Abstract

New information technologies such as distributed computing systems and the National Information Infrastructure have brought the integration for global enterprises closer to being a reality. These technologies and standards alone, however, are not sufficient to effect scalability, adaptability, and usability of integrated enterprise information systems, without which the integration can collapse of its own weight. We present a metadatabase model to address these issues in the context of global manufacturing enterprises. The problem, the model and an application solution are described for this domain.

Enterprise Information Management for Global Manufacturers

Cheng Hsu and Alan Rubenstein
Rensselaer Polytechnic Institute
Troy, N.Y. 12180-3590

1. The Vision: Scalable and Adaptive Integration

In today's global marketplace, manufacturing companies are continuously seeking new ways to address competitive pressures; the latest efforts include Manufacturing Intelligently, Agile Manufacturing, Virtual and Horizontal Corporations and other information technology based visions and models. To fully realize their promises, however, requires new fundamental results in the key areas of integration and management of enterprise information across multiple data and knowledge systems. This need can be articulated from two angles:

(1) Reality Why does this problem arise and will not go away with standards alone?

(2) Technology Why are previous results not sufficient in solving this problem?

An Analysis of the Reality Problem: Enterprise Integration is Difficult to Accomplish

On the reality side, the striving for enterprise - wide management and integration of information has triggered major industry and government sponsored initiatives. These include efforts to define standards which will enable data exchange between systems from different vendors, techniques for business process re-engineering and complete architectures for computer integrated manufacturing. However, the information systems which support modern automated enterprises exhibit several characteristics which continue to make enterprise integration difficult.

1. The introduction of information systems into manufacturing enterprises is often ad hoc and"bottoms-up".This tendancy to rely on customized and compartmentalized technology is due to technical necessity as well as the need of organizational control. Technically, the most successful environments for automation are those in which well defined business requirements can be isolated and solutions carved out to meet the specific requirements of users. Such specialized systems are not designed to meet requirements for access to information from other enterprise users.

2. Once in place, however, these specialized systems eventually must interact with each other across local and wide area networks to support many different manufacturing activities. In other words, an enterprise cyberspace environment must be created to accommodate this interaction. The implementation of information exchanges among independently designed systems in contexts that were not known or fully understood by their designers is difficult. Procedural rules and data semantics that have been designed into the systems must be understood and often must be extended to enable access to multiple databases by decision support applications.

3. Systems must continually evolve if they are to respond to changes in technology and business requirements. Unfortunately, the need to understand the complexity of integrated systems that serve many application systems and user communities in order to modify them often increases rather than reduces the expense and time required to adapt them to changing requirements. The integration of systems to meet enterprise needs rather than the requirements of their primary users can be perceived as an inhibitor to local autonomy, flexibility and performance.

An Analysis of the Technical Problem: There Needs to be an Active Heterogeneous Distributed Data Base Management System (HDDBMS) with Adaptability, Scalability and Intuitive Ease of Use.

On the technical side, modern manufacturing practices, such as concurrent engineering, require global access to information stored and processed at multiple data and knowledge systems operating over local and wide-area networks. The hallmark of these multiple systems is their defiance of traditional control models: they do not abide by any single standards, are not amenable to centralized administration, and cannot be comprehensively modelled for the entire enterprise. A single factory in a modern manufacturing enterprise could easily include over a thousand databases running at a scale of over a million transactions a day. The complexity is overwhelming:

4. Traditional approaches to sharing data among heterogeneous systems involve schema integration and serialization to insure data consistency. Designs that accomplish this in networks of heterogeneous systems are complex. They are prone to performance problems and bottlenecks due to the need for a central control point to manage schema translations and

synchronization protocols. The coupling among systems that result also inhibits the flexibility of local systems to adapt quickly to changes in the requirements of their primary users.

5. In addition to such well-known issues of heterogeneous databases as interoperability and distributed updates, this environment also faces rapid changes in its underlying (data and knowledge) technology, business processes, and applications. Thus, the unpleasant reality of legacy systems and heterogeneity will always stay. Progress on standards would hardly make the problem fade away, since today's standards will become tomorrow's legacies in the face of new cutting edge technologies, which tend to transcend any standards that require a long time to develop and take effect.

The technical nature of the problem is thus concerned with adaptive integration and the solution can be called an active heterogeneous distributed database management system (HDDBMS). In the context of established literature, such a solution simply does not presently exist.

In this paper, we present an industrial scenario to illustrate the Enterprise Information Management problem (section 2) and a technical formulation of a novel model to solve the problem (section 3). The basic elements of the solution - i.e., the metadatabase model and its further development into a prototype Enterprise Information Manager are discussed respectively (sections 4 and 5).

2. An Industrial Case for the Enterprise Information Manger

We propose to design an enterprise information manager which is the basis for a commercial product and which will provide shared access to enterprise data and knowledge through a new improved intuitive graphical visualization system, supporting distributed and scalable client/server systems, multiple database technologies and legacy systems. The rationale can best be illustrated by the specific example of a hypothetical but otherwise realistic industrial case..

Consider a multi-billion dollar world-wide enterprise who's objective is to be the marketplace leader through excellence in products and services at the lowest cost in the industry with superior service to the customer. It's products include heavy machinery, industrial systems and service activities

Employment is over 30,000 with 80% domestically and the remainder overseas..

With regard to the necessity for sharing information, consider that there are over 20 manufacturing facilities, over 75 apparatus shops, 0ver 100 sales offices, almost 200 engineering offices and almost 100 business associates/licensees in over 50 countries. These often manufacture all or parts of the product.

A recent study uncovered a number of suspected, but not previously highlighted, factors associated with the manner in which information was utilized.

- The knowledge of what information was available is limited
- Most available information exists in a format suitable to a select few.
- Current solutions address predefined needs as opposed to supporting ad hoc inquires
- Current solutions are not readily used on a casual basis
- Redundant information exists but confidence in accuracy is lacking

In short, it is apparent that a significant need exists for easy, rapid access to consistent accurate information throughout the business.

Conservatively speaking, access to required information throughout the business can easily save over five million dollars per year. With 10% usage from a population of 37,000 working 240 days per year and a savings of only 15 minutes a day, a productivity saving and corresponding cycle reduction of 222,000 hours per year is easily realizable. Additionally, a large proportion of the information access requirements involve at least one other person, thereby offering additional opportunities for saving. and increasing the savings proportionally.

From both a cycle time and dollar saving viewpoint, a system which allows shared access to timely and easily obtained information, is not only a good business investment, but is essential to achieve a competitive position in the agile manufacturing environment.

Some Requirements of the Solution Technology

- Existing applications and their data must be integrated into the framework without redesign.
- Changes in information management technology and business requirements must be incorporated easily.

- Design must be implementable in a network of distributed heterogeneous nodes.

- Design must accommodate a number of access points through different hardware.

The user access must be intuitive and suitable for users with minimal knowledge of the environment and minimal computer expertise.

3: The Solution Approach: The Notion of Enterprise-Level Metadata Independence

The unique requirements of multiple systems in enterprise information management may be summarized as follows:

Scalability	The total enterprise information integration environment must allow incremental development and be expandable, such that the integration can start with a small part of the enterprise and gradually extend to the rest (even to other organizations) over time, without losing operational continuity and structural integrity.
Adaptability	Systems that use either standard or non-standard technologies as well as new and legacy systems, can be incorporated into the integrated environment in a seamless way without causing any disruption to any existing systems.
Parallelism	The multiple systems must be able to operate concurrently while achieving synergism for the enterprise, without requiring global serialization or similar synchronization mechanisms imposed on any instance-level transactions.
Autonomy	Local systems in the integration need to have the flexibility to be designed, constructed, and administered independently by the local management alone, without having to conform, nor later convert, to a global schema.
Visualizability	The enterprise information should be represented and presented with such methods that avail the end users a cyberspace which is accessible through intuitive visualization and potentially, virtual environments for interface. This is essential in order for the system solution to be accepted and utilized by employees at large.

The above requirements are fundamentally identical to the classical concept of data independence using the three-schema architecture; the difference is the "primitive", or primary concern, in each: At the (single-site) database level, the primary concern is multiple applications processing data instances; whereas at the enterprise level, the primitive is multiple (database and knowledge-based) systems processing applications. Since the systems are modeled and hence substantiated with metadata, then it is evident that the enterprise-level problem can be formulated as a metadata problem whose conceptual complexity, in metadata terms, is similar to the traditional database-level problem. Therefore, in the spirit of data independence for scalable, adaptable, parallel, and autonomous applications against a database, we refer to the enterprise-level requirements of scalable, adaptable, parallel, and autonomous databases as **metadata independence**. The search for a solution to the enterprise information management problem is, it follows, focused on transforming the data problem into a metadata problem and bringing the proven model of databases to the enterprise level, thereby effecting a metadata independent architecture to simplify the complexity .

Traditionally, database researchers all cherished three principles: the use of a (global) data model, the reliance on an integrated schema, and the enforcement of global serialization. This tradition has been carried on throughout the myriad efforts in developing distributed databases, federated databases, and multidatabases; and is still dominating in many the latest endeavors of integrating multiple data and knowledge systems operating in heterogeneous environments across wide-area or even worldwide networks. Although a great deal of progress has been accomplished in the past decade on, e.g., interoperability, local autonomy, and open system architecture, a great deal more still remains to be accomplished; which centers especially around the issues of **concurrent processing** and **architecture adaptability**. Other aspects of metadata independence mentioned above are also based on these two technical issues. Consider the previous example of a modern manufacturing enterprise, multiplying its thousand-databases factory by a factor of ten and linking them through a global network, and the significance of these two issues becomes immediately evident. Together, they are referred to as the adaptive integration problem. The solution to this problem entails the following basic elements transforming the three database principles into enterprise-level to achieve metadata independence:

(1) An enterprise information model: this model globally represents all local data models and their contextual knowledge in the enterprise with a metadata-independent structure which, when put online, allows all local models and other metadata contained in it to be added, deleted, or

modified through ordinary metadata transactions (as opposed to a fixed global data model);

(2) An online (independent, but sharable) metadatabase: this metadatabase implements the enterprise information model, and comprises a scalable hierarchy of mini-metadatabases for any scope of local functions in a (expandable) client-server manner (as opposed to schema integration); and

(3) A concurrent architecture for execution: this architecture (including its execution model) supports concurrent processing of local systems with localized distributed control knowledge (as opposed to global serialization).

Together, they amount to a metadabase-supported, rule-oriented concurrent systems solution to the problem. On this basis, new informmation visualization technology and existing commercial technologies for distributed computing and information warehousing can be also incorporated and result in a workable EIM.

4: Metadatabase model for Scalable and Adaptive Integration

The metadatabase model is characterized by the unique solution approach that converts the integration problem from one that deals directly with data instances to one that controls through metadata, and thereby provides metadata independence for multiple systems. This approach effects concurrent processing and architecture adaptability through the following portfolio of results.

The Basic Model: A Metadatabase-Supported, Rule-Oriented Concurrent Architecture

The concurrent architecture is depicted in Figure 1. The metadatabase itself (a rigorously constructed collection of enterprise metadata) provides an integrated enterprise model for the multiple information systems, their databases, and the interactions among the different systems; i.e. the information contents and their contextual knowledge. The metadatabase approach (1) uses the enterprise model to assist end-users performing global queries free of both technical details and a hierarchy of integrated schemata; (2) distributes the contextual knowledge to empower these local systems to update data and communicate with each other without central database control; and (3) incorporates legacy, new or changed local models into its generic structure of metadata to support evolution without system redesign or recompilation. The shells in the concurrent architecture, therefore, implements the distributed (localized) knowledge which, in turn, is managed by the metadatabase.

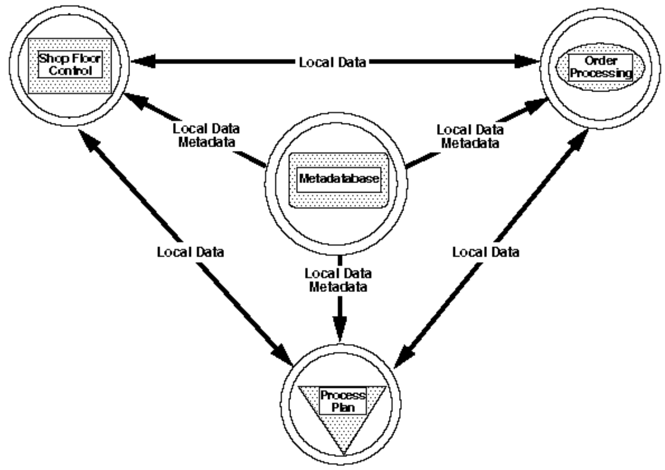

Figure 1 : The Concurrent Architecture Using a Metadatabase

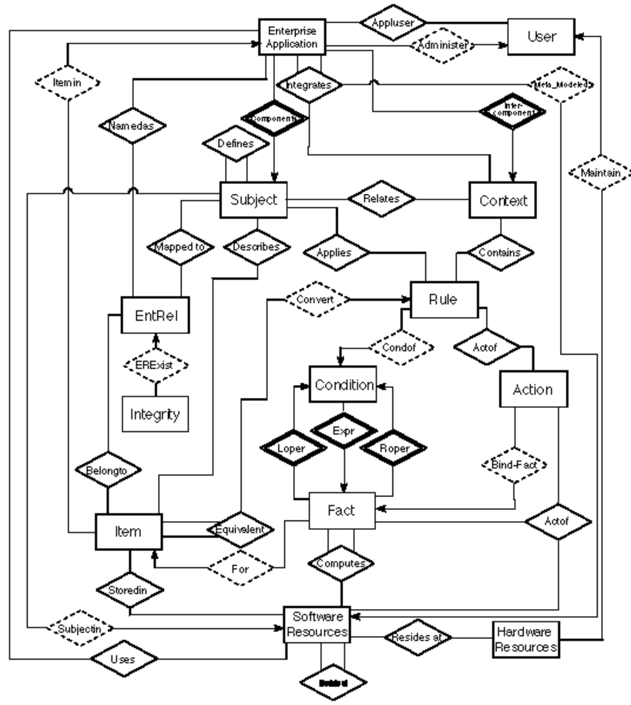

Figure 2 : The GIRD Meta-Model

The Metadatabase Structure: GIRD Model

The metadatabase itself employs a generic meta-structure, the Global Information Resource Dictionary (GIRD) model (see Figure 2), abstracting the enterprise metadata resources (i.e., models). As such, each and every local model is represented (but not duplicated, nor removed) into the metadatabase as ordinary metadata "tuples" populating the structure. The metadata independence at the model integration level is achieved since any change in enterprise models would involve only ordinary metadata transactions similar to the usual relational processing, and do not require any change to the structure itself, nor reloading/recompilation. The structure itself, as well as the representation of local models, is developed according to the Two-Stage Entity Relationship (TSER) method, which serves for this purpose as a meta-model. TSER provides at the first, functional stage two basic representation constructs: SUBJECTS (similar to objects in object-oriented paradigm) and CONTEXT (rule-based description of process and other contextual knowledge). At the second, structural stage, ENTITY (characterized with singular keys) and RELATIONSHIP (characterized with composite keys) plus two more types of associations signifying special integrity rules (functional and mandatory) are defined. All these constructs, combined with hardware and software classes of metadata, are included in the GIRD model.

A meta-model system is developed to support the modeling and creation of the metadatabase

Meta-Model : The Conceptual Schema for Integrated Multi-Model Environments

Business databases, manufacturing databases, and engineering design databases have traditionally followed different paths of evolution and espoused different paradigms, although they are all based on the same information technology. To compound the situation further, each paradigm has also prompted a number of different modeling tools since the advent of Computer-Aided Software Engineering (CASE). The integration of manufacturing functions must deal with the full scope of paradigm translation problem facing business, manufacturing and engineering design databases pertaining to the enterprise.

Therefore, the notion of using a meta-model to anchor these paradigm translations has emerged recently. A particular metamodel system using the Two-Stage Entity-Relationship (TSER) method is developed to provide a compact solution supporting the metadatabase model.

The substantiation of the meta-model concept in this system spans three levels:

(1) At the modeling (or metadata) level, the meta-model is a neutral paradigm serving as the common representation method that all paradigms are translated into (see figure 3).

(2) At the models integration (or metadata management) level, the meta-model is a generic metadata schema abstracting and structuring all models into an integrated enterprise metadatabase (see figure 4).

(3) At the information management (or data instances) level, the meta-model is the integrated enterprise model contained in the metadatabase (see figure 2).

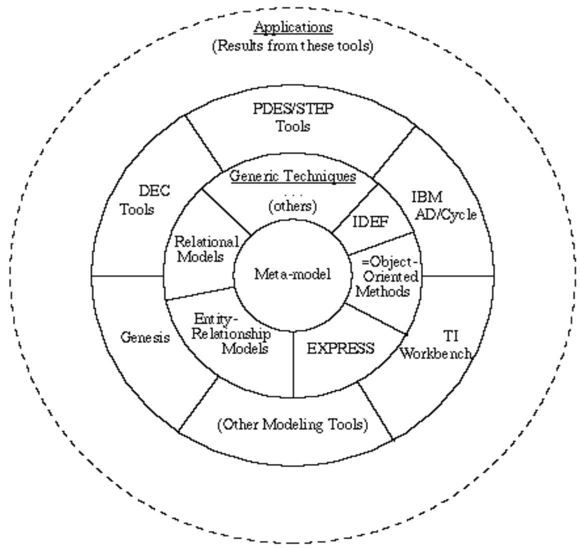

Figure 3. The Paradigm Translation Knowledge Kernel for Layered Mapping

The Metadatabase-Supported Scalable Integration Hierarchy

In large-scale systems, the metadatabase can be distributed in a recursive manner and constitutes a hierarchy of nodes, illustrated in Figure 5, where, for simplicity, only three levels are shown. At the leaf, several of the application systems are represented in a sub- or mini-metadatabase. This mini-metadatabase then can be represented in the main metadatabase system such that the meni-metadatabase becomes the gateway to the application systems it represents. There can, of course, be as many levels of sub-/mini-metadatabases as needed. The real significance of this hierarchy, however, is not its top-down construction, which is the predicament of virtually all other integration models; but rather its ability to effect bottom-up, incremental development and expansion — i.e., scalable integration: the higher level nodes can be constructed from the ones immediately below it. A key element in this bottom-up, scale-up approach is the GIRD meta-model discussed above. This structure allows new application models to be incorporated into a metadatabase without disrupting the current systems.

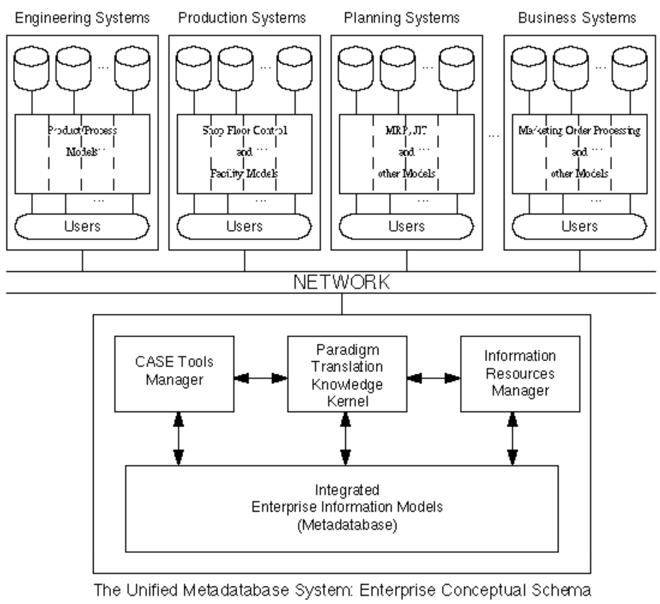

Figure 4. Enterprise Information Integration Using a Metamodel

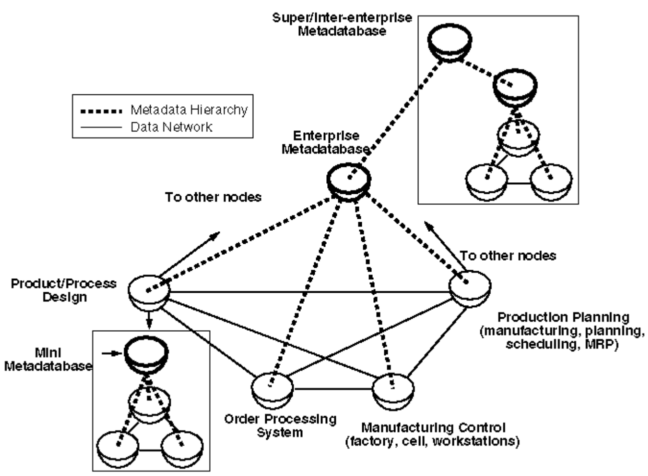

Figure 5 : Enterprise Integrated Through Distributed Metadatabase System

Thus, large scale, overwhelming integration environments can be achieved gradually, by first developing individual nodes in client (application systems) - server (metadatabase) type of clusters and then integrate them in a grand clustering of such clusters. In a similar way, the main metadatabase can be incorporated into a super or inter-enterprise metadatabase and become a mini-metadatabase of the latter. Both the clusters of metadatabases and the clusters of application systems at the leaf nodes are managed by the same active HDDBMS method.

The Metadatabase Management System Shell

A complete management system is developed for the metadatabase (signified as the center shell in Figure 1). The major elements of this system are depicted in Figure 6. It performs three major tasks: (1) management of the metadatabase itself, including metadata queries, analyses, and rulebase processing ; (2) formulation and processing of global queries, utilizing the metadatabase to provide online intelligent support to users; and (3) generation of data management rules as well as operating rules and interfacing with ROPE for event-based information flows and data management. The elements performing these tasks are self explanatory in Figure 6.

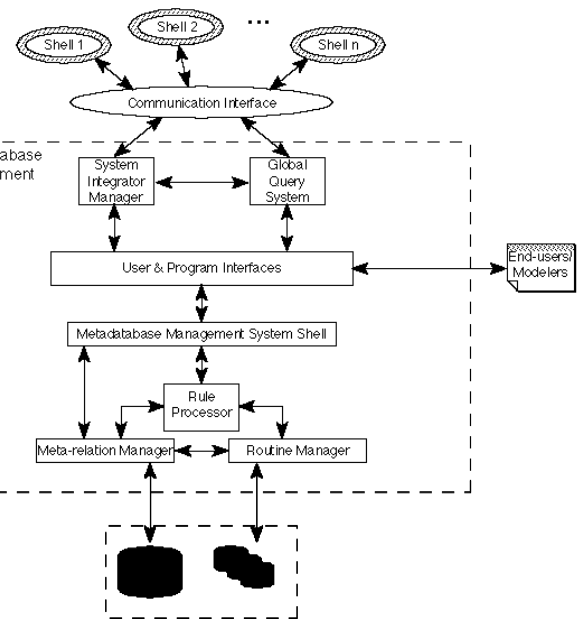

Figure 6 : The Metadatabase Management Shell Architecture

The Visual Information Universe Model

One key area that has not received adequate attention - or more accurately, has not been sufficiently understood - is the next generation user-interface for enterprise information management. We submit that this next generation technology will and must go beyond the present concept of GUI (graphical user interface) and VR (virtual reality). It should feature a new paradigm of information visualization with metaphors suitable for cyberspace type of ideas and applications; should possess contextual knowledge of the underlying information resources to assist end users acquire and assemble any pieces of information desired from everywhere needed; and should, thereby, effect a virtual environment for information access over networks where the user fly freely and query globally with little or no prior technical knowledge about the system. The new concept will be exemplified as the Visual Information Universe (VIU) in this project, using the new Globe metaphor (representing information resources), some basic VR tools (for direct spatial manipulation), and a metadatabase (containing models and other contextual knowledge). The VIU model will be further connected with network protocols and interfaces such as Mosaic to support immediately the huge user community of Internet as well as the information management needs in distributed enterprises.

The VIU Model.

Using the metadatabase for the sources of contextual knowledge as well as the object of visualization, the overall architecture of the VIU model is depicted in Figure 7 below.

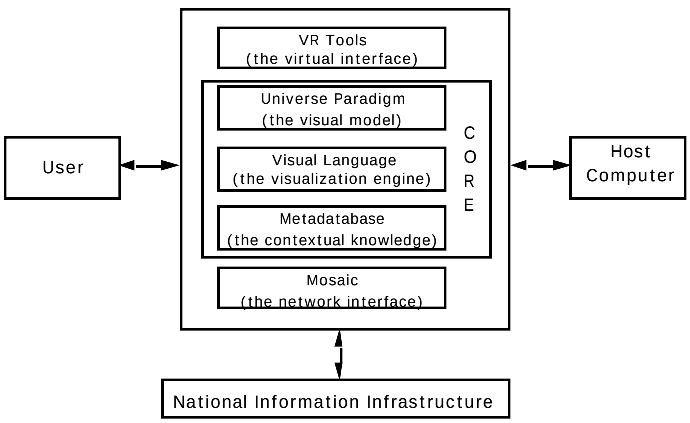

Figure 7. VIU Architecture

5: Conclusion: Preliminary Prototype Architecture

The implementation solution envisioned on the basis of the Metabase model is illustrated in figure 8 for the industrial scenario discussed in section 2. It incorporates two fundamental types of information management:

(1) Information Object Storage and Management - a collection of information stored as objects on optical discs, magnetic disks and magnetic tape. examples of objects include raster images of pre CAD engineering and manufacturing drawings and text material like old typewritten copy which were produced by means that do not lend themselves to digital representation.

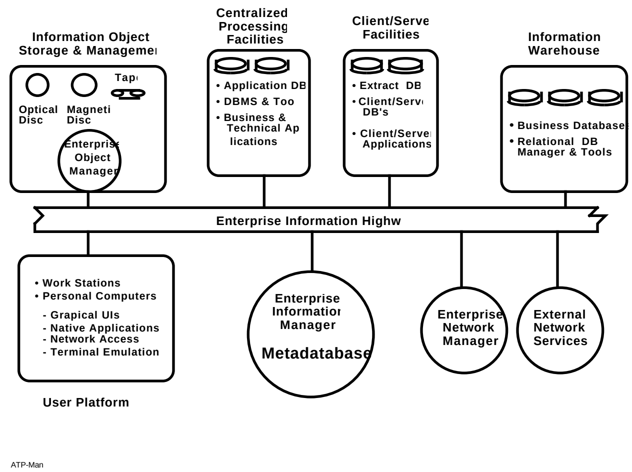

Figure 8: Preliminary Implementation Solution

(2) Information Warehouse which contains selected shareable information stored as data in an appropriate relational database. Examples include CAD system data files, analysis data, vector drawing files.

Generally, if information has been created prior to the implementation of the new system, it would be placed in (1) as a raster file. If the information is created after the implementation, it would be in an ASCII or other file which is suitable for residence in (2), the Information Warehouse.

Each of the resources contain appropriate tools and management software to allow retrieval and in some cases editing of appropriate information provided the user has detailed knowledge of the location and content of the information.

The required Enterprise Information Management system together with a new visual information system will make it possible for all designated users to have access to all of the information without detailed knowledge since the enterprise manager will contain information about location and data structures of the stored data. Most elements in the architecture shown in figure 8 are already available commercially, except for the Enterprise Information Manager. The proposed Metadatabase Technology fills in this need. A laboratory Prototype of the metadatabase exists at Rensselaer.

Acknowledgement:

The author wishes to thank Mr. John Manthorp of GE whose contributions to the industrial case and the design of the implementation architecture were very

helpful and Mr. Jay Leiserson of IBM for his articulation of the reality issues in section 1.

References

[Hsu87] Hsu, C. and Skevington, C, "Integration of Data and Knowledge in Manufacturing Enterprises: A conceptual Framework'", Journal of Manufacturing Systems, 6(4), pp277-285, December, 1987.

[Hsu90] Hsu, C. and Rattner, L., "Information Modeling for Computerized Manufacturing", IEEE Transactions on Systems, Man and Cybernetics, 20(4),pp758-776, July/August 1990.

[Elm91}Elmagarmid, A.K. and Ruzinkiewicz, M., "Critical Issues in Multidatabase Systems ", Information Sciences, 57-58, pp403-424, September - December 1991.

[Hsu91] Hsu, C., Bouzianne, Ratttner, L., and Yee, L. "Information Resources Management in Heterogeneous Distributed Environments: A Metadatabase Approach", IEEE Transactions on Software Engineering, 17(6),pp 604-625, June, 1991.

[Ozs91] Ozsu, M.T. AND Valduriez, "Distruibuted Database Systems: Where arewe now?", Computer, 24(8), pp 68-78, August 1991.

[Sil91] Silberschatz, A. M., Stonebreaker, M. and Ulman, J., " Database Systems: Achievements and Opportunities", Communications of the ACM, 34(10), pp 110-120, October, 1991.

[Bri92} Bright, M.W., Hurson, A.R., and Pakzad, S.H., " A Taxonomy and Current Issues in Multibase Systems", Computer, 25(3), pp50-60, March 1992.

[Sto92] Stonebreaker, M., "The Integration of Rule Systems and Database systems', IEEE Transactions on Knowledge and Data Engineering, 4(4), pp. 415-423, October, 1992.

[Hsu92] Hsu, C., Babin, G., Bouzianne, Cheung, W., Ratttner, L., and Yee, L., "Metadatabase Modeling for Enterprise Information Integration", Journal of Systems Integration, 2(1), PP5-39, JANUARY, 1992.

[Sto92] Stonebreaker, M., "The Integration of Rule Systems and Database systems', IEEE Transactions on Knowledge and Data Engineering, 4(4), pp. 415-423, October, 1992.

[Hsu93a] Hsu, C., L. Gerhardt, D. Spooner, and A. Rubenstein, "Adaptive Integrated Manufacturing Enterprises: Information Technology for the Next Decade," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 6, 1993.

[Hsu93b] Hsu, C., Y. Tao, M. Bouziane, and G. Babin, "Paradigm Translations in Integrating Manufacturing Information Using a Meta-Model: The TSER Approach," Engineering of Information Systems, Vol. 1, No. 3, 1993, pp 325-352.