



ALIA 2006 Biennial Conference



Australian Library and
Information Association

Refereed Paper

Dr David Wells

Acting Resources and Access Librarian
Library and Information Service
Curtin University of Technology.

Contact details

Postal: Curtin University of Technology, Library and Information Service

Email: d.wells@curtin.edu.au

Biography

David Wells is acting Resources and Access Librarian at the Library and Information Service, Curtin University of Technology.

What is a library OPAC?

Abstract

OPACs are now almost universally central to the operations of libraries of any size and represent the primary point of connection between library users and those information materials which the library owns or otherwise wishes to promote. This paper looks at OPAC operations in the light of the philosophy of information, communication theory and semiotics, and revisits fundamental questions about the nature of library OPACs and the ways in which they function. This exercise can conveniently be pursued through examination of the multiple tensions implicit in any cataloguing or indexing system: the competing requirements of organisation and retrieval, description and classification. For an OPAC to be practicably usable the potential complexity of the information it contains must be reduced to a manageable level of simplicity. OPACs purport to allow complete retrieval, even though data structures and index definitions are not transparent. They necessarily impose a framework of retrieval which conceals information at the same time as it presents it, and therefore to an extent they undermine the very concept of information literacy. In an online environment catalogues have been able easily to move beyond the scope of individual library collections and to incorporate an increasing range of non-traditional types of information, thus introducing further tensions between their roles as indexes and as information sources in their own right. The multiple tensions involved can be conceptualised in terms of indeterminacy of the communication code and indeterminacy of the linguistic sign. A clear theoretically based understanding of these issues is essential to ensuring optimal OPAC design.

Introduction

Even though the OPAC now functions against a background of alternative information-gathering technologies it is likely to remain at the centre of library operations for the foreseeable future as the primary automated point of connection between library users and those information resources which the library owns or otherwise wishes to promote. In the context of mass digitisation, internet search engines may be many people's research instrument of choice (Haigh, 2006), and the development of electronic journal publishing may have provided easier access than ever before to journal content through federated search interfaces and link resolvers, but OPACs still have a necessary role in preserving access to library collections, maintaining a consistent and authoritative form of bibliographic control and providing a targeted information environment for specific client groups (Colhoun, 2006). The OPAC is thus a crucial site of mediation, but how does this mediation actually work? What is the nature of the act of communication in which the library user engages when interrogating the OPAC? What are the structural limits of this act of communication and how can they be minimised or avoided? What, in short, from a theoretical point of view is a library OPAC?

The present paper looks at these and related questions using perspectives from the fields of semiotics, communication theory and the philosophy of information. It suggests that far from being a simple conduit of data from database to user, the OPAC, like any communication device, is in fact

a contested site, a space in which multiple discourses co-exist and serve to interrupt and dislocate the process of communication. The paper falls into three main parts. First I will trace the act of questioning the OPAC from the point of view of the theory of communication. Secondly I will highlight some of the major areas of indeterminacy which influence and undermine the communication process. Thirdly I will suggest some practical ways in which these indeterminacies can be in some degree overcome. Asking fundamental questions about the nature of the OPAC is not merely a topic of theoretical interest: it should create a grounded basis for ensuring effective OPAC development and design, and at the same time help define clear priorities for information literacy.

At the highest level, the library OPAC has at least three distinct functions. First it acts as a bibliographic database, an electronic version of the card catalogue which it replaced, acting as an index for the user in search, for example, of a particular book. As a logical extension of this, the OPAC increasingly also provides links to electronic texts, freeing the user from the necessity of physically locating material on the library's shelves. Secondly it functions as a "portal" in a way not dissimilar to a library homepage, providing links to non-bibliographical data, either relating to users themselves – information about overdue books, fines etc. – or to other library information such as opening hours. In principle this portal function could be extended indefinitely to connect to a variety of data considered to be of interest to library users. Thirdly, the OPAC functions as a promotional

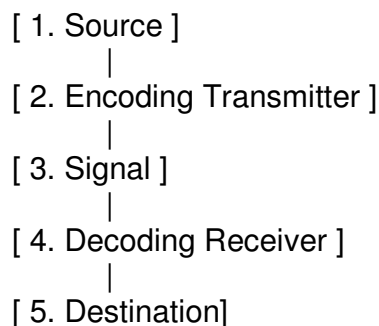
artefact, advertising the presence of the library and the services it can provide, and at the same time making a statement of authority about the communicative links that are supported and facilitated. The question of authority is an important one in the process of information creation and transmission, but it is the bibliographic function of the OPAC which will be the chief focus of what follows

The OPAC potentially has a fourth function as the management of full-text data and the management of bibliographic data converge and the bibliographic function of the OPAC itself becomes enabled for full-text searching

rather than remaining primarily an index. This technology is not widely implemented in current OPAC installations, but is likely to develop, and will add a further layer of complexity to the issues raised below.

The OPAC as Communication System

Let us look at the operation of the OPAC in terms of what is generally known as the Mathematical Theory of Communication. In simplified form (adapted from Shannon & Weaver, 1949; Eco, 1976; Floridi, 2004), this provides the following model:



The five basic steps from Source, through Encoding Transmitter, Signal and Decoding Receiver to Destination are linked by a shared Code, which is a shared system of values or meanings. Without this Code, no communication would be possible, since it would not be possible to pass meaning from one stage to the next. In terms of this model, an OPAC search constitutes a double communicative act: a question followed by an answer.

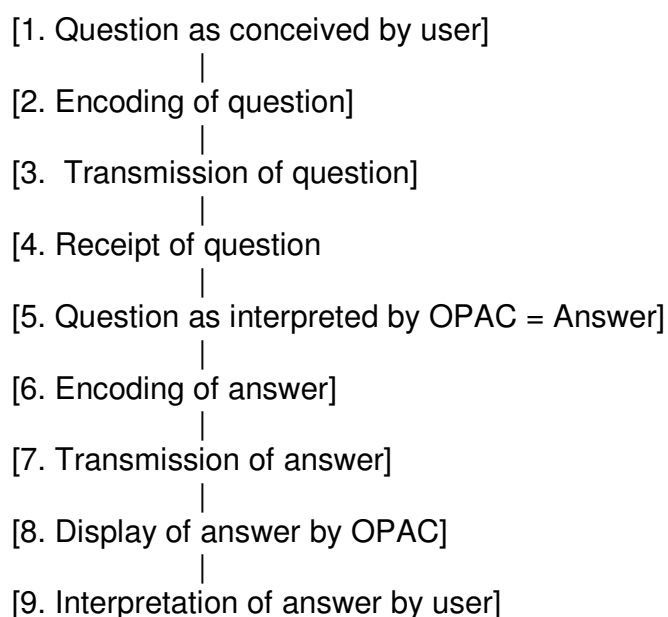
1. The initiator or source of the exchange is a library user: let us call

him George. George has a question of some kind, which he believes, on the authority perhaps of the library banner he sees on his computer screen, can be answered by the library OPAC.

2. George has to translate his question, that is, to encode it for transmission, into a form which he believes will evoke an appropriate answer. Using his knowledge of the Code by which the communication system he is using operates, George reduces his question to one or more words, and proceeds to make a choice between,

- for example, the different indexes or catalogue subsets that the catalogue may present him.
3. George's search term is accepted by the OPAC as a signal to be transmitted to the catalogue software for processing.
 4. The catalogue software receives the signal and decodes it by applying George's search term internally against the selected index to interrogate the catalogue database.
 5. As a result of this process the OPAC selects a set of records from the database and initiates the second phase of the communication process: returning the answer to George's question. The set of records returned, of course, may contain zero or one or more records. Whereas originally George was the Source and the catalogue database the Destination, in this second phase it is the catalogue which is the Source and George who is the Destination.
 6. The OPAC software then encodes the selected data into a display format for presentation. The display format generated is determined by the rules of the Code and may vary according to the type of question that was originally asked.
 7. The OPAC's response – that is the presentation of the search results – is transmitted to George and read by him.
 8. George now decodes the signal he has received, that is to say, he evaluates the OPAC's response in terms of his original question.
 9. George infers an answer to his question based on his evaluation of the OPAC's response. The information obtained by George as a result of the process comprises the difference between his understanding before and after the transaction.

Graphically, the process described above might be represented thus:



Indeterminacy of the Code

However, the flow of information is by no means as straightforward as this model would suggest, and is in fact impeded by indeterminacies at each stage. In particular, the efficiency with which the double act of communication is concluded depends to a very great extent on the degree to which the Code is consistently understood and applied by the agents adopting the role of Source and Destination. On the one hand, the Code comprises the body of cataloguing rules and standards according to which the database has been constructed (AACR, MARC, LCSH, DDC, etc.), together with the further set of rules which govern the design of the OPAC software itself and the ways in which it presents data back to the user. On the other hand, the Code consists of George's understanding of these rules – the level of his information literacy in respect of the particular OPAC he is using. Neither side of the coin is unproblematic.

In the first place, the Anglo-American Cataloguing Rules, MARC21, the Library of Congress Subject Headings, the Dewey Decimal Classification and other instruments of bibliographic control do not constitute an integrated set of instructions. There is, for example, no explicit mapping of AACR to MARC, and there is no one-to-one correspondence between LCSH and DDC. Each control mechanism was designed independently of the others and for a distinct purpose: AACR originally for the manual creation of cataloguing cards, MARC for the encoding of the same information electronically; LCSH is a pragmatic tool for subject analysis, while DDC is both a system of universal classification and a

practical mechanism for allocating numbers to facilitate both browsing and retrieval. Neither are any of these standards altogether internally consistent. AACR punctuation, for example, follows no clear pattern of hierarchical subordination; MARC is similarly fairly arbitrary in its assignment of tag numbers to data values and inconsistent in prescribing what punctuation should be included in the data and what is to be supplied by the library system software; DDC is not consistent in its association of numerical hierarchy and meaning (nor indeed could it be given its ambition to map the whole of human knowledge comprehensively to a decimal system).

Secondly, none of the standards mentioned has remained constant over time, and it would be a rare (and probably very small) library that has attempted to keep its catalogue records retrospectively up-to-date with changes to all of them. Most libraries using DDC, for example, will have books classified according to DDC22, DDC21, DDC20, etc., even though there have been major revisions from one edition to another. A case in point is the change to zoology numbers from DDC20 to DDC21 where large parts of the schedule were rewritten without reference to the previous edition. Similar changes occur regularly with LCSH. Following the dissolution of the Soviet Union in 1991, for example, it was necessary to revise headings and geographical subdivisions for the new political entities that succeeded to its territory. While these sorts of revision are certainly required to ensure that utilities such as LCSH correspond to current states of knowledge, they are not conducive to data stability and

consistency in large bibliographic databases.

Thirdly, it is unlikely that cataloguing standards are ever consistently applied by cataloguing staff, particularly over time. Not only are the rules very complex and capable of application at varying levels of detail, but individual organisations may also decide to interpret particular parts of individual standards idiosyncratically, or to ignore them altogether. This is perhaps less the case today than it has been in the past, but even with the triumph of co-operative cataloguing and the propagation of commentaries such as LCRI and Conser, localised practices persist in many libraries.

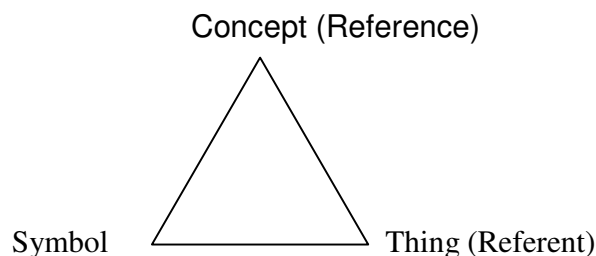
These points all relate to the bibliographic data contained within a catalogue database. There is a further set of rules that apply to the way searches are presented to the database and how results are presented to the user. Although there has been some attempt to standardise, for example, index names as a Common Command Language, and to provide guidelines for the formatting of OPAC displays (Yee, 1999; Bolin, 2000), these elements of configuration are typically left to library technical staff and are often not transparent either to users or indeed to library reference staff. Even if George is fully familiar with MARC21, he cannot normally know precisely what subfields of what fields are indexed as titles or as

keywords in the particular catalogue he is using. He cannot know what criteria are used to select or conceal particular tags from view.

Indeterminacy of the Sign

An OPAC search for a known bibliographic entity using indexes based on the traditional tracings provided by catalogue records (e.g. author, title, ISBN) will produce a relatively unambiguous result. With a keyword search – and even more with a full-text search, on the other hand, a second area of indeterminacy arises from the fact that any question on the part of a user must be mediated by language in order for it to be asked. The question originates as a thought, but must be expressed as a word in order to create a search term to interrogate a database, and the reverse happens when an answer is received back from the OPAC: a series of words (combined perhaps with certain visual information) appears on the screen and must be translated back into thought to provide an answer. The nature of this indeterminacy can be explored by examining the concept of the linguistic Sign (Raber & Budd, 2003).

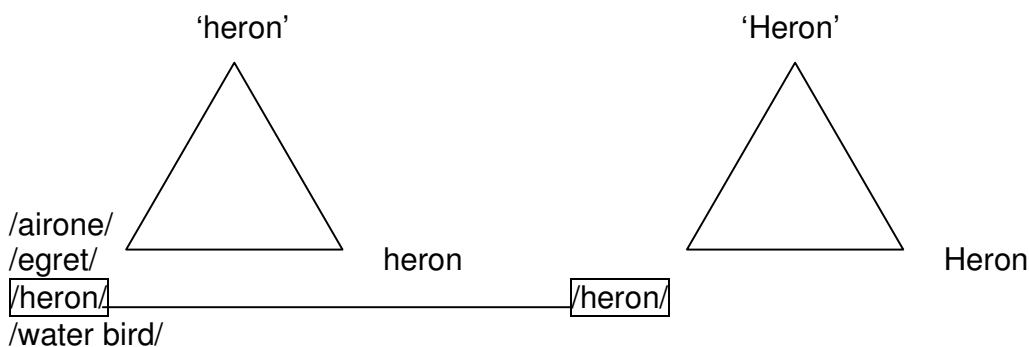
According to longstanding semantic theory (Ogden & Richards, 1927; Peirce, 1991) the Sign can be broken down into three parts, which in their simplest form can be represented schematically thus:



According to this model, the mental Concept of, for example, 'heron' is expressed through the Symbol /heron/, which consists of an arbitrary string of speech sounds or written letters. Both Concept and Symbol are in a similar relationship to the Thing or Referent, in this case an actual heron flying through the air.

In practice the three sides of the triangle are in a constant state of flux and reciprocal adjustment as thought is modified by cognitive experience of the world and by linguistic practice (Hookway, 1985). Leaving aside a probable instability in the Concept of 'heron' within any one person's mind – precisely what George conceptualises under this rubric may very well differ from day to day and may not correspond to Mary's conceptualisation – it is apparent that there can be no

one-to-one correspondence between the three sides of the triangle. The Concept 'heron', for example, could be represented in different languages by the Symbols /tsaplia/ (in Russian) or /airone/ (in Italian) or /Reiher/ (in German), or indeed by alternative terms in English such as /egret/. The number of actual herons flying around that might be referred to by the Symbol /heron/ or the Concept 'heron' is, of course, very large indeed. Conversely, the Symbol /heron/ may have more than one reference – it could be, for example, 'heron' in the sense of 'bird' or 'heron' as in Peter Heron the (hypothetical) philosopher. Moreover, the Concept 'heron' might also be evoked associatively by other concepts such as 'long beak' or 'water bird' and symbols associated with these concepts. Some of these indeterminacies are indicated in the figure below:



Relating this model of the Sign to the process of communication described earlier, it is always the Concept or Referent that is the topic of an OPAC enquiry, but always the Symbol which constitutes the search term used by the database. The difference between the two is necessarily always a potential source of ambiguity.

What is returned by the OPAC to the user as an answer is also on one level a Symbol in the sense described. Depending on the type of search conducted it typically consists of a browse list or a set of one or more bibliographic records which are words or phrases capable of being interpreted as Concepts. At the same time these

answers also act as pointers or indices to further answers – to the shelf location of books, for example, or to the electronic texts of documents. The signifying function of the OPAC response is thus multiple, requiring a correspondingly complex act of interpretation on the part of the user.

One area where the indeterminacy of the Sign interacts with the indeterminacy of the Code is the realm of classification. The classification systems commonly used in library catalogues (LCSH, DDC, LCC, etc.) necessarily impose a particular view of the world, which the user is obliged to accept in order to optimise the communication process (Paling, 2004). For example, the DDC numbers for history imply the cultural precedence of Europe over the rest of the world by assigning an earlier number (940) to Europe than to other continents (950-990). DDC also implies among other things that the Czech Republic (943.71) is part of Germany (943). Similar dependencies exist within any system of assigned subject headings – all the more so because the reference structures which explain the interconnectedness of thesaurus terms (broader term, narrower term, related term) are commonly omitted in electronic catalogues (Greenberg, 1997). In that the hierarchies involved both cannot be predicted and yet need to be known for efficient searching, this feature of classification is part of the indeterminacy of the Code. At the same time it exemplifies the indeterminacy of the Sign in that Symbols (thesaurus headings, classification numbers) may well have different References, that is, refer to different Concepts, in the mind

of George and the minds of the inventors of the classification systems.

Practical Implications

What practical lessons is it possible to draw from the above analysis? How can OPACs be designed to minimise the indeterminacies of the communication process? First, it should be possible, as least in theory, to reduce the complexity of the Code. Those parts of the Code that are concerned with description and presentation are certainly susceptible to simplification. The history of cataloguing standards, indeed, shows a clear evolutionary path in this direction. Relatively recent milestones in this area were the publication of the second edition of AACR in 1978, which among other things reconciled the separate U.K. and U.S. editions of the original AACR of 1967; and the USMARC format integration which occurred in the late 1980s. The forthcoming further revision of AACR, provisionally known as Resource Description and Access (RDA), promises further steps towards simplification as the nature of bibliographic data becomes better understood (Kiorgaard & Kartus, 2005). In particular RDA's adoption of the conceptual model proposed by the IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR) should cater better for the multiplication of versions of the same work that is characteristic of electronic publishing (Tillett, 2005). RDA's separation of the rules for description from guidelines for the presentation of data for display through OPACs or other means should also provide a welcome resolution of a currently ambiguous situation. To what extent RDA proves in practice to be backwardly compatible

with large existing databases, of course, remains to be seen.

Other simplifications of the Code applying to any specific OPAC can be (and of course often are) addressed in other ways. In this context the quality of catalogue data is of critical importance: the more accurate the data, the more consistent the Code, and hence the more efficient the process of search and retrieval. The inclusion in the data of appropriate reference structures, particularly for subject heading thesauri, though difficult to achieve systematically in practice, also significantly enhances the user's ability to interpret the Code. In the area of OPAC interface design the complexity of the Code can be minimised by a clear delineation and a clear labelling of functions – to avoid for, example, a category confusion between browsing and keyword searching, or between a catalogue subset and the catalogue as a whole, or between bibliographic searching and other functionality provided through the OPAC in its role as portal. Multiple levels of abstraction both in searching and in the display of results might allow the Code to be manipulated effectively at different levels of user competence. Clear explanations of what is actually happening at any given stage in the communication process should ensure the user's ownership and understanding of the codes necessary for searching and interpretation.

While to some extent the indeterminacy of the Code can be overcome, the indeterminacy of the Sign cannot. This is because a high degree of indeterminacy is inherent in language itself and in the process of cognition: it is this property in fact that makes it

possible to draw new conclusions from old data and to ask those questions which lead to an increase in information and hence knowledge. In order to improve the quality of OPAC communication, what cannot be overcome should be acknowledged and brought into the open. Very often OPACs are promoted as if the process of search and retrieval were transparent. This is clearly not so. In particular, two terms that are commonly used in the context of OPAC design should be avoided. The first of these is 'intuitive'. Clearly, on the model presented above, OPAC searching can never be actually intuitive; it is always dependent on the Code and on the relationship between the Symbol and the Concept, both of which are essentially arbitrary and can only be learned. The notion of 'intuitiveness' obscures this dependency and leads to a disempowerment of users, who are forced to accept a reduced view of the Code that is provided on their behalf. The second term is 'relevance'. This has a technical meaning referring to algorithms for retrieval based on the frequency of the search term or synonyms in the records received. It thus operates at the level of the Symbol. However, in ordinary language 'relevant' would normally be taken to operate at the level of the Concept, and the tension between these two meanings is liable to cause confusion. As in any case the actual mode of operation of relevance ranking is rarely obvious or discoverable, the user's interpretation of its effects is likely to be problematic.

Ultimately, effective OPAC communication is dependent on effective OPAC literacy among the user group. The better the Code and its

indeterminacies are understood the more efficiently the user will be able to exploit it. The better the indeterminacy of the Sign is understood, the more likely it is that correct interpretations will be made between thought and search term and between search result and thought, and that as a consequence information gain will be maximised. The

challenge therefore for designers of OPACs is therefore not only with the structure and appearance of systems and interfaces (although this is clearly important), but also with ensuring that appropriate levels of OPAC literacy are provided for both library staff and library users.

References

- Bolin, M.K. (2000). Catalog design, catalog maintenance, catalog governance. *Library Collections, Acquisitions & Technical Services*, 24, 53-63.
- Calhoun, K. (2006). *The changing nature of the catalog and its integration with other discovery tools [Report to the Library of Congress]*. Retrieved Sept. 19, 2006 from <http://www.loc.gov/catdir/calhoun-report-final.pdf>
- Eco, U. (1976). *A theory of semiotics*. Bloomington, Ind.: Indiana University Press.
- Floridi, L. (2004). Information. In L. Floridi (ed.), *The Blackwell guide to the philosophy of computing and information* (pp. 40-61). Malden, Mass.: Blackwell.
- Greenberg, J. (1997). Reference structures: Stagnation, progress, and future challenges. *Information Technology and Libraries*, 16(3), 108-19.
- Haigh, G. (2006). *Information idol: how Google is making us stupid*. Feb., 25-33.
- Hookway, C. (1985). *Peirce*. London: Routledge & Kegan Paul.
- Kiorgaard, D. & Kartus, E. (2005). A rose by any other name: from AACR2 to Resource Description and Access. Paper presented at *VALA Conference*.
- Ogden, C.K. & Ricards, I.A. (1927). *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism*. London: Kegan Paul, Trench, Trubner.
- Paling, S. (2004). Classification, rhetoric and the classificatory horizon. *Library Trends*, 52(3), 588-603.
- Peirce, C.S. (1991). *Peirce on signs: Writings on semiotic*. Chapel Hill: University of North Carolina Press.
- Raber, D. & Budd, J.M. (2003). Information as sign: Semiotics and information science. *Journal of Documentation*, 59(5), 507-22.
- Shannon, C. E & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Tillett, B. (2005). What is FRBR? A conceptual model for the bibliographic universe. *Australian Library Journal*, Feb., 24-30.
- Yee, M.M. (1999). Guidelines for OPAC displays. Paper presented at 65th *IFLA Conference*, Bangkok, 1999]. <http://www.ifla.org/IV/ifla65/papers/098-131e.htm>