

DESIGNING A COMPUTER-INDEX OF CLASSICAL ICONOGRAPHY

BY JOCELYN PENNY SMALL

Dr. Small is Director of the U.S. Center of the *Lexicon Iconographicum Mythologiae* at Rutgers University, New Brunswick.

SCHOLARSHIP, like history, is not immune from cycles. In the second half of the nineteenth century a number of scholarly encyclopedias, dictionaries, and compendia was produced to order the rapidly growing amounts of information that began to constitute the scholars' domain. The academic community is again embarking on another grand reconsideration of its material. Some of the process is due to the evolution of ideas and the gathering of new data. Another, and increasingly important catalyst, is the computer. The *Lexicon Iconographicum Mythologiae Classicae* bridges both worlds.

The publication, as a pictorial dictionary of classical mythology, looks back to the previous century in its concept, but is very much of this century as an international enterprise of over thirty-five countries. The U.S. Center (*US LIMC*) opened in 1973 with responsibility for cataloguing Greek, Etruscan, and Roman objects with mythological scenes in public and private American collections. The past three-and-a-half years, however, have not been spent in visits to museum collections or in the addition of new catalogue cards and photographs, but have been devoted to the development of a computer-index to the existing American material. This article describes the work of the *US LIMC*.¹

Perhaps the first question is why computerization has taken so long. The number of catalogue cards, about seven thousand, is not inordinate. To a certain extent, the small staff of two archaeologists and a part-time secretary impose limitations on the quantity of data entered. Of greater importance, however, have been and continue to be the problems of designing a system for which there are no models. Yes, RLIN and librarians have been dealing with related issues of computerization for some time, but the stress here must be on the "related." A book is a very different thing from an object, and particularly an object that could have been pro-

¹ It is with great pleasure that I acknowledge the support of Rutgers University and especially the Library where the *US LIMC* resides. The entire project would not have been possible without the generous assistance of the National Endowment for Humanities, Research Tools. For the current grant period (September 1986—August 1988) I am particularly grateful to the Lucille and David Packard Foundation for matching funds.

duced any time between 800 BC and AD 400 by any classical culture in any medium. Moreover, that object rarely survives in its pristine state. How do you deal with incomplete objects? Do you catalogue the leg of a table as a table or as a leg? What happens if it is either from a table or from a couch? What do you do with lids and handles of vases? I pass over words that have a tendency to migrate from field to field. "Column" can be part of a building and has parts itself (base, shaft, capital, etc.), be a type of vase (column-krater), or an independent entity (the column of Trajan in Rome).

The difficulties become compounded when one realizes that the *US LIMC* is providing not an index to mere titles, so-to-speak, of objects, but to subjects represented. What components comprise a scene? A statue or a gem with a single figure present fewer problems than a Roman sarcophagus with forty figures. Do you give all the figures? What about other elements like trees and bushes, architecture, animals, and "objects" like chariots, vases, and utensils? If you record the jewelry and clothes the figures wear, should you note their hairstyles (indicative of date)? Should you include the poses of the actors? As one scholar said to me, everything is important.

What about the almost total absence not just of authority lists, but even of basic definitions? Is a phiale (a dish used for libations) the same as a patera? Even if it is, should the Greek term (phiale) be used for its appearance on Greek objects and "patera" for its occurrence on Roman ones? Then what do you do with Etruscan "paterae," for which no Etruscan term is known? What about divinities who have "equivalent" names? That is, the layman may equate Ares and Mars, but the scholar knows well that the two have different, albeit overlapping, domains. What happens when several conflicting interpretations exist for a particular scene? for individual elements within that scene? for parts of those elements?

As should be evident, the problems of the computer, both hard and soft, pale in comparison to the scholarly ones. The international members of the project agreed to use IBM microcomputers for compatibility. Thus one major decision has been made. Software does vary. The *US LIMC* uses Advanced Revelation produced by Revelation Technologies, Inc.; the French and the Swiss have dBase III. Differing needs led to these choices. The *US LIMC* alone is responsible for the analytical index of the representation, while the other two computerized centers are concentrating on the production of the publication. Three of the projected seven volumes have appeared, and each double volume has about four thousand illustrations. The logistics of keeping track of photographs in the com-

puter are considerably less complex than the situation I have just outlined. In any case, Advanced Revelation provides an import and export routine for dBase. It should also be noted that the Ancient Section of the Getty Photo Archive (GPA) and the *US LIMC* had a successful exchange of data in both directions in 1987, even though the GPA uses a minicomputer with a flat file database program. The intellectual equation of fields and the data, of course, remains thorny.

The design of the *US LIMC* system did not appear whole like Athena from Zeus' head, but slowly evolved and continues to evolve. Furthermore, the software markedly affects the design. Advanced Revelation allows records up to 64K with all variable length fields, a real boon. Only display lengths are defined, and even they can be overridden either by changing the length "on the fly" in a query or by automatic tickertaping when the space on an entry screen has been filled. All fields, no matter how long (within the 64K limitation) are fully and Boolean searchable. Fields can also be repeated or multi-valued, and these too adjust dynamically to the data entered. The savings in storage compared to a program with fixed length fields is enormous.² Other characteristics of the program make the designer's and users' lives easier. Fields can be added or deleted without reloading the data, which is stored in straight ASCII. Upper ASCII is searchable, and alternate character sets can be added. Screens can be linked without programming to produce a virtual hypertext environment, which greatly speeds entry and, even more importantly, correction.³ Lastly, an efficient relational system results in a proliferation of files with a concomitant reduction of fields within each file.

Approximately a year was spent in trying various arrangements of files and fields. Since then the adjustments have been minor and limited to changes in fields, often accompanied by the addition of subsidiary classification files. Two files, "objects," which controls "scenes,"⁴ form the core

² When only one thousand catalogue cards had been entered, I did a rough calculation of total storage needs for all American catalogue cards. The worst case scenario had under six megabytes in Revelation compared to a little over forty-one megabytes in a fixed length program, which would not include all the information we record.

³ This feature is part of Advanced Revelation and not the original Revelation, with which the *US LIMC* started. In fact, some of the tweakings that the system has undergone have been due to upgrades in the program. The last change has taken several months to accomplish, partly due to the abysmal documentation. The program itself, however, is highly recommended. For more information, see my review in *Archives and Museum Information* 2(1988): 2-5.

⁴ Originally there were three core files to allow for the fact that scenes had figures. The information about figures was actually entered in the scenes file with a program to build the figures file automatically. Because experience showed that the figures file was rarely queried, the file, for the time being at least, no longer exists.

of the data with almost thirty satellite files controlling the choice of words, spell checking, defining the words, storing indexes to and classifying the data in the core files. Depending on the field, a satellite file may do one or more of these activities in addition to performing the same duties for other satellite files. Because a chart of all the connections would resemble a cobweb,⁵ I have provided only a list of the files with the current number of records in each. [Figure 1] Files which just do indexing are not included in this list.

Deciding what information to record depends on the use envisaged for the database. A research tool for ancient art historians will not include precisely the same information as one for excavating archaeologists. [Figures 2 and 3] For example, the *US LIMC* does not consider the components that make up materials like clay. On the other hand, pose, which is a significant aspect of iconography, has not been recorded with a controlled vocabulary, but only indirectly in the text descriptions of the scenes. Verbally describing poses is never easy, and becomes highly subjective when one is trying to decide between closely related movements, such as waving good-bye or waving hello. In fact, departures are frequently called arrivals by different scholars.

At the same time every figure can be broken down into an extraordinary number of distinct actions. For example, Perseus on the New York vase [Figure 3] is in the process of decapitating Medousa, sleeping on the right. He turns his head back to gaze at his protectress, Athena, and grasps Medousa's head with his extended left hand, as he slices her neck with the harpe held in his right hand. His left leg is bent at nearly a right angle so that his left foot, placed against the gentle incline, gives him sufficient leverage for the beheading. His right, weight leg is less sharply bent than his left leg with the right foot firmly on the ground. Where do you stop? To describe forty figures on a Roman sarcophagus becomes a gargantuan task. Do not even think about the challenge of developing a controlled vocabulary. A balance must be made between what is useful and what is reasonably possible.⁶

As a result the "grid" of information about the elements of a scene is now limited to three fields: the element, its type and its accoutrements. The latter used to be two fields, dress and attributes, but deciding whether the lion skin of Herakles was an item of dress or an attribute began to

⁵ This image is from Jean Aitchison, *Words in the Mind* (Oxford 1987), 64, 72-86.

⁶ In the case of poses a more effective "solution" will be the implementation of pattern matching. When the computer is able to recognize what it sees, it will also be able to search for what it knows. No programs currently exist, to my knowledge, that can achieve the degree of discrimination necessary for the art historian.

FIGURE 1:
LIST OF FILES AND THEIR SIZES

<i>FILE</i>	<i>RECORDS</i>
<i>Core Files</i>	<i>Numbers</i>
Objects	6,259
Scenes	7,374
<i>Satellite Files</i>	
Artists	1,843
Bibliography	2,402
Collections	150
Cultures	16
Dress	89
Element Types	14
Elements	1,981
LIMC Index—USA only	526
LIMC Index—Volume 1	2,319
LIMC Index—Volume 2	10,325
Materials	111
Object Terms	35
Originals	51
Parts	42
Positions	17
Proveniences	309
Purposes	12
Rolodex	284
States	57
Styles	74
Techniques	52
Titles	805
Views	20
TOTAL	35,167

resemble theological discussions of how many angels could stand on the head of a pin. Conversely the original set up had two fields to define the object: the object (Vase) and its type (Neck-Amphora). Fill-in-the-blank fields, however, only work well when the data entered is consistent. The various types of amphorae (Nolan, Panathenaic, Bail, etc.) would be separated from each other in sorts, and the field seemed clogged. As a result,

FIGURE 2:
OBJECTS RECORD

OBJECT NUMBER:	5303	
CITY:	New York	
COLLECTION:	Metropolitan Museum of Art	
INVENTORY NUMBER:	45.11.1	
EXHIBITION COLLECTION:		
DISCOVERY DATE:		
PROVENIENCE:	Gela?	
FINDSPOT:		
SUBJECT:	Vase	
TYPE:	Pelike	
SUB-TYPE:		
PURPOSE:		
CONDITION:	Repaired from fragments; incomplete.	
DIMENSIONS:		
MATERIAL:	Clay	TYPE:
CULTURE:	Greek	
REGION:	Attica	
PERIOD:		
TECHNIQUE:	Red-Figure	
STYLE:	Attic	
ARTIST:	Polygnotos	BASIS: Beazley
ORIGINAL:		
DATE:	450 BC—440 BC	BASIS:
DESCRIPTIONS:		
DECORATION:	A1—Perseus and Gorgon B1—King Polypeithes	
REMARKS:	Seen by Davies.	
BIBLIOGRAPHY:	ARV/2 Para Addenda Henle, Myths	1032 No. 55 442 155 91 fig. 43
DATE ISSUED:	03-26-88	

FIGURE 3:
SCENES Record

US NUMBER: 5303
 SCENE NUMBER: 5303-A1
 CITY: New York
 COLLECTION: Metropolitan Museum of Art
 INVENTORY NO.: 45.11.1
 PHOTOGRAPH(S): MusPh 134280
 VIEW: Side A
 PART: Body
 POSITION:
 TITLE: Perseus and Gorgon
 DESCRIPTION: Athena, leaning on her spear in her left hand, stands frontally, but looks right toward Perseus who returns her gaze, as he decapitates Medousa on the right. Medousa is in a position of sleep with her head resting on her left hand.

ELEMENTS	TYPE	DRESS/ATTRIBUTES
Athena	Clad	Peplos; Helmet, Attic; Aegis; Mantle; Spear
Perseus	Nude	Chlamys; Cap, Wings; Boots, Wings; Harpe
Medousa	Clad	Chiton, Short

OTHER TITLES:
 REMARKS:
 DATE ISSUED: 03-26-88

a third field (subtype) was added for the “modifiers” (Neck, Nolan, etc.). Thus usage has affected the design.

Outside queries have had less of an effect than the experience of the *US LIMC* staff. Until this year the amount of data was insufficient to be useful to many scholars. Those who do have requests (which include several “standing” orders) often have low expectations of success, and tend to be amazed that the system answers so many questions that are not strictly iconographical. Consider a real example. A scholar wanted all objects in American collections from Sicily. Because ordinarily only the specific provenience (Syracuse, Gela, etc.) is recorded, getting all Sicilian sites would be no easy task. Instead the *US LIMC* enters the specific site (Provenience) in the objects file, *and* classifies all sites as to region, country, and, where it applies, global area (Asia Minor, Phoenicia, etc.), in the proveniences file. [Figure 4] With the proveniences of the objects file inverted into the proveniences file, it is a simple matter to search on proveniences for all regions equal to Sicily, and thereby obtain a list of all objects from Sicilian sites.

This use of the satellite files to classify data (not just words, but phrases like “New Brunswick”) is distinctive to the *US LIMC*. Moreover, the classification schemes can be entered at any level, not just broader to narrower. If you want countries, you query for countries, if sites for sites, etc. Moreover, a modular approach allows for a gradual implementation.

FIGURE 4:
PROVENIENCES Record

PROV	Gela	OBJECTS: 18
SYNONYMS		
GLOBAL		
NATION	Italy	
REGION	Sicily	
SITE	Gela	
USAGE		
REMARKS		
REFERENCES		[1] PAGES
PECS		346-347
COMPLETE	YES	RECORDER JPS
		DATE 01/27/87

It was not possible for all fields to have fully controlled vocabularies before data entry began. As a result, certain fields or sets of fields are periodically examined, and “retrospectively” corrected by inverting them into their satellite file. In other words, the data entered in one file is used to make the records in another file. Errors are more easily corrected, because fewer entries need be proofed. The approved new entries are then filled in with their full classifications. In some cases, like proveniences, generating a list of all possible sites would have been an enormous task (a minimum of 10,000 records), and was not really part of our “brief.”

Although the core files still govern the creation and the use of the satellite files, the latter have begun to take on a life of their own. Two years ago I did a very small pilot for a bibliographical database for Classics.⁷ Seventy-four records were transferred into a separate file in the *US LIMC* system. The data in the rubric (general classification as to literature, history, etc.), the title of the work, and the full abstract were inverted into a separate index file. This file, in turn, was linked to the *US LIMC* Titles file, which controls and classifies titles of scenes. For instance, Odysseus is defined as being heroic and belonging to the Trojan cycle and to several sub-cycles (Iliad, Ilioupersis, Odyssey). [Figure 5] Without any change in the pilot data works concerning the Trojan cycle were now retrievable, even though the word “Trojan” appeared nowhere within the test records.⁸

The implications for scholarly retrieval in general are very broad. Once marble is defined as a stone, it is a stone no matter whether the period is antiquity, the Renaissance, or today. For recent times, certain materials like plastic and steel should be added. Since scholars in different fields and even sub-fields use the “same” information in different ways, overlapping modules will be necessary. Two examples should suffice. Unlike the iconographer, the classicist, as I learned in the pilot, finds “Trojan Cycle” a nearly worthless category. The geologist will certainly find the definition of marble as stone simplistic. Such an arrangement

⁷ The results were presented in a paper, “Databases for Research in Classics,” at the annual meeting of the American Philological Association in December 1986. This study was jointly conducted with Prof. Dee Clayman of Brooklyn College. Both of us are grateful to the American office of *L'Année Philologique* at the University of North Carolina, Chapel Hill, for permission to use American material already in machine readable form.

⁸ The actual process was a bit more complex, because the titles file at the moment is like the proveniences file and reflects only immediate *US LIMC* needs. That is, some entries to reflect the pilot data had to be added to the titles file. The *US LIMC* hopes to “complete” (inasmuch as that word can ever apply) the titles file over the next two years.

FIGURE 5:
TITLES Record

TITLE	Odysseus	SCENES = 2
SYNONYMS	Ulysses Ulixes	
CLASS		
Heroic	CYCLE Trojan	SUB-CYCLE Iliad Ilioupersis Odyssey
SETTING		
STATUE		
SUMMARY	Wily Greek hero married to Penelope and father of Telemachos. Often depicted bearded and wearing a pilos.	
REMARKS		
LIT REFS		
REFERENCES		[1] PAGES/PLATES
Brommer, Ody		passim
COMPLETE	No	
RECORDER	JPS	
DATE	03/27/88	

reflects not just the way scholars search for information, but also the way we store words in our brains according to recent studies.⁹

The extreme specificity of terms to each discipline also means that the scholars themselves will have to make the classification schemes, if the classifications are to be effective. In the current state of natural language understanding and artificial intelligence, I see no automatic way to create such modules. The data they classify, however, does not have to be touched. In other words, to produce a fairly dramatic improvement in retrieval of works from RLIN, for example, these modules could be used "on top of" existing indexes of words from the titles and on key word indexes.

Nor does the idea of classification modules need to be limited to bibli-

⁹ Aitchison, *op. cit.*, 198.

ography and structured data. Full texts, be they primary or secondary sources including the footnotes, will be used. An abstract, if it is good, may summarize an article or book better than an author did in the work itself. Nonetheless, the amount of time it takes and the inevitable selectivity of the abstracter increase the abstract's cost and reduce its scholarly worth. Key words have similar shortcomings. Eventually the *US LIMC* would like to combine its material, both verbal and visual (the photographs), in one computerized form that can also be linked to other scholarly resources such as the *Thesaurus Linguae Graecae*.

Thus in the four years that the *US LIMC* has been computerizing its catalogue cards it has developed a new system to analyze pictures and their components. At the same time a general search engine, not entirely anticipated, has been created which has immense significance for the production of "The Compleat Scholarly Resource." In the meantime on a less lofty, but more realistic, level the *US LIMC* welcomes inquiries.