Abstract

Purpose – The purpose of this paper is to show the value of measuring usage of cultural heritage materials held by libraries.

Design/methodology/approach – The paper studies how best to record usage patterns of cultural heritage material.

Findings – Digitization has been found to make the measurement of usage of collections more useful. Experience so far has been that where older material is concerned, digitisation of the content does not mean that people neglect the physical volumes – instead, it tends to increase usage.

Originality/value – The paper looks at the complex issue of older material consisting of numerous volumes, and how best to record usage of such material.

Keywords Usage statistics, Libraries

Paper type Case study

In a large research library there are many materials which are being kept “for posterity”, that is, as part of the cultural heritage of the community. These materials are used infrequently, and some may never be consulted. The time-scale over which the materials are used may be very long. How can we begin to understand the patterns of use of this material, and define exactly what it means to keep something “for posterity”?

There is very little literature on the subject. Most studies of collection usage concentrate on the very real problems and issues surrounding the active circulation of current stock. One notable study was the “Pittsburgh study”, carried out from October 1968 to December 1975 at the University of Pittsburgh to measure which library materials were used and the cost of such use (Schad et al., 1979). Even this study looked at usage of material newly added to a research collection, and led to essentially unresolved arguments about the longevity of materials in such a collection. I have found no direct studies of rare and old material, whose value lies precisely in the fact that it has survived the centuries. For special collections, use rate is not the primary criterion for retention.

The problem of estimating usage of cultural heritage material has a parallel in the problem of estimating how many animals of a given species exist in a specific region – how many leopards are there in the savannah? In the whole collection of books, how many are “alive”, that is, in use?
The method, based on a Bayesian statistical approach, is as follows. We know that there are leopards in the savannah, but we do not know how many. They move about, they are elusive, and not seen very often. If we try to count them we do not know if we are counting leopards that we have seen before. The answer is to set traps, and to catch some leopards. When we have caught some, we tag them in some way, and then let them go.

Some time later, we return. The leopards which we caught last time are roaming about the savannah, and we do not know where they are now. We set some more traps and catch some leopards. Some of the leopards that we catch will turn out to be tagged – that is, they are the same leopards that we caught last time. In this case it turns out that one in three leopards has a tag. This strongly suggests that last time we caught a third of the available population of leopards, and this gives us an indication of how many leopards there are in total.

Of course there are various factors which will affect the outcome. Some leopards will have died, and some born. It may be that being caught once makes a leopard wary, and so less likely to be caught a second time.

Let us now turn from leopards to books. We do know how many books we have, but we do not know how many are “alive” – how many will be used. However, we can count uses each year, and we can identify items used in successive years. Just as with the leopards, we can see how many items are requested in each of the two years, and make some calculations.

First of all let us look at what would happen if the usage pattern is entirely random – in other words, like with the leopards in the savannah, it is entirely a matter of chance whether a book is used or not. Imagine a rare books department with 10,000 volumes, and suppose that 1,000 volumes are requested each year. That means an item has a one in ten chance of being requested.

If requests are random, then in the second year:

- There will be 100 requests (1/10) for items fetched the previous year.
- There will be 900 requests for items not fetched the previous year.

So over the two-year period different 1,900 items will have been fetched. In the third year, 190 of these (10 per cent) will have been fetched before, so only 810 new items will be fetched. We can extrapolate to see how this pattern develops over succeeding years (see Table I).

Cumulatively, after 30 years 9,576 out of the 10,000 items will have been requested. In effect, almost all of the collection is in active use. But remember the leopards: the chances of being requested twice may not be entirely random. In particular, the fact

<table>
<thead>
<tr>
<th>Year</th>
<th>Requested</th>
<th>Chance of repeat</th>
<th>Repeated requests</th>
<th>New requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,000</td>
<td>0.000</td>
<td>0</td>
<td>1,000</td>
</tr>
<tr>
<td>2</td>
<td>1,000</td>
<td>0.100</td>
<td>100</td>
<td>900</td>
</tr>
<tr>
<td>5</td>
<td>1,000</td>
<td>0.344</td>
<td>344</td>
<td>656</td>
</tr>
<tr>
<td>10</td>
<td>1,000</td>
<td>0.613</td>
<td>613</td>
<td>387</td>
</tr>
<tr>
<td>30</td>
<td>1,000</td>
<td>0.953</td>
<td>953</td>
<td>47</td>
</tr>
</tbody>
</table>

Table I. New items requested over time
that a book is requested in one year may make it more, rather than less, likely that it
will be requested again.

It is time to turn to some real figures. The Bodleian Library has one of the richest
collections of old and rare material in the world. The printed books, including those in
the Special Collections department, are all available for request on its automated
catalogue, so there are good data about usage. In 2006 there were 704,515 items listed
in the Special Collections areas of its closed-access bookstack. (A number of large-scale
book moves has taken place in recent years, making comparative use of more recent
figures problematic.)

In 2005 30,417 items were requested from these areas, and in 2006, 31,075. The
overlap of items requested in both 2005 and 2006 was 4,878 items.

With these usage figures – about 30,000 items out of 700,000 items – the overall
chance of an item being selected is 4.4 per cent. If the requests were entirely random,
we should expect about 1,700 requests to be repeated in both years, i.e. 0.2 per cent. But
as we have seen, almost 4,900 requests were repeated – about 0.7 per cent. So requests
are not random – being requested in one year significantly increases the chances that
an item will be requested in the following year.

What effect does this have on cumulated use over time? This is the table of
cumulated usage calculated on the basis of these figures (Table II).

After 40 years almost everything that is requested has been requested previously
(and growth has been slow for the previous decade). More significantly, this stasis is
reached when only 27.5 per cent of the collections have been consulted. The rest is
being kept, essentially, “for posterity”.

This has been just a first attempt at analysing the problem. There are various ways
in which the analysis can be improved:

(1) Data from a third year would be very helpful. The Bodleian’s collections are in a
state of flux as large numbers of items are being moved while we re-organise
our estate, and it was not possible to run the analysis for a third year with
comparable data.

(2) We have ignored multiple requests within a year. If an item was requested three
times in the first year and twice in the second, this still counts as a single repeat
request. By choosing to analyse at shorter time intervals we could get a clearer
picture.

<table>
<thead>
<tr>
<th>Year</th>
<th>Numbers</th>
<th>Percentage of stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30,417</td>
<td>4.4</td>
</tr>
<tr>
<td>2</td>
<td>56,614</td>
<td>8.0</td>
</tr>
<tr>
<td>5</td>
<td>112,091</td>
<td>15.9</td>
</tr>
<tr>
<td>10</td>
<td>159,340</td>
<td>22.7</td>
</tr>
<tr>
<td>20</td>
<td>187,652</td>
<td>26.6</td>
</tr>
<tr>
<td>30</td>
<td>192,682</td>
<td>27.3</td>
</tr>
<tr>
<td>31</td>
<td>192,855</td>
<td>27.4</td>
</tr>
<tr>
<td>40</td>
<td>193,576</td>
<td>27.5</td>
</tr>
<tr>
<td>41</td>
<td>193,607</td>
<td>27.5</td>
</tr>
</tbody>
</table>

Table II. Cumulated usage
We have taken the bibliographical record as the unit of counting, not the physical copy circulated. This is because the Bodleian Library does not always have item records for its older material. So an eighteenth-century journal with several physical volumes counts as one item, and if different volumes are requested in successive years, this still counts as a “repeat” request.

The method has been put to use in the University of Oxford as an element in the apportionment of costs in what is known as the “123-model”. Until a few years ago the costs of the library service were met by charging the academic divisions a “tax” on the basis of the numbers of their staff and students. The University now adopts a more nuanced approach and identifies three types of library costs: those attributable directly to the departments and divisions (e.g. use by students and staff), those arising generally within the library (for example, costs of conservation), and those arising from the Bodleian’s role as a national and international research resource. The cost of maintaining that proportion of the collections which is being kept “for posterity” is not charged to the academic divisions, but borne centrally by the University.

One major factor may change in the future the pattern of use revealed by the statistics. It seems a reasonable presumption that the main reason why requests for items are likely to be repeated is that people follow references in the literature. If a work is cited, then people interested in the topic will follow the citation and request the item. Catalogue records do not tell us much about the content of a work. How can we know about the detailed content of a book unless somebody refers to it?

Mass digitization projects such as that undertaken by Google, or the Early English Books Online Text Creation Partnership, can uncover unknown content in the collections. We can now search inside the full content of works that have been neglected on the shelves for decades. Experience so far has been that where older material is concerned, digitisation of the content does not mean that people neglect the physical volumes – instead, it tends to increase usage. It is going to become a lot easier to catch our leopards!

Reference

Corresponding author
Michael Heaney can be contacted at: michael.heaney@bodleian.ox.ac.uk

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints