Bibliographic Babel: Surrogation as communication, or obfuscation?

Lynne C. Howarth and Joseph P. Cox
Faculty of Information Studies
University of Toronto
<howarth@fis.utoronto.ca>
<cox@fis.utoronto.ca>

This paper describes ongoing research on the content and display of bibliographic records in online public access catalogues (OPACs) accessible via the Internet. A total of ten surrogate records for monographs derived from twelve academic libraries in Canada were compared and evaluated. Using checklist methodology and a previously validated scoring mechanism (Howarth and Cox 1996) the records were examined to determine: 1) to what extent and how well do selected surrogates adhere to established bibliographic standards, and 2) how, and to what extent, do selected surrogates vary from records emanating from recognised national cataloguing agencies.

1. Introduction and background to the research

It is a well held tenet of library and information science that the creation of bibliographic records following established standards, such as the *Anglo-American Cataloguing Rules* (*AACR*) facilitates the identification of, and access to, different types and formats of materials in library collections. Uniformly structured, and consistently devised records are the surrogates for actual items or documents, and serve to represent their intellectual and physical characteristics, as appropriate. As we near the millennia, however, the validity of these assumptions concerning the utility and relative value of bibliographic surrogates is under increasing scrutiny. Witness the flurry of conferences, tasks groups, and articles in cataloguing literature [1] questioning the continued relevance and viability of existing bibliographic codes and standards (*AACR*, Machine-Readable Cataloguing (MARC), International Standard Bibliographic Description (ISBD), etc.) in an increasingly electronic, digital world.

Much of this debate concerning the future of bibliographic control mechanisms and tools is occurring within the cataloguing community itself. The Cooperative Cataloging Council (CCC), constituted in 1992, and comprised of representatives from the Library of Congress, the OCLC Online Computer Library Center, Inc., the Research Libraries Group (RLG), and several public and academic libraries in

the United States, appointed six Task Groups in 1993 to determine changes necessary to facilitating more effective cooperative cataloguing projects — the "better, faster, cheaper" framework. One group, the Task Group on Standards, was charged with determining a core-level standard for bibliographic records (Cromwell 1994). To this end the Task Group distributed a survey to the cataloguing community requesting feedback on a proposed model of a core record, and obtaining input on the needs and preferences of those individuals and institutions which created and/or derived bibliographic records. Based on findings from the questionnaire, the Task Group on Standards articulated a core-level standard mandating fewer data elements than a full-level record [2], but a more complete array of descriptive elements than a minimal-level record (Library of Congress 1980), along with twenty-one related recommendations. The CCC acted on the formation in early 1994, of a new cooperative cataloguing project, the Program for Cooperative Cataloging (PCC). The latter proposed that, "Two levels of program bibliographic records should be recognized for the books format: the currently defined full level record and the proposed core level record (as defined by the CCC Standards Task Group)" (Program for Cooperative Cataloging 1994, 5), and that an *AACR2* Code Revision Task Group be appointed.

Within a similar time-frame, the IFLA Study Group on the Functional Requirements of Bibliographic Records (established in 1992) was working to delineate the functions performed by bibliographic records with respect to various media, applications and user needs, and to propose basic data requirements for records that are to be created by national bibliographic agencies (IFLA 1992). The IFLA Study Group arose from the 1990 Stockholm Seminar on Cataloguing where, "the development of a model that would define entities of interest to users of bibliographic records, relationships that may exist between those entities, and attributes of entities and relationships required to carry out the functions of the bibliographic record" (IFLA Study Group 1995, 1) was proposed. [3] A draft report was issued for worldwide review in Spring 1996, with comments due November 30, 1996, and a final report anticipated in September 1997. As a consultant to the Study Group commented, "the model will provide for a rethinking of the records and the structure that we are currently using for catalogs" (Tillett 1996).

A third initiative for reexamining descriptive elements, with particular reference to electronic resources, was the OCLC/NCSA Metadata workshop, held in March 1995 in Dublin, Ohio, and organized by OCLC and the National Center for Supercomputer Applications (NCSA), to address the problem of providing metadata for materials accessible via the Internet. Attended by computer and information scientists, software developers, publishers, librarians, archivists, members of the Internet Engineering Task Force (IETF) working groups, and

researchers, the Workshop focused on defining a simple data element set and instructions that authors or publishers could provide when mounting documents on a network server. The resulting "Dublin Core", a "fairly short list of data elements . . . most useful and simple for naive users to use" (Library of Congress 1997, 3), concentrates on describing intrinsic properties of the electronic resource. Should the "Dublin Core" become an official or *de facto* standard, it can, ". . . serve as a basis for more detailed cataloguing or description when warranted by specific communities" and ". . . ensure a common core set of elements that could be understood across communities, even if more specific information was required within a particular interest group" (Library of Congress 1997, 3).

What each of the preceding considerations of the relevance and applicability of standards for bibliographic description have in common, in addition to their reliance on the cataloguing community for feedback, is their lack of empirical data on which to base future cataloguing rules (Tillet 1996), and their absence of direct input from end-users of the bibliographic record. A review of the literature confirms the paucity of research into bibliographic record content and end-user requirements for descriptive elements. While a baseline study by Howarth and Weihs (1994; 1995) determined that self-reporting academic, public, special, and school libraries across Canada were using the *Anglo-American Cataloguing Rules* for creating original bibliographic records for an extensive variety of material formats, and that this commitment to applying the code was even stronger among libraries contributing to bibliographic networks, the national survey could not confirm the nature and extent of individual record content.

As part of an ongoing study (Cherry and Howarth 1997, in progress), data derived from checklist and focus group methodologies were analysed to provide a client-focused framework for assessing content in bibliographic records, and to highlight gaps between user preferences and the availability and presentation of elements in some existing bibliographic systems. Using a second-level descriptive record framework from the *Anglo-American Cataloguing Rules* to develop a checklist of bibliographic elements, records for ten monographs in Internet-accessible OPACs of 10 public libraries across Canada were evaluated for their content relative to the prescribed standard (Howarth and Cox 1996). Concurrently with checklist data collection, focus group interviews were held with a stratified sample of public library users who were asked to rank the relative importance and usefulness of elements of bibliographic records, and subsequently to "prototype" their ideal record (Luk 1996). Findings suggested that the content and integrity of bibliographic records being created currently may be insufficient and inadequate to appropriately address even a "short list" of bibliographic requirements articulated by a sample of public library OPAC users. The authors mused further: "If the

bibliographic structures we have created to facilitate user identification of, and access to, materials have proved insufficient, incomplete, or even useless for clients in the OPAC environment, one wonders at the further alienation of bibliographic 'product' from 'consumer' in the increasingly visually oriented and graphically flexible world of electronic resources" (Howarth and Cox 1996, 48).

As the discussion, above, may suggest, while there is a "dialogue on description" proceeding apace in a number of communities, the relative lack of empirical evidence concerning the actual nature and content of records being created for, and subsequently accessed through, OPACs, combined with scant reference to empirically verified end-user requirements for bibliographic data elements, may be serving to reinforce the gaps between and among those who create standards, those who create the records based on those standards, and those who use the records that are created based on agreed standards. This "bibliographic babel" may be perpetuating long acknowledged difficulties with achieving the goal of creating records that, ". . . satisfy the needs of the user in identifying and locating bibliographic entities in the bibliographic universe" (IFLA Study Group 1995, 7).

This paper describes one part of ongoing research on the content and display of bibliographic records in OPACs accessible via the Internet. In an effort to provide for the Canadian academic library community, a profile of bibliographic record content, cataloguing practice similar to that determined for public libraries in Canada (Howarth and Cox 1996), and to add to the body of empirical research concerning the nature and content of bibliographic records in OPACs, a total of ten surrogate records for English language, non-fiction monographs derived from twelve academic libraries in Canada were compared and evaluated. Using checklist methodology, and a previously validated scoring mechanism (Howarth and Cox 1996), the bibliographic records from each site were examined within the framework of the following research questions:

1. To what extent, and how well do selected surrogates for monographs held in a subset of Canadian academic libraries adhere to the established bibliographic standard as created by cataloguing "experts" based on a second level descriptive record defined in the *Anglo-American Cataloguing Rules*, second edition, 1988 revision, (with 1993 amendments) (*AACR2R*)?

Presumably the less the discrepancy between records created by each site, and those done by experts, the greater the adherence to established cataloguing standards.

2. How and to what extent do selected surrogates for monographs held in a subset of Canadian academic libraries vary from records emanating from recognized national cataloguing agencies, as represented by the National Library of Canada (NLC), and the Library of Congress (LC), respectively?

Assuming that the national cataloguing agencies are creating bibliographic records to *AACR2R* second level description standard (i.e. full-level records), then presumably the greater the variability between site records and those created by each of the national cataloguing agencies, the less reliance on a "mark and park" policy whereby sites derive NLC or LC copy with little or no modification. High variability might also indicate that the site had created the record originally or modified a derived record to meet local requirements. For example, a site might choose to provide for their users additional notes not included in a record from NLC or LC.

2. Method

2.1 Developing the checklist

The checklist created to evaluate bibliographic displays is based on *AACR2R*, second level description. The "Material specific details" area was excluded as inapplicable to monographic materials. The Checklist of Bibliographic Elements in Items (see Table 1) illustrates which bibliographic elements are present in each record for items in the study. Elements which should be present in the record for each item are indicated with an asterisk * ; shaded cells are those elements which are not applicable for each item. The far right side column gives the total number of times each element occurs across all items, while the bottom row totals the number of elements in records for each item. Out of twenty-three possible elements (as per *AACR2R*), all were represented in at least one of the items selected for the study, for an overall total of 133 occurrences of bibliographic elements.

Table 1: Checklist of Bibliographic Elements in Items

	Items										
Element	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**	**Total**
Title proper	*	*	*	*	*	*	*	*	*	*	10
General material designation	*	*	*	*	*	*	*	*	*	*	10
Parallel title				*							1

	Items										
Other title information		*	*	*				*	*		5
First statement of responsibility	*	*	*	*	*	*	*	*	*	*	10
Subsequent statement(s) of responsibility					*						1
Edition statement					*	*	*				3
Statement of responsibility for edition						*	*				2
First place of publication	*	*	*	*	*	*	*	*	*	*	10
First publisher	*	*	*	*	*	*	*	*	*	*	10
Date of publication	*	*	*	*	*	*	*	*	*	*	10
Extent of item	*	*	*	*	*	*	*	*	*	*	10
Other physical details	*	*	*				*	*		*	6
Dimensions	*	*	*	*	*	*	*	*	*	*	10
Series title proper	*	*	*					*	*	*	6
Series statement of responsibility			*								1
Series ISSN									*		1
Series numbering	*		*								2
Subseries title	*	*								*	3
Subseries ISSN		*									1
Subseries numbering		*								*	2
Notes	*	*	*	*	*		*	*	*	*	9
ISBN	*	*	*	*	*	*	*	*	*	*	10
Total	14	16	15	12	12	11	13	13	13	14	133

2.2 Selection of monographs

A total of ten monographs were selected to evaluate displays in academic libraries. To be considered, items had to be English language, non-fiction, and owned by most, if not all, twelve sites. As well, they had to illustrate different bibliographic features in order to provide a good sampling of how OPACs organize and display the various bibliographic elements permissible according to guidelines for second level description in *AACR2R*. The following lists the items selected:

T1 Barbour, Douglas. *Michael Ondaatje*. New York: Twayne, 1993.

T2 Buck, Arthur C. *Jean Giraudoux and Oriental thought: a study of affinities*. New York: P. Lang, 1984.

T3 Cartwright, Donald G. *Official language populations in Canada: patterns and contacts*. Montreal: Institute for Research on Public Policy, 1980.

T4 Finley, E.G. *Education in Canada : a bibliography*. Toronto : Dundurn Press, 1989.

T5 Foucault, Michel. *The history of sexuality*. 1st American ed. New York: Pantheon Books, 1978-1986.

T6 Garland, Henry and Mary Garland. *The Oxford companion to German literature*. 2d ed. Oxford: Oxford University Press, 1986.

T7 Gray, Henry. *Anatomy of the human body*. 30th American ed. Philadelphia: Lea and Febiger, 1985.

T8 *Marketplace Canada: some controversial dimensions*. Toronto: McGraw-Hill Ryerson, 1982.

T9 McGee, R. Harley. *Getting it right: regional development in Canada*. Montreal: McGill-Queen's University Press, 1992.

T10 Ali, M.A., ed. *Photoreception and vision in invertebrates*. New York: Plenum Press, 1984.

Asterisks (*) in Table 2 indicate which bibliographic records for the items were available in the selected sites. Note that Site A9 did not hold item T2, that neither Site A7 nor A9 held T7, and that item T10 was unavailable in Site A8. The remaining sites held all of the items, for a total matrix of 116 bibliographic records.

Table 2 : Monograph Items held at selected Academic Libraries

	Items										
Sites	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**	**Total**
A1	*	*	*	*	*	*	*	*	*	*	10
A2	*	*	*	*	*	*	*	*	*	*	10
A3	*	*	*	*	*	*	*	*	*	*	10
A4	*	*	*	*	*	*	*	*	*	*	10
A5	*	*	*	*	*	*	*	*	*	*	10
A6	*	*	*	*	*	*	*	*	*	*	10
A7	*	*	*	*	*	*		*	*	*	9
A8	*	*	*	*	*	*	*	*	*		9
A9	*		*	*	*	*		*	*	*	8
A10	*	*	*	*	*	*	*	*	*	*	10
A11	*	*	*	*	*	*	*	*	*	*	10
A12	*	*	*	*	*	*	*	*	*	*	10
Total	12	11	12	12	12	12	10	12	12	11	116

2.3 Selection of Canadian academic libraries

Selection of academic library sites was limited to members of the Canadian Association of Research Libraries (CARL) with OPACs accessible through the Internet. The National Library of Canada and the Canada Institute for Scientific and Technical Information were removed from the list. The OPAC system and geographical location of each site were also considered, in order to ensure as representative a sample as possible. Predominantly French language institutions were not included, as there was a very low level of overlap with the collections of English language institutions. The ranking of libraries by success rates of searches for 12 sample monographs also eliminated two low-scoring English language institutions.

The following sites were selected for the study:

Memorial University	University of British Columbia
Queen's University	University of Toronto
University of Western Ontario	University of Manitoba
University of Saskatchewan	Simon Fraser University
McGill University	McMaster University
University of Alberta	Dalhousie University

2.4 Expert cataloguing

During the data collection process, it became apparent that there would be value in developing model records for items to facilitate comparing how various libraries handle different bibliographic elements. Therefore, all items were catalogued by a group of experts with instructions to create records solely on the basis of *AACR2R*, second level description.

2.5 National agency cataloguing (source records)

The researchers also recognized the importance of capturing LC or NLC cataloguing copy for comparison with and evaluation of, records from selected sites. This additional information might provide some insight into the dissemination of bibliographic information as well as trends in cataloguing practices.

3. Data collection

Once the items had been selected, two research assistants separately evaluated the bibliographic records from each of the sites. On a data collection sheet, a check mark was used to indicate an element was present; an "X" indicated the element should have been present, but was not; a dash meant the element was not applicable. In cases where more than one record for the same item was available in a catalogue, the first record was chosen. The OPAC screens were captured using Host Presenter in Windows and retained for future reference, including MARC format screens if accessible. Additionally, the research assistants made notes on any errors, omissions or anomalies in the records in the "comments" section of the data collection instrument.

4. Data tabulation

Once data were collected for each item from all sites, summaries by (1) item, (2) site, and (3) element were tabulated to show where prescribed information was

present, where data were required but not supplied, or where information was not applicable.

A scoring method previously devised and validated (Howarth and Cox 1996), was applied to every element in every item in every site, according to the criteria for variability (V1-V6) listed in Table 3. The frequency with which each type of error or variability (V1-V6) occurred by element was totalled, along with the number of times that error or variability type (V1-V6) was repeated across sites. This tabulation summarized in what elements certain types of errors or variations in records occurred, and how often those occurred.

Table 3: Scoring Mechanism for Determining Variability between Experts/LC/NLC and Sites

V1.	Typos, punctuation, diacritics, spacing
V2.	Different wording, same meaning
V3.	Differences in rule interpretation
V4.	Partially complete content
V5.	Completely incorrect content
V6.	Missing data entirely

The researchers also calculated the total number of data elements not present in expert or source records that were added by sites to their records. Most typically these included additional notes supplied to local records. With types of errors or variations and additions to records having been tallied, comparisons were made (1) between data in the expert cataloguing records and the 12 site records, (2) between the source records from the Library of Congress and the 12 site records, and (3) between the source records from the National Library of Canada and the 12 site records.

5. Data analysis and findings

5.1 Comparison of site records to expert records

Table 4 summarizes the types of errors or variations by element between the expert records, i.e. those created according to *AACR2R* second-level description (the model or "ideal" record) and academic library site records. Table 5 ranks the top ten data elements in which errors or variations occurred with greatest frequency between expert and site records, Library of Congress and site records, and National Library of Canada and site records. Ranking first for Expert records was data element "General Material Designation [gmd]", which all sites consistently omitted,

and which was scored as a difference in rule interpretation (type V3). The gmd is specified as an "optional addition" within *AACR2R*, which may be supplied by the cataloguer if deemed useful for distinguishing the same item in different material formats (e.g. *Winnie-the-Pooh* the book; the videorecording; the talking book), particularly in an integrated catalogue environment, such as an OPAC. What is noteworthy is the consistency with which all sites chose to omit the gmd, perhaps reflecting internal cataloguing policy that is devised to agree with both LC and NLC practice of excluding the designation "text" for monographs.

Table 4: Error or Variation Type (V1-V6) by Element: Site to Expert Record Comparison (Summary)

Error Type (V1-6) (frequency of occurrence/ number of occurrences) **Bibliographic Elements**	**V1**	**V2**	**V3**	**V4**	**V 5**	**V6**	**Total Errors by Element**
Title proper	3/3						3/3
General material designation [GMD]			10/ 11 6				10/ 116
Parallel title	1/9						1/9
Other title information	2/2						2/2
1st statement of responsibility	9/58	1/4				4/4	14/66
Subsequent statement of responsibility	1/6						1/6
Edition statement	3/12						3/12
Statement of responsibility for edition	2/7		1/1			2/10	5/18
1st place of publication	1/1	3/1 3	1/1				5/15
1st publisher	1/6	2/1 3	1/9				4/28
Date of publication	10/1 2	1/1	6/3 3	1/1 0	1/ 1		19/57

Error Type (V1-6) (frequency of occurrence/ number of occurrences) **Bibliographic Elements**	**V1**	**V2**	**V3**	**V4**	**V 5**	**V6**	**Total Errors by Element**
Extent of item (pagination)	3/7	1/2	2/5	6/2 8	1/ 1		13/43
Other physical details (illustrations)	2/10	3/8	3/4	2/1 3		2/13	12/48
Dimensions (size of item)	8/48		1/1			6/25	15/74
Series title proper	6/63	4/1 6				1/6	11/85
Series statement of responsibility	1/4	1/1					2/5
Series ISSN	1/1					1/6	2/7
Series numbering	1/5	1/1					2/6
Subseries title	3/9	2/7		1/1			6/17
Subseries numbering	1/4	1/8					2/12
Notes (summarized)	9/30	7/2 3		7/2 4		5/35	28/112
ISBN	8/14	5/7		4/1 6		10/6 5	27/102
Total frequency of occurrences by error type	76	32	25	21	2	31	187
Total number of all occurrences of error	311	10 4	17 0	92	2	164	843

Table 5: Top 10 Ranking of Element Variability

Rank	Experts	Rank	Library of Congress	Rank	National Library of Canada
1	General material designation	1	Additions to record	1	Notes area
2/3	Notes	2	Series title proper	2	Additions to record
	Additions to record	3	Notes	3/4	Series title proper
4	ISBN	4	First statement of resp.		First statement of resp.
5	Series title proper	5	Dimensions	5	ISBN
6	Dimensions	6	ISBN	6	Dimensions
7	First statement of responsibility	7	Extent of item	7	Other physical desc.
8	Date of publication	8	Date of publication	8	Extent of item
9	Other physical desc.	9/10	Statement of resp. for ed	9	Date of publication
10	Extent of item		Other physical desc.	10	Series statement of resp.

Errors or variations in the notes area ranked second, and included types V1, V2, V4, and V6 errors. Variations were most frequently manifested in typographical errors, or errors in spacing, punctuation, or spelling. Spelling or typographical errors would prove detrimental in OPACs that supported full-text searching on the notes field. Notes created by the sites that differed from the experts' notes area in terms of using different wording for the same meaning (type V2), or being partially complete in content (type V4), might impact a user's interpretation of the intellectual or physical nature of an item. While only 5 instances of notes included by experts but missing from sites occurred, more sites repeated this type V6 variation than repeated any other type of error. Offsetting this, however, was the incidence with which sites supplied notes that had not been included in the expert records. Notes comprised 96 (or 86.5%) of the 111 additions

made to site records. Some sites added more notes than others, with numbers ranging from a low of 2 (or on average, 1 additional note to every 5 records) (Sites A3 and A4) to 21 (or on average, 2 additional notes for every record) (A11). The relative "richness" of records afforded by including additional information about an item in the notes area, would seem to vary considerably across sites, while some sites would appear to exercise little modification to a basic standard record.

The fourth most frequently occurring error or variation was that involving ISBN. Significantly for precise record identification and retrieval, ISBN was consistently missing in records from Sites A1, A2, A3, and A4, and sometimes missing for six other sites. Only two sites (A8 and A11) consistently included ISBN in their records. For the most part, MARC records were not accessible through the OPAC displays, so it was not possible to confirm whether or not ISBN was recorded in a site's MARC record but not in its OPAC display. Regardless of whether that omission reflects cataloguer error or limited OPAC system display functionality, the fact that one third of all sites did not display ISBN, a unique identifier and precise search key for monograph titles, could reflect negatively on client search and retrieval strategies.

Ranking fifth in number of occurrences of types of errors or variability was another potential access point — series title proper. Most typically errors in typography, punctuation, diacritics, or spelling (type V1) predominated; only one instance of this data element being missing was repeated across six sites. On the other hand, relatively few errors occurred relative to a key access point, namely that of title proper. Only three type V1 variations were recorded in a total of three sites. Likewise, only one type V1 error occurred in nine sites for parallel title — another example of a potential access point. The admirably low instance of a "completely incorrect content" error (type V5) was committed by one site for data element "date of publication", and for one site with "extent of item". No "completely incorrect content" scores pertained to any access point, although the errors which were recorded might serve to confuse identification of an item to some extent.

Overall academic library sites appeared to be adhering to the standards prescribed by *AACR2R* second-level description, and as reflected in the records created by the experts. As Table 6 summarizes, the greatest variation between site and expert records is concentrated in type V1 errors (typos, punctuation, diacritics, spelling). One limitation of the scoring mechanism which could be addressed prior to subsequent data analyses, is the bundling of typographical and spelling errors with diacritics and punctuation. The former have the potential to negatively impact successful retrieval of items, whereas errors or variations with diacritics and punctuation may be of less consequence from a user's perspective. Thus, while type V1 errors predominate, their potential impact is difficult to gauge without a

more granular analysis of the nature of the error. The large percentage of type V3 errors or variation (i.e. differences in Rule interpretation) relative to percentages calculated for either of the two national agencies, is directly attributable to the omission of gmd by all sites. While, technically speaking, omission of this element could be considered a deviation from adherence to *AACR2R* standard, the rules, themselves, permit such omission through the provision of "optional addition". The 0.2% rate of error related to "completely incorrect content" is a reassuringly low figure suggesting that, somewhat consistent with the findings from a previous study of ten public libraries in Canada (Howarth and Cox 1996), academic library records were more likely to vary from *AACR2R* descriptive standards in terms of (1) errors in typography, spelling, diacritics, or punctuation, or (2) rule interpretations, or (3) missing data, than they were to include completely incorrect content — an error type most indicative of non-adherence to standard.

Table 6: Percentage Error or Variation by Type (V1-V6): Site Records to Expert, Library of Congress, and National Library of Canada Records, respectively

Error Types	% Type of error in site records when compared to the Expert records (n = 843)	% Type of error in site records when compared to Library of Congress records (n = 574)	% Type of error in site records when compared to National Library of Canada records (n = 390)
V1	36.9%	50.9%	48.7%
V2	12.3%	11.5%	15.4%
V3	20.2%	8.4%	8.2%
V4	10.9%	9.6%	6.2%
V5	0.2%	0.2%	0.3%
V6	19.5%	19.5%	21.3%

Table 7: Sample from Scoring Summary: Cataloguing of Titles T1-T10 at Sites A1-A12 Compared with LC Cataloguing

Elements	Error type V 1-V 6	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Error total	Element total	Rank
Title proper	1								2				2	4	4	19
Parallel title	1		1	1		1	1	1	1	1	1		1	9	9	16
Other title information	1								2					2	2	21
First statement of responsibility	1	9	9	5		2	9	2	5	3	8	2	5	59	68	4
	2	1			1	1		1						4		
	6								5					5		
Subsequent statement of responsibility	1	1	1	1			1				1		1	6	7	17
	6								1					1		
Edition statement	1	1	1	1			2		3	1	2		3	14	14	14
Statement of responsibility for edition	1	1	2				2				2			7	18	9
	3					1								1		
	6	1		1	1	1		1	2	1		1	1	10		
First place of publication	1						1					1		2	17	12
	2				1		1			1				3		
	3	1	2	1	1		1	1	1	1	1	1	1	12		
First publisher	1	1	1		1		1		1				1	6	16	13
	2								1					1		
	3	1			1	1	1	1	1	1	1		1	9		

5.2 Comparison of site records to Library of Congress and National Library of Canada records (source records)

From the data collected it is clearly evident that there are differences between the records created by the two national agencies considered in this study, namely, the Library of Congress and the National Library of Canada, and the display of the records of the selected academic library sites. In separate analyses, the bibliographic records created by these two agencies were compared with the bibliographic records available for the same titles at the 12 sites. There was cataloguing copy available from the Library of Congress for all of the items used in this study, and there was copy available from the National Library of Canada for items T1, T3-T4, and T8-T10 (6 items), for a total of 16 source records. Using the same scoring mechanism as identified above, each element in every site record was compared to the elements of the source records and Table 7 presents a compilation of he error types for each bibliographic element at all of the sites compared to the cataloguing created by the Library of Congress. Table 7 is an abridged example of the tables created to analyse the collected data. Similar tables were also created to compare the records of the academic sites with the records of the National Library of Canada and the Experts.

Table 7 identifies the bibliographic elements found in the source records and the site records, and indicates the type of error found in each element across all sites. The last three columns on the right give: an error type total for each element; total number of errors per element; and an overall ranking of the elements based on the gross number of errors per element. For instance, First Statement of Responsibility, shows a high degree of variation between the data found in the source record and that found in the site records. Across all sites there were 59 occurrences of V1 errors, 4 occurrences of V2 errors, and 5 occurrences of V6 errors, for a total of 68 errors. This bibliographic element thus ranked fourth within the ranking of all bibliographic elements, with the first ranked element having the greatest number of variances. In addition to identifying the variances in the site records, additional information present in the site records but absent in the source records was also tabulated. In the case of LC, there was a total of 119 pieces of additional information included in the sites that did not appear in the source record and this ranked first as the difference between the site records and the LC source records.

Table 5, above, shows the rankings of the bibliographic elements in all of the sites when compared with the elements found in the Experts cataloguing, the Library of Congress source records, and the National Library of Canada source records, respectively. From the information contained in the complete tables, of which Table 7 is an example, site records differ in many respects to source records.

The major differences between source records and the site records are identified above in Table 5. For the LC records, the major difference is the additional fields which are added to site records. There were a total of 119 additional pieces of information in site records that did not appear in LC records. The majority of additional data related to the notes fields (98), followed by 21 additions to the Other Physical Details area. The variation in sites ranged from a high of 22 pieces of additional data at site A11, to a low of 2 at site A3. The addition of data ranked second in the records when compared to the NLC records. Sites A9 and A11 added the most information (10), while A2 and A7 added the least (1). The highest ranking variation between the NLC records and the site records was in the Notes area, where there were 67 occurrences of errors. This can also be compared to the rankings of the site records to the Experts records, where Notes and Additions tied for second and third, while the greatest difference between the Experts and the sites records was in the General Material Designation area, where there were 116 occurrences of the data not being present. The lowest variability between LC and site records was in Series ISSN, for which there was a score of 1. Comparing the lowest variability between NLC and site records, one finds that there were just two elements with 0 variability, namely Subsequent Statement of Responsibility, and Subseries Numbering. When one compares how the sites ranked in variability, one finds that site A6 had the highest number of variances (92 or 13.3%, where n=693), while site A5 had the lowest variance (43 or 6.2%, where n=693).

From the data it appears that libraries are not simply deriving records for "mark and park". There is a great deal of variability between site and source records, whether the source records are from the Library of Congress or the National Library of Canada. It is most interesting that many libraries appear to make additions to source records, particularly with the addition of notes. Table 6, above, identifies the percentages of the type of errors encountered when comparing LC and NLC records with the site records. As can be seen, the percentage type of error in site records compared to both LC and NLC records is quite similar. V1 errors (typos and errors in punctuation, diacritics and spacing) in site records compared to LC records are 50.9% of all errors, while for NLC records, V1 errors are 48.7% of all errors. V2 errors (different wording, same meaning) yielded 11.5% of errors against LC records and 15.4% against NLC records. It is interesting to note that V5 errors (completely incorrect content) occurred very rarely and only accounted for 0.2% of errors against LC records and 0.3% of errors against NLC records.

Thus from the data it appears that there is indeed variability between the cataloguing data provided in site records and the data provided in source records. Academic libraries apparently see the need to have additional notes that are not

present in the source records, and the greatest number of errors in records are V1 errors, namely, typos and errors in punctuation, diacritics and spacing.

6. Discussion and conclusions

This study has investigated the nature and content of bibliographic surrogates created by a sample of academic libraries in Canada for a selection of monograph titles. In reviewing the findings, and within the framework of surrogates as "communication" versus "obfuscation", how well do the candidate sites and records measure up? For nearly a century and a half the bibliographic community has grappled with developing standards that would assist users in identifying and retrieving the information they required. This work continues, and has even accelerated with the more recent explosion of electronic resources entering the bibliographic domain.

Findings suggest that, insofar as the standard itself (i.e. *AACR2R* second level description) can be considered an appropriate vehicle for communicating the intellectual content and physical characteristics of an item, the academic library sites sampled are consistently adhering to standard for the most part. There is little evidence of incorrect content in records to mislead or misinform an OPAC user. Key data elements, such as the important title proper access point, seem relatively free of errors or variations. Moreover, sites appear to be enhancing records with additional notes, potentially providing for better communication of an item's properties to a searcher. There is some indication that much cataloguer discretion, or bibliographic judgement is being exercised, and that records are being created or, in the case of records derived from sources such as the Library of Congress, or National Library of Canada, modified to address local requirements. While the extent of this "cataloguer autonomy" varies across sites, it seems apparent that institutions are not simply deriving records, adding a local call number or other identification, and filing the copy in the OPAC database otherwise untouched — the so-called practice of "mark and park".

But while the ability to successfully communicate information about items in an academic library's collection may be supported to some degree by these observations, findings also suggest the potential for "obfuscation" — for including in, or omitting from, bibliographic records information necessary to the identification and access to materials. Evidence points to the consistent omission of a precise access key, the ISBN, across a number of sites. Spelling and typographical errors can render an item inaccessible to someone searching an OPAC. Missing data elements, such as gmd or other fields, may also detract from identification or confirmation of material content and relevance for a user. The concentration of error or variation types V1 (typos, diacritics, punctuation,

spelling), and V6 (missing data) point to the potential for obscuring access to materials or confusing users as to their content.

Findings from this study can inform us as to the nature and content of bibliographic surrogates in academic library OPACs. The piece which is missing as context for evaluating the value or relative usefulness of those records in an empirically-based profile of academic library user requirements. Catalogue code developers continue to discuss among themselves, and often in concert with those who design and maintain large bibliographic databases, database systems, and OPACs, appropriate record structure and content. It may be that the dialogue needs to be expanded to include the voice of the OPAC end-user. With a clearer understanding of how OPAC searchers interpret and use the descriptive elements in bibliographic elements to identify and retrieve different materials, it may be possible to reassess the extensive bibliographic resources that have been created, and to focus cataloguing energies on those descriptive areas that are of greatest benefit to the search process and that may require revision or further enhancement. Cataloguers, database developers, and OPAC system designers share in common a commitment to facilitating the link between information resources and end-users. This research would suggest that a sound bibliographic foundation has been created within the framework of an international standard. With user profiles to add context to this base of data elements we may have a potent combination for even more effective modes of bibliographic communication than have been achieved to-date.

Acknowledgement

The authors gratefully acknowledge the financial assistance provided by the Social Sciences and Humanities Research Council of Canada in funding this research project. They also wish to thank Trina Richard, Catalogue Librarian at Ryerson Polytechnic University Library, Toronto, for her many and invaluable contributions to this project, including the scoring of errors in the bibliographic records.

Notes

[1] Examples include the International Conference on the Future of *AACR* to be held in Toronto, Canada, October 1997; conferences on the Dublin Core (March 1995; 1996; 1997); meetings of the PCC Task Groups; meetings of the IFLA Task Group on Functional Requirements; meetings of MARBI and the ALA CCS Committee on Description and Access [CC:DA]; ongoing discussions concerning harmonizing USMARC, UKMARC, CANMARC; etc.

[2] A full-level record contains, in part, a description based on the second level of descriptive cataloguing as contained in the *Anglo-American Cataloguing Rules*, second edition, 1988 revision, 1993 amendments.

[3] The descriptive portion of the catalogue record was only one of several components assigned for Study Group consideration. Access points (name, title, subject, etc.), other "organizing" elements (classification, etc.), and annotations were also under review.

References

Cherry, Joan M., and Lynne C. Howarth. 1997. *Towards more useful bibliographic displays* (in progress).

Cromwell, Willy. 1994. The core record: A new bibliographic standard. *Library Resources and Technical Services*: 36: 415-424.

Howarth, Lynne C. and Joseph P. Cox. 1996. Facilitating access to electronic resources: Matching bibliographic record content with client preferences. In *Electronic publishing: Its impact on publishing, education, and reading*, ed. Charles T. Meadow, Maggie Weaver, and Francoise Hebert, 37-49. Canadian Association for Information Science, Proceedings of the 24th Annual Conference, June 2-3, 1996, Faculty of Information Studies, University of Toronto, Toronto, Ontario.

Howarth, Lynne C. and Jean Weihs. 1995. *AACR2R* use in Canadian libraries and implications for bibliographic databases. *Library resources and Technical Services*: 39: 85-99.

Howarth, Lynne C. and Jean Weihs. 1994. *AACR2R*: Dissemination and use in Canadian libraries. *Library resources and Technical Services*: 38: 179-189.

IFLA Standing Committee for Section on Cataloguing. 1992. *Terms of reference for a study group on functional requirements for bibliographic records.* September 4.

IFLA Study Group on the Functional Requirements of Bibliographic Records. 1995. *Functional requirements of bibliographic records: report.* Draft, July 31.

Library of Congress. 1997. *Metadata, Dublin Core and USMARC: A review of current efforts.* Discussion paper no. 99. Revised January 21. Available at: gopher://marvel.loc.gov:70 00\listarch/usmarc/dp99.doc.

Library of Congress. 1980. *National level bibliographic records — books.* Washington, D.C.: Library of Congress.

Luk, Annie T. 1996. *Evaluating bibliographic displays from the users' point of view: A focus group study.* Master of Information Studies Research Project Report. Toronto: Faculty of Information Studies, University of Toronto.

Program for Cooperative Cataloging. 1994. *Five-year strategic plan: 1994-1999.* Washington, D.C.: Program for Cooperative Cataloging.

Tillett, Barbara B. 1996. Comments derived from minutes of the meeting of the American Library Association, Association for Library Collections and Technical Services, Cataloging and Classification Section, Committee on Cataloging: Description and Access (CC:DA), New York, N.Y., July 8.