

Representation: A Common Thread in the Traditions and Practices of Information Organization

William E. Moen

Associate Professor, Texas Center for Digital Knowledge, University of North Texas,
PO Box 311068 Denton, TX 76203.
Email: <wemoen@unt.edu>

ABSTRACT

A singular activity in the practice of information organization (outside of physically organizing or arranging information objects in files, on shelves, or in rooms) is the creation of representations of information objects that can be manipulated and assist in the management of as well as user access to the objects represented. The library community has referred to bibliographic records as surrogates, things that stand in for the objects described. Tools resulting from the information organization traditions of various communities such as the eight traditions referenced by Miksa (i.e., bibliography, library cataloging, indexing, documentation, computerized information storage and retrieval, archival control, records management, and museum) typically use representations as key components for information organization. For example, a bibliography contains representations of articles, papers, reports, or books organized by a topic, person, or other criteria.

- What is the purpose of the representation?
- Who is the intended audience of a particular scheme of representation?
- What actions can be taken on or because of the representation (or component parts thereof)?

A representation is not the object represented. Salient features of an object are selected as important and are included in the representation. Other features of the object are left behind. Different communities of practice, and the audiences for which they prepare their representations, may highlight similar features or choose features to represent that are specific to their community's needs. The same object may be represented differently by different communities of practice (or traditions in Miksa's terminology), and each representation may be valuable within the context of that community.

The goal of this presentation is to explore the concept of representation and argue for its role as a component in a general approach to information organization.

Questions

1. How can the concept of representation in the context of information organization practices provide insight into community-specific traditions?
2. To what extent can the concept of representation provide a common point of reference when discussing information practices of different communities and their traditions?
3. Can the concept of representation play a key role in developing any general approach to information organization?

Introduction

About 15 years ago, when I was a student at Syracuse University, Dr. Liz Liddy gave a presentation at a brown bag lunch on some of her work on natural language processing. Not having worked with NLP and still learning about how machines acted on information stored in catalogs and other information retrieval systems, the whole idea of NLP and results from NLP systems seemed magical to me. How could the machines be so intelligent?

Yes, I knew the machine weren't intelligent, and the intelligence was really human intelligence embodied in software programs. Yet it wasn't clear to me how the programs operated on full-text, unstructured text, to analyze and make sense, to "understand" as it were, the meaning of the text.

Liz described and discussed the various types of analysis her systems conducted on the text. Those analyses and processes added new information to the text. The new information was in the form of assigning tags, tags that might indicate a piece of text was a noun phrase, a name, etc. The full result of the processes was a new version of the text, a new representation of the text, a representation that could then be acted on by other programs yielding the results that had seemed so magical to me. In a parallel manner, queries were subjected to similar analyses, with new information added to the query to give the impression the NLP system "understood" the meaning of the query. The matching component of the system then attempted to match these representations.

I tell this anecdote because of the long-lasting effect Liz's presentation had on my thinking about what we do to or with information objects to help users efficiently and effectively find relevant information or information objects through retrieval systems. Representation, I believe is a key and critical activity in the theory and practice of information organization.

A number of people have influenced how I have been thinking about representation. Hopefully I haven't mangled their ideas too much as I try to wrestle with this concept and its implications for theory and practice.

Let me read a short paragraph from one of those people, Dr. Brian O'Connor, a colleague of mine at UNT. This is from his book on indexing and abstracting:

Representation is a fundamental concept in indexing and abstracting. Most methods of retrieving documents depend on some form of representation of the collection documents, as well as the representations of the questions brought to that collection. Although we need to make an exhaustive study of concepts of representation, it will be informative to spend some time considering representation issues relevant to indexing and abstracting.

I think we still have that outstanding challenge from Brian to make an exhaustive study of concepts ore representation as it relates to our work.

I would like to share with you today some ideas that I believe point to the concept of representation as a necessary component to a general approach to information organization. I mean to stimulate discussion about this idea, but admit at the outset that there is likely much more work to do to flesh some of these ideas out and likely many things I have overlooked in my own thinking about this.

Representation and Description – Some Definitions

Library catalogers, one of the traditions mentioned by Fran, operate on the principle of describing the work in hand. Through the application of sets of rules, i.e., the Anglo American Cataloguing Rules, the engage in one process that is called descriptive cataloging.

With that legacy, why not simply use the term description rather than representation? There is definitely an interplay between the concepts of representation and description. The **Oxford English Dictionary** provides a starting point for some definitions.

Describe

2. To set forth in words, written or spoken, by reference to qualities, recognizable features, or characteristic marks; to give a detailed or graphic account of. (The ordinary current sense.)

Represent

3. a. To describe as having a specified character or quality; to give out, assert, or declare to be of a certain kind.

6. a. To symbolize, to serve as a visible or concrete embodiment of (some quality, fact, or other abstract concept).

Representation

2. a. An image, likeness, or reproduction in some manner of a thing

The American Heritage® Dictionary of the English Language has some additional definitions:

Describe

1. To give an account of in speech or writing
2. To convey an idea or impression of

Represent

- 1a. To stand for; symbolize:
- 1b. To indicate or communicate by signs or symbols
3. To present clearly to the mind

The notion of representation seems to provide a bit more abstract way of thinking about what we do in our practices of information organization. For many years, the term surrogate has been used to describe a bibliographic record. The record “stands in for” the information object. The function of the representation is to be a surrogate. But the term surrogate doesn’t really help us think through our practices as much as the term representation.

Brian O’Connor pointed to the fact that when we represent things, the representation is not isomorphic with the thing being represented. We select features of the object that we choose to represent.

Representation is a ***system for extracting or highlighting some aspects of an original concept or object, together with some explanation of how the system does this. That is, we have some form of sign (in its broadest sense), which is generated from some original referent, by means of some code. .. The purpose of the representation will strongly influence which attributes are highlighted or selected as representative.***

Instead of a generic notion of a surrogate, representation implies that we select features or attributes of an entity for our representations. This to me, gets at the sense of standing for (rather than standing in for) or to symbolize.

As we choose the features of the object to represent, we become accountable for that selection of features. Further, along with the features we choose to represent, we also create or provide data values for the elements or fields in our representations. In doing so, we are really making an assertion about the object. Clifford Lynch and Lorcan Dempsey both have made comments that relate to this. Lorcan has described metadata records as ***Schematized statements about resources***. These statements make assertions about features of objects represented. And Clifford has suggested that a ***metadata record can be considered a collection of assertions about something***.

We will return to some of the ideas shortly.

Representation Seems to be Everywhere in LIS

In our areas of professional and scholarly work, namely library and information science, representations seem to be everywhere and of many kinds.

At the low-level of the machine, we represent the presence or absence of an electrical current as a 1 or a 0. Combinations of these 0s and 1s can be used to represent symbols such as letters and numbers as defined by character sets such as ASCII or Latin-1. We have the MARC tags, subfield codes, and other mechanisms that represent concepts such as title, statement of responsibility, etc. We also have code lists, where a code represents a language, a geographical location, and the like. We can also consider authorized names as constructs that represent a bibliographic identity. A classification code associated with an object represents what the content of the object is. Some of these representations adhere to social agreements we have made – for example, the ASCII standard, the MARC standards, etc. There is definitely the social aspect regarding the representations we create and engage with. While all of these instances of representations are of interest, I'm going to focus most of my remaining remarks on examining one category of representation, namely the metadata record for bibliographic data, of which the library catalog record is a specific type.

Dr. Miksa discussed a number of traditions of information organization, and several of those have evolved particular types of metadata records that serve as representations. The purposes of those representations may vary because of the particular practices of the communities that have evolved them and the purpose that are served by the representations. From the basic citation in a bibliography to a richly detailed bibliographic record in a library catalog, the metadata records produced by these traditions contain the representations of the salient features of an information object and representations in the form of data values associated with those salient features. The resulting metadata record is the representation available to a machine or a human for action.

An information object can be represented by one or more of these traditions, and each of the representations serves the purpose of those traditions. For example, in a bibliography, there is likely not to be an element in representation that points the user to a specific location of a copy of the information object. Yet in a library catalog record, to carry out one of the functions of the catalog, there is an element, usually some form of call number, that points the user to a physical location of the object. To adapt something Lorcan Dempsey state a number of years ago, our various representations are life forms in our ecology of information organization. Representations seem to be part and parcel of information organization practices. Therefore, what might be some of the implications of asserting that representation and the practices of representation are a critical component of a general approach to information organization?

Metadata and Representation

Since the first Dublin Core meeting in 1995, which I had the pleasure of attending, and my work on the Government Information Locator Service (GILS) a few years prior to that, it became apparent that metadata as a concept could give new energy to thinking about our practices of information organization. For example, Eliot Christian, the architect of GILS, assumed that different government agencies would likely create different views (or representations) of the same information object for their own constituents or customers. What that startling statement from a non-librarian meant to me was we would be living in a world where there would not likely be a canonical or privileged or authoritative representation that would serve all potential users of the object. Related to this were ideas coming out of the DC Warwick Framework discussions about various metadata packages being associated with a single information object, and certainly we can see this carried out in the METS standard, where we can have different types of metadata records encoded within the METS structure.

The DC element set was originally guided by a specific set of requirements, namely to come up with a simple, core set of elements to help identify document like objects: as stated by Stuart Weibel, the elements would describe networked electronic resources to serve the purpose of discovering of the

resource. The GILS record had other requirements, such as enabling a citizen to be able to contact someone in a particular agency responsible for the information. Purpose of a representation guides what features of an object need to be represented to support the purpose.

The library catalog has its own stated purpose(s). Charles Cutter in the late 19th /early 20th century defined a set of objectives or functions of the catalog, and the catalog record and the data it contains support those functions:

1. To enable a person to find any work, whether [issued](#) in [print](#) or in [nonprint format](#), when one of the following is known:
 - a. The author
 - b. The title
 - c. The subject
2. To show what the library has
 - d. By a given author
 - e. On a given and related subjects
 - f. In a given kind of [literature](#)
3. To assist in the choice of a work

Fast forward to the late 20th century, and we have the Functional Requirements for Bibliographic Records, and the four user tasks that the bibliographic data in a record should support, namely: Find, Identify, Select, and Access/Obtain.

The challenge for each of these efforts was to determine the salient features of the object to represent – in the case of Simple Dublin Core, the 15 elements can be associated with features such as name of object, people involved with creation, subject terms, etc. In the case of the MARC format for bibliographic data, there are approximately 2,000 structures that can be used to hold data related to the object.

So, an initial question for any representation scheme is what purpose is to be served by the resulting representation. Then the challenge is to identify the salient features to support that purpose. Cutter's objectives, on the surface, points to a relatively limited of features to represent, at least for the Find objective: title, author, and subject. But what are the features that will assist in the choice of a work? FRBR's four user tasks also can be a guide to the features to select for the representation.

Yet, the devil is in the details, or more lyrically, the devil is incarnate in all the potential users of the object! So, while we look at the salient features to represent, we must keep in mind not only the purpose or purposes, but as importantly, the users or intended audience of a particular scheme of representation. Woe to the person creating the representation scheme or an individual representation if salient features of interest to a set of users are left behind and not included.

One other point before I leave this thread is the importance of structuring the representation in a way that serves different classes of "users", and here I am thinking of differences between machine users (or rather software programs) versus human users. Programs may want to act on a representation and the requirements for such actions will need to be understood. Are the elements in the records structured appropriately to facilitate machine processing and re-purposing? MARC provides very granular data structures through its tagging and subfields. Yet the data values in specific elements may be unstructured – free text note fields for example.

So, purposes, salient features of objects, salient features for specific audiences, anticipated or desired machine-processing, structured versus unstructured data values – these and no doubt more are considerations for appropriate representations.

Assertions and Trust

By way of closing this presentation, let me return to something I mentioned earlier. Clifford Lynch and Lorcan Dempsey have suggested (or my reading of what they have written or said) that metadata records, or representations, can be viewed as a set of assertions about an object. If we assume, and I do, that there is not necessarily a single perfect or canonical or authoritative representation, then we can say that some representations are better than others – at least in terms of their utility to users (machine and human). Further, a representation, from one of the definitions discussed earlier, specifies a character or quality of an object or declares it to be of a certain kind.

Library catalogers and indexers have been challenged in how they represent what an information object is about. Assigning subject headings or index terms is really a process of making an assertion that the information object is about x, y, or z – or all three. Whether warranted or not, the library catalog and likely indexing and abstracting systems, are privileged in a way from the perspectives of users since they “trust” the representations they find in our systems. Professionals create these representations. They are intended to be objective statements rather than biased and personally-motivated statements. Yet, sometimes those “objective statements” don’t help the users – who often are looking for a “few good” resources. Patrick Wilson discussed certain types of information – I think he referred to this as exploitative information – in representations that could better assist users in selecting items. Typically, catalogers don’t include data related to the quality or lack thereof of an information object.

We are aware that search engines have not necessarily exploited metadata on web pages precisely because of the lack of trustworthiness of the data found in those metadata elements. So, we have to confront the issue that representations may not be truthful, or at least not complete or balanced. Imagine Rush Limbaugh as a library cataloger or rather charged with creating a bibliographic record with the item in hand, *My Life* by Bill Clinton. Would you trust the subject representation or description as accurate and balanced?

While that may be exaggerating for dramatic effect, think of a real library cataloger who asserts that X is the title of an object. Certainly the rules guide the cataloger to transcribe the title exactly as found on the chief source of information, but think of the challenges of determining which title to use when dealing with a web page. One cataloger asserts that Y is the title, but another cataloger asserts that Z is the title. The differences do not stem from malicious intent, but the conflicting titles highlight that representations can differ in representing object properties as “simple” as title information.

Summary

Our endeavors to organize information for efficient and effective access are noble. Representation and representations are everywhere in our practices and systems. Our users rely on representations as they engage in their information seeking. To me, the concept of representation certainly belongs in any general approach to information organization. And I firmly agree with Brian O’Connor, *we need to make an exhaustive study of concepts of representation*, and have that study inform our theory, practices, and systems for information organization.

Thank you.