Dotting the DOIs and crossing the ESSNs: Librarians' support for the RAE 2008

At the University of Birmingham, 2008 was the first time that library staff were involved in a Research Assessment Exercise (RAE), and by ensuring the completion of all record elements and the sourcing of over 4,000 items, they became an integral part of the process. This article highlights the information retrieval and metadata skills that they brought to the task, and the lasting benefits of involvement in the project, from better quality output data to enhanced processes for the ongoing collection and validation of publication information. In addition, there is a greater awareness of the skills and expertise of librarians in this area, which will remain even more pertinent for support in the successor to the RAE, the Research Excellence Framework (REF).

ANN-MARIE JAMES
Academic Support Team
Manager
Library Services
University of Birmingham

Introduction

In 2008, all Higher Education (HE) institutions in the UK submitted up to 70 Units of Assessment (UoAs) for the Research Assessment Exercise (RAE), the results of which will determine government research funding for each university, for a number of years. Based on peer review, a national UoA panel will spend 2008 reviewing each institution's research contribution, a key element of which is the 'outputs' that researchers have published, whether in the form of journal articles, books, conference proceedings or other. It is in this area that librarians were uniquely positioned to provide support in ensuring that data for all of the outputs was complete and correct. Moreover, for RAE 2008, for the first time, the full text of all submissions was required, providing a further challenge for information professionals to source over 4,000 items.[1]

At the University of Birmingham, two librarians were directly involved in the submission for the RAE 2008.[2] This article highlights the specific knowledge and skills that they brought to this process, and the important contribution information professionals can make in this area.

University of Birmingham context

Following the RAE 2001 and the administrative strain placed upon Corporate Services, an IT programmer at the University developed bespoke software to aid data collection for 2008. This system drew on existing staff, publications, and finance databases, allowing academic leads from each of the University's 50 UoAs to select staff and their 'top 4' publications for the submission. The system was automatically programmed to know which data elements were required for each format of submission (eg. a book, journal article, conference contribution, research report) and included sophisticated functionality, such as semi-automatic checking for a digital object identifier (DOI) for journal articles, and reporting mechanisms that flagged incomplete elements. Despite this, there remained a need for manual intervention, and a small project team including two librarians led on the completion of the RA2 'outputs' section of the return, ensuring the integrity of the data and sourcing of all items.

Utilizing librarians' information retrieval skills

Whilst the RAE system could check for a DOI against a journal article's reference, this only

worked if all the reference elements were in place. Understandably, there were hundreds of instances where only partial references were entered into the publications database and hence transferred into the RAE system, which required manual searching for DOIs in these cases. This ranged from a simple check on our e-journals system by journal title and then browsing to the correct article, through to more complex searches on bibliographic indexes such as Zetoc, Medline and Web of Science. There were also many instances of incorrect references, as a result of human error when entering the data, changes at the publisher end and abbreviated or incorrect journal titles. The librarians working on the system were able to bring their understanding of the information resources available and the skills to search them effectively in order to rectify these errors and complete the information required by the Higher Education Funding Council for England (HEFCE).

Understanding of metadata requirements

In addition, HEFCE asked for the exact date of publication for all journal articles. As this is not an element usually included in references it had not been recorded, and as a result it was manually checked for c. 3,000 journal items. In the sciences this was generally straightforward as the date was usually clearly displayed when a DOI was resolved, but for many other subjects this necessitated further investigation and a 'best guess' based on the publishing schedule of the journal. Whilst HEFCE rules were clear that an online date was required when the article had a DOI, a common problem in this area was when print and online dates were years apart; in some cases the printed date preceded the online date, as publishers digitized their back-files retrospectively, and, conversely, printed publi-cation was occasionally significantly delayed.

ESSNs (electronic ISSNs) were another new area to many of the RAE administrators. Again, while the system had in-built generation of an ESSN from an ISSN, few ISSNs had been recorded. Further issues arose, from having multiple ISSNs for journals such as the BMJ, to manual error on data entry; to tackle these, alternative sources such as Ulrichs were used to find and verify ISSNs. In other cases, for items like local history journals, ISSNs simply did not exist, and as the return was dependent on all elements being completed for all

records, this meant re-categorizing the publication to a report or 'other form of accessible output' which did not use the ISSN field. Being aware of metadata and field requirements from basic cataloguing was a distinct advantage here, as it engendered a natural understanding of record compositions and required elements.

Sourcing full-text articles

Unlike previous submissions, HEFCE required the full text of all items for the RAE 2008. In the sciences most journal articles had DOIs, and through their agreement with relevant agencies, the Panels will be able to resolve these to access the full text.[3] Around three-quarters of the University's submis-sion comprised of journal articles, and approximately 80% of these had DOIs. This, however, left about 600 journal articles to find. Many of these were straightforward, and accessible and available in PDF format from our e-journals system. Inevitably there were articles from journals to which the library did not subscribe electronically or in print, and the team liased closely with the authors themselves and departmental administrators in order to source the item, which could then be scanned. Versioning and authority issues were important here, as it became apparent that many authors did not actually have a final published version of their own article. In addition, basic scanning skills were required in order to ensure a quality copy with a reasonable file size. In a handful of cases the article was purchased electronically, using reputable and known online article databases to obtain copyright-cleared items cheaply and quickly.

As well as journal articles, there were several hundred books that had been selected and needed to be loaned to HEFCE for the submission. For the vast majority of items this was straightforward as authors supplied their own copy. In a few instances, expensive items for people who had left the Univer-sity were purchased, and this required discussions with publishers both at home and abroad. To this end, a small fund was set aside, and team members monitored and reported on this budget.

Problem solving and query resolution

As the project progressed, a number of queries arose, and these were collected on Google

Documents, allowing project team members to collaborate virtually to record and resolve problems. To begin with most of these were DOI resolution issues, relating to publishers not having registered DOIs, or inadequate mechanisms for alerting CrossRef to changes. As a result of the RAE, these systems were refined and the population and upkeep of the CrossRef master database was improved. The team then began to concentrate on other areas such as items appearing outside of the census date (1 January 2001–31 December 2007). There were a number of items that had not been submitted for RAE 2001 and were published in print in January 2001, but had technically become available online in late December 2000. Towards the end of the process, in November 2007, there were still around 30 items 'pending' publication; assurance of the expected publication date was sought from publishers in all these cases, and the team carefully monitored these.

Project management

In order to manage the outputs project, the team used basic project management methodologies. Crucially, this included key milestones and deadlines, particularly throughout the Summer of 2007, which was essential to the running of the project as the central team worked with more than 50 departmental administrators, and our metadata and sourcing work depended on confirmed staff and publications selection. Processes were established for ensuring the publications for new starters were selected, and alerting the librarians to this to complete the records and source the full text. The RAE system has its own reporting mechanism, so we were able to regularly review the number of incomplete records and unsourced items for each UoA.

Transferable skills

In addition to their technical information skills, the librarians on the project team found that good communication skills were needed, and it was these softer interpersonal skills which were vital to the successful completion of the project. This included everything from negotiating with academics to get hold of an item, to clarifying a publication date with publishers, or reassuring the

Vice Chancellor that we were ahead of schedule! Written guidelines and clear communication routes were also fundamental, as the team sent out regular briefings to the UoA administrators to interpret the HEFCE guidelines and request that specific actions be completed.

Long-term benefits of supporting the RAE 2008

A number of lasting benefits have resulted from librarians' involvement in this project. Firstly, the quality of the University's publications data has now vastly improved, through the checking, correction and completion of the metadata for all of the items submitted. In addition, building on the expertise developed during the project, all journal records in the University's publications database have now been enhanced, with DOIs, dates and ISSNs/ESSNs automatically generated and added to around 25,000 records. Academics generally are more aware of the value of DOIs, and some are including them in reading lists and linking to them from e-learning courses when directing students.

Furthermore, by collaborating with academics during the project, there is a far better understanding of the value of the metadata and information retrieval skills of library staff. This is being utilized in a number of ways: academics are encouraging their students to consult their librarians to help them find articles, and are making greater use of the support available for their current research; the work of sourcing articles has continued well into 2008, as panels request PDFs for items where the DOI has failed to work; and librarians are now also involved in the pilot for the Research Excellence Framework (REF), through which the fundamental issues around the integrity of publications data, and author and institutional disambiguation become ever more apparent.[4]

Conclusions

Relatively little has been written about librarians' contribution to the RAE 2008. This article has highlighted the important technical information retrieval and metadata skills which librarians brought to their institution's RAE submission. Whilst understanding of citation analysis and

bibliometrics generally is a new and growing area for the REF, peer review of outputs will still be used. Both ends of this spectrum rely on data integrity and in this respect, librarians remain central to the process.

References and notes

1. Research Assessment Exercise 2008
    http://www.rae.ac.uk/ (Accessed 26 August 2008)

2. Ann-Marie James, a.m.james@bham.ac.uk Academic Support Team Manager and Steve Bull, s.a.bull@bham.ac.uk Subject Advisor; Library Services, University of Birmingham

3. HEFCE (2007) Accessing and submitting research outputs
    http://www.rae.ac.uk/aboutus/policies/outputs/ licence/ (Accessed 26 August 2008)

4. HEFCE: Research Excellence Framework (2008)
    http://www.hefce.ac.uk/Research/ref/ (Accessed 26 August 2008)

Article © Ann-Marie James

■ **Ann-Marie James**
Academic Support Team Manager
Library Services
University of Birmingham
Edgbaston
Birmingham B15 2TT, UK
E-mail: a.m.james@bham.ac.uk