
Considerations in the Design of Art Scholarly Databases

DAVID BEARMAN

INTRODUCTION

THE CONCEPT OF AN INTEGRATED information source containing art historical data and images has been immensely attractive to the art scholarly community. Since the creation of the J. Paul Getty Trust Art History Information Program, which has the goal of realizing this dream, it has become a possibility. Yet the very nature of the vision was that it could be all things to all people—i.e., a catalog of art objects, a biographical dictionary, histories of auctions and exhibits, thesauri and dictionaries of technical terms, the definitive compendium of art criticism, and a comprehensive image base suitable for scholarly research. Realizing such a database in a universally accessible electronic information system requires more precise definition than the vision needed, and it turns out to be very challenging to achieve a database design without sacrificing the benefits which each community of art historical professionals has imagined for itself.

The requirement to be more specific about what is meant by an art scholarly database, and how such a database could be realized, can build upon a considerable body of work—unfortunately largely unpublished—which has grown out of three major systems definition efforts which are currently underway: (1) the discussions of the Architectural Drawings Advisory Group (ADAG), (2) the intellectual framework of the Art and Architecture Thesaurus (AAT), and (3) the data architecture models of the Smithsonian Institution.

Together these projects provide a basis for defining the fundamental considerations which will need to be incorporated into art scholarly

databases. They suggest that, with modest revisions, the MARC formats for bibliographic description and authorities existing in bibliographic information networks might be carriers of *the data* which support an art scholarly (indeed of any scholarly) database. However, it is clear from the analysis that, if existing bibliographic information networks, as applications were used to support the requirements of art scholarly use, they would need to be fundamentally transformed. The author argues that such a transformation is due anyway, and that the library community has much to gain from participating in the effort to realize art scholarly information systems as capabilities of their existing networks.

THE CONCEPT OF AN ART SCHOLARLY DATABASE

The art scholarly database is an idea which has been heralded by prophets in a variety of art scholarly communities: among curators, art historians, and conservators; from the perspectives of connoisseurs, iconographers, and collections managers; and in museums, libraries, and archives. And in the world where believing is being, it supports the scholarly needs of each. To the curator, the art scholarly database has an object record at its center, a description of a work of art as an artifact, created at a specific time, titled by its creator, made by a specified technique, and accompanied by a history of prior ownership and exhibition and provenance. Surrounding the objects at the core of the database are files containing information about artists, donors, exhibitions, styles, periods, techniques, and other recurring attributes of the universe of art.

As imagined by the art historian, the art scholarly database is a vast network of assertions, made by other art historians, about the world of art. These consist of attributions of works to artists, demarcations of stylistic periods and the assignment of works to them, and assertions about influences and about the meaning of specific drawings, paintings, or sculptures. The art scholarly database supports the reorganization of these assertions and their systematic exploration by reference to all the entities to which they refer—i.e., artists and works of art, schools of art and patrons of art, symbols and forms and techniques and styles and media and anything else about which it might be interesting to reexamine the received wisdom.

As a tool designed for the conservator, the art scholarly database is a repository of chemical and physical knowledge, a history of the materials which went into a work of art, and of the conditions to which that work might have been exposed throughout its life including the degradation—precipitous or incremental—which brought the work to his attention. The art scholarly database is also a reservoir of information about all prior treatments to which the work has been subjected and a library of information about similar works or about treatments of the sort the conservator intends to apply.

From the perspective of the connoisseur, the elements of the work of art is the most important attribute, and that art lies in form, color, motion and balance, in technique, and thus not in just materials, but also in how they have been worked. Art is aesthetics and description of the work is a description of the work as art.

As envisioned by the iconographer, art is intellectual and appreciation of the work involves finding its message and its meaning. Describing what the work is about takes precedence over description of what it is or how it achieves its effect.

To the collections manager, the work of art is a responsibility, an item which must be accounted for, stored, and loaned, an object with physical characteristics and a specific set of circumstances under which it was obtained. All this governs how the work of art must be treated within the repository and the innumerable actions which will be taken with it over the course of its life. The collection itself, and the actions taken on it, provides important units of analysis through which to understand the item.

In the museum, individual works of art are vehicles for interpretation of art, understanding of an age, or appreciation of a movement. The art scholarly database is a search room, a window into the numerous collections that might potentially hold items which will stage a magnificent show and an "exhibit" in itself in which the public that "attend" the museum (perhaps by telecommunications) may participate.

At the library, the art scholarly database is a reference source for bibliographic citations to the hundreds of thousands of articles and books, films, slide collections, and now, optical discs, which present, discuss, and define art. The art object may also be published as well as representations of it.

For the archives, the art scholarly database is a pointer to primary materials of the world of art ranging from original architectural drawings to decorative ephemera, from the personal letters of artists to the records of an art gallery or publisher, from the field notes on a cave painting to the programs for a computer generated graphics display. These are all evidence of the world of art as it lived and as it was retained because, in the judgment of the archivist, it has historical value.

Can this mirage, seen by so many observers—none of whom agree on its shape—be engineered into being? The deliberations of three major projects in the art historical world suggest that they can.

THE ARCHITECTURAL DRAWINGS ADVISORY GROUP

The Architectural Drawings Advisory Group was established in 1983 at the instigation of Henry Millan, the dean of the Center for the Advanced Study of the Visual Arts (CASVA) at the National Gallery of Art in Washington, D.C. The impetus for its organization was the expectation that the J. Paul Getty Trust would support the construction

of an architectural drawings database. An organizational meeting in May 1986 attracted representatives from the American Institute of Architects, the Cooper-Hewitt Museum of the Smithsonian Institution, the Library of Congress, the National Archives, the Canadian Center for Architecture, the Avery Library, and the National Gallery of Art. At subsequent sessions, RILA/RIBA, Marburger Index, L'École supérieure des Beaux Arts (Paris), and the Public Archives of Canada were also represented.

In the summer of 1983, ADAG's earliest discussions premised the use of the MARC format for Visual Materials, AACR2, and Elisabeth Betz (Parkers') newly issued rules for cataloging graphics materials, if all three could be revised to meet ADAG needs. (The ADAG minutes were distributed to participants but not published. Subsequent references to ADAG meeting minutes are made in the text in its preferred notation—i.e., ADAG II means the second meeting.) It remained for the group to determine what revisions were required, and this opened up a lengthy discussion over the next two years.

Among the problems noted in these discussions were:

- An object catalog is only one file among the many in an art scholarly database (ADAG III).
- Whether multiple images on a single sheet are reported separately or together depends on whether the image (art) or sheet (artifact) is primary. A variant issue with the same problem is how best to describe multiple sheets which comprise a set (ADAG III).
- The representation of time is different where hours and minutes are less interesting than twilight or dawn and dates less important than cultural time periods like Lent or Advent (ADAG IV).
- What it means to be the creator of a drawing is not self-evident—e.g., whether a firm or an individual is responsible, and whether it is the same for apprentices and partners, draftsmen and designers (ADAG V and ADAG VIII which notes the conflict with AACR2).
- Buildings seem to defy unique identification, when one accepts that the same building may be built in a number of places, a building may be built in a different place from that for which it was designed, and that a building may not be built at all and yet be represented by a full set of drawings (ADAG V).
- How to account for the fact that scholars see every piece of information in the record as potentially arguable and would like to see a provision for sources for each item of information given (ADAG V).
- What we mean—if not nationality, citizenship, or place of residence—when we say that an artist or architect is French; what is the meaning of “locus of activity” (ADAG VI).
- How to accommodate the fact that a photographic reproduction of a work of art is a work in its own right and requires its own entry, credited to the photographer (ADAG VII).

- What to do with the distinction between history of ownership and administrative history in provenance (ADAG VII).
- How to treat the vernacular, whether it is a local calendar time (ADAG VII) or the language of the repository, artist, or location of the building (ADAG X).
- What to do about the fact that geographical places change their boundaries, features, and come and go (ADAG IX).
- How to reflect the relationship between drawing, model, and structure (ADAG IX).
- How an art historical database treats the multiplicity of roles which a single person may hold throughout life or even at one time, in describing that person (ADAG cataloging procedures comparison meeting, 24-25 July 1986).

These issues, discussed by ADAG, are not peculiar to architecture or even the arts but are features of historical and scholarly databases in all fields of endeavor. Historical databases all need to accommodate the fact that the world changes but that our language for representing it is contemporary and has changed along with what it represents. Thus cities are no longer what they were, nor are river valleys, styles of painting, or occupations, but our words for them do not reflect this change any more than our names for people or organizations reflect their maturation.

Scholarly databases all have to deal with the legitimacy of a large number of incomplete perspectives on the same reality. Thus people are parents and architects, school board members and gamblers, all at once. Works of art are images and artifacts, they were created by someone and produced under the creative responsibility of a corporate entity. And dawn is an important time for an art historian, just as years of great military victories are for architects and seasons are for ecologists. ADAG members, in their deliberations, uncovered the complexity of scholarly realities and identified the kinds of relationships which must exist between entities that are central to distinct worlds of scholarship. They have not yet developed solutions, but during 1986 the author was able to demonstrate that the data which ADAG wished to share, in spite of its complexity, were compatible with the structures established by the MARC format for bibliographic data and authority data and widely used in the library and archives communities for sharing information about primary and secondary reference resources (*MARC Mapping of the ADAG Skeletal Design* 1987). Thus an envelope for sharing art scholarly information may already exist if we can agree on the meaning of what we put inside it.

THE ART & ARCHITECTURE THESAURUS

In 1976, Pat Molholt and Toni Petersen submitted a proposal to the National Endowment for the Humanities (NEH) to develop an Art and Architecture Thesaurus as a way to resolve some of the problems of the

meaning of language in art historical scholarly databases (Petersen 1983). After several years of NEH support, the Getty Trust continued to fund the project which has been discovering the vocabulary we employ to represent parts of the multifaceted realities facing art scholarship. After more than a decade of effort, the AAT has defined 30,000 terms in thirty-six separate "hierarchies." Each hierarchy is an attribute or data element in the description of an object of art historical research. An item described using the AAT vocabularies could thereby be indexed according to thirty-six separate facets of description.

The hierarchies developed by the AAT do not describe the content of the work of art at all; iconographic representation and subject description is beyond the scope of the AAT. What the hierarchies do describe are physical attributes, styles and periods, agents, activities, and materials and objects (including built environments, material culture, art forms, and documents). Figure 1 lists the hierarchies in the AAT and Figure 2 illustrates some of these hierarchies with sample terms.

The power of the AAT as a descriptive language derives from the explicit genus-species and whole-part relationships it defines between terms, its definition of synonymy, the increasingly complete scope notes it provides to distinguish among terms, and its identification of the sources that provide warrant for the use of a term. The placement of terms within a hierarchy partially resolves the legitimately different requirements of different users. In an architectural drawings database, the distinctions between Corbel arches and Extradosed arches, or even between the two types of Corbel arches, Bell and Maya, may be necessary while in a database of landscape paintings we can stop with the term *arches*.

But the AAT does not address the question of how its hierarchies are related to each other in a database. It assumes that all these terms could be assigned to the description of an object of art and does not consider how each dimension serves to qualify another. Since the purpose in developing the AAT was to provide catalogers with terminology with which to describe objects, documents about objects, and object and document surrogates, this problem was seen by the project itself as one of providing implementation instructions to indexers. As such, the AAT staff eschewed defining relationships because most implementation of object and document catalogs would not be able to implement them.

In a recent theoretical analysis of the way in which the AAT represents art scholarly discourse, however, former codirector Pat Molholt (1987) has suggested that these terms form a semantic network as illustrated in Figure 3. Her exploration of the syntactic relationships among terms within a single vocabulary reveals eighteen distinctive types of linkages between related terms—i.e., not just genus/species and whole/part linkages which are the relationships

1. ASSOCIATED CONCEPTS
 - CP Associated Concepts
 - DO Disciplines
 2. PHYSICAL ATTRIBUTES
 - DE Design Elements and Attributes
 3. STYLES AND PERIODS
 - SP Styles and Periods
 4. AGENTS
 - PO People and Organizations
 5. ACTIVITIES
 - PR Processes and Techniques
 - FT Functions and Actions
 - EV Events
 6. MATERIALS
 - MT Materials
 7. OBJECTS
 - Built Environment
 - AC Architectural Components
 - BW Single Built Works and Open Spaces
 - CS Built Complexes and Areas
 - SL Settlements, Systems, and Landscapes
 - Material Culture
 - HD Hardware and Joins
 - FR Furnishings
 - CT Containers
 - CU Culinary Artifacts
 - PA Personal Artifacts
 - MD Measuring Devices
 - TE Tools and Tool Components
 - MI Musical Instruments
 - RA Recreational Artifacts
 - AM Armament
 - TA Transportation Artifacts
 - CA Communication Artifacts
 - EM Exchange Media
 - Art Forms
 - DW Drawings
 - PD Paintings
 - SC Sculpture
 - GA Graphic Arts
 - PF Photographs and Motion Pictures
 - BA Book Arts
 - CM Communications Design
 - MM Multimedia Arts Forms
 - VG Visual Genre Documents
 - DT Document Types
- Indicates active hierarchies

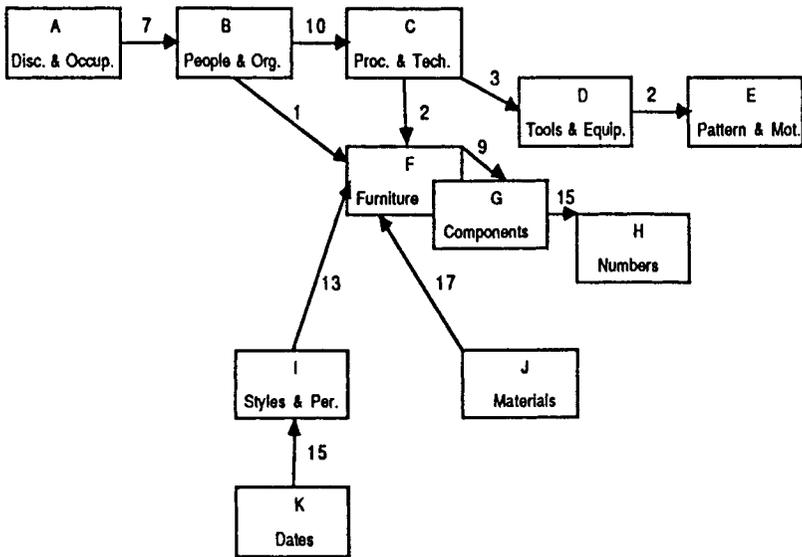
Figure 1. AAT facets and hierarchies

most commonly treated in thesauri. Molholt examines these relationships further and identifies them as operating equally between facets much in the manner in which entity-relationship modeling of the database would define them (see Figure 4).

Molholt's insight permits us to map an art historical statement to the AAT, but it does not schematically represent the universe of possible

ASSOCIATED CONCEPTS Associated Concepts Light Width Octagons Opacity Perspective Classicism	ASSOCIATED CONCEPTS Disciplines Decorative arts Design Art history Semiotics Industrial archaeology Civil engineering	PHYSICAL ATTRIBUTES Design Elements Nymph motif Oval Arabesque Chevron Egg and dart Wreathed Gabled	STYLES AND PERIODS Styles and Periods Colonial American Tudor Dutch Colonial Revival Medieval Anatolian Late Antique
AGENTS People & Organizations Artists Acoustical engineers Miniature painters Sculptors Patrons Leatherworkers	ACTIVITIES Processes & Techniques Assembling Dyeing Blasting Bricklaying Etching Contour line drawing Underpainting	ACTIVITIES Functions & Actions Abstracting Typology Bidding Budgeting Surveying	ACTIVITIES Events Auctions Architectural competitions Exhibitions Festivals Conferences
MATERIALS Materials Near cement Gold alloy Sandstone Teak Burnt sienna Silk	OBJECTS—Built environment Architectural Components Buttresses Ribbed vaults Girders Lally columns Ha-Has Pavilions		

Figure 2. Sample terms



B C E G H G K
 The cabinetmaker carved vines on the fronts of the three drawers of the 20th century
 I J F
 Art Deco Cyprus secretary.

Figure 3. Semantic network illustration (letters indicate AAT hierarchies; numbers indicate link-types)

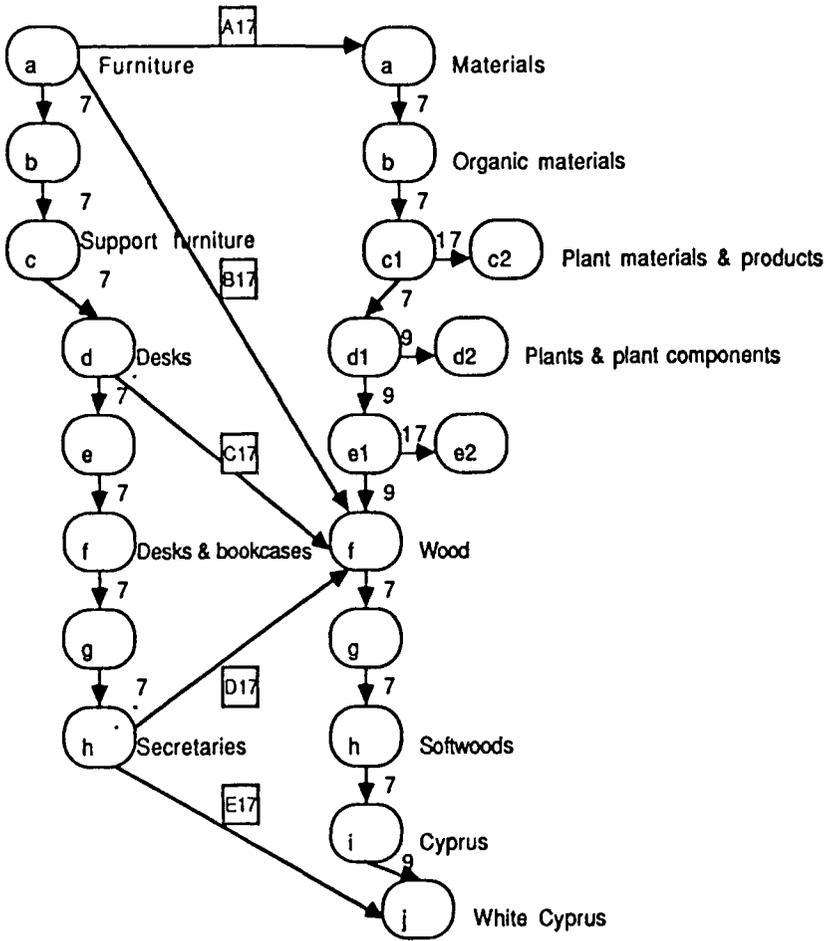


Figure 4. Links between hierarchies (small letters indicate hierarchy levels; numbers represent link-types within hierarchies; capital letters with numbers in squares represent cross-hierarchy links).

art historical statements nor give us the tools with which to determine that a given statement is within the realm of art scholarship. Thus, the schematic representation in Figure 3 links DATES as a qualifier of STYLES & PERIODS, and those in turn are linked to FURNITURE. In a different formulation of this art historical statement, DATES could be an attribute of the description of MATERIALS, TOOLS & EQUIPMENT, PEOPLE & ORGANIZATIONS, DISCIPLINES & OCCUPATIONS, or any of a number of other hierarchies.

SMITHSONIAN ART BUREAUS' DATA ARCHITECTURE

While the AAT focused on naming art scholarly entities—such as roles, materials, or periods—a data modeling effort undertaken by the Smithsonian Institution is attempting to define the attributes of each entity and define the relationships between entities required by its art “bureaus” or museums. The Smithsonian Institution includes seven art bureaus (the Archives of American Art, the Cooper-Hewitt Museum, Hirschhorn Museum and Sculpture Garden, National Museum of African Art, National Museum of American Art, National Portrait Gallery, and the National Museum of Asian Art comprised of the Freer Museum and Sacker gallery) as well as numerous collections of art in non-art bureaus. In 1986-87, as one phase in planning for implementation of the Smithsonian’s new Collections Information System, representatives of these bureaus worked with Smithsonian data administrators and contractors to define the architecture of art information including both content data and administrative (collections management) information. A report on the development of the Art Bureau data model appears elsewhere in this *Library Trends* issue; I want only to comment on its implications for design of art scholarly databases.

The most important finding of the modeling exercise is not likely to be much discussed: it is simply that art museum information is mostly museum information. The model identified *no* entities which were unique to art. (Actually, the draft with which I am working, dated 10 September 1987, seems to imply an entity named “sitters” but it appears to be left over from earlier drafts and I am confident it will not remain when it is clearly only one of many “roles” which a person might play, and “role” is a recognized entity.) The number of attributes which are unique to art are trivial compared with the number of data elements in the logical model. The largest number of entities and attributes in art museum information systems are reflections of the fact that museum holdings are acquired, stored, exhibited, and interpreted. As such, the data are about such entities as addresses, bibliographic items, educational and exhibit events, museum facilities, materials, methods of creation and care, persons and organizations, their roles and skills, and time and space.

While one is first tempted to explain this fact by noting that museum information systems exist to support collections management and only incidentally to support scholarly research and discussion, the data model constructed for the Smithsonian art bureaus forces us to consider that art scholarly discourse is also, largely, about entities other than objects of art. It too is about persons and organizations and their roles as revealed by the attribution discussions within ADAG. It too is about methods, actions, and events as revealed by the hierarchies of the AAT. Even when it is about objects of art, it is frequently about such “entity intersections” as their production (objects and methods) or reproduction (objects and A/V objects), their exhibit and publication

history (objects and events), or their provenance (objects and title transfers), etc. The entities which populate the art scholarly database, like those of other scholarly databases, are agents in our cultural world.

Indeed, art historical discourse can be about ideas abstracted from objects of art or about creators, collectors, and critics of art independent of any objects of art. Thus the art scholarly database is dependent on the development of authority reference files regarding persons, places, concepts, and events as is the political science scholarly database or the geological scholarly database. The art museum collection management database contains authority reference data on donors and collectors of art, on exhibitions and publications of art, and on methods of care and treatment as well as methods of creation. As such, its system architecture will not be significantly different from those of the purely scholarly database even if scholars will be answering different questions for themselves when they are asking similar questions of the system. Thus the scholarly query about provenance may be directed at finding a pattern in the collecting activity of an important connoisseur while the same question posed by a registrar may be intended to assist the development office in attracting another major gift.

What the Smithsonian data model is demonstrating is the intimate relationship between all the entities in the art scholarly/museum collections management database. For logical reasons, the two universes of information are interdependent. The data model identifies the interdependencies or relationships between attributes of intersecting entities which combine to form data sets used in particular museum or scholarly processes.

Interestingly, the model, which is based on a year of bottom-up (derived from actual data being collected now) and top-down (derived from internal logic of the model) work by a number of art museum staff and technical consultants, still adheres to the model the author drew up in 1982 in an invitation to vendors to bid on building such a system (Smithsonian Institution 1984). That model (illustrated in Figure 5) was intended only as a top-down framework and was illustrated with a "data dictionary" consisting of data imagined to exist based on the theoretical model.

Implications

While the data required by the art scholarly database can be shared using existing MARC formats (with minor modifications to the format for bibliographic description and extensions of some general principles across the authority formats), the systems which currently support the sharing of bibliographic data are inadequate to support scholarly databases. Their most serious limitation is their uni-centricity—one file is the focus, and the others merely elaborate on it. In this case, the bibliographic item lies at the heart of the data structure with all other information simply pointing to it.

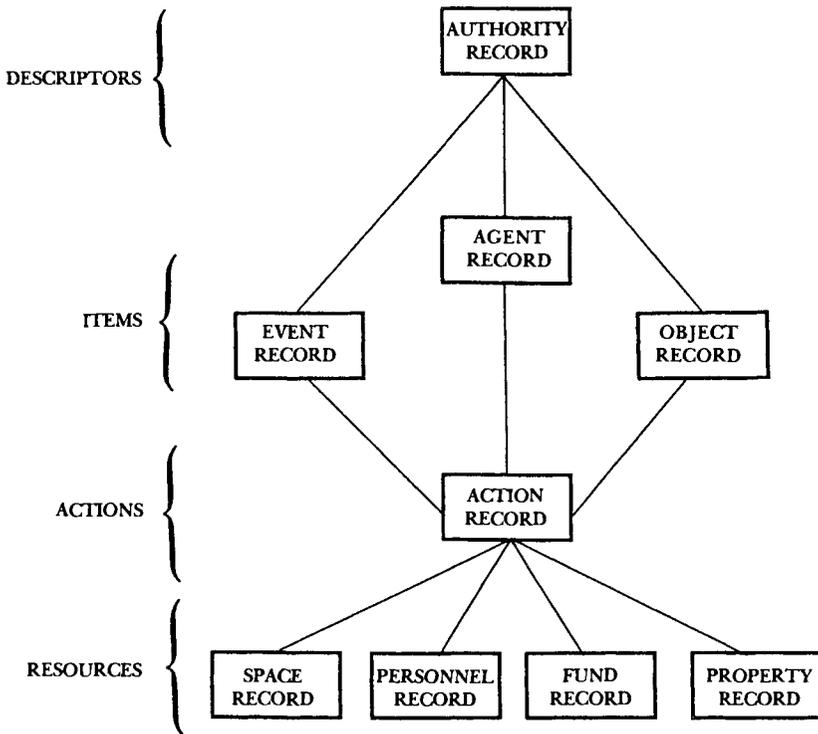


Figure 5. Records functioning as authorities

In the art scholarly database, as the Smithsonian model demonstrates, different entities take on the central role depending upon the perspective of the user. The linkages between files must be traversable in any direction; thus all files are authorities to other files and each may be linked with any other. This linked authority structure, which the author has discussed elsewhere (Bearman and Szary 1987; Bearman 1986), is demonstrated in a database developed by J. Penelope Small at Rutgers University to house the *Lexicon Iconographicum Mythologiae Classicae*, a database describing images of Greek, Etruscan, and Roman objects which illuminate ancient myths (Small. In press). Unfortunately, no bibliographic utility is yet close to being able to support such multiple-linked authority structures, although the work of the Research Libraries Group PRIMA project, which is endeavoring to add scholarly files to the Research Libraries Information Network (RLIN) will certainly face the challenge of linking at least some scholarly files to bibliographic files in a two-way connection which permits either one to serve as the organizing center of a user's query (Hume. In press).

CONCLUSIONS

The design of art scholarly databases requires that we keep in mind the complexity of scholarly realities, the potential richness of languages for describing those realities, and the value of explicitly relating entities and attributes in these databases. It also requires that we be cognizant of some political and financial dimensions of implementing national systems including constraints imposed by the most likely vehicles of the exchange of data—i.e., library bibliographic networks, MARC, and the library community.

It is important that the logical model developed by the Smithsonian Institution also demonstrates that the database must be the product of cooperative development, with values for particular attributes contributed by distinct departments in the museum and by a variety of scholarly communities including art historians. The range of information required in art scholarly discussions, as the ADAG project has shown, requires that the values for particular attributes in the database must be contributed by specialists in various disciplines. It is not sufficient, if this discourse is to be intelligible, for such databases to be constructed from commonly defined data elements. It is critical that the disciplines also accept common vocabularies for specific fields of information such as those being developed and maintained by AAT. Thus the logical design, linguistics analysis, and philosophical debate support the conclusions reached by planners of networks in which one might realize an art scholarly database—i.e., the only practical means to achieve this end is to provide to the holders of the art objects (largely museums) capabilities which will encourage them to build databases containing information which is required in the conduct of scholarship. The informational objectives of scholarly and collection management systems are distinct from one another, but their informational content overlaps, and neither function can be supported solely by the data it provides to the system; the symbiosis is complete for an intellectual point of view and is compelling as a practical matter. Only in the fusion of the needs of collections managers with the requirements of scholars will the cultural world be able to afford to construct art scholarly databases which satisfy both.

REFERENCES

- Bearman, David. 1987. *MARC Mapping of the ADAG Skeletal Design: A Report to the J. Paul Getty Trust, Art History Information Program*. New York: Archives and Museum Informatics.
- Bearman, David. 1986. "Archival and Bibliographic Information Networks." In *Archives and Library Administration: Divergent Traditions and Common Concerns*, edited by Lawrence McCrank, pp. 99-110. New York: Haworth Press.
- Bearman, David, and Szary, Richard. 1987. "Beyond Authorized Headings: Authorities as Reference Files in a Multi-Disciplinary Setting." In *Authority Control Symposium* (Occasional Paper No. 6), edited by Karen Markey, pp. 69-78. Tucson, AZ: Art Libraries Association of North America.

- Hume, Leslie. In press. "The Research Libraries Group's Program for Research Information Management: New Information Resources for Scholars." In *Proceedings of the International Conference on Databases in the Humanities and Social Sciences* (Montgomery, AL). Greenbrae, CA: Paradigm Press.
- Molholt, Pat. 1987. "Development of Term Relationships for the Enhancement of Semantic Networks and Hierarchically Structured Thesauri." Typescript.
- Petersen, Toni. 1983. "The AAT: A Model for Restructuring of LCSH." *Journal of Academic Librarianship* 9 (September), pp. 207-10.
- Small, J. Penelope. In press. "Computer Index of Classical Iconography." In *Proceedings of the International Conference on Databases in the Humanities and Social Sciences* (Montgomery, AL). Greenbrae, CA: Paradigm Press.
- Smithsonian Institution. 1984. "A Plan for the Acquisition of an Integrated, Generalized, Collections Management Information System. Request for Comment. Smithsonian Institution Collection Information System." Washington, DC: Smithsonian Institution, Office of Information Resource Management.