

# Authority Control Simply Does Not Work

F H Ayres

**ABSTRACT.** Demonstrates through case studies how authority control simply does not work. Shows how the case studies were carried out using BOPAC2 which provides facility for downloading large files and a greater range of displays than normal OPACs. Stresses that authority control is important not only to library catalogues but also to information on the Internet. Because it is so important and because it is so expensive priority action is needed to rectify the situation. Suggests a number of ways in which this might be done. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-342-9678. E-mail address: <getinfo@haworthpressinc.com> Website: <http://www.HaworthPress.com> © 2001 by The Haworth Press, Inc. All rights reserved.]*

**KEYWORDS.** Authority control, OPACs, BOPAC2, Internet, bibliographic control, cross references

I am aware that the title of this article is a very controversial statement and I do not make it lightly. What follows is based on a posting to the Discussion List for the recent Conference on the Future of Bibliographic Control. For a number of reasons it was never posted mainly because it was far too long. I was also concerned that what I had unearthed might be a small can of worms or possibly just a few red herrings. For that reason I circulated it to a number of people who might be interested in order to get some reactions. As a result I got a number of replies which I have commented on after my case studies. Apart from one, none of them were prepared to say that what I had produced was wrong. As a result I have decided to give my findings wider publicity.

---

F H Ayres, PhD, is Bibliographic Consultant, Department of Computing, University of Bradford, West Yorkshire BD7 1DP, United Kingdom.

Cataloging & Classification Quarterly, Vol. 32(2) 2001  
© 2001 by The Haworth Press, Inc. All rights reserved.

I would like start by raising a number of points on authority control. At its simplest and least controversial the hypothesis is that authority control works better in theory than in practice. The extreme hypothesis is that authority control simply does not work. I suspect that the truth lies somewhere between the two. However, I am sure that we cannot continue to rubbish search engines and highlight the chaos of the web unless we can show that the claims that we make for the library catalogue and authority control can be substantiated. We need to put the necessary changes in hand that will make authority control what we claim it to be. Information management on the Internet needs authority control but it only needs it if it works consistently and exhaustively. We spend an immense amount of money on setting up a structure to ensure that our authority control tools cater to every contingency, however unlikely. We appear to spend very little in assessing whether the end users are getting the correct answers to their input.

We have in bibliographic control a wealth of experience to contribute. However, although we have convinced ourselves that the methods we use are producing effective and efficient catalogues we have not convinced our users. They continue to find the catalogue a complex tool but what is worse, as I hope to demonstrate, they miss material that they should find. Probably in much the same way that librarians say the search engines do. The danger for librarianship is that our users are beginning to turn their backs on the library catalogue in favour of search engines.

This would be a tragic waste of all the work that has gone into bibliographic control and the complex structure that we have built up over the years with our professional expertise. The library catalogue has two important major advantages over the search engine. The first is authority control which should, if it was working properly, ensure that the user was confident that a search produced all the relevant material held by the library. The second is the display of search results in an organised way, something that the search engine cannot match. Both of these major advantages are being dissipated. In the case of authority control there is a reluctance to accept that it often simply does not work and when the deficiencies are acknowledged the action necessary to correct them are thought to be complex and long term. OPAC display is a different sort of problem. With value adding OPACs like BOPAC2 we should be moving away from the present traditional OPAC which is pre-coordinated by the cataloguer using very complex standards to the concept of post coordination where the user is able to select the type of display that is required. In addition it seems that the vendors dictate the limits within which individual OPACs have to function. With the present state of advanced database management global updates should not be a problem and cataloguers should not be expected to make the necessary changes when a heading is altered. I would now like to look at a number of key factors.

### ***CROSS REFERENCES***

Whether it has been forced on libraries by the vendors or by the cost of implementing comprehensive links the root of the problem is that the cross reference has been given a very low priority and in many important cases is missing altogether. We must ensure that the cross reference links headings which are synonymous and that only one is used as the main one. For example there are forty-one variations in the way Dostoyevsky's name can be spelt in the roman script alone. Often only one or two cross references seem to be used. What is far worse, material is likely to be missed because it is very common for more than one variation to be used as a main heading. With the computing power that is available today the catalogue users should expect every known variation of a catalogue access point to have a cross-reference to the chosen authority control access point. The claim that authority control improves the precision of searches and provides collocation is only true when it is applied exhaustively. Dostoyevsky is a good example. There are many cases when users finding entries under one variation of the name assume that this constitutes the whole of the libraries' holdings of the author's works. (See Case studies 1 and 4.) The sad fact is that the Library of Congress Authority Files although they are impressive in the amount of material they include and the cross reference and reference structure that they provide rarely achieve their potential because the reference structure seems to be used selectively and not comprehensively.

### ***SECOND INFORMATION REVOLUTION***

It is probably true to say that the Internet heralds the second revolution in information. The first revolution took place half a millennium ago with the invention of printing. This dramatically speeded up the way in which information was disseminated. However, it took mankind over five centuries to fully capitalise on the ability to transfer information and to devise methods of information retrieval. This is perhaps a misnomer as we never ever retrieved information. What we retrieved was the containers of information. The Internet, which is responsible for the second information revolution, will have an impact as great as the invention of printing but will differ from it because the effects will not take half a millennium before they are fully felt; the second information revolution is taking place almost while we watch. What is more when it works effectively, it retrieves the information itself almost immediately. In the area of bibliographic control the Internet has opened a number of doors. It is now possible to access nearly every important catalogue through the Internet and many of them belong to clumps or clusters which can be accessed as a group because they are linked by the Z39.50 protocol. What is

not clear is how or whether bibliographic control can influence information management in an Internet environment. The library catalogue is only one of a number of information sources along with search engines, internet directories and online databases. It is important that a means of integrating them all is given high priority.

### ***BOPAC2***

I am fortunate in having worked on and helped to develop a new type of tool that makes it possible to test whether authority control is working. This is BOPAC2 and it is not my purpose to use this Forum to advertise the advantages of BOPAC2 but I need to briefly describe the tool that I am using. BOPAC2 is not in itself an OPAC but it adds value to existing OPACs. It is able to download quite large retrievals. For example up to five hundred records can be easily downloaded within a reasonable time and if you are prepared to spend a little more time, a retrieval of a thousand is possible. Once downloaded the retrieval can be displayed and examined in a number of ways as I will demonstrate in the Case Studies that I give. Movement from one display to another is very fast. BOPAC2 can be accessed at <<http://www.bopac2.comp.brad.ac.uk/~bopac2/htdocs/evaluate.shtml>>.

### ***OVERALL VIEW***

I am sure that what I have written is going to be controversial as it attempts to question the viability of authority control which is one of the cornerstones of modern cataloguing. I believe that it is important to do this not because I believe that authority control is unimportant but because I believe that it is not working effectively. I believe that the profession, as a matter of priority, should be looking at ways in which authority control could be made to do the things that we claim for it. It is a long time since I catalogued a book and I am not pretending that I approach this subject with a cataloguer's viewpoint or expertise but I think that I am looking at it from the catalogue user's point of view and with twenty years of researching into various aspects of bibliographic control.

It seems that somewhere along the line the cross reference and the see also reference have been sidetracked and OPACs on the Internet rarely if at all provide links via the cross reference and the see also reference. This is in spite of the fact that our complex standards provide for their use. Why is this? I suspect that there are two reasons. The first is that the standards themselves are so complex. Cataloguers tend to be perfectionists and rules for descriptive cataloguing legislate for every possibility even when our users are not even aware that the possibility exists. The second is economic. Authority control as it is applied today is very labour

intensive and it costs money to ensure that a cross reference is in place. No one disputes that cross references are useful. There seems to be little awareness that they are essential if we are to ensure that authority control works properly. All the work that we put into authority control is wasted without a comprehensive set of cross references to ensure that our users are not misled into believing that the heading that they have chosen produces all the material that they are looking for.

The card catalogue usually provided the cross references and the see also references that enabled users to move around the library catalogue in much the same way that links enable users to surf the internet. However, having provided the structure that creates these references the attitude seems to be “there are too many of them” and “where do you draw the line.” This attitude must be wrong when at least one of the search engines is indexing over a billion web pages. Both the search engine and the library catalogue need authority control to ensure that there are no breaks in the links which join like to like and related to alternative search headings. The library profession cannot claim that they have powerful tools in the see and see also references which the search engines lack and then neglect to ensure that their own catalogues always use them. We cannot claim as many librarians often do that the Internet is a chaotic muddle of information compared with the discipline of the library catalogue. Library users are misled into thinking that the entry point that they have chosen has produced all the material held by the library. Priority must be given to putting this right.

My case studies show that the power and potential given to the library catalogue by authority control does not work unless cross references are used comprehensively and not selectively. If authority control can be made to work for the library catalogue it could be made to work for the search engine.

Many cataloguers if pressed will probably say that authority control works better in theory than in practice. I believe that these case studies have shown that it often simply does not work at all. There is a mass of literature on authority control most of it saying how valuable it is but very few examine how effective it is and how much it costs. As far as costs are concerned there is ample evidence that it is one of the most expensive components of the cataloguing operation. Librarians are often quoted as saying that the Web is chaotic and that the search engine cannot be compared with the library catalogue. This is in spite of the fact that the search engine uses relevance ranking techniques which are remarkably successful in bringing the required information to the top of the retrieval. It also searches a large amount of material very quickly and users find it much easier to manipulate than the library catalogue. However, it lacks authority control which is perhaps the most important contribution that the library catalogue has to offer to information management. But it has to be made to work properly. The links in the form of cross references and see also references must be comprehensively applied. As a matter of urgency the profession should give priority into ways in which the cross

reference and see also reference can be brought back fully into our automated catalogues.

All of my case studies can be tested fully by anyone with access to the Internet by using BOPAC2 and the methods that I have described. The evidence is there for anyone to see that authority control does not always work. My case studies I believe do provide valid evidence.

### ***POSSIBLE ACTION***

This is not meant to be a negative or destructive exercise. My object is to highlight a problem in the systems that we use to construct our catalogues. Authority control is too important to abandon. We need to put in place the changes that will make it do the things that we claim that it does. There are a number of steps that should be taken and taken as a matter of urgency. This is by no means a definitive list of the necessary action. In any case I am not the best person to provide such a list. My purpose is to alert the profession to a problem which needs urgent attention.

If you disagree with me then prove that I am wrong. If you agree then there are a number of areas of action that might be taken. The most important, however, is to ensure that authority control is applied consistently and comprehensively.

At present there are far too many cases where our standards provide options. These should not be provided where they are likely to cause conflicts with authority control.

Use the Marc format itself to automatically create cross references, For example nearly all the variations of Dostoyevsky's name are to be found in 245c.

Restructure the Marc format so that it enables the same procedure for automatically creating cross references using other parts of the bibliographic record.

Give priority to establishing methods that ensure that the cross reference is used to prevent two different headings being used for the same author or subject.

Create cross references for variant spellings of established headings.

Build up nationally and internationally a database of alternate spellings and misspellings which could be used to create cross references.

Fund research to first establish the extent of a particular problem and then the possible ways it can be righted.

There is a clear distinction where funding is required for development rather than research and the former is usually difficult to obtain. Paradoxically development funding can often be more effective than research funding. Although both are important, at this point in time it is more important to put the emphasis on what needs doing rather on what might need doing.

Global changes to authority control are feasible now with modern database management technology and implementation on all systems should be given priority.

### *CASE STUDIES*

I start from the assumption that the preferred entry point for library users is the one that they are using and that if their choice is different from that used by the catalogue they are entitled to be provided with a cross reference. All the examples that I am giving used the LC OPAC on the assumption that if anyone is likely to get it right it would be the Library of Congress. However, I hope that the Library of Congress will not feel that they have been singled out. Similar tests have been carried out on many other OPACs with similar results. Although it is not always possible to detect whether a cross reference is available on a Z39.50 retrieval, it is obvious as the case studies show when one is not available.

#### *Case 1. Dostoyevsky*

The entry in Collins English Dictionary is under Dostoevsky but three other variations are noted Dostoyevsky, Dostoevski and Dostoyevski. A search was made under Dostoyevsky and a retrieval of 80 downloaded. The retrieval was scrolled through in the MARC format display after using the FIND facility to highlight the field 245. This revealed thirteen variations to the original four in the spelling of Dostoyevsky's name. An author search under all these variations of the name resulted in the following:

Dostoyevsky 329,

Dostoevskii 223,

Dostoevski 5,

Dostoevsky 1,

Dostoevskogo 1.

The remaining twelve variations had no retrievals. If authority control is working I would expect that the preferred term would have links to all other variations and that none of the variations would be used as headings.

### *Case 2. Creativity*

Working on the assumption that if a word is used in the title it is likely to be a prime candidate for a subject approach, a search was made using title contains 'creativity.' This resulted in 1259 records. Using the 'select subject' option on a retrieval of 400 of these records showed that 398 headings were used. Of these only 2 headings and 4 records used creativity while 10 headings with 158 records used creative ability. As creative ability and creativity seem to be synonymous one should be linked with the other by a cross reference. Searching under 'subject contains' creative ability produced a retrieval of 1941 while creativity one of only 182. This seems a clear case where authority control should have used one or the other.

### *Case 3. Smoking*

'Subject contains' produced a retrieval of 1145 and 'title contains' one of 891. A retrieval of 80 of the 'title contains' was examined in the MARC format display. After using the FIND facility to highlight smoking and then scrolling through it showed that eleven records had no subject containing smoking. There were however subject entries under; cigarette habit, tobacco habit, cigarette smokers, cigar smokers, tobacco. This means that over ten per cent of the records would be missed by an approach using smoking and probably many more using the alternatives.

### *Case 4. Cummings E*

The correct form seems to be Cummings E E (Edward Estlin) Searching under Cummings E produces a retrieval of 173 and under Cummings E E 163. However under Cummings Edward Estlin only 52. This shows library users would miss a great many records if they used Cummings full name Edward Estlin and they would even miss some if they used EE instead of E.

### *Case 5. Sexual Harassment*

There seemed to be several anomalies. A search using 'subject contains' sexual harassment produces a retrieval of 549. 'Select subject' broke down the retrieval as follows: sexual harassment, sexual harassment of women, sexual harassment in education and sexual harassment in universities and colleges. Two further head-



ings are sexual harassment against women and sexual harassment in colleges and universities. These only have one retrieval and are obviously errors. One anomaly is that sexual harassment in schools is covered by sexual harassment in education. This subject heading does not cover sexual harassment in universities and colleges which has its own heading. There should be cross references and see also references to link this group together.

#### ***Case 6. Secularization***

Secularisation is an alternative spelling for secularization and should have a cross reference. However, there seems to be a distinction between secularization and secularism but it would need an expert to tell the difference and the catalogue user needs some guidance. A search using 'Title contains' secularization produced a retrieval of a 100, 'title contains' secularisation 3 and 'title contains' secularism 158.

#### ***Case 7. Codices***

One of these is Codex Vaticanus. A 'title contains' search produced a retrieval of 50. Scrolling through the Marc format display showed that 28 had no 6XX or 7XX tags. This means that a subject search would fail more than 50 per cent of the time.

#### ***Case 8. NFPA***

This is the abbreviation for National Fire Protection Association. Searching under the organisation's full name there are 395 records but none under NFPA. 141 of the 395 have a mention of NFPA somewhere in the record.

#### ***Case 9. Health Insurance***

This case study illustrates a number of points. The first is the widespread use of what I call 'librarian speak.' In the card catalogue the inverted heading served a useful purpose. In today's climate where the use of the Internet is based on direct links they only serve to confuse the user. A search under subject contains 'health insurance' produced a retrieval of 511 with over thirty inverted headings out of about 400 headings used. While the same search using 'insurance health produced 4551. At present the size of a retrieval that it is possible to download using BOPAC2 is a thousand. Examination of this retrieval using the 'select subject' display showed that nearly five hundred different headings were used and of these over fifty were inverted. In order to further test the use of inverted headings searches were made on three more inverted headings. One was under the inverted

heading only but the other two used both. Medical care, Cost had 1089 records and Cost of Medical Care 2 while Insurance, Life had 3303 records while Life Insurance had 477. The 'select subject' option display provides an alphabetical list of the subject headings and scrolling through the list reveals that any user would be hard put to find what motivates a decision to use inverted headings.

### FINAL ANALYSIS

I am grateful to a number of people who commented on an earlier version of this study. Their comments were very valuable in giving me a better picture of the present position but none of them altered my assessment of authority control as it is practised. When I started examining how authority control worked I was not sure whether I had unearthed a small can of worms or merely a few red herrings. I am now fairly sure that I have brought to light a substantial defect in catalogue construction. Nearly all those who commented made the point that authority control worked better in theory than in practice. However, if it works better in theory than in practice, then my case that authority control simply does not work is correct. For example if there are cross references to all the forty-one versions of Dostoyevsky's name and entries for his works still appear under several versions then authority control is simply not working. If it was then the cross reference should ensure that there are no entries under a particular version of Dostoyevsky's name. This is the most important point that is being constantly glossed over. Just loading the cross references is not enough. Systems should be designed to ensure that a cross reference is a cross reference and is not also used as a main heading. If it does then authority control simply does not work.

One comment was that Z39.50 did not support cross references and suggested that this had caused some of the anomalies that I had found. I agree that this is a limitation but it did not affect my findings and this can be easily checked by doing the same search direct on the LC OPAC. It was also suggested that Z39.50 still needs a lot of development to make it a fully functional method for retrieval. Z39.50 is being constantly improved but even using an older version BOPAC2 is still able to outperform the traditional OPAC. I have made the claim before that it is nearly always quicker and more effective to search the LC OPAC through BOPAC2 and I stand by this claim. I am a heavy user of LC OPAC but mainly through BOPAC2.

I accept that much of the weakness in authority control is historical mainly through the bringing together of several catalogues or the adoption of new systems. When these processes are introduced the maintenance of an important element like authority control must be a top priority. The management of large databases is at an advanced stage and alterations to large sections of the database

should not present a problem. When changes are made to authority control these should be implemented using software that is capable of making global changes. This does not seem to be happening and instead a labour intensive exercise is carried out at the local level. Research at the Department of Computing at Bradford University over the last twenty years has involved the handling of millions of MARC records and the use of quite complex operations without any major difficulties.

A number of reasons were given for the anomalies that had come about. Some of these were historical and some technical. I understand these reasons but they merely reinforce my contention that authority control simply does not work in a surprisingly high number of cases.

My general suggestions were done in some haste. I have now made considerable alterations but they are still tentative. I did however get some interesting but sometimes conflicting comments. As I have already said I approach the problem as a user and the user is only interested in one heading and that is the one that he is using. It is therefore important that it works all the time. I was also alerted to a number of relevant research initiatives. These included the simplification of AACR2 and the MARC Format, the analysis of AACR2 as a data model and the linking of our models for bibliographic control with the data models of other sectors, the development of crosswalks from MARC to emerging data standards like XML and user studies. Important as all these are none are as important as getting authority control functioning properly.

If I am wrong then I shall retire gracefully. If I am right and the evidence points to this being the case, then action is needed immediately.