The fifth element: a symbiosis of cataloguing and metadata

JAMES CAUDWELL

Head of Serials and Electronic Resources Cataloguing

The Bodleian Library, Oxford

The E-Resources Management Handbook – UKSG

This chapter addresses the mutual influence between multiple metadata standards used in both the establishment of new digital libraries and in the 'traditional' library setting. It explores the results for the user of changes in searching behaviour and perceptions of the catalogue alongside other searching tools, for better and for worse, with some reference to FRBR (Functional Requirements for Bibliographic Records). It aims to summarize the main issues of concern for those working in the field of electronic resource description and reviews the development of standards to date (including MODS and METS). Acknowledging the seismic impact of electronic resources on the library community and the inability of existing organizational structures to cope, this chapter also examines some of the automated tools for generating metadata and the formation of new workflows for managing metadata mark-up and authority control.

Thus they, believing that the primary body was something different from earth and fire and air and water, gave the name aither *to the uppermost region, choosing its title from the fact that it 'runs always' and eternally.*[1]

Introduction: terminologies for 'access'

For those who catalogue and for those who do not, 'cataloguing' ought to mean providing and enabling 'access'. Any consideration of the activity of cataloguing electronic resources should involve examining the manifold tools, human and machine, that create access to those resources. This is worth stressing because in the professional literature it has been commonplace for debate to hinge upon 'cataloguing' versus 'metadata'.

If it came to court, the question of 'The People vs. cataloguing and metadata production agencies' would more likely be: 'How do you defend the vast expenditure of public money on an activity that has demonstrably failed to grant reliable, authoritative and timely access?' than any querying of a use of terminology. For another plaintiff, though, defining metadata as distinct from cataloguing is no mere quibbling. Rather, metadata production is seen either as a byword or buzzword for cataloguing electronic media (as opposed to non-electronic), and/or variously as some lesser, cheaper, simpler, faster and more relevant occupation. These differences between 'cataloguing' and 'metadata' production have been viewed as radical, and not reducible to a synonym of one for the other. Metadata's "context of usage [is] different from library catalogues, they are typically not created by professional cataloguers, they are intended to be produced more efficiently than cataloguing records, they cover a specific kind of material (electronic resources)".[2] These fundamental oppositions, suggested in 1998, and founded chiefly on a difference of 'production paradigms', have since been corroded, although by no means eliminated.

Whereas the newness of new metadata standards has coincided with an expansion of the metadata production activity to new communities well outside libraries *per se*, paradigms within the library community have themselves shifted, 'fresh starts' have been made, and initiatives that overlap across multiple user groups continue to evolve. As was predicted – or at least hoped for – through the 1990s, each of the two 'approaches' to access provision has mutually informed the other. To a great extent, this symbiosis has been driven by the benefits that accrue from achieving interoperability between the growing numbers of alternative standards and systems. The impacts of each approach on the other are due also to the impracticality for any one institution of relying on a single means of providing metadata for, and control over, electronic resources: "no single bibliographic tool will be the magic bullet for managing online resources".[3]

Yet these metamorphoses of both the cataloguing and metadata communities' developing standards and workflows are attributable to more than simply the massively powerful effects of rapidly changing technologies. Equally influential, but less limelighted, are questions of content. At "a time when the heterogeneity of approaches to metadata for objects within a digital library has become untenable"[4] standardization of content is of increasing concern. A single standard in this field may be an illusory goal, but there may be an optimal limit to the number of times the same attribute of a resource can be recast as a different metadata element. However fast the speed of change (and the danger that travelling at such speeds entails), this question and its pressures will not change: "whether one asks by humming, drumming, nodding, knocking, blinking, speaking, or keying, the asking is for *content*".[5] Tellingly, the author of "MARC must die", who concluded in that article in 2002 "it is time to show it the long and winding road", retracted somewhat in 2004: "what must die is not MARC and AACR2 specifically, but our exclusive reliance upon those components ... We must create an infrastructure that can deal with MARC with equal facility as it deals with many other metadata standards. We must, in other words, assimilate MARC into a broader, richer, more diverse set of tools, standards, and protocols". With this acknowledgement came too the acceptance that as much attention needs to be given to "the intellectual [item]", that is, to the content.[6]

As these cross-fertilizations of approaches come into play, awareness grows of how nascent still are the resolutions to many of the issues that face the electronic resources information communities. So, it becomes increasingly impertinent to focus on the differences between 'cataloguing' and 'metadata': "distinctions ... are blurring. In the long run, a holistic approach to information access holds promise for breaking down artificial barriers between disciplines and media, and thus doing a service to the user community".[7] A recent 'how-to-do-it' manual summarizes this multiplication of solutions neatly for those working at the coal-face: "regardless of the specific tools involved, the library's bibliographic strategy for online resources must be able to grow with the collection, keep pace with evolving standards, and capitalize on emerging technologies in order to remain effective over the long term".[8]

A disappearing catalogue

There is even an acronym for the new phenomenon arising from today's use of online tools to find and locate resources: WYDSDE, or 'What You Don't See Doesn't Exist'. Before examining the incomparably more complex workflows and webs of relationships than exist for non-electronic resources, this section addresses what the user actually sees and may expect to see when searching. There are now available to many users a multiplicity of interfaces. At the same time there exist tools and applications that aim to organize search results and present these results from multiple resources (online catalogues, databases, related web sites) via one interface. In part, this array of services derives from the complexity of infrastructures supporting electronic resource delivery and their efforts to evade the potential 'mess' or 'chaos' that could result from so many competing sources of information.[9] While it is generally agreed that we can never know enough about users' experiences in their practice of information retrieval, there are definite signs that the profound climate change in users' searching behaviour (such that unfortunate consequences like 'WYDSDE' may occur) is determined precisely by the synoptic nature of the displays of search results these applications present them with.[10]

There can be no dispute that the new generation of applications, from the variously named 'metasearch', 'parallel search', 'federated search', 'broadcast search', 'cross-database search' portals[11] to the plethora of context-sensitive linking services, offer the user very real opportunities now necessary for effective resource discovery. Users at the University of Oxford, which lacks a portal at the time of writing, have spoken of their perception of a pressing need for one: "What seems to be required is a search tool as simple and user-friendly as Google, and yet able exhaustively to search at the same time both the Internet and the wealth of resources available locally"; the frustrations experienced by the lack of this functionality are not even fully conscious, presumably due to the users' innocence of such a tool: "there is virtually no awareness of the potential educational role of portals among the university staff and students at Oxford. However, staff and students seem to welcome the prospect of improving searching both across and beyond local electronic resources by implementing a portal, or a set of portals, provided with an advanced search tool".[12] Despite these anticipatory plaudits and the positive experience of those[13] who have implemented metasearching tools, there is a consensus that in comparison to the search capabilities provided by the native interfaces of most resources, a "cross-database search is a relatively blunt instrument".[14] Successful collocation is simply not yet feasible at such a scale.

It is worth, if only momentarily, pausing to consider the past. In the context of electronic resources, the traditional library catalogue appears seamlessly to blur the distinction between looking something up and going to get it from a shelf. Instead, it is now possible for the catalogue to work as both "a finding aid and a point of entry; from the user's perspective, organization is indistinguishable from access".[15] As this is such a common experience for the catalogue user, it is taken for granted. It is a feature a library catalogue and Google, for example, arguably share. In such ways a finding aid that once existed physically separate from the collections it described both gains and loses its visibility. For many, indeed, the catalogue is *de facto* marginal. Much depends on where in the academic life-cycle a user is located, but it is typical for specialist 'invisible colleges' of conference circuits and focus groups to bypass 'formalized information systems' entirely:

> "Sometimes people walk into a library ... and can put their fingers on things. They seem to sit at the metaphorical center of the social web and can tweak a vast complex and mature system of social networks to 'get stuff'. The system is quite transparent ... At the other end of the spectrum are the people who see the information system as confusing, chaotic, insurmountable, and unusable. They try to follow given directions and miss by a mile and a half. Much professional socialization concerns moving from this lost state into the state of obviousness and naturalness."[16]

Simulations of space and spacelessness

These clearly disparate and embryonic reactions to changes in searching interfaces have yet to percolate through to system designers and systems' multiple supporting infrastructures to the extent that a future information landscape cannot easily be predicted: "digital data creation is not in its infancy, but it is probably only in its adolescence, and has certainly not reached full maturity".[17] Some commentators have gone even further back into history to get a measure of the current perception of 'cataloguing' and digital library use, as far back as monastic libraries and libraries of the Baroque and the Enlightenment eras. They find striking similarities between, for example, the Baroque wall-system library and today's dedicated use of a VDU for searching:

> "The Baroque library realized a goal shared with contemporary computer screen design, namely 'first-personness', by allowing users to feel they were not only in a library but also wandering through a surveyable knowledge landscape, finding books and finding knowledge at the same time in a space common to both, leaving out the aesthetically distracting and epistemologically unnecessary intermediary of the catalog."[18]

It is a useful analogy because it exposes at one and the same time the universal experience of "library anxiety" already touched upon – the debilitating feeling of the "monstrousness" of the sheer volume of

available knowledge – and its apparently simple cure. That cure is to achieve an "access strategy" that is "*panoptic*, [where] the user could see everything", simulating the speed, visuality and seamlessness so lacking in an online catalogue: "current library catalogs are excruciatingly slow, boring, and non-transparent".[19] A vital element in achieving this simulation is to employ shapes, figures and colours rather than text, for the catalogue (and Google, it has to be said), if it is nothing else, is relentlessly text-based and this, apparently, intimidates. The essential requirement, though, reaches beyond design issues. However far an individual catalogue or a catalogue as one interface in an agglomeration of interfaces may be able to increase the user's perception of immediacy with the data, "a good OPAC or a good portal should interpret and extend the powers of intuition" of the user.[20] As stressed above, the catalogue's place in new information environments is evolving, but it is still impossible to see how any user's powers of intuition may be interpreted and extended without such basic ingredients as authority control and controlled vocabularies:

> "Subject approaches in the electronic age have become a major way of finding information ... Virtually every word in the English language has more than one meaning or sense, and many of those senses have more than one nuance; many words can be used as nouns, verbs, adjectives, or adverbs. Search systems that purport to allow the user to use 'natural language' cannot yet successfully distinguish among meanings or parts of speech in very large general systems."[21]

Like observations could, and should, be made of the names of works and their authors,[22] although it is interesting to note that OCLC reports of its pilot Open WorldCat project that "Most searches are subject searches, rather than known-item searches".[23] If what is required is "a search tool as simple and user-friendly as Google", then it would seem we are destined to reside in a state of arrested development, as evidenced by the literature that has been spawned describing how to make any headway in searching the Internet. One small example of Boolean operators shows how an interface can so easily elide the possibility for 'interpretation' or 'extension': "an almost total lack of use of some of these features would suggest that many web searchers are failing to make the best use of all that search engines have to offer".[24] In future manifestations of the catalogue it is likely to become more and more crucial to assess at what scale this elision is allowed: "Though the catalog may still be an intermediary in a formal sense, it is becoming an ever thinner and less noticeable membrane between library user and library collections".[25]

Cataloguing a diminishing percentage of value

When the computer screen becomes a "prosthetic mind"[26] and users breathe information and communications technology "like oxygen",[27] do they have any need to worry about WYDSDE? Any answer to that question must be partial unless it addresses the different needs of different users. Speak for two minutes to a biochemist or an engineer and you will most likely discover they never use a catalogue; they find the latest paper from their electronic journal gateway, full stop. No other questions need to be asked. Nevertheless, in the broadest possible sense, is the plea from the 'traditional' library community for integration of access to content, regardless of the type of publishing media (one of the main reasons for cataloguing electronic resources in OPACs at all), unfounded? Few libraries' catalogues are not now online and the walled world of a library and all it symbolically and socially represents is not what it once was; no-one writes now of "drizzling library lamps" or refuges from the "everlasting rain".[28] Sure, digital objects can be lost and technology can obsolesce (the Digital Domesday book; the 1901 census web site),[29] but thus far no-one has condemned a digital library arsonist to hell and hoped their "eyes burned out with light" as they have done for the setting fire to a physical library.[30] But the loss of digital information is a profoundly important issue. (It was one powerfully motivating force behind the Legal Deposit Libraries Act 2003.)

One doyen of traditional 'values', Michael Gorman, would not concur that there is in fact little to distinguish the library as a physical place (and the catalogue as its corollary) from the no-place or anywhere of virtual access: "the majority of electronic documents in collections of data have no value at all";[31] "I have no means of deciding the percentage of electronic resources that are of a general and lasting

value but would be very surprised if it were more than one or two percent".[32] Significantly, Gorman holds that we need "more walls not fewer".[33] Conflation of the "art" and "core value" of cataloguing with the physicality of a "library service" is another way of conceptualizing the current changes in, and variety of, models for providing access. It echoes those libraries who have reported on their experience of the "digital tsunami" in which a vast mass of material is paid for that is never used and has no long-term value, stressing that what they acquire is inimical to any traditional access strategy: "it is a collection of electronic links, of licenses to access – with levels and varying contractual conditions attached, notional grip-lines and grappling hooks slung onto a fast growing, organically evolving Gulliver travelling in cyberspace".[34] On the ground, the results are felt in a renewed need for a "recombinant library", one in which "the reconfiguration of presence and the mutual influence of physical place and network place has led to a heightened perception of the social aspect of library places, their role as a 'third place', as learning exchanges, as venues for collaboration and display".

Examples of 'adapted' catalogues

Among these shifting and recombining places, the cataloguing community seeks to deliver its product, so to speak, in a framework that is conducive for the user. A warmly welcomed example is that of RedLightGreen, the Research Library Group's (RLG) book research database.[35] Developed in consultation with student user groups, Merrilee Proffitt, RLG's program officer, explains how RedLightGreen's Google-like interface is specifically intended to attract users back to libraries' print content while including simultaneously Google search results, "an unusual innovation [that] ... can help provide additional research leads, though sometimes they may lead to false drops":[36]

> "It was clear that they want to find high-quality, relevant material that their instructors will consider appropriate; that they consider their OPACs difficult to use compared to search engines; and that they're concerned about not knowing the right words to use to find the items they need. As a site, RedLightGreen can meet these needs with search term enhancement and result ranking that reflects academic acceptance. Since results are displayed in our own easy-to-use interface, we're able to show students the top subject headings associated with their search terms. This leads students to better vocabulary and better searching."[37]

What may be less obvious is that RedLightGreen employs the Functional Requirements for Bibliographic Records (FRBR) model to organize search results.[38] For example, RedLightGreen does what OPACs conspicuously fail uniformly to do: it clusters editions of a work under one title for that work by repurposing FRBR's definition of a manifestation, "the physical embodiment of an expression of a work", as an edition.

Another example of a catalogue which aims to expand the horizons of access is the UK National Union Serials Catalogue (SUNCAT).[39] In developing this 'digital library local service' for serials, the project team has had to deal with the differing cataloguing practices of contributing libraries for dealing with electronic serials (or 'e-journals'). Of course, including e-journals in SUNCAT represents only one demand among many, the greatest of which is probably the difficulty of matching multiple records for the same serial when the quality of serials cataloguing is often below international standards (mostly because adherence to these standards has not been thought necessary for local catalogues): "many of the challenges for the SUNCAT project relate to the quality of the data from contributing libraries. Data is variable in format and often poor quality in content".[40]

Separate records for e-journals, when libraries have created them, are now in SUNCAT, and the problems with rendering an individual user's access rights to a given e-journal intelligible are being resolved by registration of each UK Higher and Further Education institution's resolver on the EDINA OpenURL Router. In this sense, SUNCAT "can be seen as a referring service".[41] SUNCAT has sought to monitor the work of the Digital Library Federation's Electronic Resources Management Initiative and its recommendations for vendors' development of electronic resource management systems. This has been seen as a necessary component in integrating e-serial records in SUNCAT since "what seems most

common is that management of information on licensing and access of e-journals originates and is managed differently from information about print".[42]

Just as influential on SUNCAT may be its future deployment of FRBR. Application of FRBR concepts to serials is still being discussed, but for the multiplicity of records for the same 'work' with which SUNCAT is trying to cope, FRBR may offer real solutions. Serials, as everyone knows, not only exist in multiple formats (print, microfiche, microfilm, direct access and remote access electronic) but they have a habit of changing title (and corporate body, entry under which can affect whether a new record is created when a change occurs). It has long been recognized that the multiple records that result, and the communication to the user of manifold and complex horizontal and chronological relationships, can seriously impede access – especially when libraries' holdings are partial and split over multiple records. One adaptation of serials records, influenced by FRBR, is to implement a 'super-record' or 'super-title' record for a serial that would group these multiple records together. This could affect cataloguing practice such that the serial work (or its expression) is catalogued once while "future records for manifestations use, or even link to, the classifications and headings in the work or expression record".[43] Obviously, there is a problem deciding among multiple titles what the 'super-title' of a serial should be. If implemented, this issue may have a system-provided solution such that the super-title would be displayed according to the user's search, with the serial's multiple records organized around that chameleon-like title.

'Interpretation' and 'extension': Functional Requirements for Bibliographic Records

Such interpretations of FRBR, however, lead us to consider FRBR's appropriateness, not for non-electronic books, but for electronic resources *per se*. FRBR may represent the most important step forward yet of the library community in defining the indispensable roles a catalogue must perform since it is based unequivocally on the user tasks: find, identify, select and obtain. Further, among experts in electronic resources cataloguing, as for serialists, it offers the potential for greatly improving and enhancing OPAC displays by resolving the frequently confusing results of maintaining separate records for multiple versions, or 'manifestations', of the same work. Despite this potential, deep problems remain in refining what is meant by the FRBR 'manifestation' and 'expression' ("the intellectual or artistic realization of a work") for electronic resources.

The mutability, the dynamic nature, of an e-resource effectively destroys our ability to describe adequately an essentially 'abstract' entity such as an e*dition*, an ability we think prevails when cataloguing from a hand-held *item* (e.g. a printed book). This is because in electronic media "the exemplar on which we base the description is perceptible only through the mediation of technologies embedded in the networks and/or devices used to transmit and serve the resource ... consequently, equating the description of the object, as perceived via a specific transmission channel at a specific point in time, with the description of the resource in its abstract sense (i.e. the abstraction we might identify as an edition) becomes problematic".[44] It would be misleading to infer that this situation is unique to electronic resources; it has existed for many years for most types of material.[45] With electronic resources, however, it becomes an acute problem, one that has led the revisers of the Italian cataloguing code, Regole Italiane di Catalogazione per Autori (RICA), to omit the concept of 'expression' entirely since it fails currently to cope with the requirements of editions understood as 'polyform manifestations'.

Gunilla Jonsson, Secretary of the Cataloguing Section of IFLA, neatly nails this issue: "it is an inevitable enigma in an ontology which describes both an abstraction and a physical entity which constitutes the embodiment of this abstraction".[46] It remains to be seen where developments to resolve these contradictions will lead, but FRBR (and the Joint Steering Committee for the Revision of AACR's Format Variation Working Group's conclusions in particular) demonstrates a concerted effort on the part of the library community to resuscitate and reinstate the catalogue in the user's information landscape (and this despite recent warnings that FRBR is not principally concerned with dictating displays in any given OPAC).

New wine in old bottles: mutual influences between cataloguing and metadata

From an anti-cataloguing stance, looking from the outside in on the developments that have been made in imposing bibliographic control over such volatile resources, it may seem that little of substance has changed. Such a perspective might be due to the impression that the cataloguing community is divided against itself. Many practitioners share the opinion that "we must be very careful indeed before jumping on this bandwagon, brightly though it is painted ... it is not within the scope of the library catalogue to include records for those remotely held documents, and it is certainly not the job of remote cataloguers to catalogue them".[47] Others have jumped, and some would even go so far as to say they would share the publishing world's 'dream vision' of implementing a *library* cataloguing service in which metadata was automatically generated and updated.[48] Although these major changes have yet to surface in practical terms, such shifts in 'production paradigms' are eagerly entertained as possibilities for future integration into library cataloguing processes. Not least pressing on the minds of library management is the inordinate cost and relative inefficiency of maintaining a purely human-based metadata production approach.

Here, 'cataloguing' is being directly informed by 'metadata' and is examining closely, for example in the research project AMeGA (Automatic Metadata Generation Applications), how applications like the UKOLN-built DC-Dot and Intology's Klarity might be remodelled for use in a library's 'access strategy'.[49] Especially influential on the 'traditional' library community, obviously, are the activities of the Open Archives Initiative and the breakthrough of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and the growing take-up in the use of eXtensible Markup Language (XML). Key to the cataloguing community's interest in automated metadata generation tools is the aim of integrating this new approach with existing workflows since at least initial reviews of the available products indicate serious doubts that all functions can – or will ever – be adequately performed. Chief among the disadvantages of the tools are their lack of support for authority control and the expression of bibliographic relationships: "sceptics asserted that attempting automatic generation of 'intellectual' fields such as subject or description is pointless or impossible"; "only a person can really grasp how items inter-relate and whether a single part is the dominant part".[50] Even so, the AMeGA researchers found no fear of 'job loss' in their interviews with practitioners, but rather a welcoming of the rationale, for example, that "automatic metadata evaluation techniques will improve the efficiency of metadata evaluation, enable human resources to be directed to metadata evaluation that automatic processing cannot adequately perform, and ultimately improve metadata quality".[51]

Future directions of research in automatic metadata generators in this context will include the development of extraction algorithms that are domain- or subject-specific (AMeGA compared the low usefulness of the same indexing algorithm applied to research resources in both blood disease and terrorism) and of algorithms that automatically and "intelligently" extract from different hierarchical levels of a web resource, making use of so-called "textual density analyses", for example. It is believed that application of FRBR (see above) may facilitate in this second requirement.[52] Thus far, libraries are urged to encourage their suppliers and vendors to develop metadata generator applications within their library management systems, and the Library of Congress is committed to create such an application for use within and beyond the library community. In this symbiosis of the two communities, 'metadata' is undergoing as radical a learning process as libraries, most perceptibly in the realm of quality control. As recently as 2004, it has been observed: "most metadata communities outside of libraries are not yet at the point where they have begun to define, much less measure, quality":

> "This process has been swifter in some communities than in others, as the early adoption of metadata as a panacea for information overload is followed all too quickly by the recognition that investments in quality are necessary for even modest gains. Furthermore, as communities of all kinds attempt to aggregate metadata (and ultimately services) via harvesting protocols – like the OAI-PMH – quality standards and measures are sorely missed."[53]

Innovating to maximize existing workflows and products

Possibly a most glaring instance exposing metadata as no 'panacea' is the qualification of Dublin Core and that standard's proliferation of application profiles. Registries of application profiles cannot by themselves manufacture quality control, although they do doubtless assist in "facilitating extensibility and interoperability in the context of networked services".[54] Similarly, web service software that can translate from one metadata schema to another via an 'interoperable core' (as in OCLC's developing MetaSwitch research project) is desperately needed throughout systems' architectures.[55] Just as important are applications that can "define views that join data from different metadata sources and reconcile conflicting and overlapping information" (as in the EDUTELLA project for heterogeneous peer-to-peer networks) which claim to enable "much richer metadata markup of resources ... that can be highly domain and resource type specific".[56]

Even so, the flexibility provided by these new generation tools (their function in reducing the ratio between rapid expansion in metadata standards and those standards' resulting 'conflict' and 'overlap') cannot answer to the vision of an agreement on metadata quality and what that may entail. What constitutes quality in 'traditional' cataloguing? 'Quality standardization' is often seen as a savoury virtue libraries keep in an aspic jelly of AACR2/MARC 21 for "top cataloguers and those on serious drugs" to relish.[57] Instead, it acts more as the fifth element, as the ether binding together the long-established *modus essendi* of collaboration and co-operation. Tangible outcomes have been the sharing of records for re-use via bibliographic utilities and the automation – as far as is possible – of authority control procedures. The Co-operative Online Resource Catalog (CORC), growing from OCLC's 'Building a Catalog of Internet Resources' (INTERCAT) project, anticipated a number of burgeoning technological developments that have since mushroomed (metadata harvesting; conversion of MARC to DC; generation of pathfinders) and integrated these with the tried and tested ('running always') standardization that depended upon inter-library co-operation.[58] This is a model the metadata community is considering adopting in some areas: "the traditional supply model for cataloguing records among libraries has demonstrated that this can become a cyclical process ... one can envisage metadata supply utilities automatically enhancing metadata and, in their turn, making this improved metadata available to information service providers".[59]

But the library community has also pioneered metadata standards of its own and in no way acted as though AACR2/MARC will suffice in any futures of metadata management. The Digital Library Federation, a partnership organization consisting chiefly of academic libraries, has instigated the development of the Metadata Encoding and Transmission Standard (METS) which goes far beyond the ambitions of MARC to "provide a method for aggregating all the metadata relevant to a given digital library object".[60] METS achieves this aggregation via its 'Structure Map' which allows the structure of a given digital object to be expressed in "nested division elements". In addition, descriptive and administrative (including technical, rights and preservation) metadata can be incorporated and associated with each other. This standard acts most efficiently as a 'wrapper' and is "designed to provide a great deal of flexibility". It comes into its own in its hospitality to multiple metadata schemes for use in the descriptive and administrative metadata sections, although METS currently 'endorses' just the three extension schemata (MARCXML, MODS, [simple] DC). A main concern in maintaining metadata for future use has been the ease of functionality for enhancing or updating over time, recognizing that the metadata will need to change as the resource changes. Already, practitioners in the digital preservation field are experimenting with METS for archived web sites, capitalizing on its accommodation of "associations between different levels of the Structure Map with the appropriate sub-parts of the descriptive metadata" by an arrangement of "root level METS, intermediary METS, [and] leaf node METS".[61]

Standing on the shoulders of the giants MARC and Dublin Core, emerges (from the Library of Congress' Network Development and MARC Standards Office) MODS – the Metadata Object Description Schema. Ideally suited for use with METS, MODS is likely to impact on both the orbits of 'cataloguing' and 'metadata' Once again, MODS exemplifies the interaction of these two communities in the field of electronic resource description, particularly in respect of the standard's extraction of the best of both worlds: it is semantically richer than Dublin Core but simpler and more syntactically sympathetic to

modern systems than full MARC. It lacks MARC's numeric tags – alien to the non-cataloguer – but regroups MARC's elements into "more logical components" (for example, it reduces the number of times one must deal, in different areas of a MARC record, with the same data element).[62] Its flexibility may "provide richer linking capability than MARC".[63] Its granularity is highly appropriate for dealing with multi-level digital objects where MARC and Dublin Core both fall down. Its being written in XML lends it a definite edge for wide-scale adoption as XML is "promoted as a possible successor to the MARC formats for the encoding of bibliographic metadata in RDBMSs (Relational DataBase Management Systems) which form the basis of the ILMS [Integrated Library Management Systems]".[64]

Where – it is reasonable to wonder – does this leave the MARC record? Presently, one clear strand is in redeploying the MARC record as an 'access-level' record, an approach that stems from a re-evaluation of the appropriateness of types of record to types (and levels) of electronic resources. As defined by Tom Delsey, the access-level record is "designed specifically to support subject and content-oriented access and to reduce the traditional emphasis on description".[65] These MARC records, it is envisaged, would provide a collection-level description to resources which at the item level are catalogued – by an automatic metadata generator for example – in MODS records. Research in this area has learned from FRBR to sift from the mass of information in those 'traditional' records those elements that may be dispensable and those which must remain. The initial responses of both cataloguing and reader services staff are persuasive. Only the lack of subtitles ("in certain instances [they] would have been helpful to 'prop up' a brief or misleading title") and the absence of a date showing when a resource was viewed for cataloguing (to help any user experiencing a broken link or locating a resource the title of which differs from the record's) have been cited as problems with sample access-level records: "In general, I feel access level is adequate as long as primary subject headings and summaries are present in the OPAC ... I don't think access level cataloguing would adversely affect the OPAC".[66]

These MARC records will often be 'integrating resource' records. That is, they will represent resources that are "added to or changed by updates that do not remain discrete and are integrated into the whole".[67] The concept of the 'integrating resource' had its genesis in the work of the Joint Steering Committee for Revision of the Anglo-American Cataloguing Rules (JSC) and in the papers of Jean Hirons and Crystal Graham, presented at the International Conference on the Principles and Future Development of AACR back in 1997.[68] Writers of a supplementary manual on cataloguing integrating resources have opined that the ability to catalogue updating web sites and updating databases in AACR2/MARC records "is perhaps the first such monumental change to be faced by catalogers of this new millennium".[69] One outcome is the disintegration of the traditional division between monograph and serials cataloguers so that both now (although not in all institutions) catalogue resources that change over time.

For serialists this idea is nothing new, but the types of resources covered by the new rules pose challenges that the 'access-level' record attempts to knock aside and seems not to address. Michael Gorman cruelly characterized these integrating resources as "more like those curious structures one sees in the streets of Paris – erected to have posters plastered on them. The posters change, become torn and overlaid with newer posters, are removed and vandalized with graffiti, so that the content and visual effect differ from week to week".[70] Despite the apparent superficiality of much content accessed via the web, librarians have embraced the new rules for cataloguing those scholarly resources which their institutions may archive or which are hidden in the deep web as yet unexposed by Google. Cataloguers also see providing OPAC access to integrating resources as valuable because they wish to reveal the relationships between electronic and non-electronic media (for example, when a print serial ceases and becomes an updating remotely-accessed web site) and they hold that by cataloguing in AACR/MARC they impose authority control (for example, by assigning subject headings) over web resources that in turn enhances access to resources regardless of their 'carrier'. Records for these resources do, however, require ongoing maintenance to keep them up to date and absorption in AACR of rules to cover their original cataloguing may be less than half the battle won. Commitment to the integration of such maintenance into libraries' workflows is patchy indeed, although much can be achieved through the mechanisms afforded by record sharing in the bibliographic utilities (as discussed above).

In conclusion, possibly the greatest challenge for the future is in maximizing the known benefits of imposing authority control over the web environment. It has been remarked that "the vast experience of

the library community in the use of authority control is not being brought to bear in the generation of metadata in ways that will prove interoperable across the global network of systems in either the short or long term".[71] In fact, the library community has not been inactive in this area; but the results of its activity have yet to bear fruit. Most significant and important to watch is the work on a Virtual International Authority File (VIAF) and also the progress of the LEAF (Linking and Exploring Authority Files) project.[72] Such future innovations that expand the application of authority control and 'access control' from libraries to the multiple domains (museums, archives, e-print repositories, etc.) of the web build again upon examples of inter-library co-operation. NACO (Name Authority Co-Operative), for instance, has successfully drawn upon traditions of sharing, documentation, training and collaboration to force as wide an application of standards (AACR2 and the MARC 21 Authority Format) as possible to a task that is inherently and intensely expensive. VIAF and LEAF both seek to introduce automated means of linking separately maintained authority files in order to facilitate a range of operations, but ultimately to improve services for the user. Barbara Tillett summarizes the accruing benefits, benefits it is always worth restating:

> "We are reminded how [authority control in the web environment] brings precision to searches, how the syndetic structure of references enables navigation and provides explanations for variations and inconsistencies, how the controlled forms of names and titles and subjects help collocate works in displays, how we can actually link to the authorized forms of names, titles, and subject that are used in various tools, like directories, biographies, abstracting and indexing services, and so on."[73]

Testing of a pilot version of a VIAF has begun in collaboration between the Library of Congress, OCLC and the German National Library by linking records of the LCNAF (Library of Congress Name Authority File) and the German Personal Name Authority File. It will be interesting to see how the project expands beyond dealing only with personal names to include, for example, corporate names, since it is well known how corporate bodies' names differ between electronic and non-electronic sources of information.[74] Further, both VIAF and the LEAF project intend or are already making use of the OAI-PMH to harvest metadata for both new and subsequently changed authority records. In sum, the VIAF – inspired to a great extent by mutual influences between 'cataloguing' and 'metadata', by 'symbioses' between hitherto distinct production paradigms and user communities – "holds long-term promise as a basic building block of the 'semantic web' – a future version of the web that will permit human-to-human, human-to-computer, and computer-to-computer communication".[75] Quite unlike the philosopher's conception, then, the fifth element, the 'quintessence', in the future of catalogues and cataloguing, will not operate separately in its sphere, but itself define the continuing interactivity of each element in the life-cycle of managing access to electronic resources.

References

1. Aristotle, *On the Heavens*, trans. Guthrie, W., 1939, Cambridge, Mass., Harvard University Press, I, 3 (270b).

2. Gradmann, S., 'Cataloguing vs. Metadata: old wine in new bottles?' Paper presented at 64th IFLA General Conference, August 16–21, 1998.
    http://www.ifla.org/IV/ifla64/007-126e.htm

3. Mitchell, A., *Cataloging and organizing digital resources*, 2005, London, Facet Publishing, p. 84.

4. Gartner, R., MODS: Metadata Object Description Schema, 2003, London, JISC, 2003, p. 8.

5. Hillman, D., *Metadata in practice*, 2004, Chicago, American Library Association, p. 191.

6. (a) Tennant, R., MARC must die, *Library Journal*, 2002, 127(17) 27–28 [and] (b) Tennant, R., A bibliographic metadata infrastructure for the twenty-first century, *Library Hi Tech*, 2004, 22(2), 175–181.

7. Borgman, C., *From Gutenberg to the global information infrastructure*, 2000, Cambridge, MIT Press, p. 206.

8. Mitchell, A., ibid., p. 71.

9. Tennant, R., (b) ibid., p. 175.

10. See especially the documentation describing the 'Sense-Making the Information Confluence: the whys and hows of college and university user satisficing of information needs' project http://imlsosuoclcproject.jcomm.ohio-state.edu/. This project aims to be "informed by a social epistemology ... an approach to knowledge mapping based not on essential categories but rather on empirical cateogories of knowledge production and use".

11. See: NISO Metasearch Initiative
    http://www.niso.org/committees/MS_initiative.html

12. Ingram, C. [et al.], CREE: Contextual Resource Evaluation Environment: CREE focus groups: results and comment, 2005, JISC, p. 32.
    http://www.hull.ac.uk/esig/cree/documents.html

13. For example: Lewis, N., 'Talking about a revolution? First impressions of Ex Libris's MetaLib', Ariadne, June/July 2002, 37.
    http://www.ariadne.ac.uk/issue32/metalib/

14. Mitchell, A., ibid., p. 79.

15. Mitchell, A., ibid., p. 73.

16. Bishop, A., *Digital library use: social design and evaluation*, 2003, Cambridge, MIT Press, p. 245.

17. Deegan, M., *Digital futures: strategies for the information age*, 2002, London, Library Association Publishing, p. 134.

18. Garrett, J., The legacy of the Baroque in virtual representations of library space, *Library Quarterly*, 2004, 74(1), p. 51; see also: Wegmann, N., Bücherlabyrinthe: Suchen und Finden im alexandrinischen Zeitalter, 2000, Cologne, Böhlau Verlag.

19. Garrett, J., ibid., pp. 45, 56–57.

20. Garrett, J., ibid., p. 60.

21. Taylor, A., *The organization of information*, 1999, Colorado, Libraries Unlimited, p. 131.

22. See, for example, Connell, T., *The future of cataloging: insights from the Lubetzky Symposium*, 2000, Chicago, American Library Association, p. 125: "There are two basic conditions upon which cataloging theory and rules are based. The first condition is that the names by which people identify themselves and by which others identify them may, and frequently do, vary. The second condition is that the names by which representations, translations, editions, etc., of a work are identified may, and frequently do, vary."

23. Chip Nilges, executive director of OCLC WorldCat Content & Global Access, quoted in: Mattison, D., RedLightGreen and Open WorldCat , *Searcher*, 2005, 13(4).

24. Stacey, A., *Effective information retrieval from the Internet: an advanced user's guide*, 2004, Oxford, Chandos Publishing, p. 221.

25. Garrett, J., ibid., p. 57.

26. Birkerts, S., Tolstoy's dictaphone: technology and the muse, 1996, St. Paul, Graywolf, p. 104.

27. OCLC, 2003 *Environmental scan: pattern recognition*, 2003, OCLC, [Social landscape section], p. 10.
    http://www.oclc.org/membership/escan/

28. Abse, D., 'Public library'. In: *New and collected poems*, 2003, London, Hutchinson; Beresford, A., 'In the library'. In: *Footsteps in snow*, 1972, London, Agenda Editions.

29. I am grateful to Alan Morrison (Collections Manager, Oxford Text Archive) for these examples.

30. Pugh, S., 'Elegy. In: *The beautiful lie*, 2002, Bridgend, Seren Books. (This poem concerns "the books lost in the fire at Canton Library, Cardiff, 1997".)

31. Anon., Updating the eternal, *Library and Information Update*, June 2005, 4(6), 19.

32. Gorman, M., Metadata or cataloging? A false choice, *Journal of Internet Cataloging*, 1999, 2(1) 5–22.

33. Anon., Updating the eternal, ibid., p. 19.

34. McCarthy, J., The print block and the digital cylinder, *Library Management*, 2005, 26 (1/2), 89–96.

35. Available at
    http://www.redlightgreen.com

36. Mattison, D., RedLightGreen and Open WorldCat, *Searcher*, 2005, 13(4).

37. Mattison, D., ibid.

38. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: final report*, 1998, München, K.G. Saur
     http://www.ifla.org/VII/s13/frbr/frbr.pdf.
     See also: Tillett, B., What is FRBR?: a conceptual model for the bibliographic universe, 2004, Library of Congress
     http://www.loc.gov/cds/downloads/FRBR.pdf.
     A useful introduction to FRBR is also given by Danskin, A.A., where they illustrate an application of FRBR to the display of editions of Tolkien's 'Lord of the Rings' in: Bibliographic records in the computer age, *Library and Information Update*, 2003, 2(9).
     A more detailed 'demonstration' and critique of FRBR is available in: O'Neill, E., FRBR: Functional Requirements for Bibliographic Records: application of the entity-relationship model to Humphrey Clinker, *Library Resources & Technical Services*, 2002, 46(4)
     http://www.oclc.org./research/publications/archive/2002/oneill_frbr22.pdf.
     The library management system VTLS has implemented FRBR and one library using VTLS's FRBR in its OPAC is the Université catholique de Louvain
     http://www.bib.ucl.ac.be/

39. See: SUNCAT
     http://www.suncat.ac.uk/

40. Burnhill, P., SUNCAT: a modern serials union catalogue for the UK, *Serials*, 2004, 17(1), 63.

41. Guy, F., 'Work package 3: requirements analysis for SUNCAT (Holdings statements and subscription information for serials in electronic format)'. Paper prepared for the SUNCAT project team, p. 13.

42. Burnhill, P., ibid., p. 65.

43. Rosenberg, F., 'The serial super record 2.0'. Powerpoint presentation made at the American Library Association Annual Conference, 2005, Chicago, IL for the ALCTS/CCS Continuing Resources Cataloging Committee
     http://www.ala.org/ala/alctscontent/serialssection/serialscomm/contresourcecata/2005annualcrcc.htm

44. Delsey, T., Functional Requirements for Bibliographic Records: user tasks and cataloguing data: part 2, *Catalogue & Index*, 2004, 151, pp. 1–4.

45. On the longevity of the compromise that equates a physical item with an entire edition, see: Heaney, M., Time is of the essence, 1997.
     http://www.bodley.ox.ac.uk/users/mh/time978a.htm

46. Jonsson, G., The basis for a record: in the light of Functional Requirements for Bibliographic Records, *IFLA Journal*, 2003, 29 (1), 41–46.

47. Oddy, P., *Future libraries, future catalogues*, 1996, London, Library Association Publishing, p. 158.

48. "In the dream vision of electronic content delivery, publishers could bypass most if not all of the intermediaries in the traditional book supply chain and supply content either directly to the end user through their own website or via an online e-tailer", Thompson, J., *Books in the digital age*, 2005, Cambridge, Polity, p. 315–316.

49. Greenberg, J., *The AMeGA project*; final report for the AMeGA (Automatic Medadata Generation Applications) Project, 2005, Chapel Hill, NC, University of North Carolina at Chapel Hill.

50. Greenberg, J., ibid., pp. 30, 31.

51. Greenberg, J., ibid., p. 53.

52. Greenberg, J., Metadata extraction and harvesting: a comparison of two automatic metadata generation applications, *Journal of Internet Cataloging*, 2003, 6(4), 59–82.

53. Hillmann, D., ibid., p. 240.

54. Heery, R., 'The MEG Registry and SCART: complementary tools for creation, discovery and reuse of metadata schemas'. Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002, Florence, Florence University Press, p. 125–132.
     http://www.bncf.net/dc2002/program/ft/paper14.pdf.

55. Godby, C., 'Two paths to interoperable metadata'. Paper presented at the 2003 Dublin Core Conference DC-2003: Supporting Communities of Discourse and Practice: Metadata Research & Applications, 2003, Seattle, Washington

    http://www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf.

56. Nejdl, W. 'EDUTELLA: a P2P networking infrastructure based on RDF'. Paper presented at the WWW2002 11th International World Wide Web Conference, 2002, Honolulu, Hawaii

    http://edutella.jxta.org/reports/edutella-whitepaper.pdf

57. Tennant, R., (a) ibid., p. 26.

58. Hunt, S., 'From INTERCAT to CORC: cooperation in the selection and cataloguing of Web resources'. Paper presented at International Conference Electronic Resources : Definition, Selection and Cataloguing, 2001, Universita' di Roma La Sapienza.

    http://eprints.rclis.org/archive/00000184/

59. Hillman, D., ibid., p. 268.

60. Cundiff, M., An introduction to the Metadata Encoding and Transmission Standard (METS), *Library Hi Tech*, 2004, 22(1), 52-63. See also:

    http://www.loc.gov/standards/mets/.

61. Myrick, L., 'Case Study: Using METS as a DIP to Navigate Archived Websites'. Paper presented at METS Opening Day, 2004, British Library Conference Centre.

    http://www.loc.gov/standards/mets/presentations/od3/myrick.ppt.

62. Gartner, R., ibid., p. 1.

63. Guenther, R., 'Rich descriptive metadata in XML: MODS as a metadata scheme: MODS, MARC and metadata interoperability'. Powerpoint presentation made at the American Library Association Annual Conference, 2005, Chicago, IL.

    http://www.loc.gov/standards/mods/presentations/ala2005-mods.htm

64. Felstead, A., 'The library systems market: a digest of current literature', *Program*, 2004, 38(2), 93.

65. Delsey, T., Defining an 'access level' MARC/AACR catalog record: project report, 2004, Library of Congress.

    http://www.loc.gov/catdir/access/report_final.pdf.

66. Reser, D., 'Defining an access level record for remote access electronic resources'. Powerpoint presentation made at the American Library Association Annual Conference, 2005, Chicago, IL for the ALCTS/CCS Continuing Resources Cataloging Committee.

    http://www.loc.gov/catdir/access/ala_crcc.ppt

67. *Anglo-American Cataloguing Rules (AACR),* 2nd ed., 2002 revision. 2002, Chicago, Ill., American Library Association (glossary definition).

68. See especially: (a) Hirons, J. and Graham, C., 'Issues related to seriality'. In *The principles and future of AACR: proceedings of the international conference on the principles and future development of AACR*, 1998, London, Library Association, pp. 180–213 [and] (b) Hirons, J., 'Revising AACR to accommodate seriality' available at

    http://www.collectionscanada.ca/jsc/docs/ser-rep.pdf.

69. *Program for Cooperative Cataloguing/BIBCO, Integrating resources: a cataloging manual*, 2003, PCC/BIBCO

    http://www.loc.gov/catdir/pcc/bibco/irman.pdf, preface

70. Gorman, M., 'Electronic resources: which are worth preserving and what is their role in library collections?'. Paper presented at International Conference Electronic Resources : Definition, Selection and Cataloguing, 2001, Universita' di Roma La Sapienza

    http://w3.uniroma1.it/ssab/er/relazioni/gorman_eng.pdf.

71. Hillmann, D., ibid., p. 10.

72. See:

    http://www.oclc.org/research/projects/viaf/

73. Tillett, B., 'A Virtual International Authority File'. Paper presented to Workshop on Authority Control among Chinese, Korean and Japanese Languages (CJK Authority 3), 2002, Karuizawa, Tokyo, Kyoto. http://www.nii.ac.jp/publications/CJK-WS3/cjk3-08a.pdf.

74. Quiang, J., Comparing and evaluating corporate names in the National Authority File (LCNAF) on OCLC and on the Web, *Cataloging & Classification Quarterly*, 2003, 36(2), 21–30.

75. Morris, S., 'Virtual international authority: library, partners form online name authority service', *The Library of Congress Information Bulletin*, Sept. 2003. http://www.loc.gov/loc/lcib/0309/authority.html

■ **James Caudwell**
E-mail: james.caudwell@bodley.ox.ac.uk
Tel: +44 (0)1865 287173
Fax: +44 (0)1865 277036

Correspondence address:
Head of Serials and Electronic Resources Cataloguing
Technical Services
Bodleian Library
University of Oxford
Oxford OX1 3BG, UK

Biographical note

James Caudwell is currently Head of Serials and Electronic Resources Cataloguing at the Bodleian Library, Oxford. Previously he has worked in a number of libraries at the University of Oxford, the Courtauld Institute of Art and the University of London Library. He is serving on the CILIP-BL Committee on AACR and has recently contributed to a special issue (on cataloguing electronic resources) of the *Art Libraries Journal*.