

Collaboration and Sharing for Art Prints and Beyond at Yale University

Print Metadata Best Practices

Jenn Riley

August 2008

Print Metadata Group, Yale University

New Haven, Connecticut

© 2008 Yale University



1. Introduction

Libraries, archives, and museums are facing a time of rapid change as the larger information environment is evolving at an astonishing rate. Collaboration and open sharing of data are among the many strategies cultural heritage institutions are using to adapt to new information models. The “Coordinating Descriptive Metadata for Print Collections in Libraries and Museums at Yale University” Project, a joint initiative of the Yale Center for British Art, The Yale University Art Gallery, The Lewis Walpole Library, and the Beinecke Rare Book and Manuscript Library, was envisioned in part to make use of these two strategies. The project has made progress in this area, and has the potential to serve as a model for future collaborative initiatives within the University.

Innovations in technology are driving factors behind changing user expectations, and libraries, archives, and museums must respond to these developments. Collaboration is an essential strategy for cultural heritage institutions during a time in which we must be expanding our services to users. By working together, we can mitigate in part the financial impact of adding new services on top of existing ones. Collaboration among repositories at a University can take many forms: they can work together to implement shared infrastructure (technological and otherwise); harmonize practices across repository-specific systems; share physical spaces; share knowledge and repository-specific practices; or maintain ongoing professional relationships, among other

possibilities. In the current environment, a mix of these collaborative strategies must be employed.

Differing missions among libraries, archives, and museums have long been cited as a barrier to higher-level collaborations between them. While a repository's mission can and should play a role in its approach to resource description, these institutions are increasingly realizing their similarities are greater than their differences. High-profile meetings and initiatives such as the 2006 American Library Association Rare Books and Manuscripts Section Annual Preconference with the theme "[Libraries, Archives, and Museums in the Twenty-First Century: Intersecting Missions, Converging Futures?](#)" and the [RLG Programs/OCLC Library, Archive and Museum Collaboration Program](#) have addressed these issues head-on, and the result is an increasing spirit of collaboration. These and similar initiatives can serve as a model for units within a single University working together to develop a shared vision for collaborative services.

Movement towards more collaboration between libraries, archives, and museums is appearing at the same time that libraries are revitalizing their interest in the description and availability of special collections. Most notably, the Library of Congress in 2007 commissioned a working group to study the future of bibliographic control for libraries in general, and the Library of Congress specifically. The working group's final report, delivered on January 9, 2008, contains recommendations in five general areas, including: "Transfer effort into higher-value activity. In particular, expand the possibilities for knowledge creation by exposing to more users rare and unique materials held by libraries that are currently hidden from view and, consequently, underused" (Library of Congress Working Group on the Future of Bibliographic Control, 2008, p. 2). The Library of Congress' response to the working group's report indicated strong support for special collections as valuable resources requiring more targeted focus (Marcum, 2008). It is clear that the current environment provides an opportunity for libraries, archives, and museums to collaborate to make special collections and their metadata more widely available.

Open sharing of metadata (and ideally content as well) is a second strategy Yale and other institutions are looking to in support of new service models. Metadata sharing activities are predicated on the realization that metadata records are not (only) a destination in and of themselves, but represent raw data on which higher-level services can be built. Expectations for these services are rapidly changing, however. Just five years ago it was considered a primary goal to build campus-level discovery services, providing students and faculty with unified access to materials from a variety of repositories across the institution, and to share at least some of this data with larger multi-institutional metadata aggregations. Often these larger metadata aggregations were discipline- or format-specific. Yet more recently cultural heritage institutions are coming to the uncomfortable realization that these approaches may not be enough. A recent report describing the work of the Cultural Materials service, a metadata and content aggregation of primary sources built and managed by the Research Libraries Group before its merger with the library organization OCLC, discusses the challenges smaller (relative to the scale of the web) metadata and content aggregations face in obtaining a critical mass of data that will draw users. "Cultural Materials was not alone in this—other initiatives have faced similar challenges in finding audiences for aggregated digital collections. Five years into this effort, we still have a lot to learn about end-user expectations and emerging uses in this rapidly transforming environment. Instinctively, we know audiences exist—finding them and engaging them is another matter entirely" (Erway, 2008, p.8).

The conclusion many libraries, archives, and museums are starting to draw is that sharing must happen even more widely than they had previously imagined.

Institution-level aggregations are still necessary to meet our users' needs, as are discipline- and format-specific aggregations. Yet to maximize the value of the investments cultural heritage institutions are making in their collections, they must make them, or at least descriptive information about them, available to the web at large. This “network-level” sharing is not simply allowing human access to a repository- or institution-specific search service, however. It instead requires providing machine-to-machine access to *structured* data. Some ways in which cultural heritage institutions might accomplish this goal include use of [XML Sitemaps](#) to make “deep web” data more accessible to web search engines, and exposing metadata for collections as part of the Semantic Web's [Linked Data movement](#). For Yale University, this means that making collections accessible in a variety of ways—through repository catalogs and search services, campus-level discipline- or format-specific aggregations, campus-level discipline- or format-agnostic aggregations, multi-institution discipline- or format-specific aggregations, and in machine-readable ways for unrestricted use by others— should be a long-term goal.

Open sharing of data allows for the building of more advanced discovery and use services, but it is also necessary economically. This may seem paradoxical, as sharing of data requires an investment in staff, expertise, and technology. Yet funding models for libraries, archives, and museums are shifting, and these institutions are increasingly being asked to demonstrate their value. Promoting the types of new services possible in an environment where data is openly shared can provide an institution with much-needed visibility and increase use of collections. Concrete demonstrations of innovative discovery and use services can serve as key elements in arguments for the ongoing and increasing funding of cultural heritage institutions.

A commitment to wide and open sharing of data represents a cultural shift on the part of many cultural heritage institutions. It requires acceptance of a certain loss of control over that data and how it will be used. It requires recognition that the context for the collections provided by the repository may be lost or changed by others. This is not a shift that has or can happen overnight; it is an ongoing conversation our institutions must have internally and collaboratively to better understand the tradeoffs we make when sharing data and to plan for them effectively.

Wide, open sharing of data is a long-term goal for most libraries, archives, and museums. It is an ideal that we must envision and work towards incrementally. For Yale University, as for most institutions, achieving this goal is best done in stages. Starting with campus-level aggregations, both format-specific and format-neutral, can set the stage for later steps in the process. The “Coordinating Descriptive Metadata for Print Collections in Libraries and Museums at Yale University” Project and the Collections Collaborative of which this project is a part can be leveraged to take concrete steps towards this long-term goal by building infrastructure and relationships.

2. Organizational and repository-specific issues

2.1. *Building on existing work*

Achieving the goals of collaboration and sharing requires a core commitment from an organization. The four Yale repositories involved in the prints project have already made a number of steps towards organizational commitment to these goals, led by concrete efforts on the part of the Yale University Art Gallery and the Lewis Walpole Library. The prints project itself exhibits an explicit recognition that metadata for print collections among the participating repositories should be aggregated together for the benefit of

users, and that working together is an essential step towards providing integrated access. The Lewis Walpole Library's work to digitize parts of its print collection and deliver these images together with brief records in the Yale University Digital Collections site provides a concrete demonstration of the value of aggregating material from multiple repositories and a willingness to expend resources for the sharing of data. The Yale University Art Gallery's participation in a [collaborative effort](#) to "create tools supporting data sharing in the art museum community, and analyze the extent to which current museum data allows for meaningful search and retrieval in an aggregate environment" represents an acknowledgment of the role of standards in metadata aggregation and openness to the exploration of new discovery models. The Yale Center for British Art's recent focus on collections cataloging and public access to the repository's descriptive data shows a realization of the benefits of accessible data. The Beinecke Library's leadership on the Yale campus of the development of digital collections and its opening up of some of these services to other repositories on campus demonstrates a commitment to collaboration.

The planning for harmonization of metadata creation practices already done by the prints project repositories is also deserving of acknowledgment. Each has shown a willingness to examine existing descriptive practices in light of new uses for this data and to compromise in cases where it is most desirable. Each has thought about local practice in light of standards in the repository's community of practice (MARC/AACR2 for the libraries, CDWA Lite/CCO for the museums). Each has realized that doing something is better than doing nothing, that perfection cannot be attained (and certainly not quickly), and that an incremental approach to sharing of data is warranted.

The work done to date by the prints project repositories can be used as leverage to invest further in collaborative and data sharing activities. Libraries, archives, and museums are still learning how best to achieve these goals, and must build on early successes. These repositories must not be afraid to take risks in this area; each new initiative represents a learning experience. Communication between repositories, especially on a campus such as Yale, can improve understanding of what works and what doesn't.

2.2. Short-term steps

2.2.1. Organizational commitment

While the prints project itself represents a commitment on the part of each of the participating repositories to collaboration, it is not enough for a handful of individuals from a repository to be involved in this process. The organization as a whole must support this same commitment. Leaders in the prints project should serve as ambassadors for the goals of the project in their individual repositories, working over time to foster a culture of sharing and collaboration. This culture should pervade throughout the organization, and all staff members from the high-level decision makers to the part-time data entry clerk must recognize what the organization hopes to achieve by sharing data. Day-to-day activities in support of creating metadata intended for sharing are more likely to be effective if all involved understand how this metadata is to be used. Recognize the accomplishments that have been made, but do not be satisfied with them. Use them to push forward with the next step.

2.2.2. Efficient metadata creation workflows

A key next step for each participating repository is to implement more efficient metadata

creation workflows. Each repository must examine its own workflow, defining who will create what data and what resources they will use. Effective metadata creation workflows put records in front of individuals exactly at the point they are needed. Use expert staff to their greatest benefit by providing them with records that have basic data already complete, only requiring them to contribute data that only they can provide. Do not hesitate to design a workflow that involves a relatively large number of individuals working on a single record. Document and make it clear to all involved who is responsible for which data elements to avoid unnecessary duplication of effort and misunderstandings. Find a place in the metadata creation process for all who want or need to be involved.

An efficient metadata creation workflow also makes good use of pre-existing metadata. Informal inventories in spreadsheets and researcher-created guides are two of many possible sources. Do not assume that all metadata that is created anew must be done in the repository's primary metadata management system. Metadata in spreadsheets, finding aids, or other sources can often (when planned for appropriately) be automatically loaded into the repository's main system. In short, choose the best tool for the job.

2.2.3. Implementing shared practices

A significant step in finalizing revised local descriptive practices is putting into production the agreements for harmonization of specific metadata elements made in principle as part of the prints project. Implementing the shared list of terms for print medium, or the common definitions of terms such as copy, state, edition, and proof, will require buy-in from a number of individuals within each repository. Leaders in the print project must provide to their repository a strong rationale behind the decision to implement the collaborative agreement, and provide a practical means for addressing challenges that will inevitably appear as a result of this decision.

2.2.4. Improve shareability of local metadata

Repositories participating in the prints project can take additional small steps towards improving the shareability of their metadata, beyond the agreements made as a result of the prints metadata project. A small number of recommendations are presented here for each of the four repositories represented in the prints project, with a recognition that some repositories are more likely to be in a position to adopt these than others. Note that the recommendations below are for when records are shared, not necessarily for how they are to be structured in their native environment.

Yale Center for British Art

- Provide date in shared records in a machine-readable form.
- Provide a place in the record for person and place depicted, to be filled in for particularly important items receiving the most complete level of description.
- Provide as many variant titles as are known and are economically feasible.
- Provide medium from controlled list in all cases; supplement with textual description when appropriate. The controlled medium term could come from the object name field if this field were in more general use.

Lewis Walpole Library

- Follow through on plan to enhance records in Orbis and carry over enhancements to online digital image delivery system.
- Include controlled term for medium in shared records. Might be able to generate this from 655 \$a in MARC?
- Provide date in shared records in a machine-readable form.
- Provide geographic subject headings in shared records in a field designed specifically for this purpose rather than (or in addition to) LCSH with geographical subdivisions.

Beinecke Rare Book and Manuscript Library

- Generate a print-level record for sharing whenever possible.
- In shared records, strip brackets from supplied title for a print from a larger book.
- Provide date in shared records in a machine-readable form.
- Provide geographic subject headings in shared records in a field designed specifically for this purpose rather than together with topical headings.

Yale Art Gallery

- Supply a machine-readable date whenever possible for shared records, even when local practice prefers not to do this.
- Provide a specific role for the artist in shared records whenever known.
- Provide a topical heading describing the subject matter of the print whenever feasible.

2.2.5. Revising local documentation

Once the metadata creation workflow is redesigned and specific decisions about how to populate key data elements have been made, detailed local documentation should be revised to reflect changes made as a result of the prints metadata collaboration. This documentation should be procedural in nature, intended as a user guide for the metadata creation process. If the workflow designed for a repository involves multiple people working on a metadata record, a section of the procedural manual should address each stage in the metadata creation process individually. This procedural manual will include data that is only useful or able to be distributed within the repository, such as passwords, menu options for a specific piece of software, and relevant contact information. This local documentation should indicate whether specific data elements are required or repeatable, provide details of how to fill in data for all fields in the repository's metadata record, indicate when this data conforms to a shared practice and when it conforms to a local one, provide information to the metadata creator on how individual fields will be used, and include examples demonstrating how the principles outlined should be put into practice.

Most, and perhaps all, of the repositories participating in the prints metadata project have existing documentation that largely achieves these goals. The Yale University Art Gallery's "TMS: object module fields data description" and specific documents on date and title, for example, includes detailed information about screen design and specific procedures such as the role of punctuation in various metadata elements, as does the Beinecke's online "Metadata guidelines and practices." It is important to note that these guidelines are applicable to all materials, rather than only prints. This is a reasonable approach as separate cataloging guidelines for the many formats each repository holds is likely to be unsustainable. An appropriate balance between generalization for all

formats and specialization to optimize metadata for a specific format must be found. Each repository should produce a new stabilized version of its local procedural documentation that takes into account the outcomes of the prints metadata project and its efforts to improve the overall shareability of its metadata, applying the lessons learned from the prints metadata project to other classes of material.

2.2.6. Explicitly define primary user base

A final and major short-term step for each repository participating in the prints project is to explicitly define its target user communities. While the core idea behind the sharing of metadata is to make it available to users and uses beyond those for which it was originally designed, the decisions we make about the creation of metadata records nevertheless inherently preference some types of users above others. The best strategy therefore is to explicitly define primary and secondary user groups for which metadata records are intended, and to document these decisions to inform later use of these records. Be creative in defining these user groups; while specialist researchers are a natural and easy group to select, they are not the only group for whom art prints are valuable resources. Special collections and primary source materials are increasingly being used in undergraduate and K-12 education, and users outside the academy are a fast-growing user group for cultural heritage institutions. In the short term, it may not be necessary to make any adjustments to planned local descriptive practices based on the user groups defined, but these groups will be used in evaluation activities (described below), and at that point potentially serve as guiding principles behind future revisions to the descriptive process.

2.3. Longer-term steps

2.3.1. Share metadata

Once revised descriptive practices are implemented with the goal of making local metadata more shareable, the next major step is to actually share that metadata. At most institutions, the metadata record in a repository's primary descriptive system is considered the "master" metadata record from which shared versions are generated, and this setup seems to meet the needs of each of the repositories participating in the prints project. As infrastructure for delivering aggregated metadata for art prints on the Yale campus develops (as described below), each repository must develop a methodology for delivering records to the shared infrastructure. This is discussed in more depth later in this report, in the Technical Issues section, but the relevant organizational issue here is a repository-level commitment to performing the extra work needed to determine the formats in which the shared data should be provided and to actually produce the data in those formats.

2.3.2. Continually refine metadata creation practices

As each repository gains practice in implementing shareable metadata practices, it can continue to refine them over time. When metadata from a repository is actually aggregated with metadata from others, the assumptions made by the prints project group about which practices would improve the shareability of their metadata can be tested to determine if they were effective and if others are needed. The success of any changes to descriptive practice already made can be demonstrated to others at the repository, and potentially serve as building blocks upon which to make further changes.

2.3.3. Expand collaboration and sharing into other formats

In a similar vein, as a repository begins to see concrete effects from a commitment to the shareability of its metadata for prints, it can expand the principles of shareable metadata to other formats as well. There is no need to wait until targeted collaborative efforts emerge on the Yale campus for other formats—once sharing of metadata becomes a core value of a repository, implementing shareable metadata practices will allow that repository to take advantage of other sharing opportunities that emerge, plan for future but as yet unrealized sharing opportunities, and even potentially take a leadership role in aggregating metadata from multiple repositories. The more experience with sharing metadata and optimizing metadata for sharing a repository gains, the better it will be able to plan for sharing into the future. A repository cannot feasibly optimize its metadata for every existing metadata aggregation, much less for future and unimagined uses. But it can over time incrementally implement principles designed to make its metadata more shareable and therefore optimized for use in unanticipated environments.

3. Communication and collaboration issues

3.1. *Building on existing work*

The very existence of the prints metadata project at Yale is a testament to the recognition by the participating repositories that collaboration is a key factor in achieving the institution's overall goals. The participating repositories have maintained open lines of communication and used these to agree on an initial set of shared practices. These agreements include a strategy for navigating the challenges of identifying and describing the medium or a print and development of common definitions of terms such as copy and impression. The prints project has also produced a detailed mapping of a number of data elements in the participating repositories' systems and selected other metadata standards. The print metadata project has additionally fostered communication between repositories at Yale through events with outside experts such as the January 2008 Print Metadata Seminar at the Lewis Walpole Library.

Collaborations between the repositories participating in the prints project are not limited to that project, however. The Collections Collaborative initiative has been a major force behind the collaborative conversations that have occurred. The work Yale has done in this area has been recognized outside of the University, including in a presentation at the 2005 RLG Members Forum by (then Yale employee) Rich Szary on the Collections Collaborative project. This and other activities at Yale have led to the University's participation in an RLG Programs "LAM" initiative – the [Library, Archive and Museum Collaboration Program](#), designed to help foster just the type of communication and collaboration Yale is currently working towards.

3.2. *Short-term steps*

3.2.1. Documentation

The most significant short-term steps that can be taken to advance the goals of the prints metadata project is to create detailed and stable documentation in a common location. This documentation should be the authoritative record of all collaboratively-

determined print metadata practices, serve as a high-level guide for metadata practices at each of the participating repositories, and provide a forum for further communication between repositories. Do not only share what you think another repository *needs* to know; share all documentation that does not contain information that has a valid reason for restriction. Even if specific details of local documentation are not relevant to other repositories, the sharing of this information promotes cross-repository understanding. A Wiki such as the one already set up for this project can serve this purpose well, although it is certainly not the only way to achieve this goal. The technology to support these collaborative functions is a tool, not a goal in and of itself. Several categories of documentation in this shared space should be short-term priorities.

Collaborative decisions

While verbal agreements to harmonize print metadata practices among the repositories participating in the prints project represent significant achievements, they must be followed up by action by each of the repositories. Perhaps the most effective way to transition from a plan to action is to put these agreements in writing in the collaborative workspace. Having agreements written down makes them more concrete and provides a reference for each repository as it works to implement these agreements. Putting them in writing also forces the participating repositories to agree on specific details, beyond the general principles involved. The decisions put in writing do not necessarily have to be permanent; rather, they represent stable positions in time that can be re-evaluated later.

The definitions of “copy,” “state,” “edition,” and “proof” collaboratively developed by the prints metadata group are a prime example of the sort of documentation that should be put in writing in the shared documentation in the near future. These definitions can then be tested and refined over time by trying them out on actual prints to be cataloged as part of a repository’s regular workflow. Revisions to these definitions should be made collaboratively, but without an undue amount of procedural weight to slow them down.

Another good candidate for documentation is the hierarchy of terms for printing methods developed by the Lewis Walpole Library staff, inspired by [TGM II](#). While full agreement on common use of these terms has not yet been reached, recording them in a common location can help to achieve this agreement. After documenting this list in the collaborative workspace, individual repositories could attempt to use it, and come back to the group with specific feedback on how it must evolve in order for it to be usable by all repositories. This list is likely to be best used as *broad* genre terms (with definitions), which individual repositories can supplement with terms from specialized vocabularies such as the [Art and Architecture Thesaurus](#) if desired. A list of this type will likely need to grow over time, and the project repositories should decide on a methodology for allowing easy additions to it. Looking to standard vocabularies for inspiration for new terms when they are warranted is useful, but for a list of this sort it isn’t necessary to follow standard vocabularies exactly. A good model for this list would be a repository selects terms from it whenever they fit, and when they don’t, the repository can add a new term to the list that is then authorized for use by all others. Repositories may wish to have a subject specialist evaluate a group of prints before they are cataloged to determine if new terms are needed, rather than having catalogers do this on an item by item basis.

Core data elements and mappings

One outcome of the prints metadata project to date has been the creation of a large

matrix of mappings between metadata elements in each repository's local system and a few relevant shared metadata standards. This is an extremely useful document, allowing each repository to see at a glance how the subset of its metadata most relevant to prints relates to that for the other repositories or a standard format. Yet this document due to its very comprehensiveness can be overwhelming. It has been referred to in the project as a core data elements list, but in its present state it does not truly serve this function. It is best used instead as a road map for a long-term collaborative strategy between repositories holding art prints at Yale, and should be available in the collaborative documentation space.

A true "core" set of data elements for prints at Yale will be a subset and slightly different view of this document, which will be useful to develop in the short term. The core elements will be those that each participating repository has agreed to include in *all* records, and will document shared practice for these fields when applicable. It should reflect the minimum metadata requirements needed to provide core discovery and delivery services for prints across the Yale campus, allowing for more robust practices at individual repositories. This document could also include preferred (but not required elements) that all repositories agree are desirable, but these should be marked as such. The difference between this "core" document and the master metadata mapping document is one of scale – the former should be selective, and the latter exhaustive. A proposal for how the core elements document might be structured can be seen in Appendix A.

An original goal of the prints metadata project was to create an "application profile" to support the coordination of metadata for prints among the participating repositories. The very definition of application profile within the metadata community has undergone a transformation since their initial formal introduction into this community in 2000 (Heery and Patel, 2000). In these early days, an application profile was understood as simply selecting metadata elements from any existing metadata formats, and mixing them together as needed. Yet this fairly loose approach began to present problems when implementers looked for ways to formally validate metadata instance documents against the constraints defined in an application profile. In 2003, the Dublin Core Metadata Initiative took a step towards more formal definition of application profiles in stating "By definition, any new term coined for use in an AP [application profile] must first be declared in a form citable in the AP" (Baker, 2003). At this time, it was still somewhat unclear what was meant by a metadata element being "citable." The Dublin Core Metadata Initiative then continued its process of further formalization, and currently defines application profiles in its space through Description Set Profiles, an "information model and XML expression" for formally expressing the constraints of an application profile (Nilsson, 2008). The Description Set Profile model and language, however, depends on elements included in the application profile being defined in metadata formats that conform to the DCMI Abstract Model (Powell et al., 2007), a constraint that few metadata formats popular in the cultural heritage sector meet. An application profile is most useful in an environment that uses a single system where its maintainers exert total control over the data formats used, but is significantly less useful in an aggregated environment where there is a strong reliance on standards-based data exports from several repositories. With these factors in mind, it does not seem necessary at this time for the Yale prints metadata project to define a formal application profile for art prints. The master mapping document and definition of core elements should be sufficient for the project needs in the medium-term, and achieve the high-level goals outlined in the project proposal.

The master mapping document should grow and evolve over time, adding or removing elements as repositories' practices change. It should also develop to reflect the metadata standard(s) used in a future institution-level aggregation of art print metadata,

the data elements that are required for that use, and the required/optional status of each element at each repository based on local practices. Recommendations for some short-term updates to the master mapping document can be found in Appendix B.

Print cataloging resources

The collaborative documentation space can also be used to share helpful print-related resources between repositories. Types of documentation to share include citations to *catalogues raisonné* from which authoritative data on prints can be found and controlled vocabularies that provide appropriate scope for specific metadata elements in the description of prints. A Wiki-based documentation system provides the appropriate low-overhead environment for the ongoing creation of this documentation, allowing individual metadata creators or repository representatives to easily add to this documentation bit by bit over time. Documentation of useful sources does not only serve to assist each repository with metadata creation, it also serves as a way to build a sense of community and trust between the contributing repositories. By contributing to shared documentation on an ongoing basis, each repository cements its commitment to the collaborative process.

Workflow and goals of each repository

The conversations between individuals in the prints metadata group have led to a more in-depth understanding of the workflow and goals of each of the participating repositories. This knowledge should not remain solely with specific individuals, however. It must be documented in writing and shared so that it can spread more widely throughout the institution. Two related sets of documentation from each participating repository can contribute in a substantive way to ongoing cross-repository understanding. These and the procedural documentation described above are likely best as separate documents. While it is tempting to include all relevant information about a single initiative in a single document, this makes for unwieldy documents that lose their utility when attempting to speak to multiple audiences simultaneously.

The first is a description of the repository's overall metadata creation workflow. This document should indicate which staff positions within the repository are responsible for which parts of the metadata creation process, and what software and other tools are used in each step of the workflow. This type of documentation can help other repositories understand what parts of the metadata creation workflow a repository most values, and at which stages subject expertise is inserted into the metadata creation process. A flowchart or similar documentation could be an effective means to convey this information.

The second is a high-level view of the repository's high-level practices and goals. This is not procedural documentation, but rather an overview of the general approach the repository takes to the description of its collections. It outlines issues such as the purpose of metadata records created by the repository; the types of resources held and described by the repository; the repository's commitment to sharing records outside of its local system; metadata formats, content standards, and controlled vocabularies used; and the repository's view on how to determine the object of the cataloging record. The Overview section of the Yale Visual Resource Center's "Cataloger's Manual," examined by the prints metadata group as part of this project, represents one model repositories might follow in creating this high-level documentation. This high-level documentation does not necessarily contain a complete list of metadata fields recorded by the repository—this information is best stored in the master mapping document, described

above. Providing this documentation in a collaborative workspace will require each repository to keep it up to date over time as its practices and goals change over time. By creating and providing access to this high-level documentation, each repository makes a commitment of time and effort to actively maintain it in support of the institution's vision of cooperation.

3.2.2. Increase understanding between repositories

At least two other activities can be undertaken to build on work increasing understanding between different repositories at Yale. The first is to follow through on the stated plan for each repository to create records for three prints from each of the others. The goals of this exercise are to pinpoint further areas of both disjoint and potential harmonization, and better understand the other repositories' points of view. This exercise shouldn't necessarily be an ongoing one; it is likely most useful at this crucial time in the ongoing development of collaborative processes within the institution. The desired outcome is *not* to plan for identical records from each repository. There are very real reasons why the repositories involved have developed the metadata creation practices they have, and creating effective shared metadata does not require complete conformance to one monolithic set of practices. The cooperative cataloging exercise should instead help to illustrate which practices need to be harmonized, and which do not.

Each repository sharing the internal analysis of its primary user base will also foster understanding among the group. There will likely be more in common among the defined user groups than there will be differences, and performing this analysis can be a significant step towards a concrete vision for a shared art print discovery service. Discussing user analyses can also help each repository focus on the material and its purpose rather than administrative issues of the repository itself. It can help bring clarity to the cooperative process—a realization that each repository is working towards a common goal.

3.2.3. Agree on shared metadata formats

A final short-term step that can be taken by the prints planning group is the selection of metadata formats to be used in an institution-level prints metadata aggregation. The selected format(s) should support basic functionality required by the defined user groups, and accommodate all adopted shared practices. This decision must be made with an understanding of the data export capabilities of the local systems used by each of the participating repositories. Each system does not necessarily have to be able to export the needed format, but the exported data does have to be automatically transformable (perhaps with some custom mapping work) into the format selected for aggregation.

3.3. Longer-term steps

3.3.1. Further harmonization of descriptive practices

As various repositories within Yale University gain experience with optimize metadata for sharing, they will likely feel more comfortable taking bolder steps towards harmonizing practices within the institution. Each repository will continue to display its own identity, and not all practices will (or will need to be) harmonized. In determining which data elements are most important to harmonize, consider functionality of an

aggregated discovery environment. Data elements that will serve as search limits (such as date and medium) or as browse indexes (such as geographical places and creator names) should be the highest priority for standardization. Data elements that will only be keyword indexed, or simply displayed to users and not indexed at all, are less likely to benefit from harmonization across repositories. It is also important to remember that the metadata useful in the shared environment does not have to be the metadata actually stored in the local environment; the former should be automatically generated from the latter, but they do not have to be the same.

Over time, repositories can also add additional data elements to their local systems, as their utility becomes evident as a result of the collaborative process. Adding data elements of course increases the time it takes to create metadata, so embarking on this task is yet another sign of a commitment a repository can make to support of the metadata sharing process.

The following data elements are good candidates for potential standardization beyond what has already been decided, or at least a collaborative agreement to include in records when feasible:

- Creator names. This element is a difficult one to harmonize, as divergent name authority lists exist, and none is likely to be comprehensive enough for use for all collections at Yale. A first step could be agreement to include this element in all records, and to take the name from a standard list whenever possible.
- Date/time frame. Dates are complex for works of art, with multiple dates often relevant to the resource, dates are frequently uncertain, and important events relevant to the art work often happen over a long period of time. Dates in particular are good candidates for splitting into separate display and indexing values. Named time periods that provide style information could also be included in a field devoted to this purpose.
- Geographic origin. Agreement on a single vocabulary for this data element across all repositories would provide a significant benefit in the aggregated environment.
- Cultural origin. Culture has long been a primary access point for museums, but receives much less attention in a library environment. For works of art, including prints, culture is a prime element for harmonization across multiple repositories.
- Title. It is unlikely all repositories would be able to agree to a shared practice for titles. A reasonable agreement for titles could be that all repositories agree to include at least one for every resource, and that titles are generated according to formally-articulated rules adopted by the repository, perhaps from a shared content standard.
- Person depicted. This data element falls into the “value-added” category, not often recorded in the traditional library or museum environment but of significant potential value to end-users. A reasonable agreement for data elements of this sort could be to include a place for it in the local repository, and fill it in when economically feasible.
- Place depicted. A “value-added” element, like person depicted, and likely to benefit from the same treatment.

Recording relationships between prints is another area in which a great deal of value could be gained, yet is extremely difficult and expensive to implement. A step towards this long-term goal could be a repository recording relationships between prints in its own collection, with plans to make relationships between prints in other repositories' collections over time as aggregated metadata for prints at Yale becomes a reality.

3.3.2. Aggregate art print metadata

The prime time to build a metadata aggregation for art prints at Yale is while momentum from the prints metadata is still strong. Repositories have made an investment in shared practices, and while shareable metadata is a long-term proposition, they will benefit from seeing the payoff of this investment realized in a concrete way. This could be achieved either through building an aggregation specifically for prints, or through providing prints metadata into a larger institutional metadata aggregation.

Both of these options have strategic value, and deciding between them depends on a variety of other institutional priorities. Building a print-specific aggregation could serve as a testbed for technology before committing to adopt it as a wider scale. Including prints metadata in a larger repository could be used as a way to push or even lead the development of Yale digital library efforts as a whole (including methods for dealing with undigitized content) or to evolve the role of the Digital Production Integration Program (DPIP) into more of a service rather than simply an advisory body. In either case, the prints project represents an ideal catalyst for significant steps forward for the institution as a whole.

3.3.3. Expand sharing principles into other formats and other repositories

The successes and lessons of the prints metadata project can be leveraged to expand collaboration and sharing activities at Yale into other types of material. Exactly which materials are best addressed next is dependent on other institutional priorities. However, natural next steps might be materials for which collaborative decisions already made can be transferred to, or other materials held by each of the repositories participating in the prints metadata project. Expanding to include closely related repositories, such as the Visual Resources Center, might also be a relatively easy transition.

Regardless of the formats or repositories included in future collaboration and sharing initiatives at Yale, the general principle of the materials driving descriptive practices to a greater degree than the administrative nature of the repository can be employed (Elings and Waibel, 2007; presentations at the Yale print metadata seminar by Waibel and O'Keefe). Yet by sharing metadata and adopting some collaboratively-developed metadata practices, repositories do not give up their individual identity or surrender their curatorial responsibility. Finding the right balance between local and shared practices is a difficult proposition, and repositories who have successfully tackled this issue are well-poised to discuss their experiences with others. The value of shared practices is not always immediately evident to new participants in this sphere, and collaborative activities can help more explicitly demonstrate the relevant tradeoffs.

3.3.4. Share your work with the community

The products of collaboration from the prints project are of use far beyond Yale University. Shared records themselves are perhaps the most visible outcome of this project. In addition to sharing records within the University, repositories participating in the prints metadata project could share their records more widely using the infrastructure proposed below in the "Technical issues" section. In addition, these repositories could take the opportunity provided to them by improving the shareability of their metadata to provide open access to digitized versions of the prints themselves as well.

In addition, the lessons learned from the prints metadata project are valuable to share with others. Few if any institutions have a coherent, institution-wide strategy for creating

and sharing metadata for a variety of resource formats. Therefore, any methodology an institution finds effective in developing these practices, even at a micro level, can provide essential information to others looking to accomplish similar goals. Documentation useful in the institution-level shared environment, such as high-level practices and workflows, the master mapping document, and the core element set, can also be useful to other institutions embarking on similar initiatives. Posting these documents in an openly-accessible location can provide much-needed guidance to others. Sharing high-level evaluative information about how to achieve institution-wide buy-in for metadata harmonization gained from the prints metadata project at conferences and in the published literature of the cultural heritage community would be a significant contribution to the state of the art in this area.

3.3.5. Implement an assessment program

While the collaborative agreements achieved through the prints metadata project are significant in and of themselves, a significant overall goal of the project is to facilitate effective metadata sharing. The planned fall meeting of representatives from the repositories participating in the prints metadata project can be used to determine how many of the agreed-upon practices each repository has implemented, and evaluate how many of the desired shared documents have actually been posted on the project Wiki. Based on the current state of putting project agreements into practice at this fall meeting, the team should articulate a detailed plan, with timelines, for moving forward beyond the end of the funded prints metadata project. This plan should be informed by the goals suggested by the user analysis performed by each repository, recommended earlier in this report. The project has achieved significant outcomes, and the momentum gained from it can be used to further collaborative and sharing practices throughout the institution.

The most effective means of evaluating the shareability of metadata, however, is to see the records themselves in a metadata aggregation. Performing this analysis on an institution-level metadata aggregation is a valuable first step; however, sharing and analysis on a multi-institution level must be done to truly understand how effective the changes made locally have been.

4. Technical issues

4.1. *Building on existing work*

Yale University already has an online presence for digitized collections and special collections in a number of ways. Destinations for end-users interested in these resources include the [Yale Collections Collaborative](#) pages, the [Primary Sources at Yale](#) site (developed as a Collections Collaborative project), and the Yale University Library [Digital Collections](#) cross-collections search. Many other search services are offered at Yale for these collections as well, with the main Orbis catalog plus a wide variety of collection-specific search engines. While making these collections and/or their metadata available online is a significant accomplishment, maintaining many separate search systems is likely to be unsustainable over the long term. Yale as an institution should be thinking strategically about what metadata should be managed in separate systems, what should be combined together, and what the relationship of all these various systems should be. Metadata for art prints very much plays a role in this discussion, raising questions about whether the metadata for undigitized content should be searchable together with digitized content, how different metadata formats might be

accommodated in larger unified systems, and how data can move seamlessly and automatically between various storage and delivery systems.

4.2. Short-term steps

4.2.1. Automate as much as possible

Automation is increasingly playing a key part in efficient metadata creation workflows. Automation can help metadata creators achieve the predictability that is a key feature of shareable metadata. Some metadata creation systems allow “pick lists” to be specified for certain fields, forcing the value in that field to be selected from a predetermined list. Many metadata creation systems allow for the creation of templates or workforms for specific types of material, which pre-fill some elements with data common to that format, reducing the time needed to create the record and reducing the potential for cataloger error.

Data export and transformation into standard formats for sharing is another area ripe for automation. Sharing of data is an activity that happens on an ongoing basis, and should require as little manual intervention as possible. Once a conversion (with an XSLT stylesheet or other appropriate technology) has been built, it can be run repeatedly with little effort. The data in TMS or Orbis in this case would serve as the “master” metadata from which all other versions are derived. Scripts would then be developed and run automatically on a predefined schedule to extract the appropriate data from the local system, run the appropriate transformations, and deposit the output into a location that a shared system can automatically pick up.

4.2.2. Build small utilities

When tools cannot be built into metadata creation systems to assist with quality review and the creation of shareable metadata, small external utilities can take their place. Simple reports showing data across records for a specific fields, deduplicated and sorted in a meaningful way, can allow for the easy location of anomalous data and a high-level view of metadata practices. A similar approach could be taken across repositories before the records can be truly aggregated, combining together metadata from predefined fields from each repository then performing the data analysis.

External tools could also be used to validate data when this is not an intrinsic feature of the metadata creation system. A small utility could analyze exported data and provide a report of fields which do not match a predetermined data type such as a date format or value expected to come from a finite list.

4.3. Longer-term steps

4.3.1. Aggregate art print metadata

Easily the most effective way to evaluate the shareability of metadata is to see it in operation in an aggregated environment. Policy issues related to actually aggregating metadata for prints on the Yale campus were discussed above, but technical issues play a significant role in this activity as well. The primary technical decision to be made related to aggregating prints metadata is to determine whether the aggregation will be for art prints only or if an existing system such as the Yale Digital Collections cross-collection search will be expanded for this purpose. If the former, a new technical

infrastructure must be developed and explicit plans for pushing the aggregated metadata out beyond this new information silo must be made—best practices for digital libraries have evolved to the point where it is not enough to think only of creating technical infrastructure to handle one specific class of material. If the latter, accommodation must be made for undigitized objects and possibly for new metadata formats.

Core infrastructure for many different material formats and many different metadata standards is a common goal among academic institutions. Developing this infrastructure is a significant challenge, but progress is being made. The general approach many institutions are taking is a single central and flexible data store allowing cross-collection search, but also allowing various “lenses” into the data to build branded services built for access to collections from a single repository or in a single medium. Art prints could represent for Yale an ideal testbed for handling this “special” kind of material in a sustainable way.

4.3.2. Share metadata beyond Yale

An institution-level aggregation of prints metadata would be a significant accomplishment, yet it is still not enough. To be truly useful, this metadata must be shared widely beyond the institution. The collaboration and technical infrastructure already developed or proposed here should be leveraged to achieve this goal. Each repository individually should not have to find its own way to share metadata widely. Instead, the common technical infrastructure of the institution-level aggregation should provide this service, sharing metadata from all participating Yale repositories simultaneously. There is no specific list of metadata sharing methods that all cultural heritage institutions must implement—each institution should choose which of the many options best fit with its needs and goals. Options include the Open Archives Initiative Protocol for Metadata Harvesting ([OAI-PMH](#)), Search and Retrieve via URL ([SRU](#)), Open Archives Initiative Object Re-use and Exchange ([OAI-ORE](#)), [OpenURL](#), [Open Search](#), [XML SiteMaps](#), [RSS](#), [Atom](#), and the Semantic Web community’s [Linked Data](#) movement.

4.3.3. Build bridges between different systems

Regardless of the technical infrastructure used for aggregating prints metadata within Yale, there will still be many different systems containing data that need to interoperate. Metadata-only systems will need to feed data to those that deliver digitized objects online and preserve them over time. Format-specific data stores will need to feed into cross-collection search systems. Digital Asset Management systems such as Portfolio, Artesia or MediaBeacon implemented by individual repositories will need to deliver data to institution-wide services. Each system does not necessarily need to interoperate with every other system, however. A coherent plan should be developed for the relationships between these systems, and only those connections crucial to the flow of data within the overall plan will then need to be built. In making this master plan, it is important to realize that shared systems may not have all the power of repository-level or format-specific systems. The latter can serve as the mechanism for access for specialists, and the former more basic access for generalist users. Efficient flow of data between them can allow both needs to be met simultaneously.

References

- Baker, Thomas. (February 11, 2003). *DCMI Usage Board Review of Application Profiles*. Retrieved from <http://dublincore.org/usage/documents/2003/02/11/profiles/>
- Elings, Mary W. and Waibel, Günter. (March 2007). "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums." *First Monday*, 12, no. 3. Retrieved from http://firstmonday.org/issues/issue12_3/elings/index.html
- Erway, Ricky. (May 2008). *Seeking Sustainability*. Retrieved from <http://www.oclc.org/programs/publications/reports/2008-03.pdf>
- Heery, Rachel and Patel, Manjula. (September 2000). "Application Profiles: Mixing and Matching Metadata Schemas." *Ariadne*, 25. Retrieved from <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>
- Library of Congress Working Group on the Future of Bibliographic Control. (January 9, 2008). *On the Record : Report of The Library of Congress Working Group on the Future of Bibliographic Control*. Retrieved from <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>
- Marcum, Deanna B. (June 1, 2008). *Response to On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Retrieved from <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>
- Nilsson, Mikael. (March 31, 2008). *Description Set Profiles: A Constraint Language for Dublin Core Application Profiles*. Retrieved from <http://dublincore.org/documents/2008/03/31/dc-dsp/>
- Powell, Andy, Nilsson, Mikael, Naeve, Ambjörn, Johnston, Pete, and Baker, Thomas. (June 4, 2007). *DCMI Abstract Model*. Retrieved from <http://dublincore.org/documents/abstract-model/>

Appendix A: Sample Structure for Core Elements Document

Element name	Definition and usage notes	Required?	Repeatable?	Data type
Medium	Term for the technique used to create the print. Use a high-level term from the Yale medium list for this element. More specific terminology describing the creation method or process should be placed in repository-specific fields.	Required	No	Select value from controlled list at http://www.yale.edu/printsmediumlist/

Cultural origin	Term(s) describing the culture influencing the design and/or content of the print. Terms for culture often overlap or are influenced by geographic place and artistic style. Provide a value for this field when known, even if it duplicates information present in another field in the record.	Optional	Yes	Select value from controlled list at http://www.yale.edu/printsulturelist/
-----------------	---	----------	-----	--

Appendix B: Updates to the Master Mapping Document

General comments

For MODS mappings, when showing hierarchy, use / between levels, rather than —.

Unique ID for digital image

Based on the mappings provided, it seems this element is intended to be a locator (URL) rather than an identifier. Map to [dc:identifier] and [mods:identifier], and change the field description to reflect the updated role of this element.

Medium

The MODS [typeOfResource] element is restricted to a closed list that will not match with data for medium recorded for prints. Map to MODS [genre] instead.

Language of title

Do not map to language elements in either Dublin Core or MODS. In each of these formats, the language elements refer to the language of the resource. While the language of the title often is the language of the resource, it is best to record these separately and not conflate the two.

Unique ID for object (call number)

From the various elements mapped to this generic label, it seems a classification number is what is desired here. There is a difference between a classification and a call number, however. The target formats are expecting just the class number rather than a full call number including the Cutter. With this in mind, only map 852 \$h (not \$i) and only \$a within 090. Update the label for this field description to read "Classification" as without the Cutter this element is no longer a unique ID. Update mapping to Dublin Core to read [dc:subject].

Role of artist/creator in the creation of this object

Map to MODS name/role, ensuring role is a child of the name element representing this person. Repeat name for multiple individuals.

Alternative titles

Include type="alternative" on MODS titleInfo/title.

Place of publication/creation and Publisher name

Include in mapping instructions that these two strings should be concatenated into a single [dc:publisher] element, as Dublin Core does not offer an element specifically for publishing place.

Creation date or earliest possible creation date

Map to MODS [dateCreated point="start"].

Last possible creation date

Map to MODS [dateCreated point="start"].

Paper/Support

Change MODS mapping to [physicalDescription][form type="support"].

Name of series/portfolio of which this object is one part

Change DC mapping to [dc:relation]. Add MOD S mapping to [relatedItem type="host"].

[BAC TMS Object Name]

Change DC mapping to [dc:format]. Map to MODS [genre].

[BRBLDL_240]

Map to MODS [titleInfo][title type="uniform"].
