The VRA Core Survey Analysis
Report dated September 1, 2011

VRA Core Survey Committee:
Robert Carter (Independent Metadata and Systems Consultant)
Jodie Double (Digital Content and Repositories Manager, University of Leeds)
Trish Rose-Sandler (Data Analyst for Biodiversity Heritage Library, Missouri Botanical Garden)
Margaret N. Webster (Independent VR Consultant, Cornell University [retired]).

Introduction: The first VRA Core survey, which was conducted from February 1, 2010 to March 5, 2010, was administered by the VRA Core Survey Committee as part of the continuing work of the Visual Resource Association's (VRA) Data Standards Committee (DSC). The survey was prominently promoted on the VRA website, and a link was sent to a wide variety of listservs including VRA-L, ARLIS-L, MCN, DIGLIB, IMAGELIB, SAA, and METADATA LIBRARIANS. The purpose of the survey was to learn both about the Core's current users as well as those who considered using it but had determined that it was not appropriate for their needs. For the first group, the committee wanted to know whether or not they had adopted Core 4.0 and, if so, to understand the challenges and issues they had in implementing it. For the latter group the committee wanted to understand the barriers to the adoption specifically of the Core 4.0 version of the VRA Core.

This inaugural survey serves as an environmental scan providing a benchmark against which to measure future changes and developments to the VRA Core. It is hoped that this is the first of a series of VRA Core surveys which will document the continuing use and evolution of this data standard. One hundred forty three people began the survey; it was completed by 103 respondents for a completion percentage of 70.5. Because a number of respondents skipped each question, the statistics given for a specific question pertain only to those who responded.

The Core survey focuses on the collection rather than the administering organization in order to accommodate multiple collections at each institution each of which might use a different version of the Core or implement the Core in different ways and to varying extents. Each respondent was invited to submit a new survey for each institutional collection he or she wished to describe. This means that the survey includes multiple responses from some institutions while other institutions only submitted one survey. It is believed that future surveys will attract responses from a greater number of institutions and that institutions will frequently submit surveys for several collections. Furthermore, the analysis of a richer, more complex data set from a larger survey and the reuse of information may, in turn, determine how future surveys are designed and conducted.

When one is choosing or considering the use of a data standard one must both gauge how widespread its adoption is as well as identify the types of organizations that have adopted it. This information helps to determine both its sustainability and interoperability as well as its fit for the particular type of organization considering adopting it (i.e. whether it is a library, museum, archive, etc.). The survey revealed that 84 collections are currently using the VRA Core data standard. Users are divided into Core 4.0 (56 collections); Core 3.0 (25 collections), and Core 2.0 and 1.0 (3 collections).The survey results indicated that the primary barriers to implementation of Core 4.0 are a lack of understanding of how to implement the relational structure of its data model into local systems as well as a lack of institutional

technical support for implimentation. While the Core data model is flexible enough to accomodate either a flat or hierachical implementation, apparently this flexibility is not being communicated well enough in the Core documentation and user support materials. The Data Standards Committee will need explore how to better assist users with this challenge.

The Core Categories for Visual Resources (VRA Core) were initially developed to fill an urgent, perceived need for a documentation standard to guide the cataloging of image collections. In 1996 VRA Core 1.0 established the first guidelines for the description of visual documents depicting works of art, architecture, and artifacts or structures from material, popular, and folk culture. It was both hoped for and assumed that the wide adoption of this standard would eventually lead to shared cataloging in a national image database or repository. Version 2.0 (October, 1997) and Version 3.0 which conflated the Work and Visual Document element sets of the previous versions into a single universally applicable element set, soon followed. The beta version of Core 4.0 which included an XML schema was released in December, 2005; the current version is dated April 9, 2007.

It was assumed when this survey was drafted that the previous versions of the VRA Core (1.0, 2.0 & 3.0) were used either exclusively or primarily as a cataloging and storage tool. Indeed, historically, this reflects the initial purpose of the Core; furthermore, the earlier versions of the Core were not accompanied by an XML schema making it difficult to use the Core for record exchange. Because Core 4.0 is the first version to include an XML schema, it was important to discover whether or not a shift had occurred (or was occurring) in the use of the Core from primarily a locally implemented cataloging utility into a medium or methodology for exchanging and sharing of Core records beyond the local context. Such a shift would fulfill an important original intention of the VRA Core developers—that of creating a mechanism for shared cataloging—if not as the originally conceived of single national image repository but rather distributed into a variety of national, regional, subject-based or institutional repositories.

The survey results also indicated that the newly available XML functionality of Core 4.0 is gradually being adopted. Although the most prevalent use of the Core continues to be as a cataloging and storage tool, a significant number of respondents indicated that they were using it as an export of delivery medium either as Core 4.0 XML or in a flattened form. An equal number of respondents indicated that they used the Core 4.0 XML Schema or that they flattened Core records for exchange. Furthermore, a few respondents indicated that they used Core 4.0 primarily as an exchange medium. It will be interesting to see whether future Core surveys document a continued shift towards the use of the Core as an exchange medium instead of or in addition to the current predominant use as a cataloging tool. It will also be interesting to determine whether collections originally cataloged using Core 1.0, Core 2.0, or Core 3.0 will migrate to Core 4.0 in order to take advantage of the Core 4.0 XML schema or whether they will find the move to Core 4.0 initially overwhelming or counterproductive.

VRA Core 4.0 added the record type “collection” to the types “work” and “image” that were introduced in VRA Core 3.0. The general adoption of this new record type is currently fairly sparse. Furthermore, many collections have only selectively implemented elements and their semantic equivalents; attributes are even more selectively implemented. It will be important to determine whether these implementation patterns are the result of constraints in institutional infrastructure or funding as opposed to either a lack of understanding or the perceived need for this type of documentation. Future surveys may document changes in the way elements, their semantic equivalents, and attributes are implemented thus providing information about what is useful and what isn’t.

The survey responses make it apparent that current and potential users would welcome additional support, training, and assistance particularly when implementing an inherently hierarchical structure mapped from their current flat implementation. It is clear from the survey responses that many users or potential users are not aware that the VRA Core 4.0 incorporates sufficient flexibility to accommodate either a flattened or a hierarchical data structure. In fact, the inherent flexibility of the Core to connect with institutional deliver—access—presentation tools (OPACs/DAPs) is poorly understood suggesting the need for more clarity in the VRA Core documentation and further training for those who desire to implement the Core.

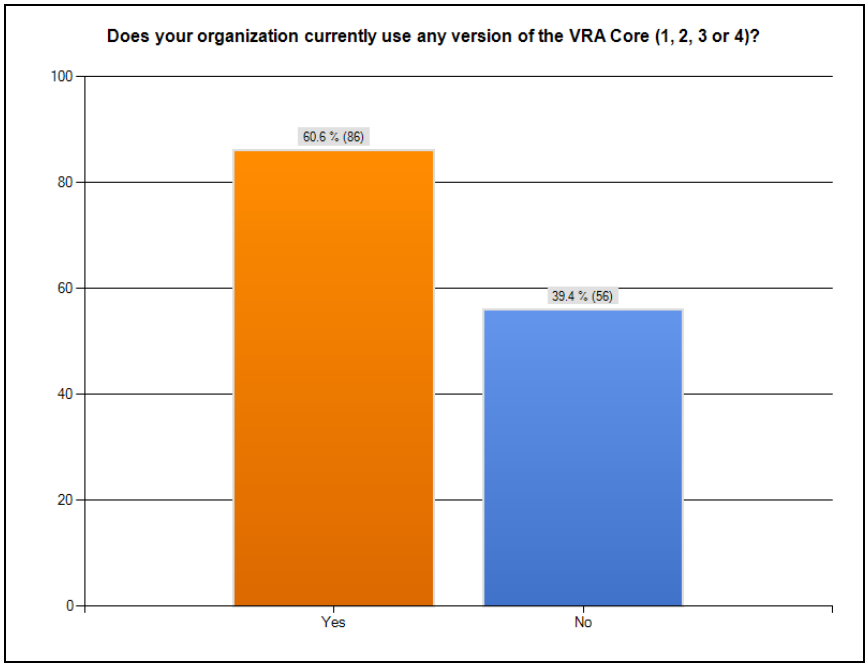
We thank the VRA Core community for taking the time to respond to this survey and share their experiences--both positive and negative. The results from the survey will be used by the VRA Core Oversight Committee to further develop and refine the Core's schemas, documentation, and user support materials. A big thank you is extended to the institutions who agreed to present their collections on the Core website in order to demonstrate their particular implementation of the VRA Core 4.0 . Many Core users and non-users have long requested this type of visible "Core in Action" and we hope these real-life scenarios will help those who are considering or currently implementing the Core.

We believe the future development of this standard should be driven by the needs of its current and potential users. Its longevity will depend upon continued active involvement of the community, and we encourage interested parties to get involved by either joining the VRACore listserv (http://listserv.loc.gov/listarch/vracore.html), adding your collection to our Implementation Registry (accesible at: http://www.vraweb.org/projects/vracore4/), or perhaps joining the Core Oversight Committee (send enquiries to vracore@vraweb.org). Together we can make the VRA Core standard a robust, nimble, and relevant standard for the future.

Question 1: Does your organization currently use any version of the VRA Core (1, 2, 3 or 4)?

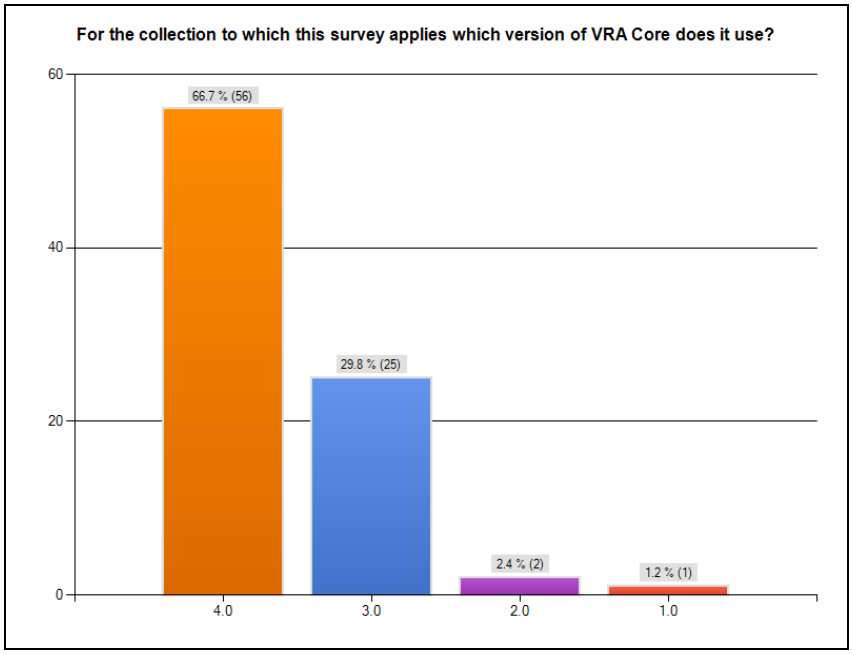
The first objectives of the committee were to determine the overall use of all versions of the VRA Core Categories and then to establish the adoption specifically of VRA Core 4.0. So, of 146 responses, 86 (60.6%) respondents used a version or some aspect of the VRA Core while 56 (39.4%) did not.

The survey focuses on the collection rather than the organization in order to accommodate multiple collections at each institution. Each respondent was invited to submit a new survey for each institutional collection he or she wished to describe.



Question 2: For the collection to which this survey applies which version of VRA Core does it use?

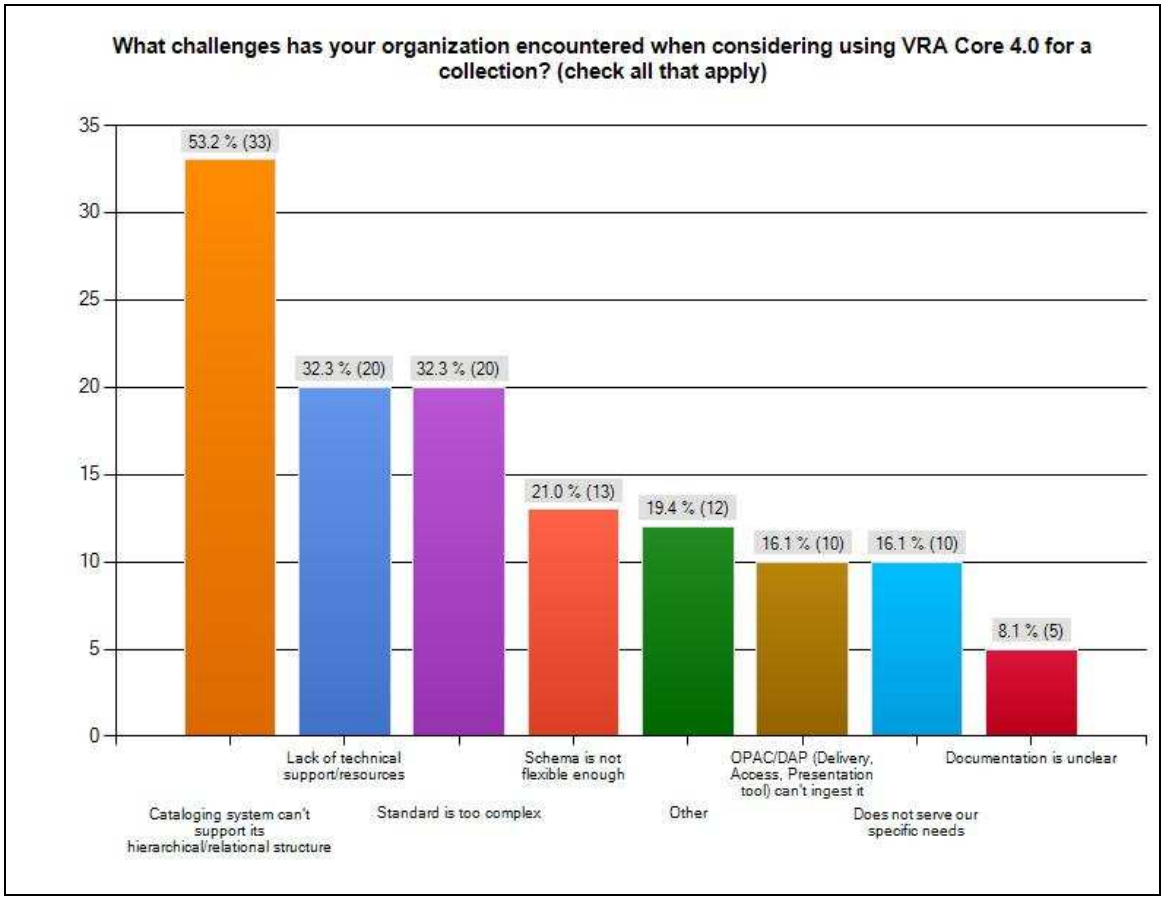
The respondents for those collections that used one or more versions of the VRA Core were then asked to indicate which version of the Core they used for the particular collection under consideration. Of the 84 responses to this question, the majority (56 or 66.7%) used Core 4.0 with another 25 (29.8%) using Core 3.0. There were a few collections (3 or 3.6%) included in this survey that reported using VRA Core 1.0 and 2.0; these are most likely legacy collections that had been digitized and cataloged when these earlier versions of the Core were first available.



Question 3: What challenges has your organization encountered when considering using VRA Core 4.0 for a collection?

The VRA Core 4.0 presents certain challenges to those who implement or would like to implement it. Respondents were invited to list multiple challenges or barriers to their implementation of Core 4.0. The most common issue reported was how to implement a hierarchical structure, such as Core 4, into the flat environment that is common among institutional cataloging systems. Respondents (33 or 53.2%) frequently asserted that their cataloging system was unable to support the hierarchical or relational structure of Core 4.0. Core 4.0 seems to be dauntingly complicated for some users (20 or 32.3%) particularly those who lack sufficient institutional technical support (20 or 32.3%).

Other important issues include a perceived lack of flexibility in the Core, an inability to connect with the institutional delivery—access—presentation tool (OPAC/DAP) to ingest Core 4.0 records, a desire for additional clarity in the documentation, and a determination that the VRA Core does not meet the specific needs of the user.



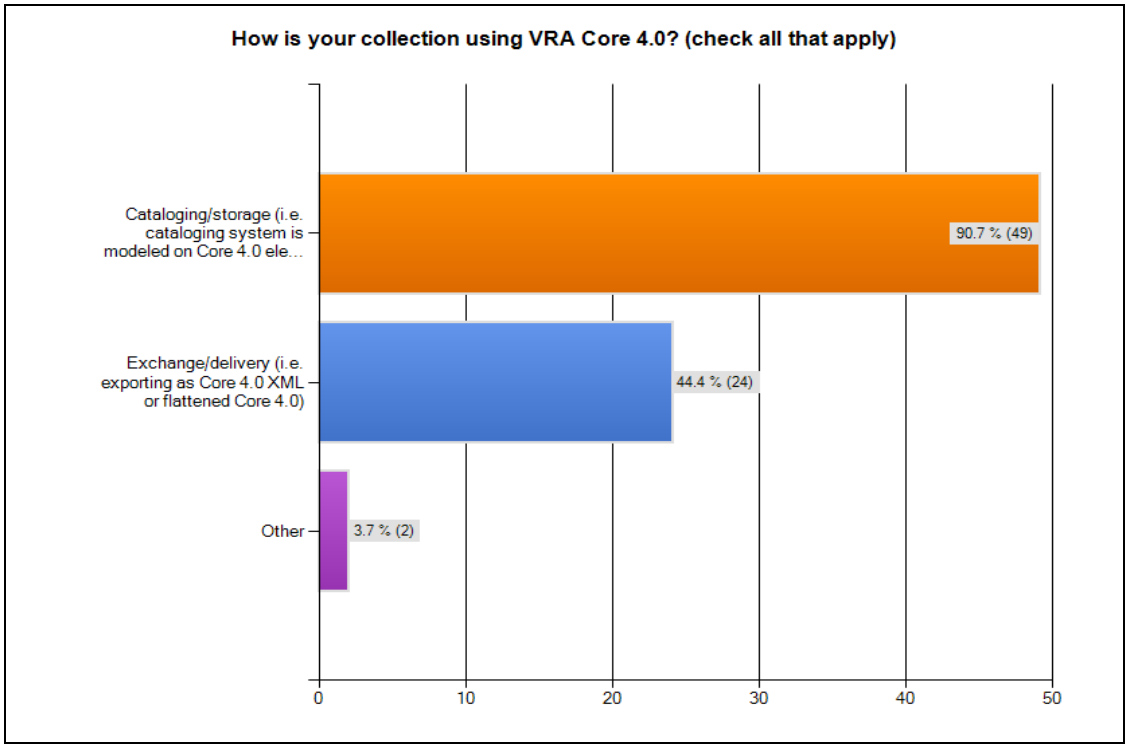
Question 4: How is your collection using VRA Core 4.0?

This question specifically addresses how the VRA Core 4.0 is being used. It was believed that the previous versions of the VRA Core were used either exclusively or primarily as a cataloging and/or storage tool. Historically this reflects the initial use of the Core; furthermore, the earlier versions of the Core were not accompanied by an XML schema. Because Core 4.0 is the first version to include an XML schema, this question sought to determine whether this addition fundamentally changed and expanded the use of the Core from that of an exclusively local use that informed cataloging practices to an exchange medium to allow the sharing of Core records outside of the local context .

When one examines all of the responses, most of the respondents continued to use Core 4.0 as a cataloging and storage tool, but a significant number of respondents indicated that they were using it as an export or delivery medium either as Core 4.0 XML or in flattened format. Forty nine respondents (90.7%) indicated that they used Core 4.0 as a cataloging or storage tool; twenty four (44.4%) said that they used it as an exchange mechanism. The most prevalent use of Core 4.0 continues to be for cataloging and storage. In some cases, those who use it for this purpose also make use of it as an export or delivery medium; furthermore, a few respondents indicated that they used it primarily for exchange.

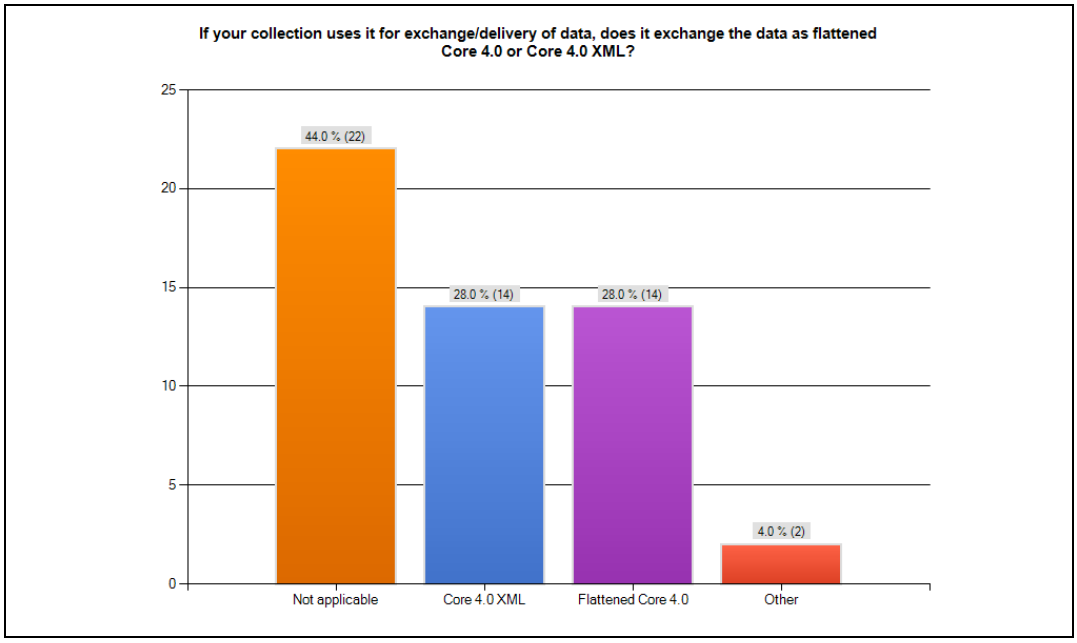
Those who indicated that they used VRA Core 4.0 in other ways mentioned the following uses:

1. Cataloging digital images with ARTstor Core, which itself is based on VRA Core 4.0.
2. Mapping collections they received (which may or may not have been cataloged using a version of VRA Core) to an in house metadata schema based on VRA Core 4 for online delivery.
3. For exhibition.
4. Indicated an intention to use it in the future for appropriate collections for display on a public interface and for exchange.



Question 5: If your collection uses it for exchange/delivery of data, does it exchange the data as flattened Core 4.0 or Core 4.0 XML?

An equal number of respondents (14 or 28%) indicated that they exchanged or delivered data using the Core 4.0 XML schema as those that flattened the data. One respondent reported using both methods although generally choosing to flatten the data; another indicated that their data was being processed through OAI harvesting. Those who did not use the Core for data exchange at all (22 or 44%) are represented in the "not applicable" column.



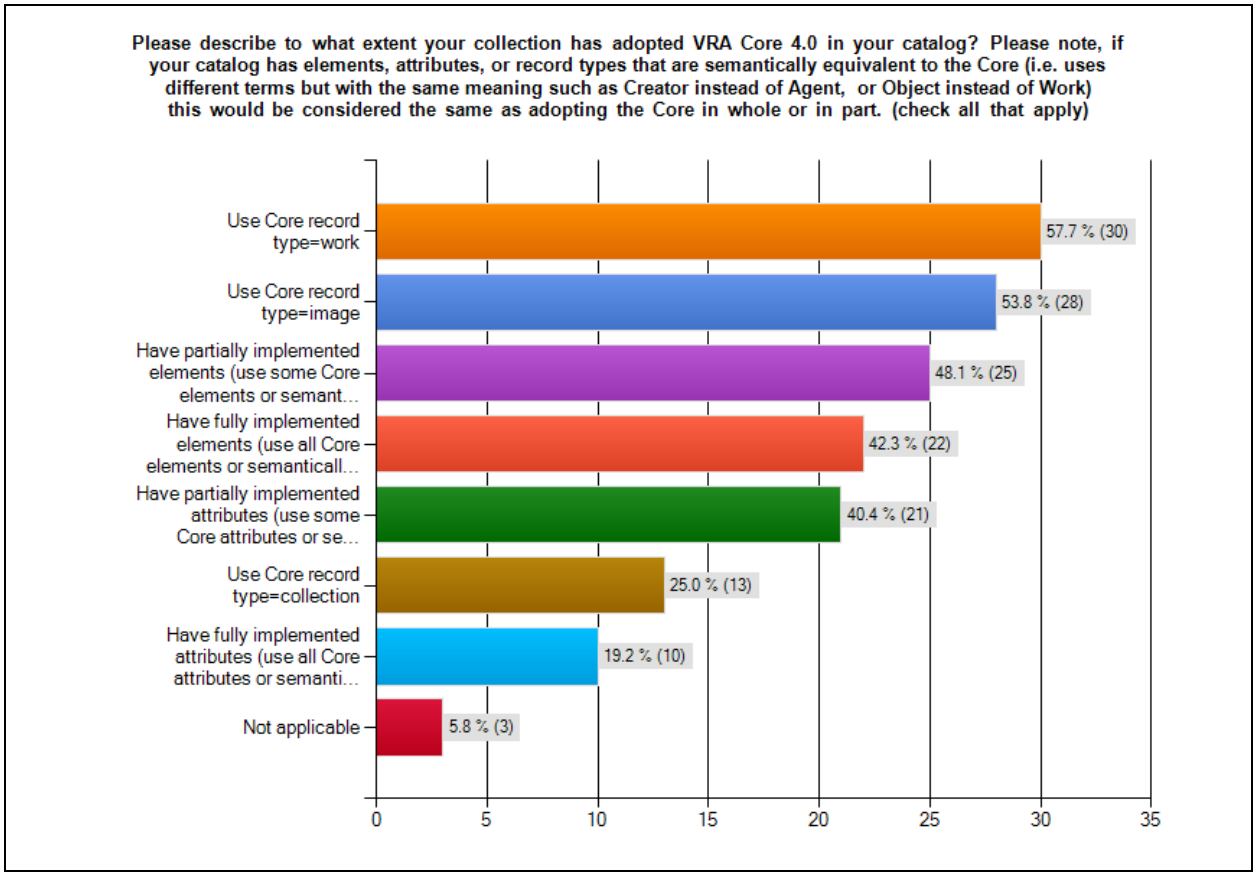
Question 6: Please describe to what extent your collection has adopted VRA Core 4.0 in your catalog

It was important to discover to what extent the components of Core 4.0 had been adopted by our respondents. For instance, how many respondents used the Core record types "Collection", "Work", and "Image." To what extent were the "elements" and "attributes" and their semantic equivalents implemented? How many respondents claimed that they had completely implemented the Core as opposed to those who had only partially implemented it?

As the following chart shows, Core record types "Work" (30 or 57.7%) and "Image" (28 or 53.8%) are most frequently implemented; the record type "Collection" (13 or 25%) is less often used. Many collections have selectively implemented elements and their semantic equivalents; attributes are even more selectively implemented.

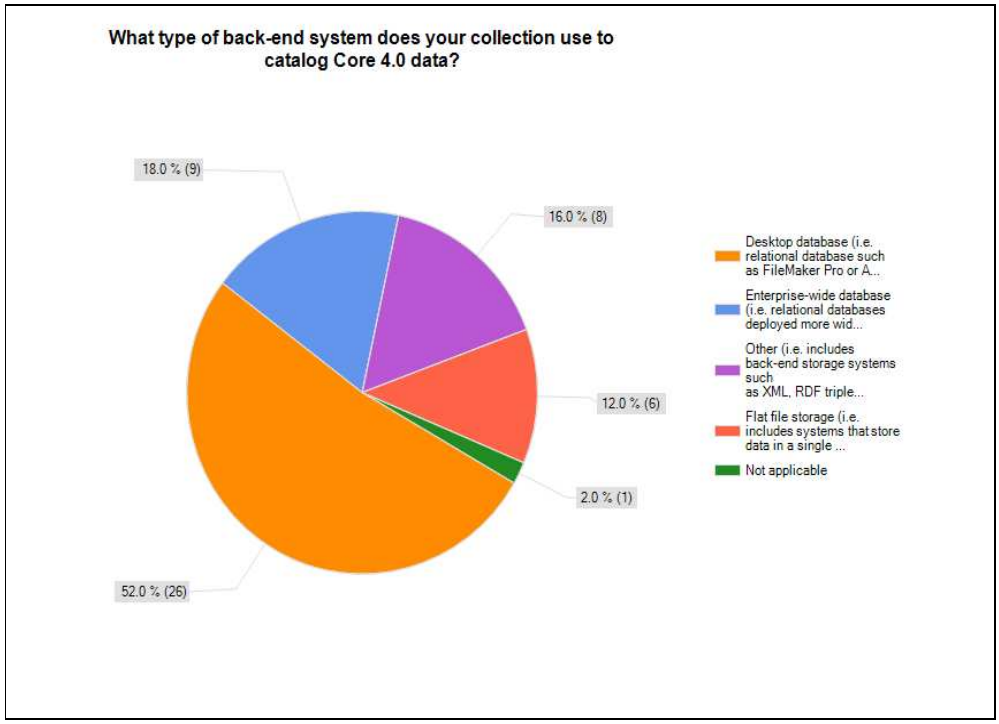
Extent of Use

- Those who have fully implemented elements (22 or 42.3%)
- Those who have partially implemented elements (25 or 48.1%)
- Those who have fully implemented attributes (10 or 19.2%)
- Those who have partially implemented attributes (21 or 40.1%)
- Use Core record type=work (30 or 57.7%)
- Use Core record type=image (28 or 53.8%)
- Use Core record type=collection (13 or 25%)



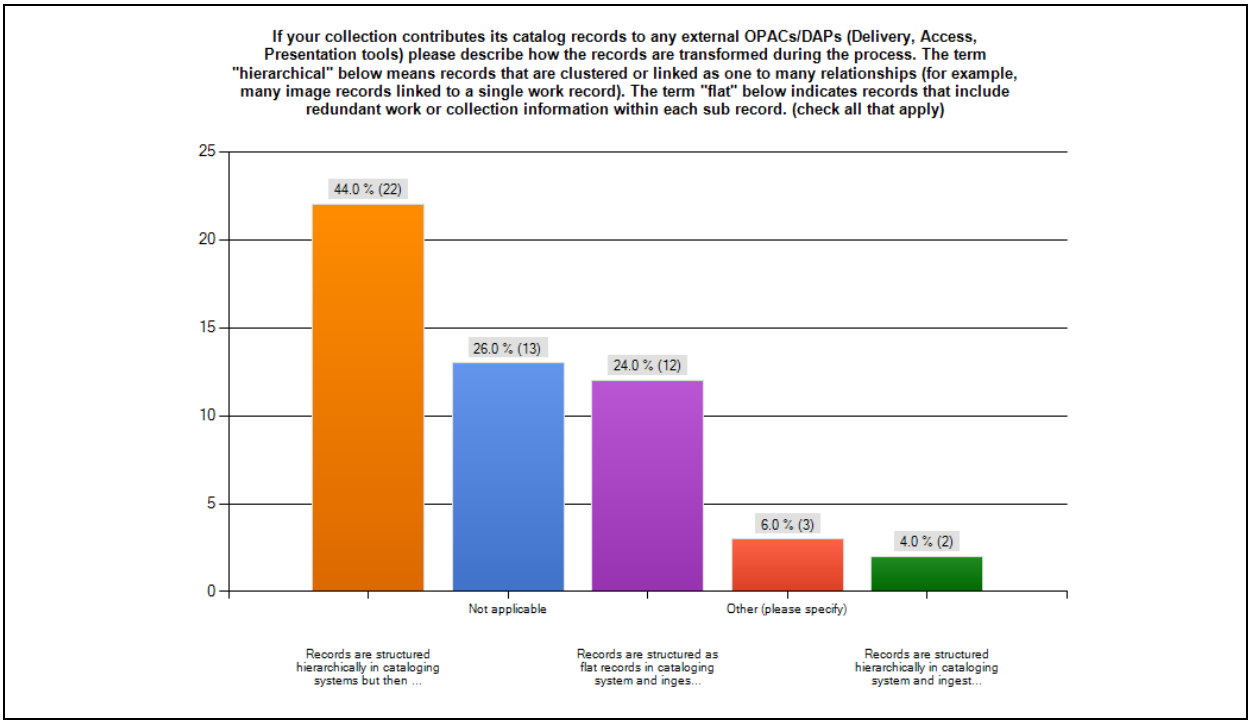
Question 7: What type of back-end system does your collection use to catalog Core 4.0 data?

It was important to discover what types of back end systems survey respondents were using with their collections. Did they use desktop, departmental, or central systems? Were these relational or were they flat? This has implications both for the ease of use and also for the critical issues of back-up and preservation of collection information. Over half of those survey respondents (26 or 52%) who answered this question reported using Core 4 on desktop database systems to create descriptive metadata; however, a number of respondents (9 or 18%) were employing enterprise-wide databases. Six respondents (12%) reported using flat file storage systems including single tables and spreadsheet; 8 respondents (16%) reported using various other systems such as XML, RDF triplestore, or object-oriented applications.



Question 8: If your collection contributes its catalog records to any external OPACs/DAPs (Delivery, Access, Presentation tools), please describe how the records are transformed during this process.

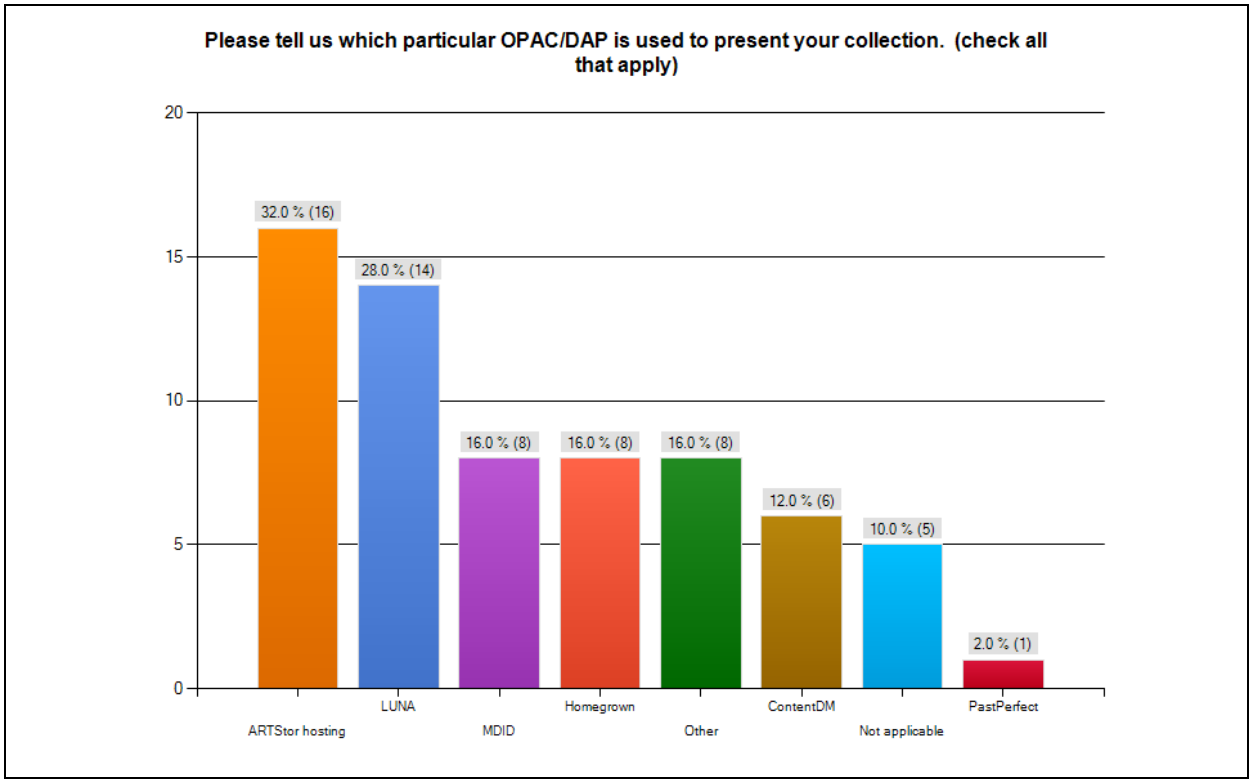
Once it was understood how the metadata was recorded and stored, it was important to know how these records are presented to the end users (i.e., which the delivery, access, and presentation tools are being used). In particular, it was essential to understand whether records that were originally structured in a hierarchical manner were flattened or whether they remained hierarchical during the ingest process. Most collections (34 or 68%) either entered the ingest process as flat records or were flattened during this procedure. Only two respondents (4%) reported keeping the hierarchical structure both as part of ingest and in presentation to end users. The need to flatten hierarchically formatted data is attributed to the limitations of the presentation software.



Question 9: Please tell us which particular OPAC/DAP is used to present your collection.

It was important at this stage to discover the identity of the external OPACs and DAPs used to present Core compliant collections. It was understood that many institutions make their collections are available in multiple ways using several different venues. For instance, a collection might be available on Luna Insight, MDID, or in a homegrown system as well as an ARTstor hosted collection (now SharedShelf). While ARTstor hosting was the most common venue (16 or 32%), Luna was almost as popular (14 or 28%), and Luna and MDID when considered together (22 or 34%) were more frequently used than ARTstor hosting. The recent transformation of the ARTstor hosting program into SharedShelf will certainly change this equation as institutions begin to use SharedShelf as both a cataloging utility and as a delivery and presentation tool.

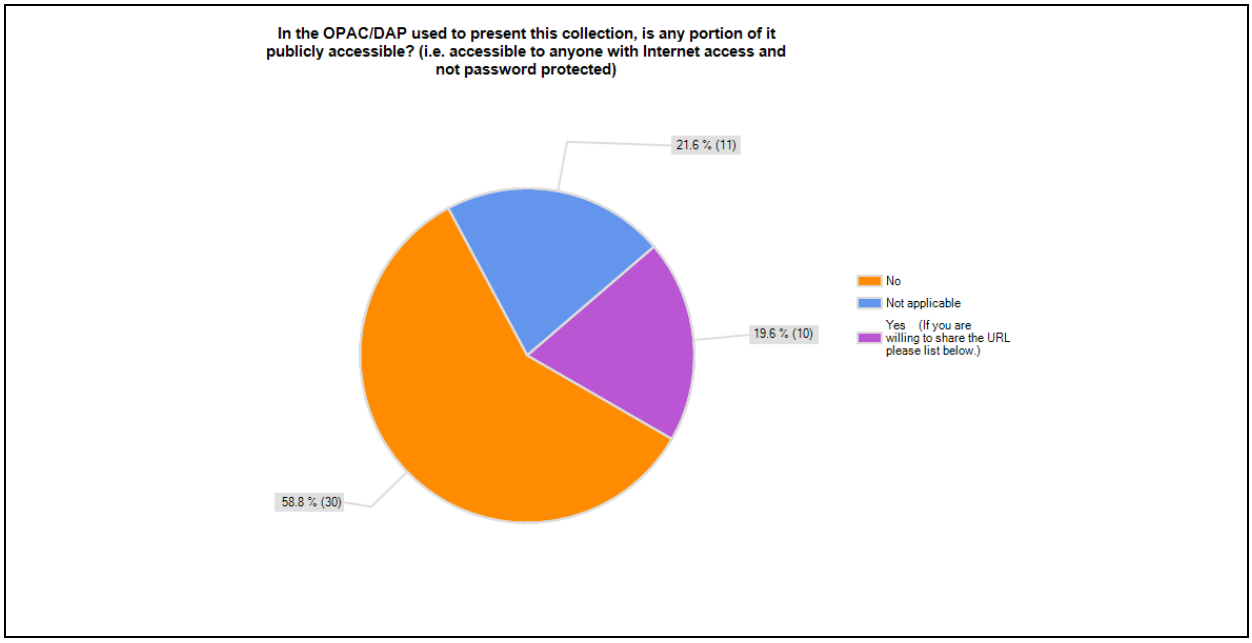
Survey respondents also mentioned using ContentDM (6 or 12%) Extensis Portfolio(2 or 3%), Fedora (2 or 3%), Voyager, WorldCat (local), DigiTool, ADLIB, and homegrown systems (8 or 16%) for this purpose.



The following questions capture information about the owner of the collections, their accessibility, and location.

Question 10: In the OPAC/DAP used to present this collection, is any portion of it publicly accessible to anyone with Internet access and not password protected?

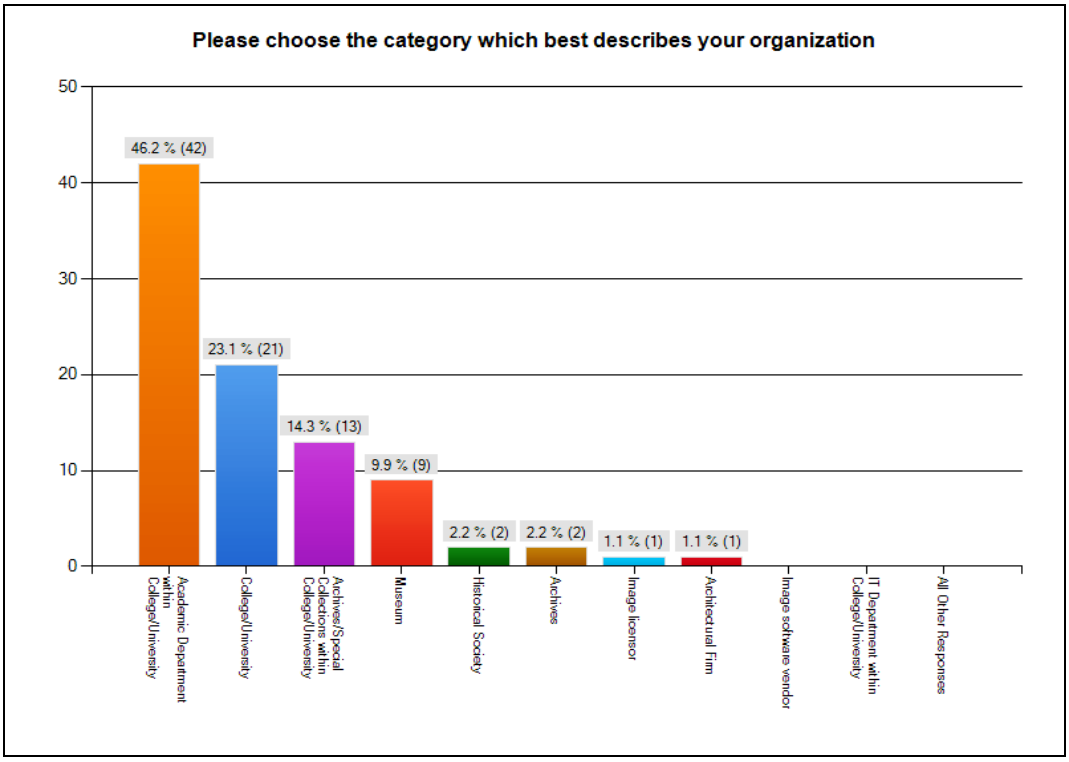
It was important to know how many collections were publicly available (10 or 19.6%), and which institutions were willing to allow information about their collections to be used to promote the use of the VRA Core. Most (30 or 58.8%) of the collections were not publicly available due to a number of factors (e.g., the rights status of the images, system could not display metadata records only, etc.).



Question 11: Please choose the category which best describes your organization.

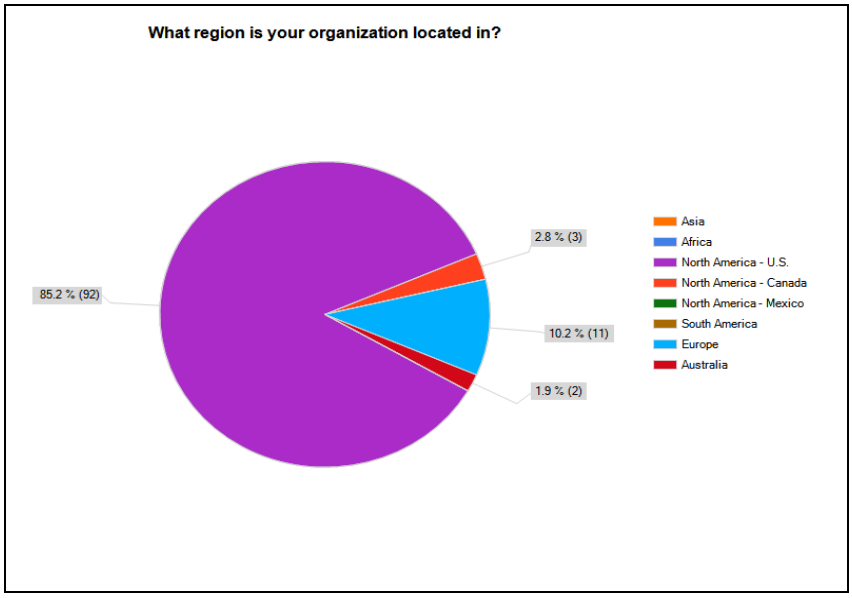
It was important to understand the institutional homes of these surveyed collections. Most were located either in a college or university often within a specific academic department or other unit within a college or university. A number of collections were located either in archives or special collections or units within a college or university or, more generally, within the library of a college or university. Others are found in a variety of jurisdictions including public libraries, state libraries, college photo labs, and archives:

- Historical Society (2 or 2.2%)
- Image licensor (1 or 1.1%)
- Architectural Firm (1 or 1.1%)
- College/University (22 or 23.3%)
- Academic Department within College/University (42 or 46.7%)
- Library within a College/University (10 or 9.5%)
- IT Department within College/University (1 or 1.1%)
- Archives (2 or 2.2 %)
- Archives/Special Collections within College/University (13 or 14.4%)
- Museum (9 or 8.9%)
- Other (21 or 23%)



Question 12: What region is your collection located in?

The over whelming majority of collections for which a survey was completed are located in North America (95 or 88.8%) specifically in the United States although responses were received from Europe (11 or 10.2%) and Australia (2 or 1.9%). It will be interesting to observe whether future surveys of Core users will reflect greater geographic diversity.


What region is your organization located in?
2.8 % (3)
85.2 % (92)
10.2 % (11)
1.9 % (2)
Asia
Africa
North America - U.S.
North America - Canada
North America - Mexico
South America
Europe
Australia