



Citation for published version:

Chapman, A 2006, 'Pitfalls and blind alleys: perspectives on quality measurements', *Catalogue and Index*, vol. 153, no. Summer, pp. 2-4.

Publication date:
2006

Document Version
Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This article was published in:
Catalogue and Index, no.153, Summer 2006, pp. 1-6

PITFALLS AND BLIND ALLEYS: PERSPECTIVES ON QUALITY MEASUREMENTS
Ann Chapman, UKOLN

Introduction

The government talks about 'education, education, education', but increasingly libraries of all kinds need to look at 'quality, quality, quality' – quality of service, quality of resources, quality of cataloguing – in order to justify funding and staffing or provide figures for reports and assessments. In this talk I am focusing on the principles of quality measurement, the various ways to measure quality, the limitations of different methods and potential hazards along the way. The objective – what you wish to find out – is the most important thing you need to remember when designing a quality study and different objectives will require different methods. Over the years UKOLN has been involved in a number of studies that have examined the quality of catalogues in different ways and with different objectives in mind, and these studies form the background to my talk.

Quality measurement

First of all, what is meant by 'catalogue quality'? A simple definition is a catalogue without errors, but a catalogue could contain no errors and still be a less than useful resource. In the 1970s when the Centre for Catalogue Research (now known as UKOLN) was set up, four factors were identified as being crucial in the quality of a catalogue, and therefore for creating a reliable, trusted, useful and usable resource.

Firstly, the data must be accurate – if a book is by a specific John Smith, the catalogue needs to record that, and not that it is by Jane Smith, John Smith (1820-1877), John B. Smith, etc. Spelling errors need to be avoided; a missing letter can change the apparent subject of a work from 'teen films' to 'ten films'.

Secondly, the data should be consistent, both within the record and in relation to other records. So in a MARC record, if the place of publication is Edinburgh in field 260, it should also have the coding for Scotland in field 008. All the records for works in a series need to use the same series title, even if some of the records also contain variant versions of the same series title.

Thirdly, catalogues have a function – to enable users to find items and to distinguish between related items; the data in individual records is there to support that function. So records need edition statements and publication dates, added entries for joint authors, the physical format of the item, etc.

Fourthly, the data needs to be up-to-date; it's all very well having a very detailed, accurate and consistent record when it is finally added to the catalogue, but how long is it reasonable to keep material off the shelves because it's not yet been catalogued? And conversely, records need to be removed when items are withdrawn from stock.

Measuring tape or string?

So in order to find out the quality of our catalogues, we need to measure them. There are various methods we can use, but the objectives of a study will determine the most appropriate methods. And a word of warning; just because you collect data for purpose

A does not mean that that data will give meaningful information in terms of a study for purpose B. For example, circulation data can give information on who's using the library and which areas of stock are most used, but gives no information on the catalogue records for any of the stock, or how the users experience the catalogue.

Surveys use statistical analysis of collected data and are designed for specific purposes. Surveys can be designed to establish levels of errors, types of errors, user difficulties, and availability of records. There is a whole science of statistics, but there are two things it is important to get right, the sample size and the sampling frame; more on this a little later.

Questionnaires are another way of collecting data. It is important when designing questionnaires that you avoid leading questions – the phrasing should be as neutral as possible. If you are using response scales (*I am satisfied with the catalogue: strongly agree, agree, don't agree, strongly don't agree*) it may on some occasions be useful to not offer a fence-sitting choice (*neither agree or disagree*). This forces the respondent to make a plus or minus judgement. On other occasions the neutral response can be useful, and it may indicate general satisfaction with the current situation.

Toolkits provide a basic methodology that others can re-use. Because toolkits are designed to be used in a wide range of situations, users do need to be aware of local circumstances that may need to be taken into account. Some toolkits have built-in options to accommodate local practice, but not all do.

Survey samples

If you are going to carry out a survey, then the size and composition of the sample is crucial. To determine these, you will need to identify what you want to find out, and all the contributing factors you need to examine.

You need a sample of at least 100 to provide meaningful results that are statistically valid. But this only works if the analysis is simple: for example, '*does the BNB have a record for this title or not?*' For more complicated analyses, larger samples are required. If we are looking at the question '*do users understand abbreviations in catalogue records?*' and we want to compare understanding by both university staff and students, we need a minimum sample of 100 staff and 100 students. In the case of the BNB Currency Survey, where we were investigating a number of factors the sample was 1440 titles over a twelve-month period. This allowed us to produce figures for public and academic libraries, and to show the proportion of records that were (a) British Library created, (b) formerly CIP, and (c) CIP.

The sampling frame is the population you are surveying. If you suspect or want to identify, for example, if there is variation in acquisitions policies depending on the type of library, then you need to include a range of each type of library. If you know you want to find out if BNB records are available when required for both academic and public libraries, you need both types of library in the sample. In the BNB Currency Survey the sampling frame for public libraries included London boroughs, metropolitan authorities, and English, Scottish, Welsh, and Northern Ireland unitary authorities in order to get a representative sample.

So what quality measures are there? Over the years UKOLN has worked on a number of studies and surveys, and I'll be using these as illustrations.

Quality = type of record

A simple measure of quality is the type of record. Although the BNB Currency Survey is primarily concerned with the availability of records, it also records which type of record is found. The 'British Library created' record and the 'formerly CIP' record are both better quality than the CIP record which has data which can change before publication. If the levels of better quality records are high, this would indicate a higher quality overall of the catalogue. In this survey we assume that the records are correct, as there is no check between book and item, although occasionally an error (e.g. spelling) has been spotted and reported.

Errors in data

To examine record quality in more detail, specific studies need to be carried out. One such study is the Legal Deposit Libraries Shared Cataloguing Programme record quality audit, which UKOLN has carried out since 1994. The sample reflects the proportion of BNB records created by each of the libraries; the British Library creates around 70% of the records and the other five libraries (Oxford and Cambridge Universities, the National Libraries of Scotland and Wales and Trinity College Dublin) each create around 6%.

A checklist of errors has been compiled and each of the 145 records in a sample is checked in detail. Errors include mismatch of 008/260 data, wrong indicators (e.g. for non-filing characters), missing added entries, etc. The sample is compiled to include several groups of related records (e.g. pupil and teacher versions of educational works, or series where several works have been published at the same time) so that consistency of treatment can also be checked. Errors are reported to the originating libraries and used to inform staff training. There are, of course, limits to what can be assessed in this way, as the physical item is not seen.

Edits = errors

Another way of assessing record quality is to look at any editing which takes place after initial record creation. In 1992-1993, UKOLN worked with Talis (then BLCMP) to look at record editing in a shared catalogue environment. Talis were able to identify the percentage of records that were edited after creation; this gave a baseline for catalogue quality.

Then a sample of 1310 'before and after editing' pairs of records was taken. These record pairs were then analysed by UKOLN. There were a number of conclusions:

- Monograph record edits are most likely to be author or imprint and physical description.
- Serial record edits are most likely to be title or imprint and physical description.
- The most frequent edit type was to add data to an existing MARC field.
- The second most frequent edit was to add a MARC field.

This analysis provided no verification of the need to edit; the records were not checked against items at any point. Therefore the final part of the study looked at small sets of edited records submitted by each member library. These records were annotated with the reasons for editing. The results of the whole study were used to revise the cataloguing guidelines for member libraries.

Quality = user understanding

Catalogues exist to help people identify, select and obtain items, and therefore record quality is of importance. But in addition to measures of accurate, consistent data and how quickly records are available, the records also have to be understandable to the end user. The Cambridge University Study in 1992 looked at this area. The 300 interviews were collected at random, but although it is known which responses were from staff and which from undergraduates and indeed which faculty area, the distribution of interviews collected was not balanced in any way. This meant that no conclusions could be drawn for the various user types, and the study only measured understanding overall.

There were three main areas of difficulty for users. Firstly, there was misunderstanding of data (e.g. there was often confusion about the holdings statements for periodicals, and also about the pagination statement where this included both Roman numerals for the introductory matter and Arabic numerals for the main text). Secondly, there was difficulty with OPAC labels (what is a uniform title?). And thirdly, abbreviations were often misinterpreted (e.g. the 'c' for circa was often identified as meaning copyright).

It would perhaps have been useful to be able to distinguish between staff and student, and / or faculty area responses; such information could have provided ideas for library induction or familiarisation sessions. However, to do this a sampling frame would be required, identifying how many people from each category needed to be interviewed. The actual sample size would depend on how detailed a breakdown was required. A staff/student breakdown could have used the 300 interviews, as long as there were equal numbers in each category. For staff, student and faculty breakdown, a larger sample would have been required.

Decide on parameters

I mentioned when talking about toolkits, the need to be aware of local circumstances, which may affect results. The Cat-Assess Tool, developed at UKOLN 1999-2001, provides libraries with a methodology to assess the quality of their cataloguing. In this toolkit, the first step in the process is to record local cataloguing policies; if the local policy is not to include pagination at all, then it is not an error when assessing that catalogue. The local policies may indicate that records in this particular catalogue are less full than ones in another catalogue, and in that sense they may be of lesser quality – though they may still have sufficient data to satisfy the needs of the users of the catalogue being assessed. The toolkit provides guidance on selecting the sample(s) and using the results to calculate error rates.

In the BNB Currency Survey it was important to only include in the sample works which came within the BNB remit; various items are excluded from BNB and if they were to be included in a sample, the results would be skewed.

With the LDLSCP Audit, there are two parameters. Firstly, only records created under the LDLSCP are assessed. (The assessment itself is limited by not having sight of the actual items being catalogued – some errors can only be picked up by comparing the item and the record.) Secondly, to assess issues of consistency, some groups of related records (e.g. pupil/teacher versions of a work, several items in a series) must be in the sample.

Re-using data

When a lot of work has been put into collecting data, it is very tempting to try and re-use

the data, and draw additional conclusions from the results. It is important to be careful when doing this; often a study designed to collect one type of information cannot yield other data.

In the case of the BNB Currency Survey, it was possible to widen its scope somewhat, though this was done partly by deciding to collect additional data. The '*acquisitions trends over time*' analysis was done by re-analysing existing data and by adding data. The *subject analyses* could be done by re-analysis since that data was already held; the *acquisition date and publication date comparison* analysis required additional data supplied by Whitaker.

For the *retention of stock* part of the acquisition trends review, again additional data was needed. Libraries were contacted to see how many items from samples submitted for the BNB Survey were still in stock ten years later.

Over a three-year period (1996-1999) additional data was collected to measure the availability of catalogue records from suppliers other than BNB. This did enable a comparison to be made between suppliers, but the supplier record profiles indicated that the comparison would not always be 'like for like'. For example, Whitaker and Book Data had both pre- and post-publication records, but CURL only contained post-publication records.

Because the survey measured BNB records over a long period, it was possible to use the results to monitor the effectiveness of changes in British Library cataloguing policy. For example, at one point it was decided that some items would only be catalogued to level 1; this was intended to speed up cataloguing and get rid of some of the backlog that had accumulated. The survey was able to identify through the raised hit rate that this was effective. The survey also showed some lowering of the hit rates at the points when first acquisitions and then cataloguing moved from London to Boston Spa.

So remember ...

So at the end of this quick tour round the world of performance measurement, if you are about to embark on a study of some aspect of record quality, there are a few things to remember.

What do you want to measure? You do need to be clear about this and record it when you start planning your study.

Do you already collect data that you can use? It is always worth looking at data you already collect. Even if it is not quite right for your study, it may be that slightly modifying some existing data collection will be sufficient.

If you don't already collect the right type of data, then look at all the different methods and choose the most appropriate.

And always, always, always write up the methodology and results. This means that other people can try the same methodology for comparison studies, and you can repeat the study at a later date to establish whether quality has improved.

This paper is based on a presentation given at Umbrella 2005, University of Manchester, Manchester, 2nd July 2005.

Author notes

Ann Chapman is an Interoperability Focus Officer at UKOLN¹, with special interest in Bibliographic Management² and Collection Description³.

¹ <http://www.ukoln.ac.uk>

² <http://www.ukoln.ac.uk/bib-man/>

³ <http://www.ukoln.ac.uk/cd-focus/>