

Structures, standards, and the people who make them meaningful by David Bade

A good way to begin any discussion of bibliographical control is to remember Ashby's *Law of Requisite Variety* is. A simple version of this is

(1) the amount of appropriate selection that can be performed is limited by the amount of information available. (2) for appropriate regulation the variety in the regulator must be equal to or greater than the variety in the system being regulated.¹

I would prefer to draw out the implications of that for our work, but we have been asked to respond to a set of questions. The questions to which we were asked to respond, like these meetings themselves, rest upon certain assumptions that I want to note and comment upon before I make any response to the particulars. The chief difficulties I had with the background paper had to do with the understanding of what goes on in libraries, why, and for whom. Additionally, the ways in which we understand users, policies and technologies and the relationships among these elements are crucial determinants for how we think about structures and standards. To start with, then, let me note three assumptions from the background paper:

- 1) There are two main users of bibliographic data and their associated use environments: consumers and management.
- 2) Structures and standards govern the creation, recording, and distribution of metadata in bibliographic control
- 3) Data are created to be processed by applications.

The first two assumptions are reflected in the first question: What kinds of structures and standards are needed to provide effective bibliographic control in the environmental spectrum spanning consumer uses and management uses? The variety of users and user purposes so noticeable in the background paper for the first meeting disappeared in the reports of that meeting and the second background paper. The third assumption indicates that the paper rests upon an understanding of what happens in libraries that is rooted in a theory of transportation rather than a theory of communication. The background paper in general assumes that what happens in libraries can be described as a technical system, not a communication system. In this model data are not created for people but for processing by applications. Supply chains, repositories, database silos, data flows, mining, manipulation and display: this is the language of the background paper, and it is not the

¹ "Law of Requisite Variety", in *Web Dictionary of Cybernetics and Systems* http://pespmc1.vub.ac.be/ASC/LAW_VARIE.html (viewed 8 May 2007). The more technical formulation as it appears in Ashby's *Introduction to Cybernetics* (1956: p. 207) is: V_D is given and fixed, $V_D - V_R$ can be lessened only by a corresponding increase in V_R . Thus the variety in the outcomes, if minimal, can be decreased further only by a corresponding increase in that of R. (A more general statement is given in S.11/9.) This is the law of Requisite Variety. To put it more picturesquely: only variety in R can force down the variety due to D; variety can destroy variety.

language of communication. My understanding of structures and standards is rooted in a very different view of what happens in libraries. So let me reorient the discussion of structures and standards, and bibliographic control in general, according to my understanding that what happens in libraries is communication, not transportation.

The philosopher H.P. Grice offered nine maxims as the general standards governing communication:

1. Make your contribution as informative as is required
2. Do not make your contribution more informative than is required
3. Do not say what you believe to be false
4. Do not say that for which you lack adequate evidence
5. Be relevant
6. Avoid obscurity of expression
7. Avoid ambiguity
8. Be brief
9. Be orderly

Grice noted that communicative exchanges are characteristically cooperative efforts in which participants recognize a common purpose or orientation. He offered a single principle—the co-operative principle—which he suggested must inform communicative activity: Make your conversation contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.²

So first of all, three things to note: communication, cooperation and common purposes. Beyond these critical matters, we notice that these standards would please most library administrators AND all users, including catalogers, reference librarians, and acquisitions personnel. Another thing that we should notice is that these standards are not the same kinds of standards as ISBD which only describes how one presents certain information. These Gricean standards are not just about how but about what information should be put forth.

The order in which I will proceed is as follows: first remarks on structures and standards for facilitating communication, then a discussion of users and uses followed by comments on cooperation and the conflicting goals that this inevitably involves. Comments on understanding technologies, in particular metadata harvesting and data-mining lead into a set of more specific observations as responses to the questions.

Structures and standards to facilitate communication, not data transport

The first task then is to understand libraries as communication systems. Grice's standards are standards for communication between and among human beings. In contrast, the theory of librarianship which informed the background paper is unmistakably a theory of industrial production, transportation and storage, in which meaning is "mined" rather than

² Grice (1975), p.45-46.

created, and information flows and is merged and manipulated *but never interpreted, evaluated or corrected*. The failure to even mention these activities is a glaring—and revealing—omission in the Working Group's statements and questions asked of the speakers at these meetings.

If bibliographical information is created in an attempt to communicate to someone, then the relationship between that someone and the creators of that information is the locus of meaning in libraries—not the structures, standards and technologies used in that communicative activity. The possibilities for the success or failure of that communication are rooted in the indeterminacy of predicting what, why and how those for whom the information is created will be searching. What can be done is to fashion the information created according to those social practices that both the library user and the library expect in each particular institution. Those practices are both fashioned by and revealed in the publications being cataloged themselves: forms of citation, tables of contents, indexing terms, abstracts, statements of authorship, readership or sponsorship, links, theoretical orientation, timeliness, genre and the like.

In addition to knowing the social practices served by the institution, another key factor for successful communication is that the person creating the bibliographical information must actually have something to communicate. Communication is successful when the hearer can infer the speaker's meaning.³ A cataloger or reference librarian who cannot read and understand the items of interest is of no more use than a machine that cannot, and sometimes probably worth less. Incomprehension of the language of the text or of the discipline makes communication impossible. The impossibility of communication in any situation of incomprehension has clear implications for all efforts to automate the creation of bibliographical information.

Moving to the more specific matters at hand, communication requires a medium and it is characteristic of all means of communication that they have differing potentials and limitations, and that they are socially regulated. There is a certain range of physical requirements and capacities associated with signing, speaking, writing and online searching with a mouse. It is equally true that all of these practices are neither innate nor intuitive but learned and socially regulated. This is just as true of Flickr as it is of LCSH, of "natural language" as it is of scientific terminology. Different persons combine and use different means of communication according to their situation and purposes to further their communicational goals. This leads us to the second matter, the subject of the first meeting: users.

Users and uses

Karen Coyle noted on her blog the inadequacies of the dichotomous model of users identified in the first meeting of the Working Group.⁴ Consumers and managers are the two user models the paper offers, and the first set of questions we have been asked to respond to follows from that identification. Unless the multiple user groups and their widely varying uses are clearly articulated, how can anyone possibly respond to the first

³ Sperber and Wilson (1986), p.23.

⁴ Coyle (2007c)

question? The goals of management can often be achieved simply by attaching a number to a book and also to a computer record. Mark it and park it: every cataloger has heard this from an administrator. Nothing else is needed to move a book from acquisitions through binding and labeling to storage, into the hands of the patron and back again. For the student who merely wants something quick Google suffices and the library need only have Internet access; students often want nothing more, even when they could benefit from more. For the historian of 13th century Jawa—not a fictional user, but me—a vast array of bibliographic tools of all sorts are needed to effectively cover literature in dozens of languages spanning 700 years.

The division between consumers and managers has certain implications. The most important implication is that consuming and managing presuppose the prior existence of information. What is left out of this dichotomy is the creation of information: who is doing that? The consumers or the managers? The answer of course is neither of these, but a third category of shadowy characters largely left out of account in the background paper's understanding of what happens in libraries. Many years ago Jean-Pierre Dupuy insisted that only if you ignore the creation of information and the meaning of information can you treat it as a thing in a technical process.⁵ This is precisely what we find in the background paper.

If consumers and managers are to consume and manage information, then someone has to create it, and in that act we can locate the essential task of the library: to provide the library's users with information about what the library has to offer. And it is precisely in this context—the individual organization, whether that is the University of Chicago, OCLC or Google Book Search—that bibliographical control is meaningful and necessary. That may mean searching other information sources for information to use for one's own purposes, but in every case that information is not simply a repetition or copying of existing information. Putting that information from whatever source into the local database is the creation of new information about location and availability, assertions of propriety and relevance, judgements about value and utility for local users. It is often also an implicit statement of approval (e.g. approval plans), of verified accuracy, and in the better libraries involves the adaptation and increase of bibliographic information with a specific institutional purpose and user community in mind.

The background paper notes that some information can serve both consumers and managers, but that most speakers at the first meeting felt that current bibliographical information and the means of accessing it were inadequate to their needs. Structures and standards do not, however, create bibliographic information: people do. The key issue facing librarians, a matter discernable in the consumer-manager dichotomy, is that of the creation of information: *there is no information, no meaning "out there" to be found, managed, mined, merged, manipulated or processed unless it has been created by someone for some purpose in a form which that creator deemed adequate for the intended users and uses.*

⁵ Dupuy (1980).

Cooperation and conflicting goals

Many types of libraries exist, serving users with vastly different goals and purposes. When we broaden our view of the bibliographic universe to include book sellers, publishers, commercial databases and other information services, the disparity of goals and purposes are more clearly seen. In cooperative databases and in any situation of autonomous data sources the highest common denominator approaches zero. The information found may not have been created for purposes consonant with one's own institutional goals, and may even make it unsuitable for the library's purposes. This leads to the final question from the Background paper:

5. Libraries now manage different flows of data, created within different regimes, much of it outside the library environment. They also want their data and services to appear in other environments. At the same time, we see more reuse and flow of data across publishers, libraries, agents, other bibliographic services, etc. What does this mean for our bibliographic structures and standards?

The use of data created outside the library requires the library to evaluate its appropriateness and adapt it as necessary to the institution's goals, the only other option being to ignore the specific nature of one's own institution along with the corresponding user needs. In a collaborative environment such as OCLC, everything matters to someone so everything ought to matter to every one. The most demanding uses of the most demanding users should determine the kind of structures and standards required. The reverse, however, is the case: because of different goals and purposes in the creation of bibliographic information, there is not a single bit of information which matters to everyone. *This is why the primary problem of information from external sources is that nothing can be assumed to be present, correct or adequate.* In other words, the problem of bibliographic control lies within each institution and is determined by its particular purposes and needs.

In the literatures of cognitive psychology, ergonomics and management a significant etiology of failure, disasters and accidents is associated with failures to articulate goals, the articulation of general or vague goals which cover a multitude of disparate and often incompatible goals, and proceeding in a manner which contradicts the stated goals. The successful use of bibliographic information generated outside the institution will always require evaluation and adaptation, and often correction in light of the particular library's goals. So we can say that

No organization can assume that anyone else is committed to serving its users and therefore each organization will have to ensure that all bibliographical information obtained from all sources be evaluated for its users.

Misunderstanding technology

When a technical system is seen as a purely technical system, it is misunderstood. A simplistic view of technology understands the functional efficacy of its objects without grasping the system of interrelations which they create. H el ene Denis (2005, p.78) observed that

[It is] cooperation between professionals that defines the technologies and their reliability, these being not any preexisting reality but a provisional achievement.

If there is only one thing that you remember from my talk, I hope it is that statement, so let me repeat it: *[It is] cooperation between professionals that defines the technologies and their reliability, these being not any preexisting reality but a provisional achievement.* Mechanical creation, mining, harvesting, and interpretation of bibliographic data are often presented as though what these techniques make technically possible is the same thing as what is realizable, that they are neither effected by implementation nor affect the system in which they are implemented. Nothing could be further from the truth. Donald Norman (1990, p.4) observed that "The behavior of an information processing system is not a product of the design specifications; it is a product of the interaction between the human and the system." While technologies understand nothing and have nothing to communicate, human beings can use a wide variety of technologies in their efforts to communicate. In order to understand how that technologically mediated communication among human beings happens, we must understand what technologies actually do, and not be confused by the misleading metaphors which abound in LIS.

Automated methods for description, indexing, subject analysis, classification and relevance ranking all share certain common features: they are based on definitions, axioms, heuristics and the statistical analysis of texts. This means they have no access to the practices and understandings of either the author of the text or the library user who might be looking for that item, with or without knowing it. Statistical statements—about meaning or anything else—are valid only over sufficiently large populations. In the face of the particular, in this unique context, statistical laws are inapplicable and statistical statements have no validity. This means nothing less than that in the automated creation of bibliographic information every bit of information created has only a certain probability of being correct. This requires the following rule for practice:

Quality assessment requires an assessment of the individual records and all machine-generated information must be evaluated by a human being capable of evaluating the results and correcting them if necessary.

To leave evaluation and correction as matters to be attended to only if someone happens to notice is to operate according to what Dörner (1997) called "a repair shop policy", this being one of the most common etiologies of disaster in the management situations which he studied.

What about metadata harvesting and data-mining?

If we look at the move from one-at-a-time cataloging to batch processing, we can see the key issue clearly: if content is a given, produced and obtained from external sources—not produced in and by the institution—then the quality of that content is the responsibility of the producers/providers, not the library. When it is assumed that the required information already exists or will be created elsewhere later, the quality, value and responsibility for the information is located externally. Nauman and Rolker (1999)

argued that since "The main source for believability is the author or creator of the information," identifying the creators of information is important for judging information quality. When information is acquired from external sources, determining the reliability of the source is critical, but they note that this cannot be done automatically, and that one must assume that information sources will be very resourceful trying to find ways to improve believability without improving the correctness of the information itself. The matter of authority and provenance you recall was also a significant issue identified by Prof. Burke at the first meeting of the Working Group.

When the work performed in the library is by policy not to be concerned with any issues of quality, complete trust has been placed in the technical system and all sources from which information may be acquired at any point in the process. The result is that the library no longer has any possibility of maintaining a mindful awareness of what is happening and loses the internal variety required to deal with the complexity of the system in its entirety. The policies and organizational structures that I have examined in my research all involve a lapse in reliability, the removal of thought, judgement and responsibility from the persons creating, manipulating and maintaining the database, and hence from the library as a whole. Marsh and Dibben remarked:

If we are told something in an electronic world, how can we trust it? Previously, a document's credibility was to some extent maintained by knowledge workers such as librarians, editors, and other intermediaries. Today, this front line of information authentication is not always in place. The problem is compounded when one considers that searching for information may be done not by humans but by automated agents. (Marsh and Dibben, 2003, p.484-485)

Trusting external sources of metadata assumes that the ends for which those sources create information are compatible with the particular institution's intended uses of that metadata. For instance booksellers describe books in order to sell them, but what serves the bookseller's purposes may cause serious problems for the bibliographer, reference librarian and library users of all sorts. The first example in the handout should sufficiently illustrate the problem.

Compare the "metadata" in the record with what appears in the book received as fulfillment of an order placed using that order record: The language of the book is French for the entries with Albanian definitions, but the metadata supplied states that the language is Russian. The book is a dictionary but this is not noted in the contents field nor in the subject fields. The title on the book is only in Albanian, but that title is misspelled in the record, and an additional English title not on the book is given. The author given has only a first initial—an incorrect one—while the second author is not given. The place of publication is given as Russia in the fixed fields and misspelled as Tirana in the imprint field. The book has a copyright date of 2004, but in the record the date is given as unknown in the fixed fields and 2000? in the imprint. The pagination is also wrong in the record.

The vendor's metadata and the vendor's book received on that order ought to match, yet

the only fields in the entire record which are both present and correct are the ISBN and the publisher's name. Every other field/subfield is either missing, contains errors or is completely wrong. For the bibliographer, cataloger, researcher who has a citation to the actual book, it may well appear that the item described by the bookseller is a different edition, and if searched by author or qualified by date or language, it will not be found at all. The cataloger and the acquisition librarian alike are forced to ask an unanswerable question: did we in fact receive what we ordered or something different?

We can see the same kind of nonsense and inappropriate data everywhere in OCLC in records supplied by libraries, booksellers and cataloging vendors. Glenn Patton wrote to me last year "The basic issue is that we cannot index data that is not there." He was referring to vernacular searching which is only available if vernacular script is entered into the record. Records without subject data in subject fields cannot be retrieved on any subject search. The same goes for every searchable element in a bibliographic record. That missing data is a massive problem in records with transliteration, in recon records, in booksellers' records and in records input by university libraries according to policies for data mining rather than communicating with users. The use of such records in libraries creates a buildup of what James Reason (1990) called "latent failures": faulty elements within the system which no one notices, but which lead to failure in certain conditions. One example: missing information makes precise searching counterproductive. "In any situation where humans use artefacts to accomplish something, the dependability of the artefact is essential" Hollnagel (2002) noted, and "if we cannot rely or depend on the artefact, we cannot really use it."

The probabilistic catalog of automatically generated metadata discussed by Markey (2007) has one particularly important characteristic: nothing in it can be trusted. Wendell Berry noted 30 years ago that the chief product of all industrial processes is garbage, and this is as true with the automatic generation of information as of any other industrial process. Machines, software, algorithms cannot evaluate and judge for they have no values, no commitments to truth, accuracy, honesty, they make no demands for evidence, they have no purposes and no practices. They follow none of Grice's 9 standards for communication. They do not communicate for they have nothing to say. The usefulness, relevance, accuracy, truth, value and potential of any machine generated information can only be determined by a human being who is both capable of making that evaluation and has the resources to do so.

When that human evaluation is not done, when the work has been relegated entirely to an automated process, whether that is bureaucratically or mechanically enforced, the values and practices of science and scholarship— attention, observation, documentation, analysis, argumentation, verification, critique, probing for error, responsibility and accountability—have been eliminated from the library. The library can indeed continue its existence as a group of managers overseeing a technical system, but can it serve the needs of a community whose values it repudiates *in toto*? When the work of libraries is assumed not to involve the building and managing of a collection for the particular needs of a particular community then the work can be performed mechanically or outsourced to the lowest bidder *because the results do not matter*. This is ultimately why Roy Tennant

could write of LibraryThing's "Also known as" mechanism that "the quality of this work is only as good as any randomly selected group of people walking down the street would be able to produce." (posting to Autocat, 24 April 2007).

Research libraries have one reason for existing: to serve the needs of research, science and scholarship. To do that it is necessary to understand those practices and the values which unite the communities of researchers, scientists and scholars. And those which divide them: in the first meeting of the Working Group Professor Burke suggested that we might have to forget backward compatibility; that is not something all of us would agree upon. We are not from Mars; like it or not, we are encumbered by history. Much of the existing debate over structures and standards betrays not only a lack of awareness of the diverse needs of science and scholarship but a repudiation of the core values of those practices. All of the characteristics of academic research have been removed from the library in the case of outsourcing, and abolished entirely in mindless copy-cataloging, original cataloging done for data mining rather than research use, and in any and all systems which relegate description, analysis and interpretation to an automated process, whether the automation is technological or bureaucratic.

What has all this to do with structures and standards? Structures and standards do not create, do not interpret, do not evaluate and do not correct information. Structures and standards know nothing of appropriateness and accuracy, of goals and purposes. Therefore structures and standards can just as easily support the creation, recording and distribution of nonsense and misinformation as anything else, something that the examples in the handout clearly demonstrate.

If we acknowledge this, the implications for all kinds of information use and all manner of information management are far reaching. For example, with the 3rd set of questions we were asked to respond to the statement: "Data are created to be processed by applications. We mine data for meaning, merge and manipulate data [etc.]" That is an admirably clear expression of the original sin of LIS. If data are created for processing by application rather than for communication with library users, then we will create and inhabit a kafkaesque bibliographic universe. The user has been abolished in theory and practice, but of course if there is to be any sense in what happens in libraries it will be the library's users that make it, not the managers or the technologies. If on the other hand we create bibliographic information for certain users in order for those users to accomplish their tasks and goals, this will require that we know what those tasks are and how we may facilitate through data structures and standards the performance of those tasks using the tools available to the user. We can choose to understand what happens in libraries according to a theory of transportation in which all data is equal and we simply move it from one place to another, or we can understand work in libraries according to a theory of communication in which we engage readers in conversations about writers and writings, musicians and music, etc. The importance of that choice lies in the fact that we can do either—with radically different consequences.

Structures and standards: questions and responses

Briefly let me respond directly to some of the questions from the background paper.

1. The first set of questions asks what kinds of structures and standards are needed. My response: *Future technological potential will depend on more structured metadata, not less.* The more you want from information technologies, the more you have to put in. The successful use of information technologies used for purposes of communication requires far more standardization than human beings need for interpretation and use. Any ambiguities or multiple uses in the definition of fields will mean that information in these fields can never be retrospectively disambiguated by automatic means. This is a significant problem for both RDA and Dublin Core.

Again, the paper asks: How can we make better use of current structures and standards in meeting both consumer and management user needs? My emphatic response: *Support rather than suppress human intelligence. Bad policies subvert structures and standards.* In many libraries contradictory policies coexist, such as PCC and below minimal level cataloging. Policies which restrict the amount of interpretation and information regardless of the items being described destroy the potential for sharing metadata. Library policies and organizational structures are largely designed to eliminate the exercise of intelligence, acts of interpretation and judgement because these take time and cost money; yet these are precisely the acts upon which all information technologies remain completely dependent because machines do not think, do not interpret, do not make judgements, make no evaluations and know no users.

Again: What relevant communities need to have input and what organizational structures would best support this? *The most demanding needs of the most demanding users must guide all decisions about structures and standards.* Those needs and users will differ from institution to institution.

2. What about controlled data? *Cross language information retrieval and collocation by subject, author or any other field require a controlled vocabulary.* Neither keyword searching nor relevance ranking collocate anything. Topicality is a judgement rooted in an interpretation and no machine has ever engaged in interpretation or in judgement. The history of science is a history of controlled vocabulary, even though that vocabulary is argued and altered through time. No known techniques can replicate the human interpretations and judgements expressed in the subject headings created by human beings. Furthermore, the problems associated with decontextualized language are to a large degree eliminated when controlled vocabularies are also constructed in precoordinated strings using a syntax that contextualizes. LCSH is such a system.

Semantic web. Reading Jacques Arzac, Hubert Dreyfus and Roy Harris would make our assessments of these projects much more realistic.

Playing tag: Del.icio.us Flickr. *Subject analysis with LCSH in MARC is tagging.* Even David Weinberger recognizes that. There is a profound lesson to be learned from the success of that other tagging—the kind not done in libraries—but to my knowledge no one has mentioned it. The value and success of tagging is that it is being done by people involved in the creation, study and use of the items being tagged, exactly like the

activities which took place in the library at Alexandria as Jochum (1999) described them, exactly like the scientists and naturalists who collect, describe, classify and study plants, stars and cuneiform tablets. In science as in Flickr, the languages and practices are not separate. In our libraries today there is almost always a split between those actively involved in a particular field (e.g. bibliographers) and those creating the metadata. For the most part—music, map and law often being the exceptions—catalogers are hired to catalog whatever comes in the library and no subject knowledge or activity in any field of endeavor is required, expected or encouraged. Catalogers are hired to be metadata specialists without knowing anything in particular about anything at all. Therefore, when they catalog any particular item they have no stake in the result, no users in mind, no knowledge of how or why anyone might be looking for that item. Wittgenstein insisted that if we have no practices in common, communication becomes impossible, and the divorce between cataloging and reading and research activity is the source of the real difficulties we have with using and making useful LCSH. The disconnect between intellectual involvement and metadata creation is the great failure of our academic libraries. Copy cataloging compounds the problem, and automatic subject analysis is the extreme limit case.

3. Data are created to be processed by applications. ... Are our structures and standards appropriate to this reality? *Tomorrow, like today, we can only search for what is in a bibliographic record, not what we hope or dream that someday software will be able to generate from that record.* Reliance upon automatic generation or harvesting of metadata must assume that either the bibliographic universe is monolingual—which it is not in our shared databases—or that the software used will be able to perform the necessary acts of interpretation and translation to and from all languages into all others. Futurologists promised that this capacity would be available in 1957 and it continues to be little more than a promise. Heck's remark should be our mantra: "Beware also of what is sometimes said on automated methodologies. Spontaneous generation of knowledge does not exist as no methodology will ever reveal knowledge that is not already somehow in the data." (Heck, 2001, p.11)

4. On mass digitization: *The more technologically dependent the resources, the more access will cost.* Comparisons between bibliographic records for print and electronic resources suggest that the metadata required for electronic resources is at least double that required for print materials. The single field for physical characteristics in a record for a print item is multiplied many times in the record for an electronic resource—especially for items in physical formats which have been digitized—since the record must include url, indications of type of file, system requirements, rights and restrictions, and a vast array of notes needed for future tasks of digital preservation, interoperability, and much else. Karen Coyle (2007a) quoted Dan Clancy in her blog from the first meeting: "The cost of asserting opinions determines value." Google relies heavily on metadata because metadata represents cost, i.e. value.

On offsite storage. *Classified shelving is a form of metadata.* Classified shelving is one of the most important structures facilitating access in a library, and it is being eliminated in many instances on the assumption that the online catalog can take over that function. Off-

site storage, closed stacks and stacks arranged by accession number severely reduce the information that would be available to the patron in an open classified stacks. Shelving by accession number renders useless half the subject information usually present in a bibliographic record: the class number. Both the information imparted by classified shelving and the learning which it enables are lost when access to bibliographic information is restricted to the online catalog record. If libraries are going to pursue offsite storage, closed stacks and accession number (or bin number) shelving, then far more information will need to be entered into the bibliographic record for *all* users, including the managers, something Andrew Pace indicated in his list of desiderata for the catalog.

5. Libraries now manage different flows of data, created within different regimes, much of it outside the library environment. ... At the same time, we see more reuse and flow of data across publishers, libraries, agents, other bibliographic services, etc. What does this mean for our bibliographic structures and standards? *It does not mean anything if you do not look at it, pay attention, and think about what information the library and its users need.* Breure (2005) referred to reusability as the Holy Grail of content engineering and noted that it "requires that digital information be well structured ... and enriched with metadata ... Most of the strategies to achieve that purpose require special, highly controlled procedures for creating content." (p. 27) Unfortunately, Breure concluded, "the majority of content is created without the strict procedures that enable reuse." (p. 47)

Conclusion

Consider again Grice's standards for communication:

1. Make your contribution as informative as is required
2. Do not make your contribution more informative than is required
3. Do not say what you believe to be false
4. Do not say that for which you lack adequate evidence
5. Be relevant
6. Avoid obscurity of expression
7. Avoid ambiguity
8. Be brief
9. Be orderly

Without human devotion to communication along the lines Grice outlined, without attention, evaluation, interpretation, correction and adaptation for the intended users of the information, all the structures and standards in the world will produce nothing but meaningless nonsense. If we want information appropriate to our users' needs and practices, we have to create it to fit those needs and practices: it does not exist otherwise.

Bibliography

(The bibliography includes not only the works cited in the paper as read May 9th, but additional material which was discussed in a longer version circulated earlier. The additional references are to studies of reliability and error in technical systems from the perspectives of ergonomics and management.)

- Ashby, William Ross (1956). *Introduction to cybernetics*. London: Chapman & Hall. Available on line at: <http://pespmc1.vub.ac.be/books/IntroCyb.pdf>
- Bainbridge, Lisanne (1987). "Ironies of automation." In: J. Rasmussen et al. *New technology and human error*, New York: Wiley. p. 271-283.
- Breure, Leen (2005). "Reuse of content and digital genres" in: Herre van Oostendorp, Leen Breure and Andrew Dillon, *Creation, use, and deployment of digital information*. Mahwah, NJ: Lawrence Erlbaum Associates. p.27-53.
- Burrows, Howard; Suresh, Ramachandran (1998). "Digital library approaches to resource discovery in earth and space science" in: Strobl, J. and Best, C. (Eds.), *Proceedings of the Earth Observation & Geo-Spatial Web and Internet Workshop '98 = Salzburger Geographische Materialien*, Volume 27. Available at: <http://www.sbg.ac.at/geo/eogeo/authors/burrows/burrows.htm>
- Calhoun, Karen (2003). "Technology, productivity and change in library technical services" *Library collections, acquisitions, & technical services* v.27 p.281-289.
- Coyle, Karen (2007a). Users and uses: Google Scholar. Available at: <http://kcoyle.blogspot.com/2007/03/users-and-uses-google-scholar.html>
- Coyle, Karen (2007b). Users and uses: Karen's summary. Available at: <http://kcoyle.blogspot.com/2007/03/users-and-uses-karens-summary.html>
- Coyle, Karen (2007c). Users and uses: official summary. Available at: <http://kcoyle.blogspot.com/2007/03/users-and-uses-official-summary.html>
- De Keyser, Véronique (1990). "Temporal decision making in complex environments" *Philosophical transactions of the Royal Society of London. Series B: Biological sciences*, v.327, no.1241, *Human factors in hazardous situations* (Apr. 12, 1990), p. 569-576.
- Denis, Hélène (2005). "Les risques et les catastrophes" In: Minguet and Thuderoz (eds.), *Travail, entreprise et société: manuel de sociologie pour ingénieurs et scientifiques*. Paris: PUF, p.68-80.
- Dörner, Dietrich (1997). *The logic of failure: recognizing and avoiding error in complex situations*. Reading, Mass: Addison-Wesley. Translation of *Die Logik des Mißlingens*.
- Dupuy, Jean-Pierre (1980). "Analyse de systèmes et critique de la société "informationnelle"." In: F. Gallouedec-Genuys, ed., *Les enjeux culturels de l'informatisation*. Fontefraud: Centre Culturel de l'Ouest. p. 183-201.
- Fallgren, Nancy J. (2007a). *Users and uses of bibliographic data: background paper for the Working Group on the Future of Bibliographic Control*. Available at: <http://www.loc.gov/bibliographic-future/meetings/docs/UsersandUsesBackgroundPaper.pdf>
- Fallgren, Nancy J. (2007b). *Users and Uses of Bibliographic Data Meeting March 8, 2007 Mountain View, CA: brief meeting summary*. Available at: http://www.loc.gov/bibliographic-future/meetings/2007_mar08.html
- Fallgren, Nancy J. (2007c). *Structures and standards for bibliographic data:*

- background paper for the Working Group on the Future of Bibliographic Control.*
Available at: <http://www.loc.gov/bibliographic-future/meetings/docs/mtg2paperfinal2.pdf>
- Gras, Alain; avec Sophie L. Poirot-Delpech (1993). *Grandeur et dépendance: sociologie des macro-systèmes techniques*. Paris: PUF.
- Grice, H.P. (1975). "Logic and conversation" in P. Cole and J. Morgan, *Syntax and semantics*, v.3: *Speech acts*. New York: Academic Press, p. 41-58.
- Heck, André (2001). "Information handling in astronomy: beyond technologies and methodologies" *High energy physics libraries webzine*, issue 3 (March). Available at: <http://library.cern.ch/HEPLW/3/papers/2/>
- Hjørland, Birger (1997). *Information seeking and subject representation: an activity-theoretical approach to information science*. Westport, Conn.: Greenwood Press.
- Hollnagel, Erik (2002). "Dependability of joint human-computer systems," in S. Anderson et al., eds., *Computer safety, reliability and security: 21st International Conference, SAFECOMP 2002, Catania, Italy, September 10-13*. (Berlin: Springer, 2002; Lecture notes in computer science, v.2434), p. 4-9.
- Hollnagel, Erik; Woods, David D. (2005). *Joint cognitive systems: foundations of cognitive systems engineering*. Boca Raton: CRC.
- Jochum, Uwe (1999). "The Alexandrian Library and its aftermath" *Library history* v.15 p.5-12.
- Landau, M.; Stout, R. (1979). "To manage is not to control, or the folly of type II errors" *Public administration review* v.39 nr.2 (March-April), p.148-156.
- Langer, Ellen J. (1989). "Minding matters: the consequences of mindlessness-mindfulness" *Advances in experimental social psychology* v.22 p.137-173.
- "Law of Requisite Variety", in *Web Dictionary of Cybernetics and Systems* http://pespmc1.vub.ac.be/ASC/LAW_VARIE.html (viewed 8 May 2007)
- Markey, Karen (2007). "The online library catalog: Paradise Lost and Paradise Regained?" *D-Lib Magazine* v. 13 nr.1/2 (January/February). Available at: <http://www.dlib.org/dlib/january07/markey/01markey.html>
- Marsh, Stephen; Dibben, Mark R. (2003). "The role of trust in information science and technology" *Annual review of information science and technology* v.37 p.465-498.
- Morel, Christian (2003). *Les décisions absurdes: sociologie des erreurs radicales et persistantes*. Paris: nrf Gallimard.
- Naumann, Felix; Rolker, Claudia (1999). "Do metadata models meet IQ requirements?" In: *Proceedings of the International Conference on Information Quality 1999 (IQ'99), MIT*. Available at: <http://hqi.de/publications.html>
- Norman, Donald A. (1990). *The design of everyday things*. New York: Doubleday.
- Orlikowski, Wanda J. (1991). "Integrated information environment or matrix of control? The contradictory implications of information technology" *Accounting, management and information technology* v.1 no.1 p.9-42.
- Poyet, Christine (1990). "L'homme, agent de fiabilité dans les systèmes automatisés." In: Leplat and Terssac, eds., *Les facteurs humains de la fiabilité dans les systèmes complexes*, p.223-240.
- Reason, James (1990). *Human error*. New York: Cambridge University Press.
- Sperber, Dan; Wilson, Deirdre (1986). *Relevance: communication and cognition*. Cambridge: Harvard University Press.

- Tennant, Roy (2007). "LibraryThing and FRBR entities." Posting to Autocat, 24 April 2007.
- Townsend, Robert (2007). *Google Books: What's not to like?* Posted 30 April on *AHA today*, the AHA blog. Available at: <http://blog.historians.org/articles/204/google-books-whats-not-to-like>
- Truitt, Marc (2006). "On 'earth-shaking matters'..." Posting to Autocat, 30 August 2006.
- Vestrucci, Paolo (1990). *Modelli per la valutazione dell'affidabilità umana*. Milano: Franco Angeli.
- Weick, Karl E.; Sutcliffe, Kathleen M. (2001). *Managing the unexpected: assuring high performance in an age of complexity*. San Francisco: Jossey-Bass.
- Woods, David D.; Johannsen, Leila J.; Cook, Richard I.; Sarter, Nadine B. (1994). *Behind human error: cognitive systems, computers, and hindsight*. Wright Patterson AFB, Ohio: CSERIAC.

Appendix/Handout

The following ten examples of bibliographic records found in OCLC (before and after correction) clearly demonstrate problems in the organizational infrastructure for creating and maintaining bibliographic information in an environment of different flows of data, created within different regimes, some of it outside the library environment. Printouts of each record as I found it in OCLC are reproduced in the pages that follow, and for each one I have also printed out the revised record.

Examples 1-3. Bibliographic records contributed to OCLC by book seller.

1. (This example is discussed in the talk.)
2. A book in the Albanian language, it is described in the fixed fields (the searchable field) as Ukrainian, and in the note field as Azerbaijani. The fixed field for country of publication has Ukraine, but the book is published in Albania. The wrong diacritic appears in the form of name. The English title does not appear on the book. The subjects given are so general as to be of no use in subject searching.
3. Again a book in the Albanian language published in Albania but described as a Russian book in both language and country of publication fields. Author's name appears in reverse order in the statement of responsibility, unlike it appears on the book. Physical description is wrong (pagination) and incomplete (dimensions). The imprint date is given as unknown in the fixed fields and 2000? in the imprint. Diacritics missing in the title and imprint, while again the English title does not appear on the book and the subjects given are so general as to be of no use in subject searching.

Examples 4-5. Bibliographic records created for data mining

4. Aside from the lack of information, the book is described as being in English and having no place of publication in the fixed fields.
5. The fixed fields are those of the default settings rather than based upon the book. The class number is for modern Russian literature in relation to Italy (Class web note: "Class here works dealing with the influence of foreign authors on Russian literature if written chiefly in the interest of Russian literature"). The title in the record is taken from the t.p. verso, not from the title page. That title says "in 2 books" and the record itself is based on volume 2 but the dates and physical description fields indicate further volumes will be published. The nature of the book, the presence of translations, Petrarch as author and subject, and vernacular data are all missing.

Examples 6-7. Bibliographic records in OCLC supplied by large research libraries

6. According to ClassWeb the classification number is for "European Union in relation to individual regions or countries" with .R8 for Russia which matches the 3rd and 4th subjects given. However the book is not about Russia at all: the title translated is Europe

without Russia: the Treaty Establishing a Constitution for Europe of 20 October 2004. It is a translation of that treaty into Russian with an essay by Václav Klaus, and should be classed KJE (European Union constitutional law). Vernacular data not included.

7. The class number and the subject headings match, but they are all wrong.

Example 8. Copy-cataloging with neither evaluation nor correction

8. Again no vernacular data. Author mistakenly identified as subject of a congress (in fixed fields), but class number based on the additional and also completely wrong subjects for Slavic peoples, ethnography (general). Records from all of the holding libraries indicate that none of these problems were identified or corrected.

Examples 9-10. PCC records in OCLC

9. Class number and subject headings for Romania—Ethnography—General works. A note mentions volume 4, which has not been published yet, and the library's catalog indicates that only volume 1 is owned. According to the introduction of this multivolume book it will be devoted entirely to reporting on the responses to a questionnaire about habitat (domestic spaces) of the living and the dead in Romania, each volume concerning a particular region. It is not folklore.

10. A guidebook to archives of Polish institutions abroad. It is not the archives themselves, as the 6th subject heading indicates. Without making corrections LC accepted this record into its catalog where a 985 field states "VENDOR LOAD". What is really interesting here is that a record created by the National Library of Poland is also available in OCLC and in that record the correct subject headings are given, albeit in Polish. LC did not use copy-cataloging or data mining to find the good record (adapting it for an English language catalog) nor to correct the bad record, but instead accepted the incorrect record and made no corrections.