platforms are not the first port of call for researchers seeking e-journal content. Only one in ten sessions on those platforms actually started there.

Our qualitative work, both interviews and observations, strongly reinforced researchers' preference for gateway or third-party sites as the means to search and browse for articles: gateway sites are preferred for discovery; publisher platforms for delivery or pick-up. Researchers use services such as Google and Google Scholar, PubMed, Scopus and Web of Knowledge because they are typically seeking a wide reach in their results, both of journal articles and other kinds of sources. No single publisher platform can provide that reach. In the humanities and some areas of the social sciences JSTOR is a highly-popular gateway service, even though the great majority of its coverage is at least 3-5 years old. It is also worth noting that some information-seeking activity takes place outside both gateways and publisher platforms: researchers use their web browser (e.g. Internet Explorer or Firefox) to find information by following links once they have made their initial search.

Searching on gateway sites is generally straightforward and easy. Researchers find that simple (even one-word) searching is effective because

- often the search terms they use are so specialised that a simple search retrieves only a small number of hits;
- it is often relatively easy to narrow the search after an initial hit-list is generated;
- it is easy to scroll through even big hit lists, especially since researchers (notably in the science-based disciplines) tend to make nearly all their relevance judgments on titles, journal names and authors alone;
- the top few entries in simple searchers are generally found to include all that is needed to obtain the information required.

Our interviews and survey show, nevertheless, that researchers do make use of advanced searching, with over a third (37%) of survey respondents saying they use advanced search 'as a matter of course' and nearly half (47%) said 'if it is necessary'. But such searching takes place on gateway sites rather than publisher platforms. All this explains why the log records for the use of publishers' platforms show searching as clipped and minimalist.

Table 1: Recorded use of search facilities on ScienceDirect e-journal platform by researchers

	percentage of ScienceDirect searches	
	Basic search	**Advanced search**
Chemistry	4.2%	0.3%
Earth sciences	4.2%	0.2%
Economics	3.1%	0.2%
Life Sciences	1.5%	0.2%
Physics	4.7%	0.2%
Mean	**2.3%**	**0.2%**

Source: CIBER deep log analysis, January to April 2007

Yes [I read online] increasingly. I don't like it, but I keyword within the article, I can scroll easily and I can switch to other tasks - like look at my own paper that I am writing at virtually the same time.
Life scientist, aged 40 - 49

2. Reading articles

Reading journals online

Once researchers have located an interesting article, how much do they read, and in what medium, print or electronic? The answer depends on what we mean by `reading'. Our survey and interviews indicate that 'reading' may cover a range of activities from quickly skimming abstracts, or searching a paper simply for images or tables, to reading the full article, perhaps more than once. It is striking that when researchers were asked about the last *important* article they had read (see Table 2), over 40% of them indicated that they had not read the article in full.

Table 2: Self-reported article reading behaviour

	percentage
I read the whole article thoroughly, once	33%
I read several sections (e.g. methodology, conclusions) thoroughly	29%
I read the whole article thoroughly, several times	26%
I read the article briefly to get the main points	9%
I skim read it to find a specific fact or reference	1%
I read only one section, briefly	1%
I read the abstract only	1%
Total	**100%**

Source: CIBER survey of researchers (n=153). Responses relating to the last 'important article' they read.

'Power browsing' is the consumption method of choice: researchers view e-journal content strategically, seeking and focusing on key messages rather than reading documents in a linear fashion. Researchers have probably always operated in similar fashion, but online search and access now makes it much easier to do so, and this is a key part of the explanation for the rapid and continuing increases in volumes of use.

3. Referencing articles

Has wider access to the literature impacted referencing behaviour?

Electronic full text access to journal articles is a relatively recent phenomenon. Older researchers can remember a different world where information-seeking probably meant a visit to the library. Searches of abstract and indexing databases required the use of complex command line languages, and were often undertaken by library intermediaries. These were followed by form-filling and time delays while inter-library-loan requests were fulfilled.

Online searching, browsing, and access have changed that world, and researchers' practices, fundamentally. The question addressed in this section is whether, and if so how, these changes have had an impact on how researchers as authors refer to other researchers' earlier work.

We therefore gathered and analysed data deriving from the reference lists at the end of journal articles, using Thomson Reuters' citation indexes over the period 1990 to 2007. We used this data to investigate whether the widened availability of journal articles, and ease of access to them, have led researchers as authors to cite more work from a broader range of journal sources than in earlier years. In order to do so, we counted not only the number of references, but also the number of different sources (unique journal titles, conference proceedings, reports, etc.) from which they came.

We restricted our analysis to the six subject areas already mentioned, and considered three key temporal reference points in particular:

- 1990 (pre-web, pre-digital library in any meaningful sense)
- 1995 (early digital library)
- 2007 (mature digital library, post Big Deal)

We collected data for all authors worldwide and for the UK only to elucidate any trends or patterns.

Key findings

The worldwide picture

- the number of articles published worldwide has increased in all our six subject areas
- the average number of references included in each article has risen too, although not as fast as the numbers of articles produced each year
- the number of unique sources (journal titles, reports etc) from which those references are drawn has risen too, at roughly the same rate as the increase in the average number of references
- *so researchers worldwide have more to read, and those of them who write journal articles are citing more papers, from a wider range of sources*

The UK experience

- the number of UK-authored articles has increased in all our subject areas, but not as fast as worldwide production, and so the UK share of all articles published across the globe has fallen
- the average number of references included, and the number of unique sources cited, in UK-authored articles is higher than the world average, and this lead is being sustained over time
- the rapid rise in article downloads, taken together with the rise in the number of references and of unique sources cited by UK researchers, suggest that they are keeping pace with increases in the volume and range of literature published worldwide, even if the average number of references they include in each article reflects a diminishing proportion of the global literature.

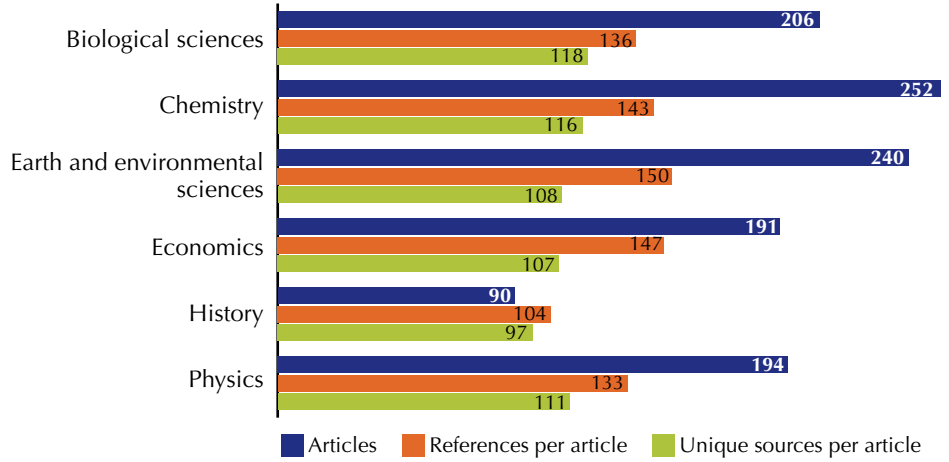
Has authors' referencing behaviour changed alongside the development of easier access to journal articles?

Worldwide experience

a) Numbers of published articles

It is commonly asserted that the number of scholarly journal articles published each year has been growing at a rate of 3-3.5% a year, and that this represents a well-established trend over many decades.[1] Growth in the number of articles in journals covered in the Thomson-Reuters Web of Knowledge databases has since 1990 been at an even higher rate in our six subject areas, of between 4% and 5% a year. Over the period 1990-2007, the number of papers shown in the Thomson Reuters databases has risen sharply in all of our six disciplines except history.

Figure 1: Growth in numbers of articles published, references per article, and unique sources per article by discipline, 1990-2007; worldwide (index 1990=100)



Source: Online analysis of Thomson Reuters citation databases

1. M Ware and M Mabe, *The STM Report: An overview of scientific and scholarly journal publishing*, STM 2009, pp18-19 (http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf)

b) Numbers of references

One of the key results of the growth in numbers of articles published worldwide is that researchers have more to read, and to cite in their own publications. It is therefore not surprising that the average number of references included in each article published worldwide rose over the same period, in all our subjects except history (where referencing behaviour is quite different from that in the sciences or most of the social sciences, and where the average number of references tended to remain static). As can be seen from Table 3, the number of references varies significantly by subject, but in the five subject areas other than history, numbers rose by between a third and a half from 1990 to 2007, though it is notable that this rise is much smaller than that in the number of articles published. Thus although researchers are citing more in absolute terms per article, those citations represent a smaller proportion of the global output of publications.

Table 3: Average number of references per article published in journals covered by the Thomson Reuters databases, 1990-2007

	1990	1995	percentage increase on 1990	2007	percentage increase on 1990
biological sciences	29.63	34.45	16.3%	40.18	35.6%
chemistry	21.31	23.40	9.8%	30.55	43.4%
earth and environmental science	23.17	25.35	9.4%	34.88	50.5%
economics	21.12	24.30	15.1%	30.95	46.5%
history	34.46	32.77	- 4.9%	35.89	4.2%
physics	18.99	20.62	8.6%	25.30	33.2%

c) Number of sources

The question then arises whether these references are drawn from a wider pool of sources. It is possible to identify unique sources in the reference lists at the end of articles so that, for example, multiple references to papers in the *Journal of Insignificant Studies* are counted as a single source. Other document types, such as monographs, reports, theses and other grey literature are treated similarly. There are big differences between disciplines when we look at the average number of sources cited per paper, as shown in Table 4.

Table 4: Average number of sources per article published in journals covered by the Thomson Reuters databases, 1990-2007

	1990	1995	percentage increase on 1990	2007	percentage increase on 1990
biological sciences	2.37	2.80	18.0%	3.36	41.8%
chemistry	1.84	2.14	16.01%	2.46	33.8%
earth and environmental science	4.38	4.72	7.6%	7.66	74.8%
economics	7.57	8.13	7.4%	9.03	19.3%
history	20.97	20.29	- 3.2%	23.95	14.2%
physics	1.89	2.10	11.0%	2.57	35.6%

It is also notable that the average in all six disciplines has grown significantly since 1990, though there is no clear pattern in the relationship across the disciplines between the rates of increase for sources and for the number of references.

When we take together the rise in the number of references and the number of unique sources cited per article, it seems clear that researchers are reading and citing more papers and other literature from a wider range of sources than they were two decades ago.

This reflects, no doubt, the growth in the volumes of journal articles and other papers, and also of journal titles.[2] Further research is required to ascertain the extent to which it reflects also greater ease of access to journal articles and other sources.

Has authors' referencing behaviour changed alongside the development of easier access to journal articles?

UK experience

There are notable similarities and differences between the experience and performance of UK authors as compared with the growth in worldwide publications and average numbers of references and sources. The rapid rise in the number of downloads by UK researchers suggests that they are at least keeping pace with the growth in the volume of literature published worldwide.

a) Numbers of articles published

While the numbers of UK-authored articles have risen significantly, the rate of growth has been slower than that for the world as a whole. As has been well-reported,[3] the UK share of the global output of research papers has therefore fallen in recent years.[4] This has been the result, in large part, of the rapid rise in papers produced in China, India and other countries, notably Brazil and Iran. The decline is not so evident in environmental sciences, economics and history, where on some measures there has been an increase in the UK market share; but this may simply reflect the characteristics of the coverage of the Thomson Reuters databases.

b) Numbers of references

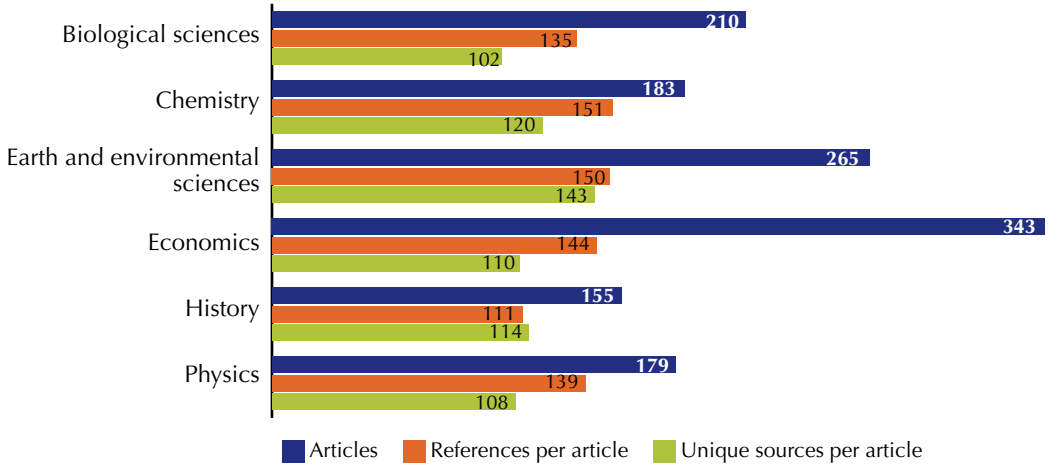
UK authors across all of our six subject areas tend to include marginally more references in their articles than the worldwide average. Thus in chemistry the average number of references in UK-authored papers in 2007 was 33.2, compared with the worldwide average of 30.6. In parallel with experience across the world, UK authors have increased the number of references in their articles over the past two decades, and hence they have sustained – and in physics, chemistry and history enhanced - their characteristic of including more than the worldwide average number of references.
It is notable, however, that while they are citing more references in absolute terms, their citations are not growing as fast as the global output of articles. As with authors in the rest of the world, the citations they include in each article thus represent a diminishing proportion of the global output of publications.

c) Numbers of sources

Just as with the number of references, UK researchers in all six subject areas also tend to cite from a wider range of sources than the worldwide average. Thus in economics, the average number of unique sources cited in UK-authored papers in 2007 was 13.8, compared with the worldwide average of 9.0. Like their colleagues in the rest of the world, UK authors have also increased the range of sources from which they cite over the past two decades, though in biological sciences and physics the increase has been small (and in biological sciences there was actually a significant fall between 1990 and 2004).

2. See M Ware and M Mabe, *The STM Report: An overview of scientific and scholarly journal publishing,* STM 2009, pp18-19 (http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf)
3. See, for example, the latest of the series of reports for the Department of Business Innovation and Skills on the *International Comparative Performance of the UK Research Base,* September 2009 (http://www.bis.gov.uk/assets/biscore/corporate/migratedd/publications/i/icpruk09v1_4.pdf)
4. Although the downward trend is clear, calculations of the precise UK share of the worldwide production of articles depend on the coverage of the database on which the calculation is based and, critically, on whether an integer or a fractional counting method is used. For an explanation of the technical details, which can lead to dramatically different results, see *UK Share of World Research Outputs: an Investigation,* RIN 2009 (http://www.rin.ac.uk/system/files/attachments/UK_share_research_output_REPORT.pdf)

Figure 2: Growth in numbers of articles published, references per article, and unique sources per article by discipline, 1990-2007; UK integer counts indexed to 1990=100



Source: Online analysis of Thomson Reuters citation databases

Table 5: Average number of references and sources per article, UK and worldwide

	References		Sources	
	1990	**2007**	**1990**	**2007**
biological sciences				
worldwide	29.63	40.18	2.37	3.36
UK	30.84	41.56	3.72	3.80
chemistry				
worldwide	21.31	30.55	1.84	2.46
UK	22.04	33.19	2.63	3.14
earth and environmental sciences				
worldwide	23.17	25.35	4.38	7.66
UK	26.34	27.00	7.67	10.96
economics				
worldwide	21.12	24.30	7.57	9.03
UK	24.30	26.29	12.55	13.83
history				
worldwide	34.46	32.77	20.97	23.95
UK	41.81	41.99	29.88	34.03
physics				
worlwide	18.99	20.62	1.89	2.57
UK	20.25	21.29	3.07	3.32

Data source: Thomson Reuters Citation Indices

d) Overall findings

UK researchers are producing more articles, with more references, from a wider range of sources. Analysis of their referencing behaviour indicates that they are keeping up with, and sustaining a lead over, their colleagues in the rest of the world. The rise in the number of unique sources they cite, across the growing number of articles they publish, exceeds the long-term growth rate in the number of journal titles published each year. This suggests that they are keeping pace with increases in the volume and range of literature published worldwide, even if the average number of references *in each article* reflects a diminishing proportion of the global literature.

"You have to remember that in history books are still very very important."
History researcher, aged 50-59

4. Subject differences

Differences in information-seeking and usage between subjects

Our earlier log analyses highlighted a number of significant differences by subject discipline, in relation to:

- extent of e-journal usage and time spent online
- use of gateways and advanced search facilities
- concentration of reading in top *n* titles

Extent of e-journal usage

Responses to our survey indicated that between three-fifths and four-fifths of researchers across all six disciplines use e-journals 'most' or 'every' working day. Life scientists are the most likely (50%) and historians least likely (16%) to use them every day. But when they are online, historians spend more time on each session. This is in part because the greater length and more discursive nature of articles in history as compared with the sciences mean that it is less easy to scan a full-text article for a single fact or figure that is not present in the abstract. Moreover, the nature of the discipline and the greater variety in the languages used make it more difficult immediately to locate relevant material.

The extent to which electronic and hard copy resources sit alongside each other also varies by discipline. While the life sciences have moved essentially to a wholly digital world, even for journal back files from the 19th century, not all journals in history are as yet available electronically. Similarly, the use of books alongside e-journals also varies by discipline. In geology, for example, researchers may need to consult 'classic' books in hard copy; and historians still rate monographs as the gold standard in scholarly works.

Use of advanced searching

Our log analysis showed that researchers across all disciplines use gateway services, with use particularly high in the life sciences (probably associated with PubMed) but notably lower in economics. Our survey, interviews and observations show that Web of Knowledge, Google, Google Scholar, Google Books, PubMed, and JSTOR are all popular as gateways.

Researchers across all disciplines also indicate that they make use of the advanced search facilities available in such services, much more than they do on publisher platforms. It is notable in particular that economists, whom the logs show to be the least likely of our six subject areas to use gateway services, nevertheless stress the importance of advanced searching: none of the economists we surveyed or interviewed indicated that they could obtain good results from a simple search.

In our observation of techniques used in advanced searching, we also found some notable disciplinary differences. Historians, for example, tend to use more words in their search strings, and to search within their results by refining or adding to their search terms, taking account of issues such as synonyms and variant spellings. We also noted examples of classic 'funnelling' behaviour, with general searches being gradually narrowed by delimiters.

Concentration of usage

Our log analysis showed high concentrations of downloads and page views on the top 5% of journals in all disciplines, with the rates particularly high in economics and chemistry. Our survey and interviews, however, suggested a rather different pattern. When we asked researchers roughly from how many different journal titles they had

'Any decent scholar can be expected to learn four or five languages at a reading level – especially German. Having said that, it's not always easy to find the material in some languages.'
Historian, aged 50-59

"I can't ever remember reading a book [in my work]."
Chemistry researcher, aged 30-39.

read at least one article over the previous four weeks, physicists indicated that they were the most concentrated in their reading, with only 11% of respondents reading ten or more titles, whereas 53% of chemists did so. Further work is required to investigate disciplinary differences in the numbers and distribution of titles that researchers read.

Barriers to accessing the literature

We also asked researchers about any barriers they faced in securing access to the articles they need. Researchers now expect immediate access to the full text, and they are frustrated when they find that their university does not have the necessary subscription, or that they are asked for a password they do not have, or that they are asked to pay for a download. Over a third of our survey respondents reported such problems, though they tend to regard them as irritants more than as a barrier to their work.

Physicists have the fewest access problems, which may be because so much literature in physics is available open access, particularly through the ArXix repository. Open access repositories are also available in other subject areas, including biosciences and economics, but they do not as yet seem to have the same level of reach within their subject communities as ArXiv does in physics. Historians, on the other hand, seem to face the most problems with access, partly in relation to currency (the most recent material in JSTOR – the database of choice for many historians – is around five years old) and partly to language (foreign language material can pose particular difficulties).

It's frustrating when you are told 'no full text available', but I never bother taking up the 'see where holdings are' option. I just go to the next entry on the hit list. All the main journals are there.
Life sciences researcher, aged 30-39

5. Institutional differences

Researchers in top rated research institutions behave differently

Our analysis of log data and SCONUL statistics in Phase I of the study pointed to large differences in the level of e-journal use between institutions. Research intensity appears to be the main driver: per capita use of e-journals is nearly three times as high in the Russell Group as in the new universities. We also found differences in information-seeking behaviour between researchers in the most-highly-rated institutions and those in less-highly-rated institutions.

Researchers in top-rated universities find information quickly

Our interviews and observations confirmed that researchers in top-rated universities tend to spend less time on publisher platforms. They use gateway sites for their initial searching, and their knowledge of the literature enables them to scan hit-lists and make relevance judgments extremely rapidly. Their engagement with the publisher platform is both clinical and pragmatic, fast and efficient.

Researchers in top-rated universities use abstracts on gateway sites, if at all

Abstracts are almost impossible to avoid on publisher platforms. But top-rated researchers use them hardly at all. If they use abstracts, they do so on gateway sites and then go straight to the full text on the publisher platform.

Researchers in top-rated universities read high impact journals

Researchers in top-rated institutions tend to view journals with a high impact factor. This may be associated with their focus on pure as distinct from applied research: impact factors tend to be highest for journals focusing on pure research. It may also be associated with their publishing habits. If you are looking to publish papers in *Nature*,

"I used to take papers home - many years ago. Although I still do that - I now follow that up by looking for other stuff online. Sometimes I just read online at home. The library is in my home now."
Earth sciences researcher, aged 50-59

you will expect to be able to cite – and therefore to have read – other papers from high-impact journals. If you are looking to publish in lower-impact journals, the drive to read and cite such papers may be less.

Patterns of work

Worldwide 24/7 access to e-journals means that it is easier than hitherto for researchers to work whenever and wherever they want. Publisher logs show that nearly a quarter of journal use by university researchers takes place outside the traditional 9-5 working day, and that a sixth takes place at weekends.

Our survey and interviews indicate that researchers often find it more convenient and effective to work away from the office, or at home. Time on-campus and during the day is often spent dealing with students or administration, with relatively little time for research. But 24/7 access, from anywhere in the world where the internet is available, has removed the barriers to working effectively beyond the university and the normal working day.

Whether actual working hours have changed is not clear. Academics have always taken material home to read or to mark, whether it be student assignments or journal articles. Email, which has been around for much longer than e-journals, enabled a wider range of work to be undertaken remotely. Now, however, virtually all academic duties except face-to-face meetings and use of laboratory or other specialist equipment can be undertaken at home or another remote location.

Government labs are different

Researchers in government laboratories tend to behave differently from their colleagues in universities. Some of this is the result of their size and research focus, but some of it is clearly down to different rhythms and requirements of the job. Thus laboratories are generally more interested in current material, in a narrower range of journals. Most of their information-seeking takes place during office hours, since they tend to work on site in order to use the laboratory equipment they need. Less of their use of e-journals takes place out of office hours or at the weekend.

6. Student use of e-journals

Comparing e-journal use by researchers and students

It is difficult, often impossible, to distinguish from log records alone between researcher and student use of e-journals. Moreover, there are no figures in the public domain regarding the levels of use of e-journals by students and researchers respectively, and it seems unlikely that any librarians or publishers know this with any confidence. Nevertheless, in order to address fully the questions at the heart of this study, we have attempted to quantify the level of student use in UK universities using two different approaches: a student survey and secondary analysis of data from the Society of College, National and University Libraries (SCONUL), and the Higher Education Statistics Agency (HESA).

Survey findings

E-journals are used in both teaching and learning, not least because they are more easily available than many hard-copy resources, A survey of over 500 students in the case study institutions shows that they use journals regularly, and that frequency of use increases as they progress their way through the academic system (Table 6). These results suggest that student usage of e-journals is substantial.

"Ninety percent of my reading is journal articles. I only read books if I want to get a general idea of a subject or if I am looking for an easier explanation of what I am reading. Sometimes I just look at Wikipedia."
3rd year undergraduate life sciences student

Table 6: Self-reported frequency of journals use: students and researchers compared

	Every day	Most days	2-3 times per week	Less often	Irregular or never
Undergraduate (yrs 1 and 2)	2.6	12.8	15.4	37.0	32.2
Undergraduate (yrs 3 and 4)	20.2	25.8	23.6	18.6	11.8
Taught postgraduate	26.3	10.5	21.1	26.3	15.8
PhD students	36.8	38.2	9.2	5.3	10.5
Researchers	45.3	30.7	6.7	9.4	7.9

Source: CIBER surveys of researchers (n=153) and students (n=512) [row percentages]

The survey also asked whether respondents had undergone training in the use of e-journals. An important finding for librarians is the clear evidence of a positive relationship between library training in e-resources and student use of advanced search facilities for finding journal articles (Table 7).

Table 7: Self-reported student use of advanced search facilities and library training in the use of e-resources

	Trained	Not trained
Yes, as a matter of course	29.0	19.2
Yes, if it necessary	56.5	47.7
No, I get good results from a simple search	14.5	33.2

Source: CIBER survey of students (n=512) [column percentages of students agreeing with each of the statements above]

Estimating student use from SCONUL and HESA data

Another way to estimate student use is to make inferences from a statistical model. Data were collected on total QR[5] funding (HESA), student and academic staff numbers (HESA) and full text downloads (SCONUL , based on COUNTER data) for 118 UK universities for the five years 2003-04 to 2007-08. Cases were scanned for outliers and the data were then averaged over the five years.

We ran a weighted least squares regression model in SPSS using the following variables: full text downloads, research postgraduate, teaching postgraduate, undergraduate, and academic staff FTE. QR funding was used as a weighting variable, and as a proxy for research activity. The model fits the data well and suggests that undergraduates and taught postgraduates probably account for around 23.7% of all e-journal use in UK universities. Researchers (academic staff and postgraduate research students) then account for around 76.3% of total usage.

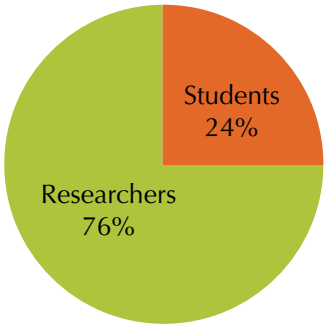
Student use of e-journals is clearly substantial, and this represents a powerful argument for sustained long-term spending on them. E-journals play a major role in supporting learning and teaching, as well as research.

Table 8: Estimates of full text article downloads per capita per year for 118 UK universities

	Full text downloads
Undergraduates	13.2
Taught postgraduates	10.2
Research postgraduates	319.2
Academic staff	314.6

Source: CIBER modelling of SCONUL and HESA returns for 118 UK universities, 2003-04 to 2007-08

Figure 4: Estimated share of downloads by university status



CIBER weighted least squares model (n=118 UK universities)

Further investigation is required to test and understand these figures more fully, and in particular how they relate to usage of e-journals in individual universities and groups of universities.

5. QR (quality-related) funding is distributed to universities by the Funding Councils by a formula based principally on their numbers of research-active staff and their ratings in the Research Assessment Exercise

7. Libraries and e-journals

Key UK trends over the period 2003/04 to 2007/08

In the first phase of our research, we found evidence of strong positive correlations between levels of expenditure and use of e-journals, and research outcomes for 2007-08. The data we used was derived from SCONUL and HESA, and some additional data supplied by Elsevier. In the second phase, we collected data for a five-year period, 2003-4 ("2004") to 2007-8 ("2008") so that we could look at trends over that period. We indexed the data against 2004 and used the Treasury GDP deflator to express all monetary values at 2008 prices. Key findings follow with 2008 index values in brackets.

Library spending [6]

- Total university spending increased substantially faster than the growth of the UK economy (121.9).
- Net library expenditure rose over the period (113.4) but at a slower rate than university spending in general. Russell Group institutions (121.3) spent more on their libraries than the rest of the sector.
- Library spending on `information content' (journals, books and database subscriptions) rose modestly in real terms (115.4) but fell back sharply as a percentage of total university spending (92.9).
- Libraries shifted their spending patterns on resources to favour electronic content (journals, books and databases) over print. Spending on e-resources rose very significantly over the period (165.7), a shift led by the Russell Group universities (189.2).
- Spending on electronic (e-only and e+print) journals rose (141.7) while spending on print journals declined (80.9).
- Spending on interlibrary loans fell back sharply (77.3).

Journal choice

- By 2008, users were able to access a much larger selection of journal titles in all formats at their institution (131.8), with pre-1992 universities being the main beneficiary (146.2).
- Large falls in print-only titles (68.6) were matched by an enormous increase in electronic-only titles (170.9) as a result of publishers' bundling deals.

Usage

- There was spectacular growth in the numbers of full-text article downloads (262.1). Usage at the new universities doubled (208.7) and more than tripled (356.7) at Russell Group institutions.
- Increases in usage meant that, excluding overheads, the average direct cost per full text article download fell sharply in real terms (58.8). Cost per download fell most sharply at Russell Group universities (38.2).
- Volumes of interlibrary loans declined (66.7) over the period but at a slower rate (74.5) in Russell Group libraries.

Research outcomes

- Income from research grants and contracts increased substantially in real terms (124.6), especially in Russell Group universities (128.5).
- The number of articles published by UK researchers rose substantially (122.4) with the new universities showing particularly marked growth (129.5).
- As measured by Elsevier Scopus, UK research rose in terms of its share of world citations (108.0 by 2007, no 2008 figures available). Again, new universities showed the greatest relative improvement (112.5). Papers published by authors in Russell Group universities attracted in 2007 49% more citations than the world average.
- The number of PhD awards increased steadily across the sector (107.6).

6. For a fuller analysis of UK library expenditure over the decade 1999-2009 see the RIN report *Trends in the finances of UK higher education libraries, 1999-2009* (http://www.rin.ac.uk/system/files/attachments/library_trends-report_screen.pdf)

Overall conclusions

- Library spending on information content failed to keep up with general spending on universities.
- Library users nevertheless have access to a much larger range of titles than ever before
- Library expenditure on e-resources and training has brought spectacular success in driving a massive rise in usage and sharp falls in the average cost per download
- UK universities across the sector improved their performance in winning research grants and contracts, in increasing the numbers of papers published, in sustaining high levels of citation impact, and in supporting growing numbers of doctoral students.

Table 9: Cost per download in UK universities, 2004-2008

Direct cost per download

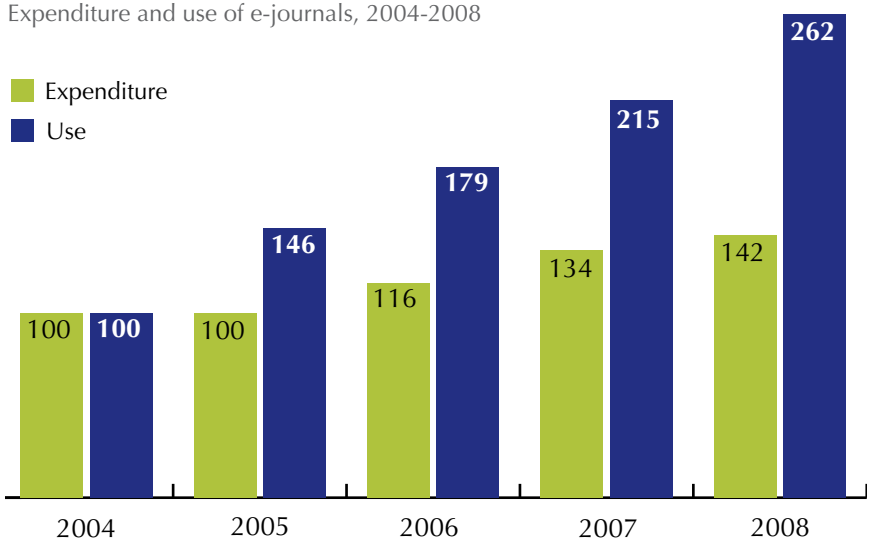
	2004	2005	2006	2007	2008
Russell Group	£1.73	£0.99	£0.82	£0.74	£0.66
Pre-1992 institutions	£1.20	£0.96	£0.98	£0.91	£0.81
Post-1992 institutions	£1.01	£0.85	£0.73	£0.68	£0.65
Whole sector	**£1.19**	**£0.91**	**£0.83**	**£0.77**	**£0.70**

Source: CIBER analysis of SCONUL annual returns for 112 UK universities [direct cost per download at constant 2008 prices using the Treasury GDP deflator]

Major efficiency savings in e-journal provision

As we have seen, expenditure on e-journals has risen sharply in real terms over the past few years but it remains a tiny fraction (less than 1%) of total university spending. At the same time, usage has increased dramatically, with 1.1 million full text article downloads in 2007-08 supporting the sector's mission of delivering high quality research, learning and teaching. Universities have thus been able to exploit new technologies and services to bring huge benefits and increased efficiencies: as shown in Table x, the average cost per download fell in real terms by nearly a third across the sector overall, with even higher savings in Russell Group institutions.

Figure 5: Expenditure and use of e-journals, 2004-2008



Expenditures on e-journals (e- and e+print combined) at constant 2008 prices versus COUNTER downloads, both indexed 2004=100

Do e-journals make a difference?

The use of e-journals and research outcomes: are they related?

The rapid growth in the use of e-journals is interesting in its own right, but it tells us nothing about whether enhanced access to the literature has had any impact on research outcomes. Does it make a difference?

In the first phase of the study, we began to explore this question by using SCONUL and HESA data for 2007-08 and building a model. Our aim was to investigate whether there are any relationships between levels of expenditure and usage of e-journals, and research outcomes.

We have taken that analysis a stage further in Table 10, which shows a number of partial correlations between aspects of library provision (the rows) and research outcomes (the columns) for 2007-08 at 112 UK universities. We control here for the varying sizes and subject profiles of the institutions, using total university spending and undergraduate users as proxies for size, and the proportion of research activity (based on RAE 2008 data) in science, technology, engineering and medicine (STEM) subjects.

We find that article downloads correlate positively, with few outliers, with all four measures of research success. The correlations are highly significant and independent both of institutional size and the balance of STEM research activity. None of the other aspects of library provision shows the same consistent relationship to outcomes, except in isolated cases (for example, between the volume of inter-library loans and numbers of PhD awards, and between the consumption of e-book materials and income from research grants and contracts).

These findings are intriguing. But correlations do not necessarily imply causation; and even if there is some causal relationship, there is no indication in which direction cause and effect might run.

Table 10: Library provision and research outcomes

	PhD awards	RGC income	Articles published	Citation impact
Database subscriptions	0.169	0.101	0.136	0.089
Book loans	0.183	0.080	0.168	- 0.056
E-book accesses	0.390**	0.470**	0.454**	0.278**
Interlibrary loans	0.416**	0.257**	0.290**	0.236**
Article downloads	0.724**	0.687**	0.721**	0.447**

*** Significant at the 1% level*

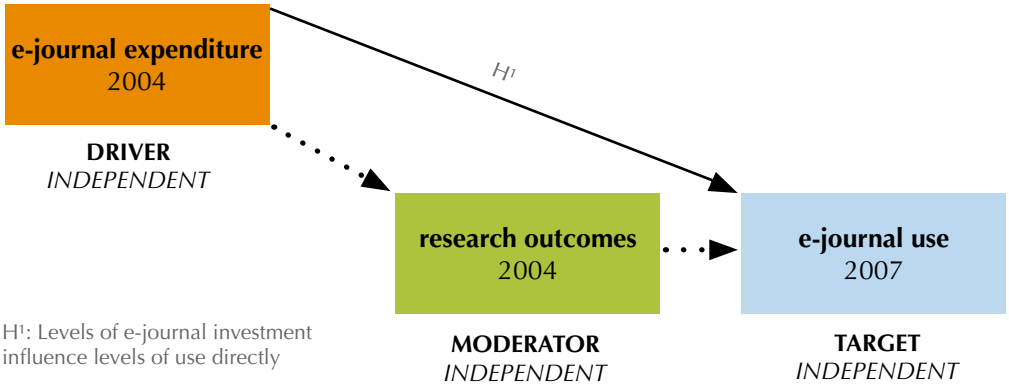
Source: CIBER analysis of SCONUL returns, 2007-2008

Hence we have attempted to build a more dynamic model, using data from a five-year period rather than a single year, to test a series of six hypotheses:

H1: Spending drives use (as in Figure 6)
H2: Use drives research success
H3: Spending drives research success
H4: Use drives spending
H5: Research success drives use
H6: Research success drives spending

We tested these hypotheses using a structural modelling technique, introducing a time lag of three years so that we could ask the question (as in hypothesis 1) `Is spending on e-journals in year one a good predictor of research outcomes in year three'?

Figure 6: 'Spending drives use' hypothesis



H¹: Levels of e-journal investment influence levels of use directly

A positive answer to this question still would not necessarily imply cause and effect. But it would nevertheless imply a much stronger relationship than a simple correlation. For it would indicate that if there is a change in the driver (in this case expenditure) in year one, there is a strong likelihood that there will be a change in the target (in this case usage) in year three. That clearly takes us some way further than the simple within-year correlations shown on the previous page. And because we can test the reverse hypothesis - that use drives spending (H4) - we can get a bit closer to understanding directionality as well.

In order to test the six hypotheses, we created three `baskets' into which we placed our statistical variables. We then used structural modelling techniques to reveal how good the levels shown by the data in each basket were at predicting subsequent levels in the other baskets, using all six hypotheses.

The results of our modelling are summarised in Figure 7 (overleaf). It shows that there are three strong driving relationships.

First, expenditure drives use. Indeed, expenditure is a precondition for use, since purchase of a licence or some other payment is required in order to gain access to any content that is not open access. The reverse hypothesis, that use drives subsequent levels of library spending, is not supported. The relationship is weak, probably because of the bundled nature of journal purchasing.

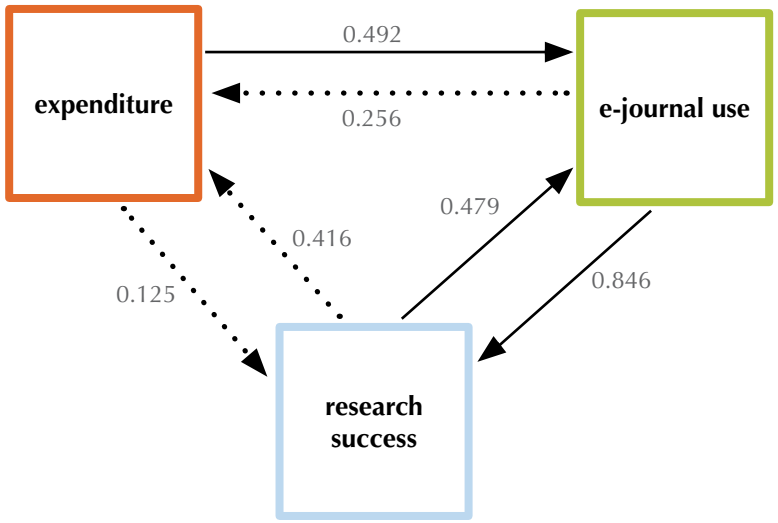
Second, and most powerfully according to the model, the use of e-journals drives subsequent research success.

Third, research success drives more usage of e-journals in the future. There is thus a strong positive feedback loop between levels of usage and research outcomes: they each feed off each other.

Other linkages are much weaker. Thus any direct relationship between expenditure on e-journals and subsequent research success is weak. This is probably because decisions to spend additional money on e-journals are unlikely to lead to tangible improvements in research performance in as little as three years' time. The reverse relationship is somewhat stronger, and it seems plausible that at least some decisions on expenditure are related in some way to a university's success, for example, in winning research grants and contracts.

None of this is to suggest that cause and effect have been conclusively established. There are many factors in the wider environment that are not included in the model, and it may be that some third element is at work as we demonstrate that levels of usage are a strong predictor of future research success.

Figure 7: Relationships between levels of expenditure and usage of e-journals, and research outcomes



Note:

The numbers above are `path coefficients' and indicate the degree to which earlier data (say for e-journal use) are a good predictor of subsequent outcomes (say for research success). The numbers range between 0 (no predictive power) and 1 (total predictive power).

Solid lines indicate that the relationship is highly statistically significant, dotted lines that it is not.

More detailed modelling and testing, for individual universities and groups of universities, and over different time periods, are required to test a range of hypotheses. Nevertheless, both libraries and universities should consider this evidence carefully in reaching decisions on the future development of their collections of e-content, and their services to support the effective use of e-journals.

Further reading

More information about this study, including detailed findings, aims and objectives and research methods, may be found on the RIN and CIBER websites.

CIBER working papers:

Aims, scope, methods and research context

Journal spending, use and research outcomes: a UK institutional analysis

Bibliometric indicators for the case study institutions

Information usage and seeking behaviour: subject and institutional profiles

Has wider access to the literature impacted upon breadth of citation?

RIN: www.rin.ac.uk/use-ejournals (further hard copies of this report can be ordered via contact@rin.ac.uk)

CIBER: www.ucl.ac.uk/infostudies/research/ciber/value

About the Research Information Network

Who we are

The Research Information Network has been established by the higher education funding councils, the research councils, and the national libraries in the UK. We investigate how efficient and effective the information services provided for the UK research community are, how they are changing, and how they might be improved for the future. We help to ensure that researchers in the UK benefit from world-leading information services, so that they can sustain their position as among the most successful and productive researchers in the world.

What we work on

We provide policy, guidance and support, focusing on the current environment in information research and looking at future trends. Our work focuses on five key themes: search and discovery, access and use of information services, scholarly communications, digital content and e-research, collaborative collection management and storage.

How we communicate

As an independent voice, we can create debates that lead to real change. We use our reports and other publications, events and workshops, blogs, networks and the media to communicate our ideas. All our publications are available on our website at www.rin.ac.uk

Research Information Network
96 Euston Road
London
NW1 2DB
UK

Telephone +44 (0)20 7412 7946
Fax +44 (0)20 7412 7339
Email contact@rin.ac.uk