N91-23037

DISTRIBUTED ACCESS VIEW INTEGRATED DATABASE (DAVID) SYSTEM

Barry E. Jacobs
Code 934
Goddard Space Flight Center
Greenbelt, MD 20771

Introduction

Due to the diversity of computers, operating systems, management systems, network protocols, etc., NASA space scientists have to learn many different access methods in order to obtain data. For example, in NASA's Space Astrophysics Program, Astrophysics observatories such as the International Ultraviolet Explorer (IUE), Hubble Space Telescope, Gamma Ray Observatory (GRO) etc, generally have one or more data centers for the analysis and distribution of data. The heterogeneity of its data centers makes Astrophysics multi-mission research almost impossible.

The Distributed Access View Integrated Database (DAVID) System, which is the subject of this paper, is a solution to the problem. See Figure 1. The DAVID System has been adopted by NASA's Astrophysics Division for their Astrophysics Data System and represents one of the most extensive national testbeds of heterogeneous distributed computing systems. Although the DAVID System was originally built to uniformly access databases, it is currently being generalized to other object types such as spreadsheets, images, manuscripts,, graphics, etc. The DAVID software is in the process of being rebuilt by the private sector for adoption by other NASA disciplines.

The DAVID approach to the heterogeneous distributed systems problem is at four levels. See Figure 2. At the lowest level, we develop universal object type management systems to provide uniform access to heterogeneous database, images, spreadsheet, manuscript, etc. management systems. At the second level, we develop "book" and "kit" management systems to provide uniform access to aggregate sets of data objects. At the third level, we develop "libraries" to provide uniform access to a local area networks of computers containing "books", "kits", and other data objects. At the fourth level, we develop consortiums of libraries to provide access to sets of libraries.

An outline of the paper is as follows. In Section 1, we outline the heterogeneous components of the Astrophysics program. In Section 2, we describe the Library and Library Consortium levels of the DAVID approach. In Section 3, we discuss the Books and Kits level. In Section 4, we describe the Universal Object Type Management System level. In Section 5, we explain the relation of the DAVID project with the Small Business Innovative Research (SBIR) Program. We conclude the paper in Section 6, with a summary.

The Heterogeneous Astrophysics Program

NASA's Astrophysics program provides a good example of the system heterogeneity problem. We briefly outline some of the heterogeneity.

The Infrared Processing and Analysis Center (IPAC) at California Institute of Technology manages data from infrared sensors such as the Infrared Satellite (IRAS). IPAC has Cyber and Sun computers, Cyber and Unix Operating Systems, IM/DM database management system, and Internet and Dec Net communications. The data includes IRAS observation logs (databases), IRAS products (databases), IRAS imagery (images), IRAS spectra (spreadsheets), IRAS documentation (manuscripts), and IRAS software.

The Smithsonian Astronomical Observatory (SAO) at Harvard University manages data from the high-energy sensors such as the Einstein or High-Energy Astronomical Observatory (HEAO). SAO has VAX, Sun and Cyber computers, VAX/VMS and Unix Operating Systems, Ingres database management system, and Internet and Dec Net communications. The data includes Einstein observation logs (databases), Einstein products (databases), Einstein imagery (images), Einstein spectra (spreadsheets), Einstein documentation (manuscripts), and Einstein software.

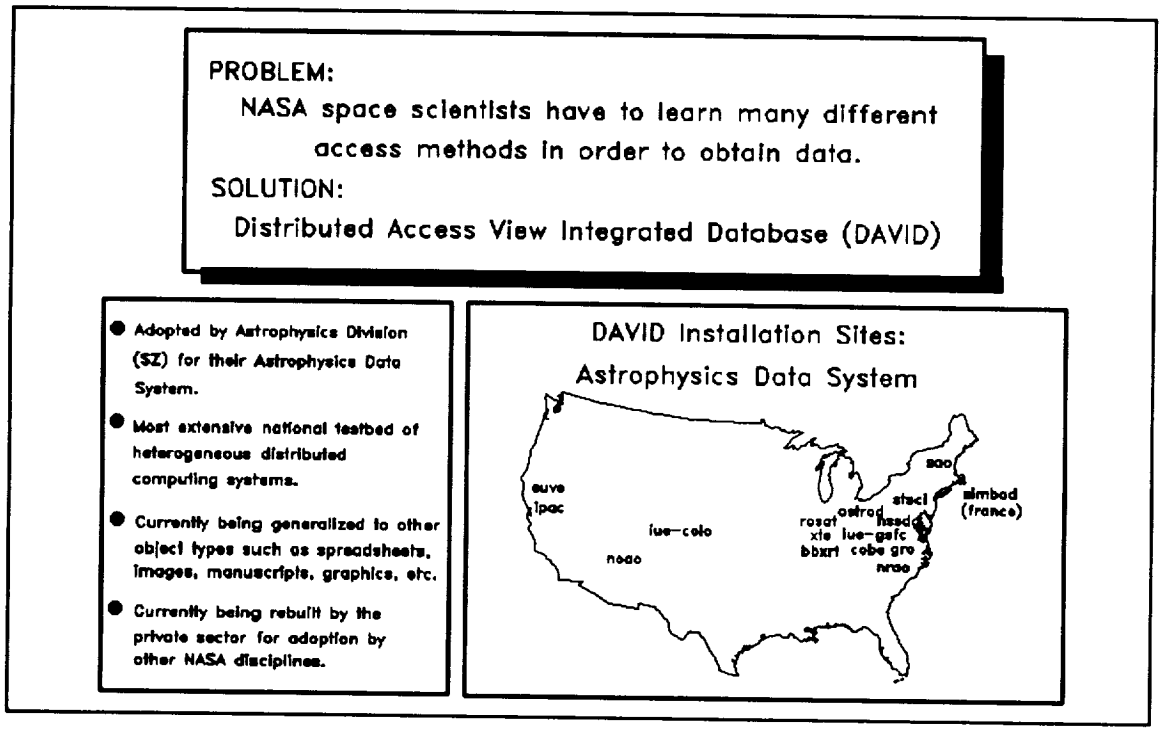

Figure 1.

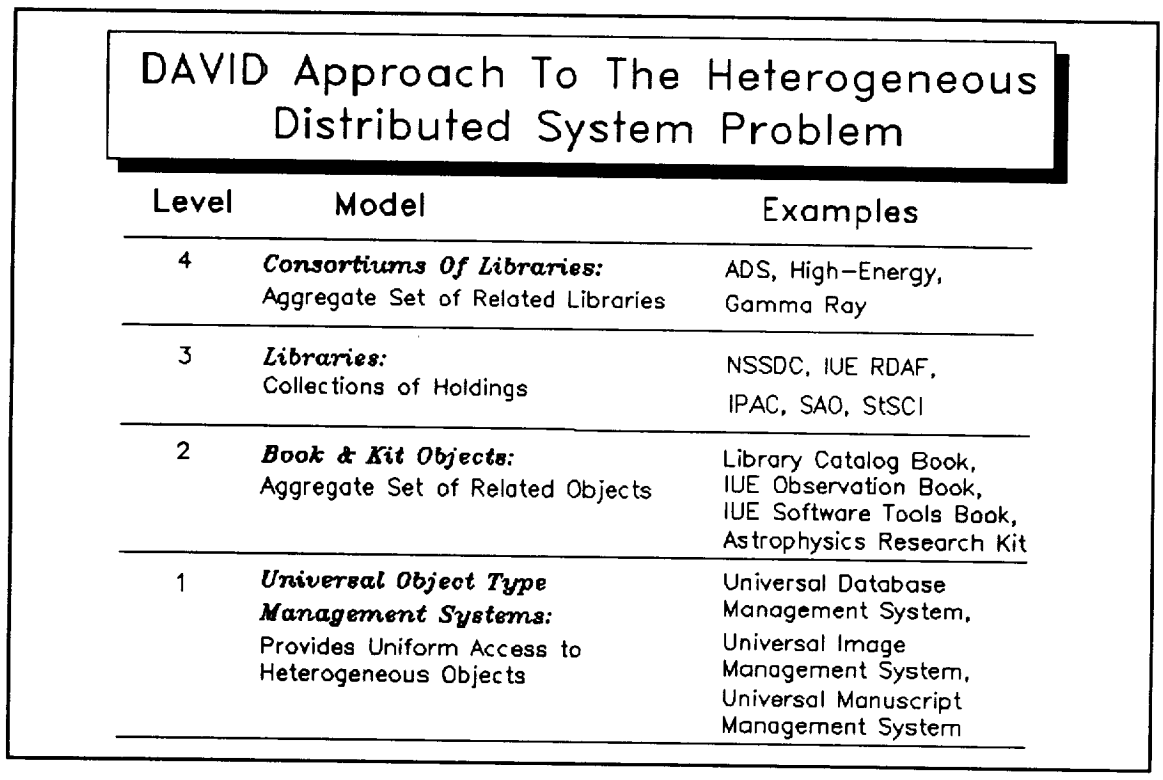

Figure 2.

The International Ultraviolet Explorer (IUE) Regional Data Analysis Facilities (RDAF) at the Goddard Space Flight Center and at the University of Colorado manage data from the International Ultraviolet Explorer mission. The RDAFs have VAX, Sun and IBM computers, VAX/VMS, Unix, and IBM Operating Systems, Ingres, IBM database management system, and Internet and Dec Net communications. The data includes IUE observation logs (databases), IUE products (databases), IUE spectra (spreadsheets), IUE documentation (manuscripts), and IUE software.

The Space Telescope Science Institute (STScI) at the Johns Hopkins University manages data from the Hubble Space Telescope (HST). The STScI has VAX and Sun and Cyber computers, VAX/VMS and Unix Operating Systems, Sybase and Britten-Lee database management systems, and Internet and Dec Net communications. The data includes HST observation logs (databases), HST products (databases), HST spectra (spreadsheets), HST documentation (manuscripts), and HST software.

The National Space Science Data Center (NSSDC) at the Goddard Space Flight Center manages data from a number of astrophysics and non-astrophysics missions. The NSSDC has VAX, Sun, IBM, and Cyber computers, VAX/VMS, IBM and Unix Operating Systems, Ingres, Oracle, Sybase and Britten-Lee database management systems, and Internet and Dec Net communications. The data includes multi-mission observation logs (databases), data products (databases), spectra (spreadsheets), documentation (manuscripts), and software.

Libraries and Library Consortiums

The DAVID approach, models a local area network as a library and a set of local area networks as a consortium of libraries. See Figure 3.

Each local area network of computers is modeled as a **library** in which each of the machines is a "room". One particular room, the "main room", serves as a gateway into the library. When one logs on to the main room and enters the command "dlib", one gets the Library Directory menu which leads the user to the Main Room (described below), Tutorials, Reading Rooms (computers for library holdings), Personal Library Rooms (computers for user holdings), and Management Rooms Menus (computers for staff holdings).

When the Main Room option is chosen, the user sees the Main Room Menu. This consists of submenus for Administrative Desk (library cards, bulletin board, mail, phone, suggestions, etc.), Reference Desk (names, phone and forms to reference librarians), Library Catalogue (database of local library holdings), Circulation Desk (forms for borrowing off-line holdings), Reproduction (forms for reproducing off-line holdings and software for reproducing on-line holdings), Union Library Desk (access to union catalogues and facilities for consortiums), and Remote Libraries Desk (access to remote libraries).

A local library may be a member of one or more **consortiums of libraries.** A library consortiums is a organization of participating libraries which share each other's resources. One of the functions of a library consortiums is to provide a **union catalogue** of all of the holdings over all of the member libraries. A user can access a union catalogue of a consortium and determine which member has the holding. He can then ask his local library to help him obtain obtain access to it.

For example, in the Astrophysics example, the library representing the Center For Astrophysics (CFA) could belong to several consortiums, for example, "X-Ray", "High-Energy", "Astrophysics", and "Multi-Disciplinary". Then if a user at CFA wants a holding not at CFA he can consult one or more of the union catalogues of the different consortiums. He would then locate the holding and ask his local librarian to obtain access to the holding.

Books and Kits

The DAVID approach, models an aggregate of relate data as a book and a kit.

A **book** is an aggregate data structure which is divided into logical units called **chapters** and physical units called **components.** An example of a book is given in Figure 4. for the Book of IUE Observations. Here each chapter represents one of the 80,000 observations taken by IUE. The IUE book has as components- a title page (manuscript), a preface (manuscript), a table of contents (database), chapter components including- observation descriptions (manuscript), proposals (manuscripts), literature references (database), raw image (images), extended-line-by-line spectra (spreadsheets), observation scripts

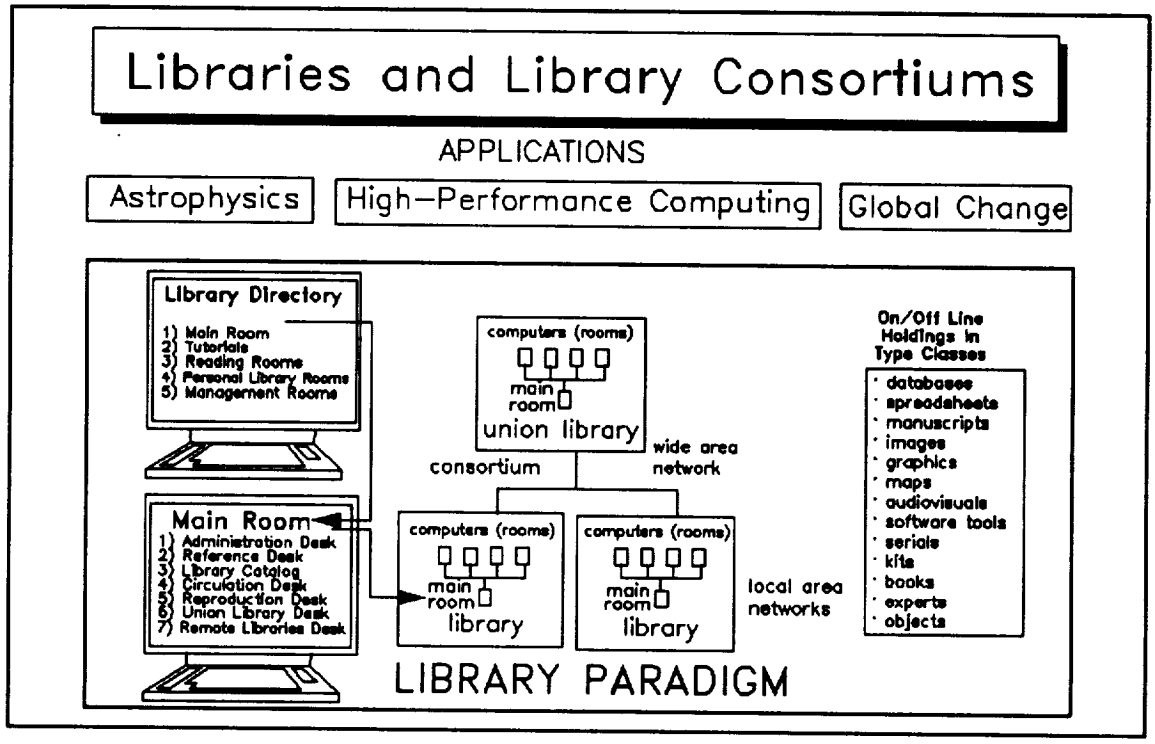

Figure 3.

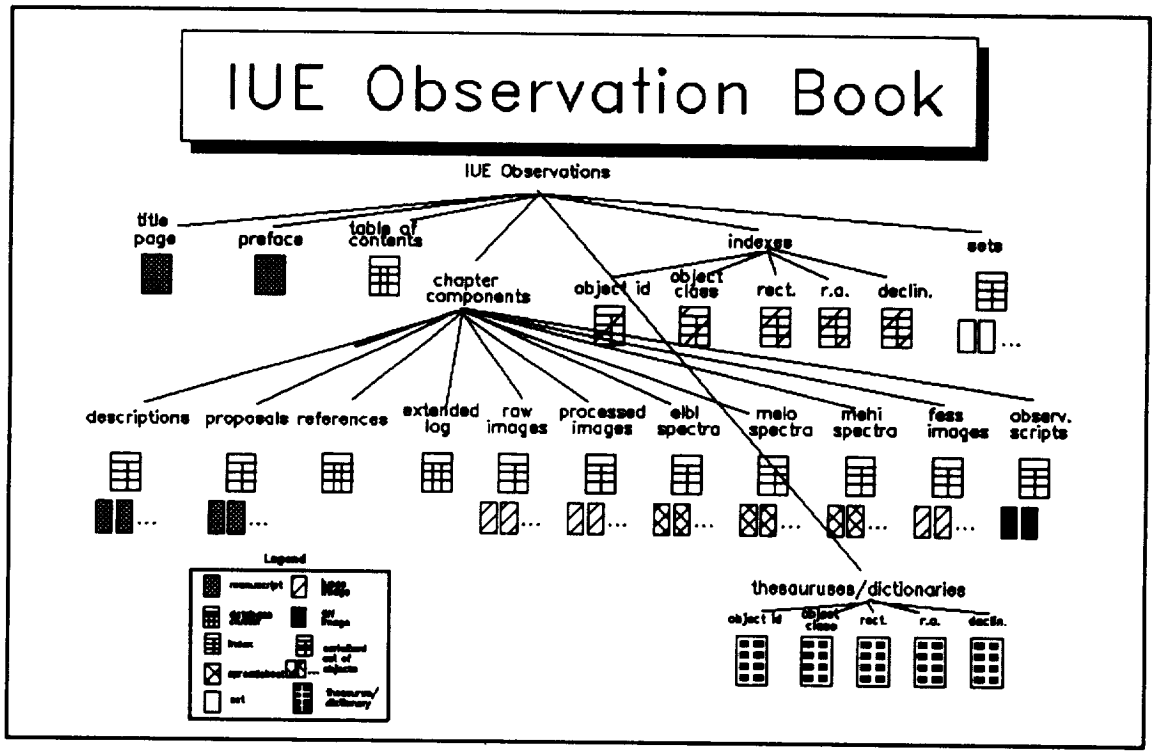

Figure 4.

(images), indexes including object id, object class, rectangle. In addition, each of the index components has a thesaurus/dictionary database. Another example of a book is the IUE Software Tool Book where each chapter represent one of 900 software programs written expressly for IUE data. A third example is the Local Library Catalogue Book where each chapter represents one of the holdings at the local area network.

A **kit** is an aggregate data structure which is just divided into physical units called **components**. An example of a book is given in Figure 5. for the IUE Software Tools kit for managing IUE software. The IUE Software Tools kit has as components- a title page (manuscript), a preface (manuscript), requirements document (manuscript), users guide (manuscripts), high-level design document (manuscript), system administrators guide (manuscript), users comments (manuscript), reference manual (book), low-level internals (book), installation (book), source code (book), object code (book), testing plan (book), and software tool (book). Another example of a kit is an Astrophysics Research Kit which represents an Research problem in Space Astrophysics and whose components are subkits each representing a sub-problem of the original problem.

Universal Object Type Management Systems

The DAVID approach, make use of universal object type management systems.

A **Universal Object Type Management system** allows one to retrieve, read and operate on object of a particular object type independent of the subtype and internal format. Figure 6. contains examples of object types for which universal object type management systems are important. For example, a universal database management system allows one to operate on databases independent of the type (e.g., relational, hierarchical, and network) and internal format (e.g., Ingres, IM/DM, Sybase, IMS, etc.). A universal manuscript management system allows one to operate on manuscripts independent of the type (e.g., structured, unstructured) and internal format (e.g., ASCII, Tex, Troff, Word Perfect, etc.). A universal image management system allows one to operate on images independent of type (e.g., level 0, level 1) and internal format (e.g., TIFF, FITS, IRAF, etc.).

DAVID And the SBIR Program

The DAVID project has made strong use of the Small Business Innovatve Research (SBIR) Program. A number of successful SBIR projects are described here.

Heterogeneous Network Access by Digital Analysis Corporation of Reston, Virginia. The purpose of this project was the development of a uniform layer that can be put on top of heterogeneous network alternatives such as TCP/IP, DEC Net, Bit Net, etc. Each subnet provides basic network services and hops between subnets are handled by servers at the gateways.

Heterogeneous Interface Building by K. Wanderman and Associates of Staten Island, New York. The purpose of this project is the development a software that will facilitate development of interfaces of database management systems and arbitrary files into the NASA developed Universal Database Management System. The software will be playing a central role in the establishment of major repository of Astronomical Catalogues at the NSSDC.

Universal Index Management Systems by Advanced Communication Technology Incorporated of Silver Spring, Maryland. The purpose of this project is the development of a universal index management system. The software will allow users and system builders to construct indexes of different types (one, two, three dimensional, etc) and different formats (B-trees, R-trees, hash, etc.). The software will be playing an important role in speeding up access of data in the Astrophysics Data System.

Universal Software Tool Management Systems by HSA Incorporated of Sugar Land, Texas. The purpose of this project is the development of a universal software tools management system for managing software over distributed systems. Each software tool is represented as a kit in a similar way that the IUE Software Tools Kit above was described. The software will be playing an important role in the management of software in the Astrophysics Data System.

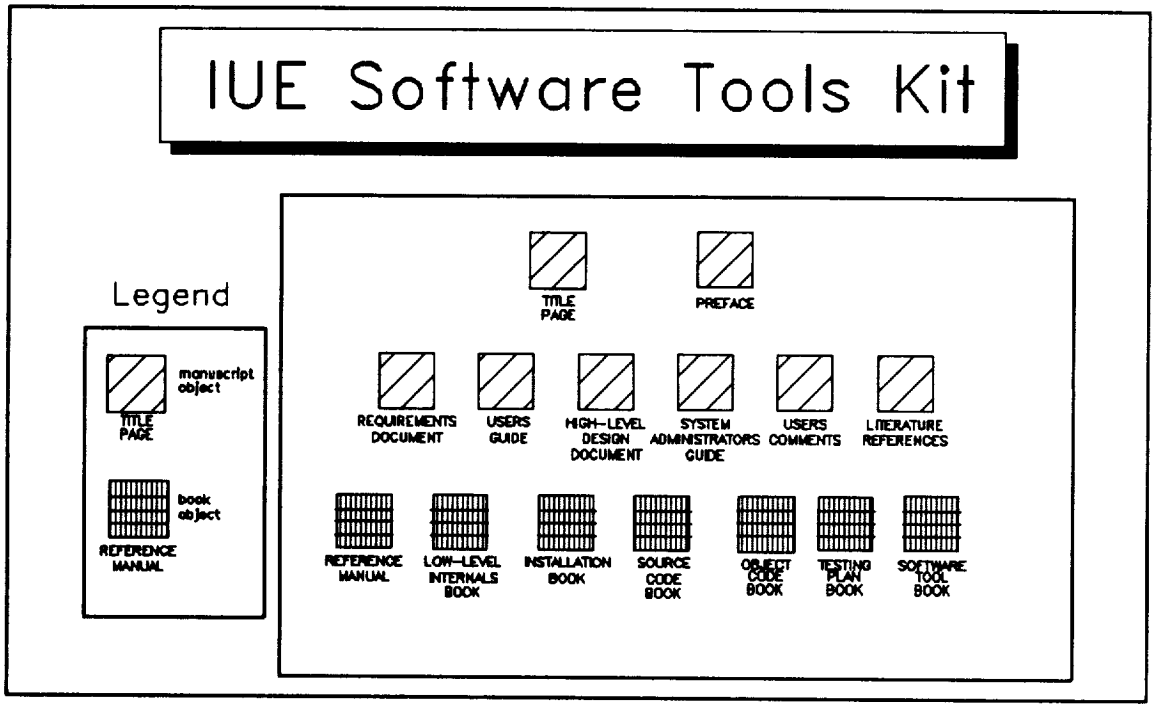

Figure 5.

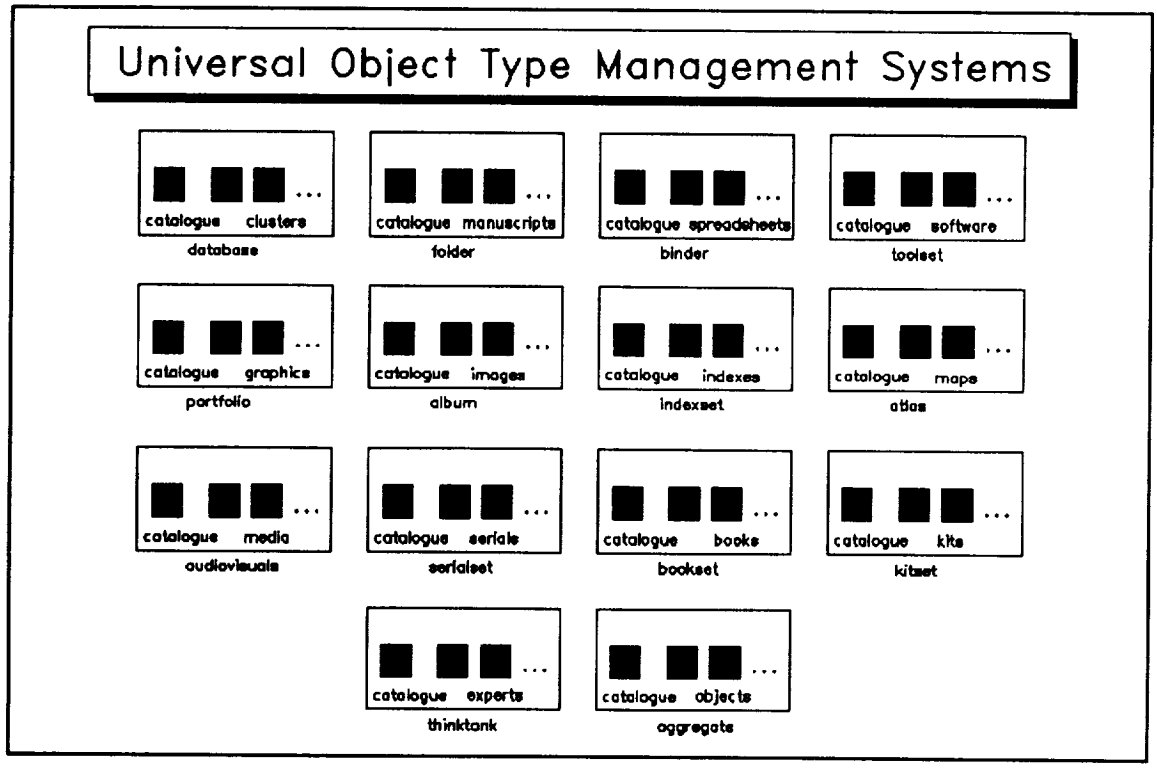

Figure 6.

Universal Object Type Management Systems by REI Associates of McLean, Virginia. The purpose of this project is the development of a management system for building other management systems. Each management system is represented as a kit which contains retrieval software tools and request server tools. The software will be playing an important role in the development of other Universal Object Type Management Systems.

Universal Book Management Systems by Advanced Applications Corporation of Potomac, Maryland. The purpose of this project is the development of a universal books management system for managing book objects. Each book is represented in a similar way that the IUE Observation Book above was described. The software will be playing an important role in the management of data in the Astrophysics Data System.

Summary

The Distributed Access View Integrated Database (DAVID) System, which has been adopted by the Astrophysics Division for their Astrophysics Data System, is a solution to the system heterogeneity problem. In Section 1, we outlined the heterogeneous components of the Astrophysics problem. In Section 2, we described the Library and Library Consortium levels of the DAVID approach. In Section 3, we discussed the Books and Kits level. In Section 4, we described the Universal Object Type Management System level. In Section 5, we explained the relation of the DAVID project with the Small Business Innovative Research (SBIR) Program.